Technische Universität München
Zentrum Mathematik
Lehrstuhl für Finanzmathematik (M13)

Estimation of factor models with incomplete data and their applications

Franz Hubert Ramsauer

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:		Prof. Donna Ankerst, Ph.D.
Prüfer der Dissertation:	1.	Priv.-Doz. Dr. Aleksey Min
	2.	Prof. Dr. Hajo Holzmann Philipps Universität Marburg
	3.	Prof. Rustam Ibragimov, Ph.D. Imperial College Business School (nur schriftliche Beurteilung)

Die Dissertation wurde am 21.02.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 12.07.2017 angenommen.

Zusammenfassung

Diese Arbeit betrachtet die Schätzung von Faktormodellen mit unvollständigen Daten, die sich aus dem Fehlen von einzelnen Beobachtungen sowie der Kombination von gemischt-frequenten Zeitreihen ergeben. Die gleichzeitige Berücksichtigung von Finanzmarktdaten und makroökonomischen Kennzahlen soll dabei ein möglichst umfassendes Bild der zu untersuchenden Märkte liefern. Der Gebrauch von Faktormodellen bietet sich in diesem Zusammenhang an, um das breite Spektrum an Informationen in Form von wenigen, unbeobachteten Faktoren zu bündeln. Mit Hilfe von approximativen Faktormodellen bilden wir zudem Abhängigkeiten zwischen verschiedenen Zeitreihen ab. Für zwei unserer Modelle bedeutet dies, dass nicht nur die Faktoren, sondern auch die idiosynkratischen Fehler querschnittlich korreliert sein können. Die Schätzverfahren, die wir hier vorschlagen, basieren auf zwei Expectation-Maximization Algorithmen, die im Wechsel benutzt werden bis ein gegebenes Abbruchkriterium erfüllt ist. Aufgrund der Tatsache, dass sich die Faktoren nicht beobachten lassen, müssen deren Erwartungswerte und Kovarianzmatrizen geschätzt werden. Neben den klassischen Kalman Filter und Smoother, verwenden wir hierzu eine analytische Lösung beziehungsweise leiten modifizierte Varianten des Kalman Filters und Smoothers her, die den zugrundeliegenden Modellspezifika explizit Rechnung tragen. In einem nachfolgenden Schritt dienen die geschätzten Faktoren als exogene Variablen bei der Modellierung von Indexrenditen. Auf Basis der Faktorverteilungen können wir sowohl Punkt- als auch Intervallschätzer für die Vorhersage zukünftiger Renditen herleiten. Zusätzlich zur Vorhersage erwarteter Renditen decken wir deren treibende Kräfte auf. Wir beantworten somit die Frage, welche Informationen uns zu diesem Ergebnis führten. Im Rahmen der Intervallschätzung stellen wir dynamische Handelsstrategien vor, die die vorhergesagten Renditeintervalle in konkrete Anlageempfehlungen übertragen. Abschließend zeigen wir, welchen Beitrag unser Ansatz im Bereich der Portfoliooptimierung leisten kann.

Abstract

This thesis considers the estimation of Factor Models with incomplete data arising from the absence of single observations and the combination of mixed-frequency time series. Thereby, the joint use of financial data and macroeconomic indicators is supposed to provide a picture, as comprehensive as possible, of the markets to be analyzed. In this context, we apply Factor Models to bundle a broad range of information by a few unobservable factors. Furthermore, with the help of Approximate Factor Models we map dependencies between different time series. This means for two of our models that not only the factors, but also the idiosyncratic errors may be cross-sectionally correlated. The estimation methods, which we propose here, involve two Expectation-Maximization Algorithms that are alternately applied until a given termination criterion is met. Due to the hiddenness of the factors, their means and covariance matrices have to be estimated. Besides the standard Kalman Filter and Smoother, we deploy a closed-form solution for this purpose or derive modifications of the Kalman Filter and Smoother that explicitly take the assumed model characteristics into account. In a next step, the estimated factors serve as exogenous variables for modeling index returns. Based on the factor distributions we determine point and interval estimates for forecasting returns of future periods of time. In addition to return predictions, we reveal their drivers. Hence, we answer the question of which information guides us to this conclusion. Within the scope of interval estimation we suggest dynamic trading strategies converting the forecasted return intervals into specific asset allocation recommendations. Eventually, we demonstrate which contributions our approach may provide in the area of portfolio optimization.

Acknowledgments

At first, I gratefully acknowledge my supervisor PD Dr. Aleksey Min for his excellent support during my whole doctoral phase. He was literally always available, not only for research related questions, but also for the general ups and downs, which a dissertation involves. His valuable feedback significantly improved the quality of this thesis.

Furthermore, I am very grateful to Prof. Dr. Rudi Zagst for the opportunity to do my doctoral studies at his chair and to present my work at several international conferences. With his comprehensive knowledge, especially in the area of asset and risk management, he raised fruitful discussions, answered many of my questions and gave the final touch to our papers.

Without the generous support of Pioneer Investments this thesis would not have been possible. Therefore, the same gratitude is due to Evi Vogl, Francesco Sandrini, Ph.D., Lorenzo Portelli and Monica Defend, who developed a project for the monitoring of financial markets, which set the cornerstone of this thesis, started a cooperation with the Technical University of Munich and eventually, extended the project plan by another year such that I was able to finish my work without any urgency.

In the course of the cooperation with Pioneer Investments, I got the chance to gain insights in the business of a globally operating asset management company in the form of a short trip to the Milan branch as well as internships in the Munich and Dublin offices. I particularly thank Thomas Kruse and Ali Chabaane, whose teams I was able to join as intern. Moreover, I thank all Pioneer colleagues for the great experiences in Dublin, Milan and Munich.

A doctorate is a great and unique experience, in particular, when there are nice and awesome colleagues. This is why I thank all my colleagues at the chair of finanical mathematics for the constructive discussions, events and fun we had together.

I warmly thank Prof. Dr. Hajo Holzmann, Prof. Rustam Ibragimov, Ph.D. and Prof. Donna Ankerst, Ph.D. for serving as referees of this thesis and chair of the examination committee, respectively

Eventually, I particularly thank my family and friends for their support and patience during my doctorate. In this context, I most sincerely thank my parents for their love, their outstanding support and their never ending patience. The same gratitude is due to my girlfriend Michaela for her love, her incredible patience and her extraordinary support.

Contents

Chapter 1

Introduction

We start with the reasons behind this thesis. Thereby, we state some general questions serving as attractive research opportunities from our point of view. Thereafter, we list the main objectives and contributions of this work. In the second section, we describe the structure of the thesis in more detail.

1.1 Motivation and Objectives

At any point in time, knowledge of the current financial and economic conditions is important for investors, asset managers, central bankers, politicians and many more. Whenever decisions are taken, a picture of the present situation, as comprehensive as possible, is particularly precious. On the one hand, the interactions between capital markets are not always the same. On the other hand, the market participants' perception of risk and hence, their risk appetite are changing. In this context, they focus more on financial or economic information. For consistency reasons, a realistic market model includes finanical and economic data, since the timely detection of the transition from one source to another appears impossible. This is why models, which are restricted to financial or economic time series, do not map the whole environment and so, cover only a small part of the bigger picture.

Among other things, information of macroeconomic, valuation, technical and flow nature moves financial markets. Unfortunately, changing dependencies between such signals may cause contrary indications of the current and future market conditions. To be precise, in the literature they distinguish between *nowcasting* and *forecasting*. For instance, Bańbura et al. (2011) regard nowcasting as the "problem of predicting the present, the very near future and the very recent past". Especially in case of macroeconomic variables, which are quarterly reported with long delays, monitoring the present market conditions is challenging. As stated in Bańbura and Rünstler (2011, ECB working paper, p. 23, Table A.1), there are also monthly indicators with a publication lag of more than two months. Many countries publish their Gross Domestic Product (GDP) figures once per quarter such that nowcasting of GDP became quite popular in research. In this sense, Liu and Hall (2001), Giannone et al. (2008), Lahiri and Monokroussos (2013) and Aastveit et al. (2014) focus on nowcasting of United States (US) GDP, while Schumacher and Breitung (2008) and Marcellino and Schumacher (2010) consider German GDP. For Norwegian GDP, see Aastveit and Trovik (2012) and Luciani and Ricci (2014). Although forecasting covers the prediction of future periods of time and was thoroughly discussed in past papers, the transition from short-term forecasting to nowcasting is rather smooth. Therefore, further contributions to the now- and forecasting of GDP are stated in Giannone et al. (2008, 2009), Barhoumi et al. (2010), Angelini et al. (2010, 2011) and Bańbura and Rünstler (2011). Instead of proceeding with a detailed, but lengthy literature overview or a short, disordered and most likely non-exhaustive stringing together of papers, we do not state further references in the field of forecasting now. By contrast, we refer to the preliminaries in Chapter 2, where we review all articles that are most relevant for this thesis.

Modern information technologies admit the collection, storage and processing of huge amounts of data. On the one hand, the abundance of data supports the construction of new models and the extension of existing ones. On the other hand, Boivin and Ng (2006) showed that more data does not always improve forecasting results. In their setting, Factor Analysis restricted to 40 out of 147 time series outperformed Factor Analysis based on the overall 147 time series. Hence, the identification and extraction of relevant information from big data remains an important issue. Sometimes, published data is revised afterwards, e.g., preliminary GDP values, such that data revisions are another burden. Even though the magnitude of their impact on forecasting results depends on the respective application, Croushore (2006) confirmed that data revisions can affect forecasts. Finally, big data may cover a variety of diverse frequencies ranging from time series updated every second to quarterly ones. In addition, holidays, publication conventions, trading suspensions, etc. also cause the absence of single observations. Consequently, an overall data set might be incomplete representing another topic to be addressed.

From a practical perspective, state-of-the-art market models should tacle from the previous tasks as many as possible within a reasonable period of time. Hence, a model's implementation also matters. Irrespective of whether the current or future market conditions are considered, an appropriate model detects how all inputs contribute to the outcome. In this way, it reveals possible sources of risk and indicates how reliable the findings are. Thereby, uncertainties or instabilities caused by the models themselves, e.g., from model selection and parameter estimation, should be taken into account. As soon as the drivers of an expected market environment are known, subsequent investigations can be triggered. For instance, if a model shows that a loose monetary policy of the central bank is responsible for a stock rally, an investor would be more concerned about a change in the monetary policy than poor GDP growth rates. With this in mind, he could timely prepare a hedging strategy that fits to his risk attitude.

This thesis aims at constructing now- and forecasting frameworks for financial markets. Thereby, we estimate Factor Models with incomplete panel data, address potential identification and model selection issues, discuss uncertainties caused by factor and parameter estimation, trace outputs back to our input data and propose dynamic trading strategies, i.e., asset allocations based on our findings. Eventually, we develop a similar approach for analyzing the impact of monetary policy actions on financial markets and the real economy.

With a view to the existing literature main contributions of our work are as follows: First, we apply Factor Models (FMs) for mixed-frequency panel data to support portfolio optimization. That is, we are among the first to transfer FMs, which are well-known in statistics and econometrics for modeling macroeconomic data, to the field of asset and risk management. Besides theoretical considerations, we provide algorithms and illustrative examples based on real-world data. This makes our approach attractive for practitioners.

Second, we determine the conditional means and covariance matrices of the latent factors in Approximate Dynamic Factor Models (ADFMs) in closed form. In the literature, both are usually estimated by a run of the standard Kalman Filter (KF) and Kalman Smoother (KS). In a Monte Carlo (MC) simulation study, our two-step method for estimating ADFMs with incomplete panel data based on the conditional means and covariance matrices in closed form performed better than the same two-step approach using means and covariance matrices of the standard KF and KS. Hence, the usage of Kalman Filtering techniques in such cases becomes optional.

Third, Doz et al. (2012) showed that the factors in ADFMs with cross-sectionally and serially correlated idiosyncratic shocks can be consistently estimated in the maximum likelihood framework. In the sequel, Bańbura and Modugno (2014) derived an estimation method for Exact Dynamic Factor Models (EDFMs) with incomplete panel data. Referring to Doz et al. (2012), they argued that cross-sectional dependence of the idiosyncratic shocks can be neglected to justify the validity of their estimation method for ADFMs. Since Doz et al. (2012) provided asymptotic results, we present an alternative two-step estimation method for ADFMs with incomplete panel data, which admits cross-sectionally correlated shocks. In a MC study, this method dominates the approach of Bańbura and Modugno (2014) for incomplete panel data of small sample size with cross-sectionally correlated errors. So, we show that cross-sectional dependencies matter in such scenarios.

Fourth, we design a two-step procedure for the selection of the factor dimension and autoregressive order. In doing so, we keep general factor dynamics of order $p \geq 1$ and do not consider the simple case of $p = 1$. This is why our two-step estimation method for ADFMs simultaneously performs parameter estimation and model selection.

Fifth, we propose single-market trading strategies, which convert prediction intervals into concrete actions. Moreover, we break point forecasts of returns for future periods of time down into the single contributions of the panel data. This approach supports plausibility assessments of the obtained results and reveals the main drivers and so, the main sources of risk, of the expected market behavior.

Sixth, we derive a modification of the standard KF, which takes into account that factors in case of Factor-Augmented Vector Autoregression Models (FAVARs) are partially observed. For the sake of completeness, we verify that the standard KS remains valid. With the new KF, we estimated the FAVARs of Bernanke et al. (2005) for incomplete panel data. In contrast to Bork (2009) and Marcellino and Sivec (2016), we do not treat FAVARs as specific ADFMs. This is why our estimation method admits, but does not require that the observable factor components are also part of the panel data.

Seventh, to prevent our estimation method for FAVARs from parameter ambiguity we first include the rotations in Bai et al. (2015) in our model preparations. As an alternative to common loadings constraints, we determine restrictions for the coefficients of the factor dynamics to remove left degrees of freedom. In this manner, we gain flexibility with regard to parameter constraints.

1.2 Thesis Structure

Within this thesis we alternately apply two Expectation-Maximization Algorithms (EMs) for estimating FMs. Besides factor dynamics, some model formulations admit cross-sectionally correlated idiosyncratic shocks and partially observed factors, which call for modifications of the original estimation procedure.

In Chapter 2, we mathematically define the considered FMs and provide a non-exhaustive list of common estimation techniques. That means, we describe the basics behind Principal Component Analysis (PCA), Probabilistic Principal Component Analysis (PPCA), Maximum-Likelihood Estimation (MLE), EMs and the standard KF and KS. Next, we revive the reconstruction formula of Stock and Watson (1999a, 2002b), which in Chapters 3-5 supports parameter estimation with incomplete panel data. Finally, we briefly state alternative approaches for the treatment of data incompleteness. In all sections, we review relevant publications in the respective areas.

In Chapter 3, we estimate Exact Static Factor Models (ESFMs) in the framework of Tipping and Bishop (1999) and allow for incomplete panel data using the reconstruction formula of Stock and Watson (1999a, 2002b). Then, the dynamics of a multivariate return process is supposed to obey a Vector Autoregression Model (VAR) with the estimated factors as exogenous variables, when we determine prediction intervals, empirical means and covariance matrices for returns of subsequent periods of time with the help of MC simulations. Finally, these forecasted return moments as well as their historical counterparts enter classical mean-variance and marginal-risk-parity portfolio optimizations, respectively, to demonstrate how our approach may support asset and risk management decisions.

In Chapter 4, Approximate Dynamic Factor Models (ADFMs) admitting homoscedastic, cross-sectionally correlated errors are considered. For incomplete panel data, two EMs are alternately applied for parameter estimation, where the inner EM is a modification of the EM in Bańbura and Modugno (2014), since it explicitly deals with cross-sectionally correlated idiosyncratic shocks. By contrast, Bańbura and Modugno (2014) follow the argumentation in Doz et al. (2012) and so, actually prove their findings for EDFMs. Another distinguishing feature to the ansatz in Bańbura and Modugno (2014) is the fact that we estimate the latent factor moments in closed-form instead of using the standard Kalman Filter and Smoother. The outer EM derives complete data panel from the observations and latest parameter estimates. Thereby, it reuses the reconstruction formula of Stock and Watson (1999a, 2002b). For the dynamics of a univariate return process we assume an Autoregressive Extended Model (ARX), when we break the predicted returns down into the contributions of the input data. For this purpose, we need closed-form expressions for the conditional factor means and covariance matrices instead of the KF and KS solutions. All in all, we aim at market monitoring.

In Chapter 5, we alter our fully-parametric two-step approach in the form of two EMs for estimating the FAVAR in Bernanke et al. (2005) with ragged panel data. As in Bai et al. (2015), we first investigate the implications of the partially observable factors for the uniqueness of the model parameters. Furthermore, we simplify the original FAVAR formulation. In contrast to Bork (2009) and Marcellino and Sivec (2016), who rearranged the data and deployed specific loadings restrictions such that they were able to apply standard techniques of ADFMs for parameter estimation, we derive new Kalman Filter and Smoother equations, which take the observability of factor components into account. For identification reasons, we allow for parameter constraints. In doing so, the loadings matrix as well as the coefficent matrices of the factor dynamics can be linearly constrained.

In Chapter 6, we conclude the main findings of this thesis and provide directions for the future research. For reasons of comprehensiveness, we repeat important definitions and results in Appendix A. If applicable, we also provide alternative proofs in Appendix A to preserve the clarity of Chapters 3-5. Appendix B lists sources and descriptions of the data our empirical studies are based on and so, supports the replication of our results. In the remainder of this thesis, we have overviews of used acronyms, nomenclature, figures and tables. Moreover, we summarize publications and working papers, which arose during my doctorate, and general references.

Chapter 2

Preliminaries

Factor Models (FMs) were thoroughly investigated in the literature and are the backbone of this thesis. Therefore, this chapter discusses diverse FM specifications, but it also addresses their classification and estimation. With regard to the latter point, we restrict ourselves to the most common estimation methods. Besides the non-parametric Principal Component Analysis, we explain parametric estimation procedures in a maximum likelihood framework. The given overview is not exhaustive, since we omit, e.g., Bayesian approaches. Furthermore, we introduce some notation, which is fundatmental for FMs, in the first part.

The focus of the second section is on the sophisticated treatment of data incompleteness. Besides missing observations, the inclusion of mixed-frequency information causes gaps in panel data. For instance, when monthly and quarterly times series are taken into account and the underlying time horizon is monthly, each quarterly time series offers one third of the amount of monthly observations. In this context, there are different modeling approaches for incomplete data available. Finally, we briefly mention alternative solutions for the treatment of mixed-frequency and missing data.

2.1 Factor Models

2.1.1 Classification of Factor Models

Before we repeat the mathematical definitions of FMs, let us start with some notation. An overview of all abbreviations and expressions is given in Appendices "Acronyms" and "Nomenclature", respectively. Let $\mathbf{0}_K \in \mathbb{R}^K$ denote the K-dimensional zero vector. Furthermore, let $O_K \in \mathbb{R}^{K\times K}$ and $I_K \in \mathbb{R}^{K\times K}$ stand for the square zero matrix and identity matrix, respectively, of dimension K. Finally, $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ refers to the multivariate normal distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^K$ and covariance matrix $\Sigma \in \mathbb{R}^{K\times K}$. In this thesis, lower case letters serve as running indices, whereas capital letters express dimensions. For instance, the index t with $1 \leq t \leq T$ picks a single element of a time series of length T. Because of this, the previous vector and matrix dimensions were written with capital letters.

In econometrics, a distinction is made between cross-sectional data and longitudinal data. Thereby, cross-sectional data describes, e.g., a population at a single point in time, while longitudinal data maps the evolution of individuals over time. If both dimensions are linked, e.g., a population is considered over time, econometricians call such a sample panel data. Hence, panel data constitutes cross-sectional, longitudinal data (Ruppert, 2011, p. 361). In Definition 2.1.1, we formally define what the previous description means. However, to emphasize that there are not any missing observations we speak about complete panel data. For clarity reasons, we highlight vectors in bold, e.g., $\boldsymbol{X}_t \in \mathbb{R}^N$.

Definition 2.1.1 (Complete Panel Data)
For any $1 \leq t \leq T$, the vector $\boldsymbol{X}_t = (X_{t1}, \ldots, X_{tN})' \in \mathbb{R}^N$ collects the panel data at time t, while the vector $\boldsymbol{X}^i \in \mathbb{R}^T$ contains the univariate time series of each input signal $1 \leq i \leq N$. The total data sample is covered by the matrix $X = [\boldsymbol{X}^1, \ldots \boldsymbol{X}^N] = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T]' \in \mathbb{R}^{T \times N}$.

To prevent us from lengthy expressions in subsequent sections, we introduce additional notation. Let the operator $\otimes$ denote the Kronecker product as in Definition A.1.9, let $\mathbb{1}_T \in \mathbb{R}^T$ be a vector of ones only and let $X' \in \mathbb{R}^{N \times T}$ be the transpose of the matrix $X \in \mathbb{R}^{T \times N}$. Furthermore, we use the hat symbol to refer to parameter estimators. For instance, the vector $\hat{\boldsymbol{\mu}} \in \mathbb{R}^N$ and the matrix $\hat{\Sigma} \in \mathbb{R}^{N \times N}$ stand for the estimators of the mean $\boldsymbol{\mu} \in \mathbb{R}^N$ and the covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$.

Definition 2.1.2 (Empirical Moments of Complete Panel Data)
Let the vector $\boldsymbol{\mu}_{\boldsymbol{X}} \in \mathbb{R}^N$ and matrix $\Sigma_{\boldsymbol{X}} \in \mathbb{R}^{N \times N}$ be the time-invariant mean and covariance matrix of the complete panel data $X \in \mathbb{R}^{T \times N}$ from Definition 2.1.1. Then, we deploy the empirical mean $\hat{\boldsymbol{\mu}}_{\boldsymbol{X}} \in \mathbb{R}^N$ and covariance matrix $\hat{\Sigma}_{\boldsymbol{X}} \in \mathbb{R}^{N \times N}$, which are given below, as estimators of $\boldsymbol{\mu}_{\boldsymbol{X}}$ and $\Sigma_{\boldsymbol{X}}$:

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{X}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{X}_t = \frac{1}{T} \left(X' \mathbb{1}_T \right), \tag{2.1}$$

$$\hat{\Sigma}_{\boldsymbol{X}} = \frac{1}{T} \sum_{t=1}^{T} \left(\boldsymbol{X}_t - \hat{\boldsymbol{\mu}}_{\boldsymbol{X}} \right) \left(\boldsymbol{X}_t - \hat{\boldsymbol{\mu}}_{\boldsymbol{X}} \right)' = \frac{1}{T} \left(X - \left(\mathbb{1}_T \otimes \hat{\boldsymbol{\mu}}'_{\boldsymbol{X}} \right) \right)' \left(X - \left(\mathbb{1}_T \otimes \hat{\boldsymbol{\mu}}'_{\boldsymbol{X}} \right) \right). \tag{2.2}$$

A time series $\{\boldsymbol{X}_t\}$ is a sequence of observations $\boldsymbol{X}_t \in \mathbb{R}^N$ over time and can be interpreted as realization of a stochastic process (Ruppert, 2011, p. 201). In the sequel, we will use the terms times series, stochastic process and process for $\{\boldsymbol{X}_t\}$ synonymously. Unless stated otherwise, we assume such a process $\{\boldsymbol{X}_t\}$ as complete, that is, there are no missing elements. In econometrics, a usual assumption for a process $\{\boldsymbol{X}_t\}$ is stationarity. For instance, Ruppert (2011, p. 202) calls a process "strictly stationary if all aspects of its behavior are unchanged by shifts in time". If its mean, variance and covariance are independent of time, he denotes it as weakly stationary. Here, we make some stationarity assumptions, too. For this purpose, Definition A.2.1 recalls that a process $\{\boldsymbol{X}_t\}$ is said to be covariance-stationary, if its first and second order moments are time invariant (Hamilton, 1994, p. 258). In successive derivations, the alternative conditions for covariance-stationarity of Vector Autoregression Models in Lemma A.2.3 may prove very useful. With the above notation and concepts in mind, we are ready to introduce Static Factor Models (SFMs).

Definition 2.1.3 (Static Factor Model)
Let the covariance-stationary vector $\boldsymbol{X}_t \in \mathbb{R}^N$, which collects all observations at time t, be driven by a common, covariance-stationary factor $\boldsymbol{F}_t \in \mathbb{R}^K, 1 \leq K \leq N$, and an idiosyncratic error $\boldsymbol{\epsilon}_t \in \mathbb{R}^N$. The latent factors and idiosyncratic errors are supposed to be identically and independently distributed (iid) Gaussian and independent of each other, i.e., $\boldsymbol{F}_t \perp \boldsymbol{\epsilon}_s \; \forall \, t, s$. Then, for data $\boldsymbol{X}_t$, latent variables $\boldsymbol{F}_t$ and shocks $\boldsymbol{\epsilon}_t$ a Static Factor Model obeys:

$$\boldsymbol{X}_t = W \boldsymbol{F}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \; \boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\boldsymbol{0}_N, \; \Sigma_{\boldsymbol{\epsilon}}\right) iid, \; \boldsymbol{F}_t \sim \mathcal{N}\left(\boldsymbol{0}_K, I_K\right) iid, \tag{2.3}$$

with vector $\boldsymbol{\mu} \in \mathbb{R}^N$ *and matrices* $W \in \mathbb{R}^{N\times K}$ *and* $\Sigma_{\boldsymbol{\epsilon}} \in \mathbb{R}^{N\times N}$ *as constants. If* $\Sigma_{\boldsymbol{\epsilon}}$ *is a diagonal matrix, the shocks are cross-sectionally uncorrelated and the model in* (2.3) *is called an Exact Static Factor Model (ESFM). Otherwise, we refer to it as an Approximate Static Factor Model (ASFM).*

Note, the relation in (2.3) provides for the covariance matrix of $\boldsymbol{X}_t$ given the parameters $\Theta = \{W, \boldsymbol{\mu}, \Sigma_{\boldsymbol{\epsilon}}\}$: $\mathbb{V}\text{ar}_\Theta[\boldsymbol{X}_t] = WW' + \Sigma_{\boldsymbol{\epsilon}}$. If the matrix $\Sigma_{\boldsymbol{\epsilon}}$ has uniformly bounded eigenvalues, Chamberlain and Rothschild (1983) called the vector $\boldsymbol{X}_t$ in (2.3) to have an approximate K-factor structure. In Definition 2.1.3, the hidden factors are supposed to be iid multivariate Gaussian. Since there is no time-dependent factor dynamics, the above FMs rank among the static ones. If the unobservable factor dynamics satisfies a Vector Autoregression Model of order $p \geq 1$, we receive the Dynamic Factor Models (DFMs) in Definition 2.1.4.

Definition 2.1.4 (Dynamic Factor Model)
The covariance-stationary vector process $\{\boldsymbol{X}_t\}$ *gathers all observations over time. Thereby, let the vector* $\boldsymbol{X}_t \in \mathbb{R}^N$ *be affected by a common factor* $\boldsymbol{F}_t \in \mathbb{R}^K, 1 \leq K \leq N$, *and an idiosyncratic shock* $\boldsymbol{\epsilon}_t \in \mathbb{R}^N$. *The stochastic process* $\{\boldsymbol{F}_t\}$ *is supposed to be zero-mean, covariance-stationary and autoregressive, i.e., it obeys a VAR(p) of order* $p \geq 1$. *Thus, for data* $\boldsymbol{X}_t$, *latent factors* $\boldsymbol{F}_t$ *and shocks* $\boldsymbol{\epsilon}_t$ *a Dynamic Factor Model is given by:*

$$\boldsymbol{X}_t = W\boldsymbol{F}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t,\ \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}_N, \Sigma_{\boldsymbol{\epsilon}})\, iid, \tag{2.4}$$

$$\boldsymbol{F}_t = \sum_{i=1}^{p} A_i \boldsymbol{F}_{t-i} + \boldsymbol{\delta}_t,\ \boldsymbol{\delta}_t \sim \mathcal{N}\left(\mathbf{0}_K, \Sigma_{\boldsymbol{\delta}}\right) iid, \tag{2.5}$$

with constant vector $\boldsymbol{\mu} \in \mathbb{R}^N$ *and matrices* $W \in \mathbb{R}^{N\times K}$, $\Sigma_{\boldsymbol{\epsilon}} \in \mathbb{R}^{N\times N}$, $A_i \in \mathbb{R}^{K\times K}, 1 \leq i \leq p$, *and* $\Sigma_{\boldsymbol{\delta}} \in \mathbb{R}^{K\times K}$. *The errors in* (2.4)-(2.5) *are supposed to be independent, i.e.,* $\boldsymbol{\epsilon}_t \perp \boldsymbol{\delta}_s\ \forall\, t, s$. *Let the matrix* $\Sigma_{\boldsymbol{\epsilon}}$ *be diagonal, then, the model in* (2.4)-(2.5) *is called an Exact Dynamic Factor Model (EDFM). Otherwise, it belongs to the Approximate Dynamic Factor Models (ADFMs).*

The ESFM in Definition 2.1.3 coincides with the FM in Tipping and Bishop (1999), if the matrix $\Sigma_{\boldsymbol{\epsilon}}$ is a constant times the identity matrix, i.e., all of its diagonal elements are the same, and so, describes an isotropic error model. The derivation of the estimation procedure in Bańbura and Modugno (2014) relies on an EDFM as in Definition 2.1.4 and hence, their idiosyncratic errors $\boldsymbol{\epsilon}_t$ in (2.4) are cross-sectionally uncorrelated at first glance. However, this restriction is not essential due to the work of Doz et al. (2012) such that their results remain valid for ADFMs with weakly cross-sectionally correlated errors. Since their errors $\boldsymbol{\epsilon}_t$ can be serially correlated, their model is more general in another direction. At this point, we ignore serial correlation of the errors $\boldsymbol{\epsilon}_t$ such that we can later on estimate the moments of the latent factors in closed form instead of using the Kalman Filter or Smoother. For generating forecasts this does not matter, but tracing forecasts back to the original input data is far easier with closed-form solutions for the factor moments.

After a comparison between Definitions 2.1.3 and 2.1.4, we can conclude: On the one hand, a distinction is made between SFMs and DFMs. In the first case, all factors are supposed to be iid, while the latter assume a VAR of order $p \geq 1$ for the factor dynamics. On the other hand, FMs are classified as exact and approximate, respectively. The idiosyncratic errors of exact FMs are not admitted to be cross-sectionally correlated such that their covariance matrix is diagonal. By contrast, approximate FMs permit cross-sectional correlation of the idiosyncratic errors and thus, assume a full covariance matrix. However, the division into exact and approximate FMs is not as strict as the distinction between SFMs and DFMs due to the results of Doz et al. (2012). They "treat[ed] the exact factor model as a misspecified approximating", when they derived the consistency of the estimated factors.

For the diverse FM specifications, there is abundant literature available. In this context, problems ranging from model selection to parameter estimation were discussed in detail. Neither theoretical questions nor empirical challenges remained untouched. This is why we give a brief overview of some well-known papers in this field, but cannot guarantee for the comprehensiveness of our summary. If there is any work missing, we apologize for this and ask the respective authors for their indulgence.

Although Dempster et al. (1977) did not particularly focus on FMs, they significantly influenced research in this field. They suggested an EM for parameter estimation, when the underlying data set is incomplete. To be precise, they replaced the log-likelihood function by its expectation conditioned on the observations and latest parameter estimates. In this way, they integrated out all missing elements from the objective function. Then, they searched for the (global) optimum of this conditional expectation to update previous parameter estimates. In the sequel, Rubin and Thayer (1982) used the EM of Dempster et al. (1977) for estimating SFMs, while Shumway and Stoffer (1982) applied it to DFMs. In both articles, the unobserved factors took the role of the missing elements in Dempster et al. (1977). The work in all three papers laid a first cornerstone in research, since their results were reused, applied and extended in, e.g., Tipping and Bishop (1999), Reis and Watson (2007), Bork (2009), Giannone et al. (2009), Jungbacker et al. (2009), Bork et al. (2010), Bańbura et al. (2011, 2013, 2014), Doz et al. (2011, 2012), Modugno (2011), Stock and Watson (2011), Bańbura and Modugno (2014) and Luciani (2014). In this context, Bork (2009) recommended a hybrid solution, which starts with an EM, but changes to the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method after a while. As justification for this transition, they argued that EMs quickly find the neighborhood of a maximum, but quasi-Newton approaches, e.g., the BFGS method, then outperform EMs in precisely locating the parameters of the maximum. Besides this, Bork (2009) and Bork et al. (2010) allowed for linear constraints of the loadings matrix to tacle potential identification issues. Jungbacker et al. (2009) devoted themselves the computational efficiency of parameter estimation in the presence of missing observations, whereas Bańbura and Modugno (2014) paid attention to arbitrary patterns of data incompleteness.

A second milestone in the area of Factor Analysis (FA) was set by Stock and Watson (1999a,b, 2002a,b). Among other things, they proved for Approximate Factor Models with time-dependent loadings that the factors can be consistently estimated. In addition, they benefited from the properties of the conditional normal distribution, when they derived an own EM for parameter estimation. Thereby, the EM generated estimates for missing elements and so, provided balanced panel data. Based on this full data, an ordinary PCA eventually estimated the factors and unknown parameters. Bernanke and Boivin (2003) applied the approach in Stock and Watson (1999b, 2002b) to Federal Reserve data and confirmed that large data sets can improve forecasts of economic times series. Further applications and extension of the seminal work of Stock and Watson (1999a,b, 2002a,b) are Artis et al. (2005), Boivin and Ng (2005), Angelini et al. (2006), Breitung and Eickmeier (2006), Bai and Ng (2008b), Giannone et al. (2008), Hogrefe (2008), Schumacher and Breitung (2008), Bork (2009), Stock and Watson (2009, 2011), Barhoumi et al. (2010), Doz et al. (2011), Aastveit and Trovik (2012), Bańbura and Modugno (2014), Barigozzi et al. (2014) and Luciani (2014). For further information about DFMs see, e.g., Stock and Watson (2011) and Luciani (2014).

A third crux of the matter was marked by Bai and Ng (2002). For Approximate FMs and large data sets, they developed a couple of information and panel criteria for the selection of the factor dimension. But they noted that their panel criteria may behave differently in case of finite samples, although they are asymptotically equivalent. To highlight the importance of the criteria in Bai and Ng (2002) we refer to the subsequent papers, which tested, relied on or extended at least one of them: Breitung and Eickmeier (2006), Amengual and Watson (2007), Reis and Watson (2007), Bai and Ng (2008b, 2013), Bork (2009), Stock and Watson (2009, 2011), Angelini et al. (2010, 2011) and Barigozzi et al. (2014). Since the size of their data set was small, Barhoumi et al. (2010) did not use the criteria of Bai and Ng (2002) on purpose.

Forni et al. (2000) introduced Generalized Dynamic Factor Models (GDFMs), which admit infinite factor dynamics and hence, opened another research direction. Applications, the theoretical background, comparisons and extensions of GDFMs were discussed in Forni and Lippi (2001), Forni et al. (2004, 2005, 2009), Bai and Ng (2008b), Altissimo et al. (2010) and Luciani (2011). GDFMs do not represent a main concept of this thesis, this is why we are brief regarding this topic.

A disadvantage of Vector Autoregression Models is the fact that only a limited number of time series can be included. Factor Analysis supports the treatment of big data sets. Especially, the inherent dimension reduction condenses large panel data in the form of a few factor time series. To benefit from the advantages of both Bernanke et al. (2005) developed the Factor-Augmented Vector Autoregression Models (FAVARs). Stock and Watson (2005) also link VARs and FMs, but their focus is on the implications, if dynamic factor models are put into VAR form. In both papers, FMs have additional terms, i.e., exogenous variables or lagged panel data. Besides the two estimation methods in Bernanke et al. (2005), Bork (2009) presented a third (fully parametric) procedure for estimating the FAVARs in Bernanke et al. (2005). In Chapter 5, we take a closer look at FAVARs. Thereby, we modify the estimation procedure in Bork (2009) such that it explicitly allows for the partially observed factors.

2.1.2 Principal Component Analysis

Although there are many differences between PCA and FA, both concepts are sometimes treated equally. For this purpose, we capture the definition of PCA (Jolliffe, 2002, pp. 1-6, Section 1.1) in the successive lemma, before we discuss its advantages and disadvantages compared to FA. Both techniques can provide the same results, however, this remains valid only under specific conditions, which we also address. Finally, we state a non-exhaustive list of papers estimating FMs using PCA. In the sequel, let $\mathbb{R}_+$ be the positive real line and $\|\boldsymbol{u}\|_2 = \sqrt{\boldsymbol{u}'\boldsymbol{u}}$ denotes the Euclidean norm or 2-norm of the vector $\boldsymbol{u} \in \mathbb{R}^N$.

Lemma 2.1.5 (Principal Components)
Assume $\boldsymbol{X}_t \in \mathbb{R}^N$ as random vector, where $\lambda_1 > \ldots > \lambda_N \in \mathbb{R}_+$ are the descendingly ordered eigenvalues of its covariance matrix $\Sigma_{\boldsymbol{X}} \in \mathbb{R}^{N \times N}$ with orthonormal eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N \in \mathbb{R}^N$. This means, for $1 \leq l < k \leq N$ the eigenvectors satisfy: $\boldsymbol{u}_k'\boldsymbol{u}_l = 0$ (orthogonal) and $\|\boldsymbol{u}_k\|_2^2 = \boldsymbol{u}_k'\boldsymbol{u}_k = 1$ (normal). Then, the k-th principal component $\boldsymbol{u}_k'\boldsymbol{X}_t$ maximizes the variance in the elements of $\boldsymbol{X}_t$, that is, $\boldsymbol{u}_k'\Sigma_{\boldsymbol{X}}\boldsymbol{u}_k$, and is uncorrelated to all previous principal components $\boldsymbol{u}_l'\boldsymbol{X}_t$ with $1 \leq l \leq k-1$. Furthermore, it follows for the variance of the k-th principal component: $\mathbb{V}ar\left[\boldsymbol{u}_k'\boldsymbol{X}_t\right] = \lambda_k$.

Proof:
The method of Lagrange multipliers with Lagrange multiplier λ and normalization constraint $\|\boldsymbol{u}_1\|_2^2 = 1$ provides for the first principal component the following maximization problem:

$$\boldsymbol{u}_1'\Sigma_{\boldsymbol{X}}\boldsymbol{u}_1 - \lambda\left(\boldsymbol{u}_1'\boldsymbol{u}_1 - 1\right).$$

Now, the partial derivatives with respect to the vector $\boldsymbol{u}_1$ and searching for the zeros of the arising system of linear equations yield:

$$\left(\Sigma_{\boldsymbol{X}} - \lambda I_N\right)\boldsymbol{u}_1 = \boldsymbol{0}_N,$$

which is solved by all eigenvalues and their associated eigenvectors. Because of $\boldsymbol{u}_1'\Sigma_{\boldsymbol{X}}\boldsymbol{u}_1 = \lambda\boldsymbol{u}_1'\boldsymbol{u}_1 = \lambda$, which we shall maximize, λ has to be the largest eigenvalue. Next, the fact that the principal components $\boldsymbol{u}_2'\boldsymbol{X}_t$ and $\boldsymbol{u}_1'\boldsymbol{X}_t$ are uncorrelated arises from the assumed orthogonality of the vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ in the following manner: $\mathbb{C}\mathrm{ov}\left[\boldsymbol{u}_2'\boldsymbol{X}_t, \boldsymbol{u}_1'\boldsymbol{X}_t\right] = \boldsymbol{u}_2'\Sigma_{\boldsymbol{X}}\boldsymbol{u}_1 = \lambda\boldsymbol{u}_2'\boldsymbol{u}_1 = 0$. Using Lagrange multipliers λ and ϕ with the orthonormality of the vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ the method of Lagrange multipliers results in the subsequent maximization problem for the second principal component:

$$\boldsymbol{u}_2'\Sigma_{\boldsymbol{X}}\boldsymbol{u}_2 - \lambda\left(\boldsymbol{u}_2'\boldsymbol{u}_2 - 1\right) - \phi\ \boldsymbol{u}_2'\boldsymbol{u}_1.$$

The partial derivatives with respect to $\boldsymbol{u}_2$ cause the following equation system:

$$2\Sigma_{\boldsymbol{X}}\boldsymbol{u}_2 - 2\lambda\boldsymbol{u}_2 - \phi\boldsymbol{u}_1 = \mathbf{0}_N.$$

By multiplying $\boldsymbol{u}_1'$ from the left to both sides of the above equation we receive $\phi = 0$ and end up with:

$$\Sigma_{\boldsymbol{X}}\boldsymbol{u}_2 - \lambda\boldsymbol{u}_2 = \mathbf{0}_N.$$

By similar reasoning as before, we conclude that λ is the second largest eigenvalue of $\Sigma_{\boldsymbol{X}}$ and $\boldsymbol{u}_2$ is its normalized eigenvector. An interative application of this procedure eventually proves the statement for all principal components $\boldsymbol{u}_k'\boldsymbol{X}_t$ with $3 \leq k \leq N$. □

Note, Lemma 2.1.5 assumes all eigenvalues of the covariance matrix $\Sigma_{\boldsymbol{X}}$ as distinct and positive. For n equal eigenvalues with $2 \leq n \leq N$, the n-dimensional space spanned by their eigenvectors is unique, but the eigenvectors themselves are exchangeable and thus, are not clearly identifiable (Jolliffe, 2002, p. 27, Section 2.4). The normlization $\boldsymbol{u}_k'\boldsymbol{u}_k = 1$ in Lemma 2.1.5 ensures to reach the maximum for finite $\boldsymbol{u}_k$, but it is only one, perhaps the most common one, of serveral alternatives (Jolliffe, 2002, p. 5, Section 1.1). In empirical studies, the covariance matrix $\Sigma_{\boldsymbol{X}}$ is usually replaced by the empirical covariance matrix $\hat{\Sigma}_{\boldsymbol{X}}$ in (2.2).

For distinguishing features between PCA and FA, we follow Jolliffe (2002, pp. 150-161, Sections 7.1-7.3). First, FA assumes an underlying model as in Definitions 2.1.3 and 2.1.4, whereas PCA is a non-parametric approach and does not assume such a model. Second, for the same panel data the number of factors and principal components might be different. Guess there is a time series that is uncorrelated to the remaining ones of the panel data. Then, in PCA this time series likely becomes a principal component, but no factor in FA. In case of PCA, it specifies an own eigenvector $\boldsymbol{u}_i \in \mathbb{R}^N$ of the covariance matrix $\Sigma_{\boldsymbol{X}}$. In the end, it depends on the total number of principal components K and the variation covered by $\boldsymbol{u}_i'\Sigma_{\boldsymbol{X}}\boldsymbol{u}_i$, whether the principal component $\boldsymbol{u}_i\boldsymbol{X}_t$ is chosen or not. If a time series behaves indenpendently to the remaining ones, FA assigns this individual nature to an idiosyncratic shock instead of a factor, since the factors cover communalities of the panel data. This fact highlights the third characteristic. In PCA, the focus lies on the diagonal elements of the covariance matrix $\Sigma_{\boldsymbol{X}}$, while in case of FA the off-diagonal entries matter more. Fourth, especially in empirical studies, the true number of factors or principal components is unknown and therefore, has to be estimated. If the number of principal components increases from K_1 to K_2, $K_2 - K_1$ new principal components are added to the original K_1 ones. By contrast, if the number of factors increases from K_1 to K_2, K_2 new factors are determined, which not necessarily comprise the former K_1 ones. Fifth, principal components arise from an exact linear function of the panel data, that is, $\boldsymbol{u}_k'\boldsymbol{X}_t$, while factors are a linear combination of the panel data and errors. Due to these differences Jolliffe (2002, p. 150, Chapter 7) assessed the use of PCA as part of FA as "bending the rules that govern factor analysis".

Despite the differences between PCA and Factor Analysis, PCA often provides initial parameter estimates for FA. Similar to Jolliffe (2002, p. 157, Eq. 7.2.3) we have:

Remark 2.1.6 (PCA for Parameter Initialization in Factor Analysis)
Assume the SFM in Definition 2.1.3 and let $\lambda_1 > \ldots > \lambda_N$ be the descendingly sorted eigenvalues of the covariance matrix $\Sigma_{\boldsymbol{X}}$ with orthonormal eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N \in \mathbb{R}^N$. Then, it holds for the parameters of a Static Factor Model initialized using PCA:

$$\begin{aligned}
\boldsymbol{X}_t &= \begin{bmatrix} \boldsymbol{u}_1 & \cdots \boldsymbol{u}_K | \boldsymbol{u}_{K+1} \cdots \boldsymbol{u}_N \end{bmatrix} \left[\begin{array}{c} \boldsymbol{u}_1' \boldsymbol{X}_t \\ \vdots \\ \boldsymbol{u}_K' \boldsymbol{X}_t \\ \hline \boldsymbol{u}_{K+1}' \boldsymbol{X}_t \\ \vdots \\ \boldsymbol{u}_N' \boldsymbol{X}_t \end{array}\right] \\
&= \begin{bmatrix} \boldsymbol{u}_1 & \cdots \boldsymbol{u}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1' \boldsymbol{X}_t \\ \vdots \\ \boldsymbol{u}_K' \boldsymbol{X}_t \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_{K+1} \cdots \boldsymbol{u}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_{K+1}' \boldsymbol{X}_t \\ \vdots \\ \boldsymbol{u}_N' \boldsymbol{X}_t \end{bmatrix} \\
&= W \boldsymbol{F}_t + \boldsymbol{\epsilon}_t,
\end{aligned}$$

which coincides with the ASFM in Definition 2.1.3. In general, we cannot assume that the idiosyncratic shocks are cross-sectionally uncorrelated such that the conditions of an ESFM might be violated.

Under certain conditions PCA and FA can be reconsiled. For the ESFMs in Definition 2.1.3 with isotropic shocks, i.e., we have $\Sigma_{\boldsymbol{\epsilon}} = \sigma_{\boldsymbol{\epsilon}}^2 I_N$, Tipping and Bishop (1999) showed how to determine principal components using MLE. To highlight the underlying probabilistic framework they introduced the term Probabilistic Principal Component Analysis (PPCA). In Section 3.1.1, we will reapply their estimation procedure. This is why we repeat their MLE parameter estimates in Theorem 3.1.3. A similar idea pursued Schneeweiss and Mathes (1995) by analyzing how small deviations between factors and principal components can be. For further reading on the reconsilement of PCA and FA see, e.g., Jolliffe (2002, pp. 158-161, Chapter 7.3).

PCA and FA share an important feature, namely, both techniques admit a reduction in dimension, when panel data is condensed by a few principal components or factors. Since PCA is a well-known concept in the literature, especially for now- and forecasting applications, we review some work in this area. Stock and Watson (2002a,b) forecasted univariate time series based on factors, which obey an Approximate FM and are estimated using principal components. In addition, they suggested the combination of PCA and an EM for parameter estimation with incomplete panel data, which is revived in Schumacher and Breitung (2008) and Marcellino and Schumacher (2010). The two-step estimation method for the FAVARs in Bernanke et al. (2005) first extracts factors from panel data using PCA and then, applies an Ordinary Least Squares Regression (OLS) for estimating the coefficient matrices of the factor dynamics. Bai and Ng (2002, 2006, 2008a,b) derived panel and information criteria for model selection, proved consistency and asymptotic intervals of predicted variables, provided a general overview and considered non-linear or targeted predictions, when factors are estimated using PCA. As in Bernanke et al. (2005), De Mol et al. (2006) compared Bayesian and PCA based estimation methods. Doz et al. (2011) proposed a two-step estimation method for ADFMs, which first combines PCA and OLS. In the second step, the factors are reestimated by the KS. This approach was applied or modified in Giannone et al. (2004, 2008), Hogrefe (2008) and Angelini et al. (2010). Bai and Ng (2013) studied conditions such that PCA provides asymptotically unique factor estimates, that is, they aimed to remove the uniqueness except for rotation. Finally, Stock and Watson (2011) summarized recent developments regarding FMs. Thereby, they collected contributions and results of PCA in this field.

2.1.3 Expectation-Maximization Algorithm

Let $\mathcal{L}(\Theta|X)$ be the log-likelihood function of a model with parameter set Θ for a complete data sample X as in Definition 2.1.1. Then, the idea behind Maximum-Likelihood Estimation (MLE) is to find parameter estimates $\hat{\Theta}$ such that the sample X occurs most likely. That is, the maximum likelihood estimates $\hat{\Theta}$ satisfy the subsequent optimization problem:

$$\hat{\Theta} = \arg\max_{\Theta} \mathcal{L}(\Theta|X). \tag{2.6}$$

But what happens, if the sample X is incomplete, for instance, due to missing observations? As a solution, Dempster et al. (1977) introduced Expectation-Maximization Algorithms (EMs), which integrate out all missing data from the log-likelihood function, before MLE is applied.

Definition 2.1.7 (Expectation-Maximization Algorithm)
Let $\mathcal{L}(\Theta|X)$ be the log-likelihood function for a model with parameters Θ given a complete data sample X, but let the set X_{obs} collect all actually observed data. Furthermore, for loop $(l) \geq 0$ the set $\hat{\Theta}_{(l)}$ contains the current maximum likelihood estimates. Then, an Expectation-Maximization Algorithm moves forward to the next loop $(l+1)$ in the form of two steps:

1. ***Expectation Step:*** *Update the expected log-likelihood function based on the observed data and latest parameters, i.e.:*

$$\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}(\Theta|X)|\, X_{obs}\right].$$

2. ***Maximization Step:*** *Update parameter estimates, that is:*

$$\hat{\Theta}_{(l+1)} = \arg\max_{\Theta} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}(\Theta|X)\,|X_{obs}\right].$$

The above definition makes clear how the name Expectation-Maximization Algorithm came up and why EMs rank among the iterative schemes. In addition, it shows need for a termination criterion to stop the overall routine. As soon as the change in the expectation of the log-likelihood function or in the estimated parameters falls below a prespecified threshold, EMs usually stop. At least, these were the most popular conditions in recent articles such as Schumacher and Breitung (2008), Doz et al. (2012) and Bańbura and Modugno (2014).

After the fundamentals, we look back on the impact of EMs on the following research. Rubin and Thayer (1982) were among the first to transfer the EM of Dempster et al. (1977) to Factor Analysis. Thereby, the latent factors took the role of missing data. In the meanwhile, Shumway and Stoffer (1982) estimated the hidden process of a state-space representation using an EM together with the Kalman Filter. Watson and Engle (1983) also combined an EM with the KF and KS for parameter estimation, when the underlying model comprised unobservable components. After a comparison with a scoring based estimation method, they concluded that EMs have an advantage at the beginning, since they rapidly find the region of the maximum. However, for the precise location of the maximum scoring can be a better choice.

Wu (1983) discussed conditions such that EMs yield converging sequences of the model parameters and likelihood function. He also highlighted that obtained parameter estimates may mark a local maximum of the likelihood function, but in special situations it is a saddle point only. Tipping and Bishop (1999) actually aimed at reconsiling FA and PCA. In doing so, they derived closed-form solutions and an EM for parameter estimation. Stock and Watson (1999a, 2002b) suggested an EM for estimating FMs based on incomplete panel data, which is applied in Schumacher and Breitung (2008), Boivin et al. (2010) and Marcellino and Schumacher (2010).

In the sequel, mixtures of EMs and Kalman Filtering techniques gained in importance in the literature. E.g., the one in Reis and Watson (2007) supported the decomposition of changes in prices of consumption goods. Hogrefe (2008) preferred a two-step approach, where PCA and an EM initialized a mixed-frequency DFM, before KF and KS reestimated the hidden factors. Among other things, Bork (2009) promoted a hybrid estimation method, which was introduced in Jungbacker and Koopman (2008). To improve the convergence of the overall scheme, Jungbacker and Koopman (2008) started with an EM to quickly find the neighborhood of a maximum, but then switched to the BFGS method, i.e., a quasi-Newton ansatz, converging more rapidly to the precise location of the optimum. Next, Bork et al. (2010) extended the EM to include linear parameter restrictions, while Jungbacker et al. (2009, 2011) accelerated the estimation of high-dimensional DFMs with incomplete data by reformulating its state-space representation.

Doz et al. (2011) proposed a two-step estimation method, which first applies PCA and OLS to initialize the factors and model parameters of ADFMs. In the second step, the factors were reestimated using the KS. In Doz et al. (2012), this two-step approach was modified, roughly spoken, it was iteratively applied such that an EM was received. This EM was compared with a Bayesian framework for modeling large European data in Bańbura et al. (2014). Thereby, Bańbura et al. (2014) concluded that both techniques provide reasonable, in particular, similar results. Eventually, Bańbura and Modugno (2014) extended the ansatz of Doz et al. (2012) to allow for incomplete data. Their estimation procedure entered, for instance, the analyses of Modugno (2011), Bańbura et al. (2011) and Kuzin et al. (2013).

2.1.4 Kalman Filter and Smoother

Let us recall the definition of DFMs. As shown in Definition 2.1.4, our DFMs consist of four stochastic processes, i.e., the observed panel data $\{\boldsymbol{X}_t\}$, the hidden factors $\{\boldsymbol{F}_t\}$, the idiosyncratic errors $\{\boldsymbol{\epsilon}_t\}$ and the shocks $\{\boldsymbol{\delta}_t\}$. The VAR$(p), p \geq 1$, in (2.5) models the factor dynamics and hence, describes for each point in time t the current state of the factors $\boldsymbol{F}_t$. Because of this, (2.5) is called state equation (Hamilton, 1994, p. 372). In opposition to (2.5), which only comprises the latent processes $\{\boldsymbol{F}_t\}$ and $\{\boldsymbol{\delta}_t\}$, Equation (2.4) incorporates an observable process, i.e., the panel data $\{\boldsymbol{X}_t\}$. On the one hand, the panel data is observed and so, only the process $\{\boldsymbol{X}_t\}$ provides information about the DFM in Definition 2.1.4. On the other hand, only (2.4) maps the relation between observed and hidden processes. Therefore, Equation (2.4) is called observation equation. If a time series $\{\boldsymbol{X}_t\}$ is completely defined by the observation equation in (2.4) and the state equation in (2.5), it has a state-space representation (Brockwell and Davis, 2002, p. 261, Definition 8.1.1). Unfortunately, a model can have several state-space representations, which may influence the computational efficiency of estimation methods such as the Kalman Filter (KF) or Kalman Smoother (KS). For instance, see Crone and Clayton-Matthews (2005), Jungbacker and Koopman (2008) and Jungbacker et al. (2009, 2011). This is why the choice of the state-space representation really matters.

The main idea originates from the work in Kalman (1960) such that both methods were named after him. In general, the Kalman Filter and Smoother gradually update linear projections of a model, which is cast into one of its state-space representations. Here, we restrict ourselves to the state-space representation of DFMs in (2.4)-(2.5), when we explain the two estimation methods and add references for the respective proofs. For detailed explanations, extensions and modifications, we propose classical textbooks as Hamilton (1994, pp. 372-408, Chapter 13) or Brockwell and Davis (2002, pp. 271-277, Section 8.4). As shown in Lemma A.2.2, we can rewrite any K-dimensional VAR(p), $p \geq 1$, as pK-dimensional VAR(1). Moreover, we define the matrix $\tilde{\mathbb{W}} \in \mathbb{R}^{N \times pK}$ as follows: $\tilde{\mathbb{W}} = \left[W, O_{N\times(p-1)K}\right]$. Then, the state-space representation of DFMs in (2.4)-(2.5) is equal to:

$$\boldsymbol{X}_t = \tilde{\mathbb{W}}\tilde{\boldsymbol{F}}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \tag{2.7}$$

$$\tilde{\boldsymbol{F}}_t = \tilde{\mathbb{A}}\tilde{\boldsymbol{F}}_{t-1} + \tilde{\boldsymbol{\delta}}_t, \tag{2.8}$$

where $\tilde{\boldsymbol{F}}_t \in \mathbb{R}^{pK}, \tilde{\boldsymbol{\delta}}_t \in \mathbb{R}^{pK}$ and $\tilde{\mathbb{A}} \in \mathbb{R}^{pK\times pK}$ are given in Lemma A.2.2. Hence, without loss of generality, we consider KF and KS equations for DFMs with factor dynamics of order one. For clarity reasons, let $\Theta = \{\tilde{\mathbb{W}}, \boldsymbol{\mu}, \Sigma_{\boldsymbol{\epsilon}}, \tilde{\mathbb{A}}, \Sigma_{\tilde{\boldsymbol{\delta}}}\}$ collect all model parameters. Neither the KF nor the KS estimate Θ, which requires additional techniques such as MLE, EMs or quasi-Newton methods. For instance, in a maximum likelihood framework the alternating application of the KF, KS and estimating Θ ensures that the estimated factors and model parameters optimize the (expected) log-likelihood function. With this in mind, the KF satisfies:

Lemma 2.1.8 (Kalman Filter for Dynamic Factor Models)
Assume the state-space representation in (2.7)-(2.8) *of the DFM in Definition 2.1.4 with known model parameters Θ. Furthermore, for complete panel data $X \in \mathbb{R}^{T\times N}$ as in Definition 2.1.1, the set Ω_t contains all observations up to time $0 \leq t \leq T$ as follows:*

$$\Omega_0 = \emptyset,$$
$$\Omega_t = \{\boldsymbol{X}_1, ..., \boldsymbol{X}_t\} \;\forall\; t > 0.$$

For clarity reasons, we introduce:

$$\hat{\boldsymbol{F}}_{t|t-1} = \mathbb{E}_\Theta\left[\tilde{\boldsymbol{F}}_t|\Omega_{t-1}\right] \in \mathbb{R}^{pK},$$
$$\hat{P}^F_{t|t-1} = \mathbb{V}ar_\Theta\left[\tilde{\boldsymbol{F}}_t|\Omega_{t-1}\right] \in \mathbb{R}^{pK\times pK}.$$

Then, for $1 \leq t \leq T$ the Kalman Filter consists of two steps:

$$\textit{Prediction Step:}\quad \begin{aligned} \hat{\boldsymbol{F}}_{t|t-1} &= \tilde{\mathbb{A}}\hat{\boldsymbol{F}}_{t-1|t-1}, \\ \hat{P}^F_{t|t-1} &= \tilde{\mathbb{A}}\hat{P}^F_{t-1|t-1}\tilde{\mathbb{A}}' + \Sigma_{\tilde{\boldsymbol{\delta}}}, \end{aligned}$$

$$\textit{Update Step:}\quad \begin{aligned} \hat{\boldsymbol{F}}_{t|t} &= \hat{\boldsymbol{F}}_{t|t-1} + \Gamma^{KF}_t\left(\boldsymbol{X}_t - \boldsymbol{\mu} - \tilde{\mathbb{W}}\hat{\boldsymbol{F}}_{t|t-1}\right), \\ \hat{P}^F_{t|t} &= \hat{P}^F_{t|t-1} - \Gamma^{KF}_t\left(\tilde{\mathbb{W}}\hat{P}^F_{t|t-1}\tilde{\mathbb{W}}' + \Sigma_{\boldsymbol{\epsilon}}\right)\left(\Gamma^{KF}_t\right)', \end{aligned}$$

with Kalman Filter Gain $\Gamma^{KF}_t \in \mathbb{R}^{pK\times N}$ defined by:

$$\Gamma^{KF}_t = \hat{P}^F_{t|t-1}\tilde{\mathbb{W}}'\left(\tilde{\mathbb{W}}\hat{P}^F_{t|t-1}\tilde{\mathbb{W}}' + \Sigma_{\boldsymbol{\epsilon}}\right)^{-1}.$$

Proof:
The above solutions are derived in Hamilton (1994, pp. 377-381, Section 13.2). For the readers convenience, all results are summarized in Hamilton (1994, p. 394, Eq. 13.6.5-13.6.8). □

The prediction and update steps in Lemma 2.1.8 are also stated without proof in Ghahramani and Hinton (1996), Särkkä et al. (2004) and Bork (2009, Appendix C.1). As shown in Lemma 2.1.8, the Kalman Filter progressively generates linear projections of the expectation and covariance matrix of the hidden factors. At any point in time $1 \leq t \leq T$, information gained up to time $t-1$ enters the prediction step, while all observations up to time t affect the update step. For all points in time, the Kalman Filter Gain Γ_t^{KF} controls the reliability of the latest predictions. In case of bad predictions, i.e., for large deviations $\boldsymbol{X}_t - \boldsymbol{\mu} - W\hat{\boldsymbol{F}}_{t|t-1}$, major adaptations are required, whereas for good ones the opposite holds. Before the interaction of predictions and updates starts, the moments $\hat{\boldsymbol{F}}_{0|0}$ and $\hat{P}^F_{0|0}$ are required. For this purpose, the unconditional mean and covariance matrix in Lemmata A.2.6 and A.2.7 may serve for initialization.

Eventually, Lemma 2.1.8 highlights two drawbacks of the KF, which numerical inaccuracies in empirical studies can produce. On the one hand, the ranks of the matrices W and $\Sigma_{\boldsymbol{\epsilon}}$ are at most K and N. Hence, for high-dimensional FMs of sufficiently large lag length p, this might cause that the matrix inverse part of the KF Gain does not always exist. On the other hand, in the presence of numerical noise the updates of the covariance matrix $\hat{P}^F_{t|t}$ do not necessarily ensure its semi-positive definiteness. To tacle the latter problem the Joseph form for updating the covariance matrices as in Haykin (2002, p. 8, Eq. 1.24) can be used.

Next, we consider the KS, which again assumes known model parameters Θ and is given by.

Lemma 2.1.9 (Kalman Smoother for Dynamic Factor Models)
Assume the setting and notation in Lemma 2.1.8. Then, we have $\Omega_T = X$, i.e., Ω_T covers the whole data sample, and for all points in time $1 \leq t \leq T$, the Kalman Smoother applies the following updates:

$$\begin{aligned}\hat{\boldsymbol{F}}_{t|T} &= \hat{\boldsymbol{F}}_{t|t} + \Gamma_t^{KS}\left(\hat{\boldsymbol{F}}_{t+1|T} - \hat{\boldsymbol{F}}_{t+1|t}\right),\\ \hat{P}^F_{t|T} &= \hat{P}^F_{t|t} - \Gamma_t^{KS}\left(\hat{P}^F_{t+1|t} - \hat{P}^F_{t+1|T}\right)\left(\Gamma_t^{KS}\right)',\end{aligned}$$

where the KF in Lemma 2.1.8 provided the means $\hat{\boldsymbol{F}}_{t|t}$ and covariance matrices $\hat{P}^F_{t|t}$. The matrix $\Gamma_t^{KS} \in \mathbb{R}^{pK\times pK}$ denotes the Kalman Smoother Gain, which is defined by:

$$\Gamma_t^{KS} = \hat{P}^F_{t|t}\tilde{\mathbb{A}}'\left(\hat{P}^F_{t+1|t}\right)^{-1}.$$

Proof:
See Hamilton (1994, pp. 394-397, Section 13.6). □

The above expressions are stated without any proof in Ghahramani and Hinton (1996) and Särkkä et al. (2004). As shown in Lemma 2.1.9, the KS always requires a run of the KF in advance. Otherwise, there are no estimates for the means $\hat{\boldsymbol{F}}_{t+1|t}$ and covariance matrices $\hat{P}^F_{t+1|t}$. In contrast to the KF, the KS updates the means $\hat{\boldsymbol{F}}_{t|T}$ and covariance matrices $\hat{P}^F_{t|T}$ based on the full information Ω_T. Thereby, the KS starts at the sample end, i.e, $t = T$, and goes back in time until the beginning at $t = 1$ is reached. Hence, the KS moves backward, while the KF moves forward in time. Like the KF Gain, the KS Gain includes a matrix inverse, which possibly causes problems due to numerical errors.

Besides the expectation and covariance matrix of the factors, some papers smooth the autocovariances between factors of different points in time. For instance, De Jong and Mackinnon (1988) and De Jong (1989) contributed to this topic. In the sequel, we deploy the following lag-one autocovariance smoother.

Lemma 2.1.10 (Lag-One Autocovariance Smoother for Dynamic Factor Models)
In the setting of Lemmata 2.1.8 and 2.1.9, for any point in time $1 \leq t \leq T-1$, the lag-one autocovariance smoother provides for the covariance matrix between the factors at times $t+1$ and t:

$$\hat{P}^F_{(t+1,t)|T} = \mathbb{C}ov_\Theta \left[\tilde{\boldsymbol{F}}_{t+1}, \tilde{\boldsymbol{F}}_t | \Omega_T \right] = \hat{P}^F_{t+1|T} \left(\Gamma^{KS}_t \right)' \in \mathbb{R}^{pK \times pK},$$

with KS Gain $\Gamma^{KS}_t \in \mathbb{R}^{pK \times pK}$ as in Lemma 2.1.9.

Proof:
See De Jong and Mackinnon (1988). □

The lag-one autocovariance smoother also calls for the KS gain, this is why it is often determined together with the KS instead of being separately estimated later on. In Chapter 4, we also link the KS and lag-one autocovariance smoother within a single routine, when we compare our closed-form solutions for the factor means and covariance matrices with their counterparts provided by Kalman Filtering techniques. We similarly proceed in Chapter 5, when we develop a KF and KS, which take the partial observability of factors in case of FAVARs into account, and compare them with the standard KF and KS.

Finally, we give a brief, non-exhaustive overview of articles considering KFs and KSs in the area of FA. Shumway and Stoffer (1982) and Watson and Engle (1983) estimated models with latent components using Kalman Filtering techniques and the EM in Dempster et al. (1977). Doz et al. (2011) pursued a two-step ansatz for estimating ADFMs. Thereby, they applied PCA and OLS, before the KS reestimated the hidden factors. This estimation procedure was used, modified and extended, among other things, in Giannone et al. (2008, 2009), Angelini et al. (2010, 2011), Bańbura and Rünstler (2011), Bańbura et al. (2014) and Doz et al. (2012). The EM in Bork (2009) and Bork et al. (2010) admits linear parameter constraints. With this in mind, Bańbura and Modugno (2014) extended the method in Doz et al. (2012) such that they were able to estimate ADFMs with incomplete panel data. For applications of this approach see Modugno (2011), Bańbura et al. (2011) and Kuzin et al. (2013). Mariano and Murasawa (2003) constructed a composite leading indicator based on monthly and quarterly data. Thereby, a quasi-Newton method and the Kalman Filter estimated the model parameters and hidden factors. Within the scope of suitable state-space representations, Nunes (2005), Proietti and Moauro (2006), Aruoba et al. (2009), Aruoba and Diebold (2010), Mariano and Murasawa (2010) as well as Kuzin et al. (2011) also derived composite indicators based on mixed-frequency information.

2.2 Incomplete Data and Temporal Aggregation

Technological advances supported the collection, storage and processing of vast amounts of data. In the scope of this progress, models accommodating big data received more attention, but data abundance also raised some questions. For instance, how to treat missing elements? Shall we fill gaps during the preprocessing, e.g., using interpolation or are there alternatives for estimating missing values? In particular, how to preserve cross-sectional dependencies in doing so? Furthermore, does more data necessarily improve the (forecasting) performance of a model? Do monthly time series such as production and wholesale indicators enhance the forecasting of quarterly indices like GDP? Finally, do models based on mixed-frequency information perform better than those restricted to a single time horizon? The above questions and many more were extensively discussed in the past. Therefore, we briefly summarize some well-known articles in this area, before we define what incomplete data means for us and how it is treated within this thesis.

Challenges arising from incomplete data have a long history in the literature and must be clearly separated from the factor hiddenness in Section 2.1.1. After the seminal work of Dempster et al. (1977), Harvey and Pierse (1984) estimated Autoregressive Integrated Moving Average Models (ARIMAs) in the presence of missing observations or temporal aggregation. Thereby, the expression *missing observations* denotes the absence of single values due to, e.g., public holidays, operational interruptions and trading suspensions. For clarity reasons, they call each time series with missing observations a *stock variable*. By contrast, they talk about *temporal aggregation*, when an observation results from the aggregation of some more frequent, but hidden analogs. For instance, quarterly GDP growth rates (annualized, in logarithmic terms) can be interpreted as published averages of monthly growth rates, which are not available. In terms of temporal aggegration, there is a distinction between *flow variables* and *change in flow variables*. We call a time series a flow variable, if each of its observations serves as average or sum of some more frequent, hidden analogs. If the observed values represent the differences between two successive elements of a flow variable, we denote such a time series a *change in flow variable*.

The classification as stock, flow and change in flow variables or the concept of missing data and temporal aggregation are common in research. For more insight on this topic we refer to, e.g., Abeysinghe (1998), Stock and Watson (1999a, 2002b), Liu and Hall (2001), Mariano and Murasawa (2003, 2010), Evans (2005), Nunes (2005), Angelini et al. (2006, 2010, 2011), Proietti (2006), Proietti and Moauro (2006), Hogrefe (2008), Hyung and Granger (2008), Schumacher and Breitung (2008), Wohlrabe (2008), Aruoba et al. (2009), Aruoba and Diebold (2010), Marcellino and Schumacher (2010), Kuzin et al. (2011, 2013), Bańbura et al. (2011, 2013), Bańbura and Rünstler (2011), Modugno (2011), Marcellino et al. (2013), Bańbura and Modugno (2014), Luciani (2014) as well as Schorfheide and Song (2015).

In this thesis, we consider FMs based on incomplete panel data. Thereby, we aim to cover two scenarios: *(i)* public holidays, trading suspensions, etc. cause the absence of single values, *(ii)* mixed-frequency data, e.g., monthly and quarterly information, results in systematically missing observations. As a solution, we apply the above concepts of stock, flow and change in flow variables to obtain balanced data without any gaps. That means, we introduce for each irregular times series an artificial counterpart of higher frequency and define an appropriate relation between both. Unlike publication delays, we have permanent gaps in *(i)* and *(ii)*, since they cannot be filled by any future observations. Besides publication conventions, signal types and the chosen time horizon, calendar irregularities such as the numbers of (trading) days, public holidays and weeks per month affect the pattern of missing data and thus, shall be addressed.

Definition 2.2.1 (Observed and Artificial Time Series)
For a sample of length T, let the counter $1 \leq t \leq T$ map each point in time, when new data arrives. Hence, it covers the most frequent time horizon. For N time series the vector $\boldsymbol{X}^i_{obs} \in \mathbb{R}^{T(i)}$ with $T(i) \leq T$ collects all observations of signal i with $1 \leq i \leq N$. Let $\boldsymbol{X}^i \in \mathbb{R}^T$ be its artificial, high-frequency counterpart, then, we assume a linear relation between $\boldsymbol{X}^i_{obs}$ and $\boldsymbol{X}^i$ as follows:

$$\boldsymbol{X}^i_{obs} = Q_i \boldsymbol{X}^i, \tag{2.9}$$

with constant matrix $Q_i \in \mathbb{R}^{T(i)\times T}$ of full row rank, i.e., $T(i)$.

With a view to Definitions 2.2.1, any complete time series satisfies $T(i) = T$ implying that the matrix Q_i is given by the T-dimensional identity matrix. For any time series, which is incomplete or less often updated, the full rank condition removes unnecessary zero rows from the matrix Q_i. Furthermore, the full rank condition ensures that the inverse matrix $(Q_i Q_i')^{-1}$ exists and thus, the matrix $Q_i' (Q_i Q_i')^{-1} \in \mathbb{R}^{T\times T(i)}$ is well-defined. The matrix $Q_i' (Q_i Q_i')^{-1}$ is also known as the unique Moore-Penrose Inverse of Q_i, which obeys: $Q_i Q_i' \left(Q_i Q_i'\right)^{-1} = I_{T(i)}$ (Rao and Toutenburg, 1999, pp. 372-373, Definition A.64). This is essential for the reconstruction formulas in Lemma 3.1.7, Equation (4.25) and Lemma 5.1.19.

Next, we formalize the previous distinction between stock and flow variables.

Definition 2.2.2 (Stock and Flow Variables)
Assume the notation in Definition 2.2.1 and let the integers $(n_j)_{1 \le j \le T(i)}$ count the high-frequency periods between two successive observations of signal $i, 1 \le i \le N$, such that $o_j = \sum_{k=1}^{j} n_k$ captures the point in time, when the j-th observation $\boldsymbol{X}^i_{obs,j}$ is made. Then, we have for $1 \le i \le N$ and $1 \le j \le T(i)$:

- *For stock variables, each observation coincides with its high-frequency analog: $\boldsymbol{X}^i_{obs,j} = \boldsymbol{X}^i_{o_j}$.*
- *For the sum formulation of flow variables, it holds: $\boldsymbol{X}^i_{obs,j} = \sum_{k=0}^{n_j-1} \boldsymbol{X}^i_{o_j-k}$.*
- *For the average represenation of flow variables, we obtain: $\boldsymbol{X}^i_{obs,j} = \frac{1}{n_j} \sum_{k=0}^{n_j-1} \boldsymbol{X}^i_{o_j-k}$.*

For illustrative purposes, in case of stock variables the linear relation in (2.9) may look like the following:

$$\boldsymbol{X}^i_{\text{obs}} = \underbrace{\begin{pmatrix} 1 & & 0 & 0 & \cdots & 0 & \cdots & 0 \\ & \ddots & & \vdots & & & & \vdots \\ 0 & & 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 0 & 1 & & \\ \vdots & & & & \vdots & & \ddots & \\ 0 & \cdots & 0 & \cdots & 0 & & & 1 \end{pmatrix}}_{Q_i} \boldsymbol{X}^i.$$

Similarly, examples for sum and average formulations of flow variables are given by:

$$\boldsymbol{X}^i_{\text{obs}} = \underbrace{\begin{pmatrix} 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 1 \end{pmatrix}}_{Q_i} \boldsymbol{X}^i \qquad \textit{(sum)},$$

$$\boldsymbol{X}^i_{\text{obs}} = \underbrace{\begin{pmatrix} \frac{1}{n_1} & \cdots & \frac{1}{n_1} & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \frac{1}{n_{T(i)}} & \cdots & \frac{1}{n_{T(i)}} \end{pmatrix}}_{Q_i} \boldsymbol{X}^i \qquad \textit{(average)}.$$

In total, the above examples confirm for complete time series the claim that the choice between stock and flow variables does not matter, since we always get an identity matrix. In addition, all examples show the flexibility of the presented modeling approach. Whenever we would like to switch from one series type to another, we just have to accordingly adjust the shape of the matrix Q_i. To allow for irregularly missing observations, we just remove the respective rows of an identity matrix. In case of calendar irregularities, e.g., the number of (trading) days per months is varying over time, we can map each possible day-week or day-month structure through the matrix Q_i. Note, calendar irregularities are a priori known and hence, are deterministic.

Eventually, we mathematically derive the relation between observed and artificial data for change in flow variables. As for flow variables, there is a sum and average variant, respectively.

Lemma 2.2.3 (Change in Flow Variables)
Assume the notation in Definition 2.2.2. Then, the average version of a change in flow variable satisfies:

$$\Delta \boldsymbol{X}^i_{obs,j} = \sum_{k=0}^{n_j-1} \frac{k+1}{n_j} \Delta \boldsymbol{X}^i_{o_j-k} + \sum_{k=0}^{n_{j-1}-1} \frac{n_{j-1}-1-k}{n_{j-1}} \Delta \boldsymbol{X}^i_{o_{j-1}-k}.$$

For the special case that all low-frequency periods comprise the same number of high-frequency intervals, i.e., $n_j = n \ \forall\ 1 \leq j \leq T(i)$, a similar solution for the sum formulation of a change in flow variable exists and is given by:

$$\Delta \boldsymbol{X}^i_{obs,j} = \sum_{k=0}^{n-1} (k+1)\, \Delta \boldsymbol{X}^i_{nj-k} + \sum_{k=0}^{n-1} (n-1-k)\, \Delta \boldsymbol{X}^i_{n(j-1)-k}. \tag{2.10}$$

Note, the equality $n_j = n \ \forall\ 1 \leq j \leq T(i)$ provides: $o_j = \sum_{k=1}^{j} n_k = nj$. That means, the point in time, when the j-th observation is made, is given by the product of j and n.

Proof:
We obtain for the average formulation of change in flow variables from the respective flow specification in Definition 2.2.2, by inserting a telescoping sum and replacing o_j through $o_{j-1} + n_j$, the following:

$$\begin{aligned}
\Delta \boldsymbol{X}^i_{\text{obs},j} &= \boldsymbol{X}^i_{\text{obs},j} - \boldsymbol{X}^i_{\text{obs},j-1} \\
&= \frac{1}{n_j} \sum_{l=0}^{n_j-1} \boldsymbol{X}^i_{n_j+o_{j-1}-l} - \frac{1}{n_{j-1}} \sum_{l=0}^{n_{j-1}-1} \boldsymbol{X}^i_{o_{j-1}-l} \\
&= \frac{n_{j-1} \sum_{l=0}^{n_j-1} \boldsymbol{X}^i_{n_j+o_{j-1}-l} - n_j \sum_{l=0}^{n_{j-1}-1} \boldsymbol{X}^i_{o_{j-1}-l}}{n_j\, n_{j-1}} \\
&= \frac{n_{j-1} \sum_{l=0}^{n_j-2} (l+1)\, \Delta \boldsymbol{X}^i_{n_j+o_{j-1}-l} + n_j\, n_{j-1} \boldsymbol{X}_{o_{j-1}+1} - n_j \sum_{l=0}^{n_{j-1}-1} \boldsymbol{X}^i_{o_{j-1}-l}}{n_j\, n_{j-1}} \\
&= \frac{n_{j-1} \sum_{l=0}^{n_j-1} (l+1)\, \Delta \boldsymbol{X}^i_{n_j+o_{j-1}-l} + n_j \sum_{l=0}^{n_{j-1}-1} (n_{j-1}-1-l)\, \Delta \boldsymbol{X}^i_{o_{j-1}-l}}{n_j\, n_{j-1}} \\
&= \sum_{l=0}^{n_j-1} \left(\frac{l+1}{n_j}\right) \Delta \boldsymbol{X}^i_{n_j+o_{j-1}-l} + \sum_{l=0}^{n_{j-1}-1} \left(\frac{n_{j-1}-1-l}{n_{j-1}}\right) \Delta \boldsymbol{X}^i_{o_{j-1}-l}.
\end{aligned}$$

If $n_j = n \ \forall\ 1 \leq j \leq T(i)$ holds, we receive by multiplication with n the sum version in (2.10). Finally, we verify why this equality is essential. For this purpose, we assume $n_j = n_{j-1} + 1$, i.e., one period exceeds its prior by a single (high-frequency) interval. Then, the above procedure results in:

$$\Delta \boldsymbol{X}^i_{\text{obs},j} = \sum_{k=0}^{n_j-1} (k+1) \Delta \boldsymbol{X}^i_{o_j-k} + \sum_{k=0}^{n_{j-1}-2} (n_j-1-k) \Delta \boldsymbol{X}^i_{o_{j-1}-k} + \boldsymbol{X}^i_{o_{j-2}+1}.$$

The last term is the artificial signal itself and so, the observed change is not equal to a pure combination of high-frequency increments. By similar reasoning, the same holds for any $n_j \neq n_{j-1}$. The special case in (2.10) for quarterly and monthly data is stated in Bańbura et al. (2013, ECB working paper, p. 10). □

For illustrative purposes, we also give examples of the matrix Q_i in Definition 2.2.1 for sum and average variants of change in flow variables. Thereby, let the vectors $\Delta \boldsymbol{X}^i_{\text{obs}} = [\Delta \boldsymbol{X}^i_{\text{obs},1}, \ldots, \Delta \boldsymbol{X}^i_{\text{obs},T(i)}]' \in \mathbb{R}^{T(i)}$ and $\Delta \boldsymbol{X}^i = \left[\boldsymbol{X}^i_1, \ldots, \boldsymbol{X}^i_T\right]' \in \mathbb{R}^T$ contain the observed and artificial increments. Then, we have:

$$\Delta \boldsymbol{X}^i_{\text{obs}} = \underbrace{\begin{pmatrix} 1 & \cdots & n-1 & n & n-1 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & n-1 & n & * & \cdots & * & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & * & \ddots & * & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & * & \cdots & * & n & n-1 & \cdots & 1 \end{pmatrix}}_{Q_i} \Delta \boldsymbol{X}^i \qquad \textit{(sum)},$$

$$\Delta \boldsymbol{X}^i_{\text{obs}} = \underbrace{\begin{pmatrix} \frac{1}{n_1} & \cdots & \frac{n_1-1}{n_1} & \mathbf{1} & \frac{n_2-1}{n_2} & \cdots & \frac{1}{n_2} & 0 & 0 & \cdots & 0 & 0 \cdots 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \frac{1}{n_2} & \cdots & \frac{n_2-1}{n_2} & \mathbf{1} & \frac{n_3-1}{n_3} & \cdots & \frac{1}{n_3} & 0 \cdots 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \frac{1}{n_3} & \cdots & \frac{n_3-1}{n_3} & * \cdots * & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & * \ddots * & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & * \cdots * & \frac{n_{T(i)}-1}{n_{T(i)}} & \cdots & \frac{1}{n_{T(i)}} \end{pmatrix}}_{Q_i} \Delta \boldsymbol{X}^i \qquad \textit{(average)}.$$

For completely observed change in flow variables, the matrix Q_i again becomes the T-dimensional identity matrix. Hence, we have a consistent shape for all variable types, when the observed time series is complete. Since the choice of a variable type only affects the matrix Q_i, we can keep all remaining parts of an overall model unchanged. This is a nice feature, in particular, when real-world data is used. On the one hand, we can easily switch between different variable types, which possibly happens several times during the preprocessing phase. On the other hand, it reduces the risk that a change in the data type is not properly taken into account, e.g., in a model's state-space representation. For instance, any change in the variable type calls for a new state-space model in Bańbura and Modugno (2014).

Besides FMs and VARs, there is another well-known concept for the treatment of mixed-frequency time series called Mixed-Data Sampling (MIDAS). Although we do not apply it within this thesis, the MIDAS approach sometimes serves as benchmark of mixed-frequency FMs and VARs. For reasons of completeness, we therefore spend a few words on its history and how it works. In Ghysels et al. (2004, 2005, 2006, 2007) MIDAS was introduced, nicely explained, extended and applied to return volatilities. In a nutshell, MIDAS denotes a regression framework, which deploys more timely information to describe a variable of interest that is less often available. For instance, monthly data models quarterly GDP growth rates. Extensions of MIDAS, applications and comparisons with alternative estimation methods are stated in Hogrefe (2008), Clements and Galvão (2008), Ghysels and Wright (2009), Andreou et al. (2010, 2011, 2013), Armesto et al. (2010), Marcellino and Schumacher (2010), Francis et al. (2011), Kuzin et al. (2011, 2013), Chiu et al. (2012), Bai et al. (2013), Galvão (2013), Lahiri and Monokroussos (2013) and Ghysels (2015).

Chapter 3

Mixed-Frequency Information Supporting Asset Allocation Decisions

We use ESFMs for incomplete panel data to determine optimal portfolios in a mean-variance or marginal-risk-parity framework. For this purpose, we condense the information in large, ragged panel data by a few hidden factors. Thereby, we estimate for all points in time the factor means and covariance matrices. In a next step, a Vector Autoregression Model with Exogenous Variables (VARX) describes the dynamics of a multivariate return process, where the factors serve as exogenous variables. In this way, we derive empirical means, covariance matrices and prediction intervals for returns of future periods of time. Irrespective of whether we calculate means, covariance matrices or prediction intervals, we use samples randomly drawn from the factors' distribution instead of their estimates to allow for uncertainties arising from factor estimation. Similarly, we take estimation risks inherent in the VARX parameters into account. Eventually, the predicted means and covariance matrices enter our portfolio optimizations. For comparison reasons, we run a backtest with US data. In doing so, we construct a portfolio, which is denoted in United States Dollar (USD) and consists of Gold and three stock indices, i.e., Nasdaq Composite (NASDAQ), Standard & Poor's 500 (S&P500) and Dow Jones Industrial Average (DJIA). Besides financial risk and performance figures, our analysis provides statistical point and interval measures to examine from diverse perspectives that the combination of financial and macroeconomic data, which often causes incomplete panel data, can improve forecasts and hence, really pays off.

3.1 Mathematical Background

At first, we apply the method of Tipping and Bishop (1999) for estimating the parameters of the ESFM in Definition 2.1.3 with complete panel data. At this stage, we address the selection of the factor dimension. Thereafter, the reconstruction formula in Lemma 3.1.7 enables us to expand the estimation method to incomplete panel data. Next, we specify a VARX for the dynamics of a multivariate return process, where the factors take the role of exogenous variables. At the end of this section, we repeat the mean-variance and marginal-risk-parity, respectively, optimization problems to be solved.

3.1.1 Estimation of ESFMs with Complete Panel Data

In this section, we only consider complete panel data. Before we explain parameter and factor estimation, we discuss the assumptions of the ESFM in Definition 2.1.3. Thereby, we investigate how restrictive those are and whether there remain some identification issues.

Remark 3.1.1 (Covariance Matrix of Factors)
The ESFM in Definition 2.1.3 assumes iid standard normal factors, that is, $\boldsymbol{F}_t \sim \mathcal{N}(\mathbf{0}_K, I_K)$ iid for all $1 \leq t \leq T$. Let the positive-definite matrix $\Sigma_{\boldsymbol{F}} \in \mathbb{R}^{K\times K}$ be the general version of the covariance matrix of the covariance-stationary factor process $\{\boldsymbol{F}_t\}$. Although the equality $\Sigma_{\boldsymbol{F}} = I_K$ seems restrictive at first glance, it is very useful and supports the factors' uniqueness. Unfortunately, it does not solve all issues, since the factors are unique except for rotation.

To verify the statement in Remark 3.1.1, let $D^{-1} \in \mathbb{R}^{K\times K}$ be the inverse matrix of a non-singular matrix $D \in \mathbb{R}^{K\times K}$. Then, the factor $\breve{\boldsymbol{F}}_t = D\boldsymbol{F}_t \in \mathbb{R}^K$ and loadings matrix $\breve{W} = WD^{-1} \in \mathbb{R}^{N\times K}$ also meet the observation equation in (2.3):

$$\boldsymbol{X}_t = W\boldsymbol{F}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t = \breve{W}\breve{\boldsymbol{F}}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \tag{3.1}$$

and do not affect the dependencies between factors and idiosyncratic shocks:

$$\mathbb{C}\mathrm{ov}_\Theta\left[\breve{\boldsymbol{F}}_t, \boldsymbol{\epsilon}_s\right] = D\ \mathbb{C}\mathrm{ov}_\Theta\left[\boldsymbol{F}_t, \boldsymbol{\epsilon}_s\right] = O_{K\times N}. \tag{3.2}$$

But the equality requires:

$$\mathbb{V}\mathrm{ar}_\Theta\left[\breve{\boldsymbol{F}}_t\right] = D\ \mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{F}_t\right] D' = I_K,$$

and so, removes some degrees of freedom. For any orthonormal matrix $R \in \mathbb{R}^{K\times K}$ with $R'R = RR' = I_K$, the rotated factor $\bar{\boldsymbol{F}}_t = R'\boldsymbol{F}_t \in \mathbb{R}^K$ and loadings matrix $\bar{W} = WR \in \mathbb{R}^{N\times K}$ obey (3.1) and (3.2), while the equality $\Sigma_{\boldsymbol{F}} = I_K$ is still preserved. This is why the factors are unique except for rotation.

The essential assumption in Remark 3.1.1 is the positive definiteness of the covariance matrix $\Sigma_{\boldsymbol{F}}$. If the rank of $\Sigma_{\boldsymbol{F}}$ is smaller than K, less factors describe the panel data X just as well such that the dimension of the factor space can be reduced. In this thesis, we aim at having as few factors as possible to model the panel data X for a desired level of covered variation. This is why this condition is not really restrictive from our perspective. Moreover, in our empirical applications we usually have $K \ll N$ and so, achieve a noticeable dimension reduction, when migrating from the panel data to the factor space.

For the ESFMs in Definition 2.1.3, the covariance matrix of the idiosyncratic errors is a diagonal matrix admitting distinct elements. However, to be in accordance with Tipping and Bishop (1999), we consider an isotropic error model instead. This is, we have: $\boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\mathbf{0}_N, \sigma^2_{\boldsymbol{\epsilon}} I_N\right)$ iid for any $1 \leq t \leq T$. Thus, the errors at a fixed point in time do neither interact with each other nor with the errors at another point in time, but their variance is the same for each time series.

The assumptions in Definition 2.1.3, the equality $\Sigma_{\boldsymbol{F}} = I_K$ and the isotropic error model call the matrix W to fulfill the following tasks: First, it maps the cross-sectional dependencies of $\boldsymbol{X}_t$. Second, for any point in time the matrix W balances the data variance, which is not covered by the idiosyncratic errors, and the factor variance.

In the sequel, we present the PPCA approach of Tipping and Bishop (1999) for estimating the parameters $\Theta = \left\{W, \sigma^2_{\boldsymbol{\epsilon}}\right\}$ and factors of the ESFM in Definition 2.1.3 with isotropic error model. Regarding the mean

$\boldsymbol{\mu}$ we use the empirical mean $\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}$ in Definition 2.1.2 for parameter estimation, as we subtract the empirical means from the sample X during the data preprocessing. Therefore, we do not add $\boldsymbol{\mu}$ to the set Θ. By contrast, Tipping and Bishop (1999) treated mean $\boldsymbol{\mu}$ as part of the parameters Θ in their MLE, but also received $\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}$ as estimator for $\boldsymbol{\mu}$. Tipping and Bishop (1999) actually aimed at reconsiling PCA and FA, but they also derived estimates for Θ, which are useful for our purposes. The idea behind the derivation of their estimation method is as follows: Before they can determine maximum likelihood estimates of Θ, they need the log-likelihood function $\mathcal{L}(\Theta|X)$ for a given data sample $X \in \mathbb{R}^{T\times N}$. Because of iid factors $\boldsymbol{F}_t$ and idiosyncratic errors $\boldsymbol{\epsilon}_t$, the panel data $\boldsymbol{X}_t$ is iid, too. Hence, the overall log-likelihood function $\mathcal{L}(\Theta|X)$ can be decomposed as sum of the log-likelihood functions $\mathcal{L}(\Theta|\boldsymbol{X}_t)$. In doing so, $f_\Theta(\boldsymbol{X}_t)$ stands for the probability density function of the random variable $\boldsymbol{X}_t$ and parameters Θ. Let $|\cdot|$ denote the matrix determinant as in Definition A.1.6 and let $tr(\cdot)$ refer to the matrix trace in Definition A.1.1. Furthermore, let $ln(x)$ be the natural logarithm of any $x \in \mathbb{R}_+$ and let $exp(y)$ be the value of the exponential function for any $y \in \mathbb{R}$. Then, it holds:

Lemma 3.1.2 (Log-Likelihood Function of ESFMs with Isotropic Error Model)
Let $\mathcal{L}(\Theta|X)$ be the log-likelihood function of the ESFM in Definition 2.1.3 with an isotropic error model and parameters Θ for complete panel data $X \in \mathbb{R}^{T\times N}$ as in Definition 2.1.1. Then, $\mathcal{L}(\Theta|X)$ is given by:

$$\mathcal{L}(\Theta|X) = -\frac{T}{2}\left[N\ ln(2\pi) +\ ln(|C|) +\ tr\left(C^{-1}\hat{\Sigma}_{\boldsymbol{X}}\right)\right], \tag{3.3}$$

with matrices $C = WW' + \sigma^2_{\boldsymbol{\epsilon}} I_N \in \mathbb{R}^{N\times N}$ and $\hat{\Sigma}_{\boldsymbol{X}} \in \mathbb{R}^{N\times N}$ as in Definition 2.1.2.

Proof:
The factors $\boldsymbol{F}_t$ and shocks $\boldsymbol{\epsilon}_t$ are iid Gaussian. Thus, the vector $\boldsymbol{X}_t$ is also iid Gaussian with expectation $\mathbb{E}_\Theta[\boldsymbol{X}_t]$ and covariance matrix $\mathbb{V}\mathrm{ar}_\Theta[\boldsymbol{X}_t]$ as follows:

$$\mathbb{E}_\Theta[\boldsymbol{X}_t] = \boldsymbol{\mu} \quad \text{and} \quad \mathbb{V}\mathrm{ar}_\Theta[\boldsymbol{X}_t] = WW' + \Sigma_{\boldsymbol{\epsilon}}.$$

The isotropic error model results in $\mathbb{V}\mathrm{ar}_\Theta[\boldsymbol{X}_t] = C$ as defined in Lemma 3.1.2. With this in mind, we can conclude for the log-likelihood function:

$$\begin{aligned}
\mathcal{L}(\Theta|X) &= ln\left(f_\Theta(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_T)\right) = ln\left(\prod_{t=1}^{T} f_\Theta(\boldsymbol{X}_t)\right) \\
&= \sum_{t=1}^{T}\ ln\left[(2\pi)^{-N/2}|C|^{-1/2}\ exp\left(-\frac{1}{2}(\boldsymbol{X}_t-\boldsymbol{\mu})' C^{-1}(\boldsymbol{X}_t-\boldsymbol{\mu})\right)\right] \\
&= -\frac{T}{2}\left[N\ ln(2\pi) +\ ln(|C|)\right] - \frac{1}{2}\sum_{t=1}^{T}\ tr\left((\boldsymbol{X}_t-\boldsymbol{\mu})' C^{-1}(\boldsymbol{X}_t-\boldsymbol{\mu})\right).
\end{aligned}$$

Finally, cyclical permutation inside the brackets of the matrix trace, see *(vi)* from Lemma A.1.2, and the definition of $\hat{\Sigma}_{\boldsymbol{X}}$ yield the statement. □

Based on the log-likelihood function in Lemma 3.1.2, MLE in case of the parameters Θ is straightforward. In this context, let $Z^{1/2}$ denote the matrix square root of a matrix $Z \in \mathbb{R}^{K\times K}$, i.e., $Z^{1/2}Z^{1/2} = Z$.

Theorem 3.1.3 (MLE for Parameter Estimation in ESFMs)
For the ESFM in Definition 2.1.3 with an isotropic error model, let $\lambda_1 \geq \ldots \geq \lambda_N \geq 0$ be the descendingly

sorted eigenvalues of the covariance matrix $\hat{\Sigma}_{\boldsymbol{X}}$ *in Definition 2.1.2, whose orthonormal eigenvectors are given by* $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N \in \mathbb{R}^N$*. Then, MLE provides the subsequent parameter estimates:*

$$\hat{W} = U_K \left(\Lambda_K - \hat{\sigma}^2_{\boldsymbol{\epsilon}}\ I_K\right)^{1/2} R, \tag{3.4}$$

$$\hat{\sigma}^2_{\boldsymbol{\epsilon}} = \frac{1}{N-K} \sum_{i=K+1}^{N} \lambda_i, \tag{3.5}$$

where the diagonal matrix $\Lambda_K \in \mathbb{R}^{K \times K}$ *consists of* $\lambda_1, \ldots, \lambda_K$ *and the columns of the orthonormal matrix* $U_K \in \mathbb{R}^{N \times K}$ *contain the vectors* $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K$*. The matrix* $R \in \mathbb{R}^{K \times K}$ *stands for any rotation matrix.*

Proof:
See Tipping and Bishop (1999). □

The matrix $\hat{\sigma}^2_{\boldsymbol{\epsilon}}\ I_K$ is diagonal, thus, the matrix $\left(\Lambda_K - \hat{\sigma}^2_{\boldsymbol{\epsilon}} I_K\right)$ is also diagonal. Due to (3.5), $\hat{\sigma}^2_{\boldsymbol{\epsilon}}$ averages the variation lost through dimension reduction. Since all eigenvalues are sorted in descending order, we have: $\lambda_i \geq \hat{\sigma}^2_{\boldsymbol{\epsilon}},\ 1 \leq i \leq K$. Hence, all diagonal elements of $\left(\Lambda_K - \hat{\sigma}^2_{\boldsymbol{\epsilon}} I_K\right)$ are non-negative and the matrix square root is well-defined. For reasons of uniqueness, we usually set for the rotation matrix $R = I_K$.

So far, the choice of the factor dimension K was ignored, however, in empirical studies the true factor dimension is unknown and therefore, has to be estimated. If PCA is used for factor estimation, Abdi and Williams (2010) provide diverse criteria for choosing K. Besides PCA, Bai and Ng (2002) minimize the average variance of the model residuals to obtain loadings and factors. In this manner, they receive further information and panel criteria. Here, we pursue a simpler method, that is, we take the smallest K such that K factors still capture a prespecified percentage ζ of the overall variation. Formally, we have:

Definition 3.1.4 (Selection of Factor Dimension in SFMs)
Let $\zeta \in (0,1)$ *be the percentage of data variation we intend to cover. Then, the optimal* $K = K(\zeta)$ *meets:*

$$K(\zeta) = \min_k \left\{ \left[\sum_{i=1}^{k} \frac{\lambda_i}{\left(\sum_{j=1}^{N} \lambda_j\right)} \right] \geq \zeta \quad \textit{and} \quad 1 \leq k \leq N \right\}, \tag{3.6}$$

with $\lambda_1 \geq \ldots \geq \lambda_N \geq 0$ *being the descendingly sorted eigenvalues of* $\hat{\Sigma}_{\boldsymbol{X}}$ *in Definition 2.1.2.*

In Theorem 3.1.3, we obtained estimates of the parameters Θ for fixed factor dimension K. In addition, Definition 3.1.4 offered a criterion for the choice of K. All in all, both results yield the following algorithm for model estimation and selection:

Algorithm 3.1.1: Estimate ESFMs with isotropic errors based on complete panel data

Determine the eigenvalues and eigenvectors of $\hat{\Sigma}_{\boldsymbol{X}}$ in (2.2);
For a chosen $\zeta > 0$ derive the optimal number of factors K according to (3.6);
Set $R = I_K$ and determine $\hat{W}$ and $\hat{\sigma}^2_{\boldsymbol{\epsilon}}$ as in (3.4) and (3.5);

Algorithm 3.1.1 estimates the parameters Θ, but does not provide any further information on the factors $\boldsymbol{F}_t$. In Definition 2.1.3, the factors $\boldsymbol{F}_t$ were supposed to be iid standard normal, however, this marginal distribution ignores any information inherent in the panel data X. The factors $\boldsymbol{F}_t$ and idiosyncratic errors $\boldsymbol{\epsilon}_t$ are iid Gaussian such that the panel data $\boldsymbol{X}_t$ is iid Gaussian, too. Furthermore, this implies that only $\boldsymbol{X}_t$ offers information on $\boldsymbol{F}_t$, which is shown in Lemma 3.1.5.

Lemma 3.1.5 (Conditional Distribution of Factors in ESFMs)
For the ESFM in Definition 2.1.3 with an isotropic error model, the distribution of the hidden factors $\boldsymbol{F}_t$ given the panel data $\boldsymbol{X}_t$, i.e., $\boldsymbol{F}_t|\boldsymbol{X}_t$, is multivariate Gaussian and defined as follows:

$$\boldsymbol{F}_t|\boldsymbol{X}_t \sim \mathcal{N}\left(M^{-1}W'\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right), \sigma_{\boldsymbol{\epsilon}}^2 M^{-1}\right) = \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}, \Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right), \tag{3.7}$$

with symmetric, non-singular matrix $M = W'W + \sigma_{\boldsymbol{\epsilon}}^2 I_K \in \mathbb{R}^{K\times K}$.

Proof:
As in Tipping and Bishop (1999), we obtain by virtue of the Bayes' theorem:

$$\begin{aligned}
f_\Theta\left(\boldsymbol{F}_t|\boldsymbol{X}_t\right) &= \frac{f_\Theta\left(\boldsymbol{F}_t, \boldsymbol{X}_t\right)}{f_\Theta\left(\boldsymbol{X}_t\right)} = \frac{f_\Theta\left(\boldsymbol{X}_t|\boldsymbol{F}_t\right) f_\Theta\left(\boldsymbol{F}_t\right)}{f_\Theta\left(\boldsymbol{X}_t\right)} \propto f_\Theta\left(\boldsymbol{X}_t|\boldsymbol{F}_t\right) f_\Theta\left(\boldsymbol{F}_t\right) \\
&\propto exp\left(-\frac{1}{2}\left(\boldsymbol{X}_t - W\boldsymbol{F}_t - \boldsymbol{\mu}\right)' \sigma_{\boldsymbol{\epsilon}}^{-2} I_N \left(\boldsymbol{X}_t - W\boldsymbol{F}_t - \boldsymbol{\mu}\right)\right)\ exp\left(-\frac{1}{2}\boldsymbol{F}_t'\boldsymbol{F}_t\right) \\
&= exp\left(-\frac{1}{2\sigma_{\boldsymbol{\epsilon}}^2}\left[\left(\boldsymbol{X}_t - W\boldsymbol{F}_t - \boldsymbol{\mu}\right)'\left(\boldsymbol{X}_t - W\boldsymbol{F}_t - \boldsymbol{\mu}\right) + \sigma_{\boldsymbol{\epsilon}}^2 \boldsymbol{F}_t'\boldsymbol{F}_t\right]\right) \\
&= exp\left(-\frac{1}{2\sigma_{\boldsymbol{\epsilon}}^2}\left[\boldsymbol{F}_t'\left(W'W + \sigma_{\boldsymbol{\epsilon}}^2 I_K\right)\boldsymbol{F}_t + \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)'\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)\right]\right) \\
&\quad \cdot exp\left(\frac{1}{2\sigma_{\boldsymbol{\epsilon}}^2}\left[\boldsymbol{F}_t' W'\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right) + \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)' W\boldsymbol{F}_t\right]\right) \\
&\propto exp\left(-\frac{1}{2\sigma_{\boldsymbol{\epsilon}}^2}\left[\boldsymbol{F}_t' M\boldsymbol{F}_t - \boldsymbol{F}_t' M M^{-1} W'\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right) - \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)' W M^{-1} M\boldsymbol{F}_t\right]\right) \\
&\propto exp\left(-\frac{1}{2}\left(\boldsymbol{F}_t - M^{-1}W'\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)\right)'\left(\sigma_{\boldsymbol{\epsilon}}^{-2} M\right)\left(\boldsymbol{F}_t - M^{-1}W'\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)\right)\right),
\end{aligned}$$

which is proportional to the probability density function of the multivariate Gaussian distribution with mean $\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} = M^{-1}W'\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right) \in \mathbb{R}^K$ and covariance matrix $\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} = \sigma_{\boldsymbol{\epsilon}}^2 M^{-1} \in \mathbb{R}^{K\times K}$. The symmetry of the matrix M follows from its definition. For $K < N$, we have: $\sigma_{\boldsymbol{\epsilon}}^2 > 0$ such that the matrix M is non-singular. For $K = N$, the trivial solution $W = I_N$ and $\sigma_{\boldsymbol{\epsilon}}^2 = 0$ also ensures the non-singularity of the matrix M. □

Hence, at any point in time the factor $\boldsymbol{F}_t$ given the observation $\boldsymbol{X}_t$ is Gaussian. If we replace the matrix M by $W'W$, i.e., we neglect the term $\sigma_{\boldsymbol{\epsilon}}^2 I_K$, the mean of $\boldsymbol{F}_t$ in (3.7) coincides with the OLS estimator. For $K < N$ and any positive-definite covariance matrix $\Sigma_{\boldsymbol{\epsilon}}$, it follows: $\sigma_{\boldsymbol{\epsilon}}^2 > 0$ and thus, M causes a bias compared to OLS. If dimension reduction does not take place, i.e., $K = N$, there is no bias for the trivial solution: $W = I_K$ and $\sigma_{\boldsymbol{\epsilon}}^2 = 0$.

For us, the conditional distribution $\boldsymbol{F}_t|\boldsymbol{X}_t$ in Lemma 3.1.5 is an intermediate result on our way to estimate ESFMs with incomplete data in Section 3.1.2. In case of missing observations and mixed-frequency data, we pursue an interative scheme that, among other things, derives artificial, complete panel data from the latest parameter and factor estimates. To be precise, this is achieved in the following manner:

Lemma 3.1.6 (Optimal Reconstruction of Panel Data)
Let $\hat{\Theta} = \{\hat{W}, \hat{\sigma}_{\boldsymbol{\epsilon}}^2\}$ denote the parameter estimates from Theorem 3.1.3 and let $\hat{M} \in \mathbb{R}^{K\times K}$ be the estimated matrix M from Lemma 3.1.5, i.e., $\hat{M} = \hat{W}'\hat{W} + \hat{\sigma}_{\boldsymbol{\epsilon}}^2 I_K$. If the vector $\hat{\boldsymbol{\mu}}_{\boldsymbol{F}_t|\boldsymbol{X}_t} = \hat{M}^{-1}\hat{W}'(\boldsymbol{X}_t - \hat{\boldsymbol{\mu}}_{\boldsymbol{X}}) \in \mathbb{R}^K$ stands for the estimated mean $\boldsymbol{F}_t|\boldsymbol{X}_t$ in Lemma 3.1.5, the vector $\boldsymbol{X}_t^{rec} = \hat{W}\hat{\boldsymbol{\mu}}_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \hat{\boldsymbol{\mu}}_{\boldsymbol{X}} \in \mathbb{R}^N$ is not an orthogonal projection of $\boldsymbol{X}_t$. By contrast, the below expression for $\boldsymbol{X}_t^{rec}$ minimizes the reconstruction

error of the panel data $\boldsymbol{X}_t$ based on $\hat{\Theta}$ and $\hat{\boldsymbol{\mu}}_{\boldsymbol{F}_t|\boldsymbol{X}_t}$:

$$\boldsymbol{X}_t^{rec} = \hat{W}\left(\hat{W}'\hat{W}\right)^{-1}\hat{M}\ \hat{\boldsymbol{\mu}}_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \hat{\boldsymbol{\mu}}_{\boldsymbol{X}} = \hat{W}\left(\hat{W}'\hat{W}\right)^{-1}\hat{W}'\left(\boldsymbol{X}_t - \hat{\boldsymbol{\mu}}_{\boldsymbol{X}}\right) + \hat{\boldsymbol{\mu}}_{\boldsymbol{X}}. \tag{3.8}$$

Here, the upper index "rec" highlights that we have "reconstructed" panel data.

Proof:
See Tipping and Bishop (1999). □

In Lemma 3.1.6, the vector $(\hat{W}'\hat{W})^{-1}\hat{W}'(\boldsymbol{X}_t - \hat{\boldsymbol{\mu}}_{\boldsymbol{X}})$ represents the fitted value of an OLS applied to (2.3) for known matrix $\hat{W}$. Therefore, (3.8) reconstructs panel data as an OLS. However, we benefit from the ansatz of Tipping and Bishop (1999), since the matrix $\hat{M}$ captures uncertainties caused by lost variation. For instance, when we estimate prediction intervals for returns of future periods of time in Section 3.1.3, we use random factor samples drawn from the distribution in (3.7) instead of the means $\hat{\boldsymbol{\mu}}_{\boldsymbol{F}_t|\boldsymbol{X}_t}$ to allow for estimation risks. Within a Bayesian framework De Mol et al. (2008) apply a similar approach for the estimation of latent factors, although they do not incorporate the bias caused by $\hat{M}$ instead of $\hat{W}'\hat{W}$.

3.1.2 Model Estimation Based on Incomplete Panel Data

Before, we considered complete panel data. Now, we permit data incompleteness arising from the inclusion of mixed-frequency and missing information. For this purpose, the linear relation in Definition 2.2.1 offers a comfortable way for modeling returns, yields, spreads, growth rates, etc. This comes from the fact that many data transformations can be addressed by (2.9). In this context, we first modify the reconstruction formula of Stock and Watson (1999a, 2002b) for ESFMs with an isotropic error model.

Lemma 3.1.7 (Reconstruction of Panel Data From Observations)
Assume the ESFM in Definition 2.1.3 with isotropic error model, where $X \in \mathbb{R}^{T\times N}$, $F = [\boldsymbol{F}_1, \dots, \boldsymbol{F}_T]' \in \mathbb{R}^{T\times K}$ and $\epsilon = [\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_T]' \in \mathbb{R}^{T\times N}$ collect complete panel data, hidden factors and idiosyncratic shocks. With respect to Definition 2.2.1, for $1 \leq i \leq N$ let the vector $\boldsymbol{X}^i_{obs} \in \mathbb{R}^{T(i)}, 1 \leq T(i) \leq T$, summarize the actually observed values of $\boldsymbol{X}^i \in \mathbb{R}^T$, which is the i-th column of X. Then, the vector $\boldsymbol{X}^i$ given F, $\boldsymbol{X}^i_{obs}$ and the parameters $\Theta = \{W, \sigma^2_{\boldsymbol{\epsilon}}\}$ is multivariate Gaussian with mean and covariance matrix as follows:

$$\mathbb{E}_\Theta\left[\boldsymbol{X}^i|F, \boldsymbol{X}^i_{obs}\right] = \left(F\boldsymbol{W}_i' + \mu_i \mathbb{1}_T\right) + Q_i'\left(Q_iQ_i'\right)^{-1}\left[\boldsymbol{X}^i_{obs} - Q_i\left(F\boldsymbol{W}_i' + \mu_i\mathbb{1}_T\right)\right],$$
$$\mathbb{V}ar_\Theta\left[\boldsymbol{X}^i|F, \boldsymbol{X}^i_{obs}\right] = \sigma^2_{\boldsymbol{\epsilon}_i}\left[I_T - Q_i'\left(Q_iQ_i'\right)^{-1}Q_i\right],$$

where $\boldsymbol{W}_i$, μ_i and $\boldsymbol{\epsilon}^i$ denote the i-th row of W, the i-th element of $\boldsymbol{\mu}$ and the i-th column of ϵ, respectively.

Proof:
Rearranging (2.3) in matrix form and focussing on the i-th time series provide for (2.3) and (2.9):

$$\boldsymbol{X}^i = F\boldsymbol{W}_i' + \mu_i\mathbb{1}_T + \boldsymbol{\epsilon}^i,$$
$$\boldsymbol{X}^i_{\text{obs}} = Q_iF\boldsymbol{W}_i' + Q_i\mu_i\mathbb{1}_T + Q_i\boldsymbol{\epsilon}^i.$$

Because of $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}_N, \Sigma_{\boldsymbol{\epsilon}})$ iid, for all $1 \leq i \leq N$ we get $\boldsymbol{\epsilon}^i \sim \mathcal{N}(\boldsymbol{0}_T, \sigma^2_{\boldsymbol{\epsilon}_i}I_T)$ resulting in:

$$\left.\begin{pmatrix}\boldsymbol{X}^i\\ \boldsymbol{X}^i_{\text{obs}}\end{pmatrix}\right|_{F,\Theta} \sim \mathcal{N}\left(\begin{pmatrix}F\boldsymbol{W}_i' + \mu_i\mathbb{1}_T\\ Q_iF\boldsymbol{W}_i' + Q_i\mu_i\mathbb{1}_T\end{pmatrix}, \sigma^2_{\boldsymbol{\epsilon}_i}\begin{pmatrix}I_T & Q_i'\\ Q_i & Q_iQ_i'\end{pmatrix}\right).$$

Finally, the conditional mean and covariance matrix of the multivariate Gaussian distribution (Greene, 2003, pp. 871-872, Theorem B.7) yield the statement. □

The matrix $Q_i' \left(Q_i Q_i'\right)^{-1} \in \mathbb{R}^{T \times T(i)}$ in Lemma 3.1.7 represents the unique Moore-Penrose Inverse of the matrix Q_i (Rao and Toutenburg, 1999, pp. 372-373, Definition A.64) and satisfies: $Q_i Q_i' \left(Q_i Q_i'\right)^{-1} = I_{T(i)}$. Regarding the EM, which we will later on develop, the uniqueness eliminates undesired degrees of freedom, while the condition $Q_i Q_i' \left(Q_i Q_i'\right)^{-1} = I_{T(i)}$ ensures that the linear relation in Definition 2.2.1 holds, when the EM terminates. Note, Lemma 3.1.7 requires the idiosyncratic shocks to be iid such that the presented isotropic error model is just a special case of a more general result.

In a next step, we combine Lemmata 3.1.6 and 3.1.7 to obtain a comprehensive solution for deriving complete panel data from observations, estimated factors and parameter estimates. The matrix representation for the reconstruction of the panel data $X \in \mathbb{R}^{T \times N}$ in (3.8) is given by:

$$\begin{aligned} X^{rec} &= \begin{bmatrix} (\boldsymbol{X}_1^{rec})' \\ \vdots \\ (\boldsymbol{X}_T^{rec})' \end{bmatrix} = \left[\boldsymbol{X}^{rec,1}, \ldots, \boldsymbol{X}^{rec,N}\right] \\ &= \left(X - \left(\mathbb{1}_T \otimes \hat{\boldsymbol{\mu}}_{\boldsymbol{X}}'\right)\right) \hat{W} \left(\hat{W}'\hat{W}\right)^{-1} \hat{W}' + \left(\mathbb{1}_T \otimes \hat{\boldsymbol{\mu}}_{\boldsymbol{X}}'\right), \end{aligned} \tag{3.9}$$

and so, supports the following update formula.

Corollary 3.1.8 (Update Complete Panel Data)
Assume the setting in Lemma 3.1.7, where $\hat{\Theta}$ comprises the parameter estimates in Theorem 3.1.3, and let the matrix $X^{rec} \in \mathbb{R}^{T \times N}$ be the reconstructed panel data in (3.9). Then, for $1 \leq i \leq N$ the conditional expectation of vector $\boldsymbol{X}^i \in \mathbb{R}^T$ in Lemma 3.1.7 can be estimated for observations $\boldsymbol{X}^i_{obs} \in \mathbb{R}^{T(i)}, T(i) \leq T$, factor means $\hat{\boldsymbol{\mu}}_{\boldsymbol{F}|\boldsymbol{X}} = [\hat{\boldsymbol{\mu}}_{\boldsymbol{F}_1|\boldsymbol{X}_1}, \ldots, \hat{\boldsymbol{\mu}}_{\boldsymbol{F}_T|\boldsymbol{X}_T}]' \in \mathbb{R}^{T \times K}$ and parameter estimates $\hat{\Theta}$ as follows:

$$\mathbb{E}_{\hat{\Theta}}\left[\boldsymbol{X}^i | \boldsymbol{X}^i_{obs}, \hat{\boldsymbol{\mu}}_{\boldsymbol{F}|\boldsymbol{X}}\right] = \boldsymbol{X}^{rec,i} + Q_i' \left(Q_i Q_i'\right)^{-1} \left[\boldsymbol{X}^i_{obs} - Q_i \boldsymbol{X}^{rec,i}\right]. \tag{3.10}$$

Proof:
Replace $\left(F\boldsymbol{W}_i' + \mu_i \mathbb{1}_T\right)$ by $\boldsymbol{X}^{rec,i}$ in Lemma 3.1.7. □

At this point, we briefly explain what we have so far. Theorem 3.1.3 provides parameter estimates for the ESFMs in Definition 2.1.3 with an isotropic error model. For model selection the criterion in Definition 3.1.4 is used. In a next step, Corollary 3.1.8 updates the artificial, complete panel data, before the overall procedure starts again. Hence, two questions arise: On the one hand, how to initialize artificial, complete panel data? On the other hand, when does the total approach stop? Regarding the initialization of panel data, gaps can be filled with random numbers, interpolations, zeros, etc. At the beginning, the linear relation in Definition 2.2.1 does not necessarily have to hold, since this will be automatically reached by the updates in Corollary 3.1.8. With respect to the second question, we define a termination criterion.

Similar to Doz et al. (2012) and Bańbura and Modugno (2014), the termination of the updates in Corollary 3.1.8 relies on the change in the log-likelihood function $\mathcal{L}\left(\Theta|X\right)$. To be more accurate, let $\mathcal{L}(\hat{\Theta}_{(l)}|X_{(l)})$ be the log-likelihood function gained from the estimated model parameters $\hat{\Theta}_{(l)}$ and data sample $X_{(l)}$ of loop (l) of the total routine. Then, our updates stop as soon as it holds:

$$\frac{abs\left(\mathcal{L}\left(\hat{\Theta}_{(l)}|X_{(l)}\right)-\mathcal{L}\left(\hat{\Theta}_{(l-1)}|X_{(l-1)}\right)\right)}{\frac{1}{2}\left(abs\left(\mathcal{L}\left(\hat{\Theta}_{(l)}|X_{(l)}\right)\right)+abs\left(\mathcal{L}\left(\hat{\Theta}_{(l-1)}|X_{(l-1)}\right)\right)\right)}<\xi, \tag{3.11}$$

with $abs(\cdot)$ denoting the absolute value of a real number. We stop in (3.11), when the absolute value of the relative change in the log-likelihood function is smaller than a prespecified limit $\xi > 0$. By contrast, Doz et al. (2012) and Bańbura and Modugno (2014) omit the absolute value in the numerator of (3.11), that is, they consider the relative improvement in $\mathcal{L}(\Theta|X)$. On the one hand, this reflects the theoretical convergence properties of the EM in Wu (1983), since each EM delivers a non-decreasing sequence of log-likelihood functions. On the other hand, numerical inaccuracies in case of real-world data may cause a few tiny declines, before the nearest local optimum is reached. Then, the approach of Doz et al. (2012) and Bańbura and Modugno (2014) might stop too early.

As in Section 3.1.1, we summarize all steps as algorithm. For the initialization of $\boldsymbol{X}_{(0)}$ diverse approaches provide a first set of complete panel data. In general, an EM detects a local optimum of the log-likelihood function. Therefore, the application of various initialization methods, which offers different starting values, improves the chance of reaching a global maximum. In opposition to the choice of $\boldsymbol{X}^i_{(0)}$, the shape of the matrices $Q_i, 1 \leq i \leq N$, matters, since Algorithm 3.1.2 does not adjust it later on.

Until Algorithm 3.1.2 stops, the estimated factor dimension K may change several times. To prevent its termination behavior from changes in K, the criterion ξ controls changes in $\mathcal{L}(\Theta|X)$ instead of changes in Θ. For instance, the convergence criterion in Schumacher and Breitung (2008) does the latter. Moreover, ξ checks for relative changes instead of absolute ones to ensure that neither the dimension of the parameter space nor the sample size have any impact on the termination of the overall routine.

The construction of complete panel data from the latest parameter estimates and observations in (3.10) guarantees that the equality in (2.9) holds at convergence. In each loop, the second term on the right-hand side of (3.10) punishes any deviations from the observed signals. For all complete time series we have $Q_i = I_T$ and thus, the simplified version of (3.10) is given by: $\boldsymbol{X}^i = \boldsymbol{X}^i_{obs}$. That is, these times series are kept in total without any adjustments. For any time series with $T(i) < T$, we benefit from the unique Moore-Penrose Inverse $Q_i'\left(Q_iQ_i'\right)^{-1}$ satisfying: $Q_iQ_i'\left(Q_iQ_i'\right)^{-1} = I_{T(i)}$.

The advantages of the presented framework are as follows: First, the underlying FM reduces the dimension from the panel data space N to the factor span K. In our empirical studies, we have $K \ll N$ causing a significant dimension reduction. Second, the distribution of the factors in Lemma 3.1.5 takes uncertainties arising from lost variation into account. This will be important, when we construct empirical prediction intervals for returns of future periods. Third, Algorithm 3.1.2 admits the inclusion of mixed frequencies and the estimation of high-frequency analogs for low-frequency time series (nowcasting).

Algorithm 3.1.2: Estimate ESFMs with isotropic errors based on incomplete panel data

Initialization
Define level of variation $\zeta > 0$ to be covered;
Choose termination criterion $\xi > 0$;
Set loop index $(l) = 0$;

for $i = 1$ **to** N **do**
  Initialize $\boldsymbol{X}^i_{(l)}$ (if necessary, fill gaps);
  Specify matrix Q_i;
end

Estimate ESFM with $X_{(l)}$ using Algorithm 3.1.1 for variation level ζ and store parameters $\hat{\Theta}_{(l)}$;
Determine log-likelihood $\mathcal{L}(\hat{\Theta}_{(l)}|X_{(l)})$ in (3.3);

for $i = 1$ **to** N **do**
  Derive updated panel data $\boldsymbol{X}^i_{(l+1)}$ from (3.10) and model parameters $\hat{\Theta}_{(l)}$;
end

Estimate ESFM with $X_{(l+1)}$ using Algorithm 3.1.1 for var. level ζ and store parameters $\hat{\Theta}_{(l+1)}$;
Determine log-likelihood $\mathcal{L}(\hat{\Theta}_{(l+1)}|X_{(l+1)})$ in (3.3);

Alternating reconstruction and reestimation
while $\frac{abs\left(\mathcal{L}(\hat{\Theta}_{(l+1)}|X_{(l+1)})-\mathcal{L}(\hat{\Theta}_{(l)}|X_{(l)})\right)}{\frac{1}{2}\left(abs\left(\mathcal{L}(\hat{\Theta}_{(l+1)}|X_{(l+1)})\right)+abs\left(\mathcal{L}(\hat{\Theta}_{(l)}|X_{(l)})\right)\right)} > \xi$ **do**
  Set loop index $(l) = (l+1)$;

  for $i = 1$ **to** N **do**
    Derive updated panel data $\boldsymbol{X}^i_{(l+1)}$ from (3.10) and model parameters $\hat{\Theta}_{(l)}$;
  end

  Estimate ESFM with $X_{(l+1)}$ using Algorithm 3.1.1 for var. level ζ and store parameters $\hat{\Theta}_{(l+1)}$;
  Determine log-likelihood $\mathcal{L}(\hat{\Theta}_{(l+1)}|X_{(l+1)})$ in (3.3);
end

3.1.3 Portfolio Optimization

After extracting information from large, possibly incomplete panel data by a few factors, we process the output of Algorithms 3.1.1 and 3.1.2. Thereby, we describe the dynamics of a multivariate return process as VARX, where the hidden factors take the role of exogenous variables. Unlike the FAVAR of Bernanke et al. (2005), which we consider in Chapter 5, the factors in a VARX affect the returns, whereas the reverse relation does not matter. For reasons of simplicity, the factors and returns are simultaneously updated to avoid new mixed-frequency iusses. Moreover, the sample lengths of the panel data and returns are the same. Hence, the sample lengths of the factors and returns coincide, too. Although we can easily drop the last assumption by dealing with the intersection period of both, it improves the clarity of subsequent calculations. With this in mind, we define the return VARX as follows.

Definition 3.1.9 (Vector Autoregression Model with Exogenous Variables)
For any point in time $t \geq 1$, the vector $\boldsymbol{r}_t \in \mathbb{R}^H$ comprises returns gained over the period $(t-1\,,\,t]$. If the integers $\tilde{q}, \tilde{p} \geq 1$ refer to the autoregressive orders of the returns and factors, respectively, the return

VARX($\tilde{q},\tilde{p}$) with the factors of the ESFM in Defintion 2.1.3 as exogenous variables is given by:

$$\boldsymbol{r}_t = \boldsymbol{\gamma} + \sum_{i=1}^{\tilde{q}} A_i \boldsymbol{r}_{t-i} + \sum_{i=1}^{\tilde{p}} B_i \boldsymbol{F}_{t-i} + \boldsymbol{\delta}_t, \tag{3.12}$$

where the vector $\boldsymbol{\gamma} \in \mathbb{R}^H$ and matrices $A_i \in \mathbb{R}^{H\times H}, 1 \le i \le \tilde{q}$, and $B_i \in \mathbb{R}^{H\times K}, 1 \le i \le \tilde{p}$ are constants. The errors $\boldsymbol{\delta_t}$ are iid Gaussian with zero mean and covariance matrix $\Sigma_{\boldsymbol{\delta}} \in \mathbb{R}^{H\times H}$, i.e., $\boldsymbol{\delta}_t \sim \mathcal{N}(\mathbf{0}_H, \Sigma_{\boldsymbol{\delta}})$. In the sequel, we set $\tilde{m} = \max\{\tilde{q},\tilde{p}\}$ and define the coefficient matrix $\Theta \in \mathbb{R}^{H\times(1+\tilde{q}H+\tilde{p}K)}$ by:

$$\Theta = [\boldsymbol{\gamma}, A_1, \dots, A_{\tilde{q}}, B_1, \dots, B_{\tilde{p}}]. \tag{3.13}$$

Definition 3.1.9 specifies the return VARX, however, for parameter estimation using least-squares regression a sample based formulation is more appropriate. For this purpose, Definition 3.1.10 introduces such a representation, before we state in Lemma 3.1.11 the asymptotic distribution of the VARX coefficients.

Definition 3.1.10 (Matrix Representation of VARX)
Assume the VARX($\tilde{q},\tilde{p}$) from Definition 3.1.9 for return and factor samples of length T. In addition, we have for known lag lengths $\tilde{p},\tilde{q} \ge 1$: $\tilde{m} = \max\{\tilde{q},\tilde{p}\}$ and keep the coefficient matrix $\Theta \in \mathbb{R}^{H\times(1+\tilde{q}H+\tilde{p}K)}$ as in (3.13)*. Then, the model in* (3.12) *can be rewritten as:*

$$\underbrace{[\boldsymbol{r}_{\tilde{m}+1}, \dots, \boldsymbol{r}_T]}_{\mathcal{R}} = \Theta \underbrace{\begin{bmatrix} 1 & \cdots & 1 \\ \boldsymbol{r}_{\tilde{m}} & \cdots & \boldsymbol{r}_{T-1} \\ \vdots & & \vdots \\ \boldsymbol{r}_{\tilde{m}+1-\tilde{q}} & \cdots & \boldsymbol{r}_{T-\tilde{q}} \\ \boldsymbol{F}_{\tilde{m}} & \cdots & \boldsymbol{F}_{T-1} \\ \vdots & & \vdots \\ \boldsymbol{F}_{\tilde{m}+1-\tilde{p}} & \cdots & \boldsymbol{F}_{T-\tilde{p}} \end{bmatrix}}_{\mathcal{G}} + \underbrace{[\boldsymbol{\delta}_{\tilde{m}+1}, \dots, \boldsymbol{\delta}_T]}_{\mathcal{D}}. \tag{3.14}$$

Let the operator $vec(\cdot)$ be the matrix vectorization in Definition A.1.12, that is, the successively stacked columns of a given matrix. With this in mind, we receive the following approximative distribution of the coefficient matrix Θ:

Lemma 3.1.11 (Least-Squares Estimation of VARX Parameters)
In the setting of Definition 3.1.10, let the matrix $\hat{\Theta} \in \mathbb{R}^{H\times(1+\tilde{q}H+\tilde{p}K)}$ be the least-squares estimator of the coefficient matrix Θ. Under certain regularity assumptions the vector $vec(\hat{\Theta})$ is asymptotically Gaussian, that is, it holds:

$$vec(\hat{\Theta}) \sim \mathcal{N}\left(vec\left(\mathcal{R}\mathcal{G}'\left(\mathcal{G}\mathcal{G}'\right)^{-1}\right), \left(\mathcal{G}\mathcal{G}'\right)^{-1} \otimes \hat{\Sigma}_{\boldsymbol{\delta}}\right), \tag{3.15}$$

where $\hat{\Sigma}_{\boldsymbol{\delta}}$ represents the empirical covariance matrix of the iid errors in (3.12).

Proof:
See Lütkepohl (2005, p. 74, Proposition 3.1). □

As mentioned in Remark 3.1.1, the factors are unique except for rotation. Therefore, we address this issue right now, before we later on discuss new topics such as model selection.

Remark 3.1.12 (Impact of Factor Ambiguity on Returns)
We discussed in Remark 3.1.1 that the hidden factors of the ESFM in Definition 2.1.3 are unique except for rotation. However, this does not affect the returns $\boldsymbol{r}_t$ of the VARX$(\tilde{q},\tilde{p})$ in Definition 3.1.9.

Proof:
For an arbitrary rotation matrix $R \in \mathbb{R}^{K\times K}$ the vector $\bar{\boldsymbol{F}}_t = R'\boldsymbol{F}_t \in \mathbb{R}^K$ stands for the rotated factor in Remark 3.1.1. Then, Equation (3.12) coincides with:

$$\boldsymbol{r}_t = \boldsymbol{\gamma} + \sum_{i=1}^{\tilde{q}} A_i \boldsymbol{r}_{t-i} + \sum_{i=1}^{\tilde{p}} B_i \underbrace{RR'}_{I_K} \boldsymbol{F}_{t-i} + \boldsymbol{\delta}_t = \boldsymbol{\gamma} + \sum_{i=1}^{\tilde{q}} A_i \boldsymbol{r}_{t-i} + \sum_{i=1}^{\tilde{p}} \bar{B}_i \bar{\boldsymbol{F}}_{t-i} + \boldsymbol{\delta}_t,$$

with matrices $\bar{B}_i = B_i R \in \mathbb{R}^{H\times K}, 1 \leq i \leq \tilde{p}$. Hence, whenever rotated factors enter (3.12), the corresponding coefficient matrices are reversely transformed such that there is no impact on the returns $\boldsymbol{r}_t$. □

Least-squares estimation in Lemma 3.1.11 requires known autoregressive orders $\tilde{p},\tilde{q} \geq 1$. Hence, we have to estimate them, too. For this purpose, we apply the standard Akaike Information Criterion (AIC). That is, we estimate the return VARX$(\tilde{q},\tilde{p})$ in Definition 3.1.9 for diverse orders $\tilde{p},\tilde{q} \geq 1$ and choose the pair (q^*,p^*) that minimizes the AIC in Lemma 3.1.13.

Lemma 3.1.13 (Model Selection for Return VARX)
Assume the VARX in Definition 3.1.9 based on a data sample of length T. Then, the optimal autoregressive orders (q^*,p^*) in the sense of the Akaike Information Criterion satisfy:

$$(q^*,p^*) = \underset{(\tilde{q},\tilde{p})}{\arg\min} \left\{ -2\tilde{\mathcal{L}}\left(\hat{\Theta}(\tilde{q},\tilde{p})\,|F,r\right) + 2\left(H + \tilde{q}H^2 + \tilde{p}KH\right) + H(H+1) \right\}, \tag{3.16}$$

where the matrix $\hat{\Theta}(\tilde{q},\tilde{p}) \in \mathbb{R}^{H\times(1+\tilde{q}H+\tilde{p}K)}$ collects the estimated parameters of the VARX$(\tilde{q},\tilde{p})$ in (3.12) and $\tilde{\mathcal{L}}\left(\hat{\Theta}(\tilde{q},\tilde{p})\,|F,r\right)$ is its log-likelihood function for factors $F \in \mathbb{R}^{T\times K}$ and returns $r \in \mathbb{R}^{T\times H}$ given the first $\tilde{m} = \max\{\tilde{p},\tilde{q}\}$ return observations, i.e.:

$$\tilde{\mathcal{L}}\left(\hat{\Theta}(\tilde{q},\tilde{p})\,|F,r\right) = -\frac{1}{2}\left[(H(T-\tilde{m}) + TK)\ \ln(2\pi) + (T-\tilde{m})\left(H + \ \ln\left(\left|\hat{\Sigma}_{\boldsymbol{\delta}}\right|\right)\right) + \sum_{t=1}^{T} \boldsymbol{F}'_t\boldsymbol{F}_t \right].$$

Proof:
In general, we have for AIC (Akaike, 1987):

$$(q^*,p^*) = \underset{(\tilde{q},\tilde{p})}{\arg\min} \left\{ -2\tilde{\mathcal{L}}\left(\hat{\Theta}(\tilde{q},\tilde{p})\,|F,r\right) + 2\,(\text{number of estimated parameters}) \right\}.$$

The coefficient matrix Θ has $H(1+\tilde{q}H+\tilde{p}K)$ parameters, while the covariance matrix of the shocks $\Sigma_{\boldsymbol{\delta}}$ comprises $H(H+1)/2$ parameters (the symmetry of $\Sigma_{\boldsymbol{\delta}}$ matters here), which results in the number of parameters in the second summand of (3.16). For the log-likelihood function, it holds:

$$\begin{aligned}
\tilde{\mathcal{L}}\left(\hat{\Theta}(\tilde{q},\tilde{p})\,|F,r\right) &= \ln\left(f_{\hat{\Theta}(\tilde{q},\tilde{p})}\left(\boldsymbol{r}_T,\ldots,\boldsymbol{r}_{\tilde{m}+1},\boldsymbol{F}_T,\ldots,\boldsymbol{F}_1|\boldsymbol{r}_{\tilde{m}},\ldots,\boldsymbol{r}_1\right)\right) \\
&= \sum_{t=\tilde{m}+1}^{T} \ln\left(f_{\hat{\Theta}(\tilde{q},\tilde{p})}\left(\boldsymbol{r}_t|\boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_{t-\tilde{q}},\boldsymbol{F}_{t-1},\ldots,\boldsymbol{F}_{t-\tilde{p}}\right)\right) + \sum_{t=1}^{T} \ln\left(f_{\hat{\Theta}(\tilde{q},\tilde{p})}\left(\boldsymbol{F}_t\right)\right).
\end{aligned}$$

Because of $\boldsymbol{r}_t|\boldsymbol{r}_{t-1},\ldots,\boldsymbol{r}_{t-\tilde{q}},\boldsymbol{F}_{t-1},\ldots,\boldsymbol{F}_{t-\tilde{p}} \sim \mathcal{N}\left(\underbrace{\boldsymbol{\gamma}+\sum_{i=1}^{\tilde{q}} A_i\boldsymbol{r}_{t-i}+\sum_{i=1}^{\tilde{p}} B_i\boldsymbol{F}_{t-i}}_{\boldsymbol{\mu}_{\boldsymbol{r}_t}},\Sigma_{\boldsymbol{\delta}}\right)$, we obtain the following, after replacing the model parameters Θ by their estimates $\hat{\Theta}$:

$$
\begin{aligned}
&\tilde{\mathcal{L}}\left(\hat{\Theta}\left(\tilde{q},\tilde{p}\right)|F,r\right)\\
&=\sum_{t=\tilde{m}+1}^{T} ln\left[(2\pi)^{-H/2}\left|\hat{\Sigma}_{\boldsymbol{\delta}}\right|^{-1/2} exp\left(-\frac{1}{2}\left(\boldsymbol{r}_t-\hat{\boldsymbol{\mu}}_{\boldsymbol{r}_t}\right)'\hat{\Sigma}_{\boldsymbol{\delta}}^{-1}\left(\boldsymbol{r}_t-\hat{\boldsymbol{\mu}}_{\boldsymbol{r}_t}\right)\right)\right]+\sum_{t=1}^{T} ln\left[f_{\hat{\Theta}}\left(\boldsymbol{F}_t\right)\right]\\
&=-\frac{T-\tilde{m}}{2}\left(H\ ln\left(2\pi\right)+ln\left(\left|\hat{\Sigma}_{\boldsymbol{\delta}}\right|\right)\right)-\frac{1}{2}\sum_{t=\tilde{m}+1}^{T}\left(\boldsymbol{r}_t-\hat{\boldsymbol{\mu}}_{\boldsymbol{r}_t}\right)'\hat{\Sigma}_{\boldsymbol{\delta}}^{-1}\left(\boldsymbol{r}_t-\hat{\boldsymbol{\mu}}_{\boldsymbol{r}_t}\right)+\sum_{t=1}^{T} ln\left[f_{\hat{\Theta}}\left(\boldsymbol{F}_t\right)\right]\\
&=-\frac{1}{2}\left(H\left(T-\tilde{m}\right)+TK\right)\ ln\left(2\pi\right)-\frac{T-\tilde{m}}{2}\left[\ ln\left(\left|\hat{\Sigma}_{\boldsymbol{\delta}}\right|\right)+\ tr\left(\hat{\Sigma}_{\boldsymbol{\delta}}^{-1}\hat{\Sigma}_{\boldsymbol{\delta}}\right)\right]-\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{F}_t'\boldsymbol{F}_t.
\end{aligned}
$$

Thus, $tr\left(\hat{\Sigma}_{\boldsymbol{\delta}}^{-1}\hat{\Sigma}_{\boldsymbol{\delta}}\right)=H$ proves the claim. □

Next, we introduce an algorithm that deals with the uncertainties in the estimation of the hidden factors and VARX$(\tilde{q},\tilde{p})$ parameters, when it predicts returns of future periods of time. The current formulation of Algorithm 3.1.3 forecasts returns of the next period of time. However, after some minor adjustments it also generates return samples for any s-step ahead forecast with integer $s \geq 1$. Similar to Algorithm 4.2.1, Algorithm 3.1.3 can be modified such that the drivers of the predicted returns (e.g. autoregressive return behavior, factor impact and error add-ons) are detected. This is important, if we are interested in a decomposition of our forecasts. One of the main features of Algorithm 3.1.3 is that it captures a high level of uncertainty. It samples from the distribution of the factors instead of relying on their estimates. Regarding model selection, it restricts itself to the estimates of the VARX$(\tilde{q},\tilde{p})$ parameters. However, as soon as the autoregressive orders $(\tilde{q},\tilde{p})$ are fixed, a sample is randomly drawn from the distribution of the VARX$(\tilde{q},\tilde{p})$ coefficients to generate the return forecast.

Finally, we process the output of Algorithm 3.1.3. For instance, we construct empirical prediction intervals for returns of the next period of time. In addition, we calculate empirical means and covariance matrices of the predicted return samples to determine mean-variance and marginal-risk-parity, respectively, optimal portfolios. For this purpose, we first define empirical prediction intervals as follows:

Definition 3.1.14 (Empirical Prediction Intervals)
For $1 \leq i \leq H$ and a sample $\left[\boldsymbol{r}_{T+1}^1,\ldots,\boldsymbol{r}_{T+1}^V\right]\in\mathbb{R}^{H\times V}$ of predicted returns, let $\boldsymbol{r}_{T+1,i}^{(1)}\leq\ldots\leq\boldsymbol{r}_{T+1,i}^{(V)}$ be the order statistics of the univariate time series of its i-th element. Then, the ν-prediction interval with $\nu\in[0,1]$ is given by:

$$
\left[\boldsymbol{r}_{T+1,i}^{(\lfloor V(1-\nu)/2\rfloor)},\boldsymbol{r}_{T+1,i}^{(\lceil V(1+\nu)/2\rceil)}\right], \tag{3.17}
$$

where $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ represent the floor and ceiling functions, respectively.

In case of mean-variance portfolio optimization, there are two options: First, we specify a target return, e.g., 10%, and minimize the variance of the total portfolio. That is, we choose from all portfolios offering the target return, e.g. 10%, the one with the lowest variance. Second, we fix an upper threshold for the volatility of the overall portfolio, e.g., 15%, and maximize the expected return. Hence, we take from all

Algorithm 3.1.3: Sample of predicted VARX($\tilde{q}, \tilde{p}$) returns in (3.12) for next period of time

```
### Initialization
Define number V > 0 of returns to be predicted;
Set upper limits of autoregressive orders q̄ ≥ 1 and p̄ ≥ 0;
Estimate factor distribution (3.7) using Algorithm 3.1.2;

### Generation of Return Sample
for c = 1 to V do
  | Draw sample F_1^c, ..., F_T^c from estimated factor distribution (3.7);
  | Initialize coefficient vector θ̂ = vec(Θ̂) = ∅ ;
  | Reset overall AIC value AIC_ov = ∞ (or any sufficiently large number);
  |
  | # Model Selection for Return VARX
  | for q̃ = 1 to q̄ do
  |   | for p̃ = 0 to p̄ do
  |   |   | Estimate mean of temporary coefficient vector θ̃ in (3.15) based on returns r_1, ..., r_T,
  |   |   |  sampled factors F_1^c, ..., F_T^c and autoregressive orders q̃ and p̃;
  |   |   | Determine temporary AIC value AIC_tmp for θ̃ using q̃, p̃ and (3.16);
  |   |   |
  |   |   | if AIC_tmp < AIC_ov then
  |   |   |   | Renew overall AIC value by AIC_ov = AIC_tmp;
  |   |   |   | Update overall coefficient vector by θ̂ = θ̃;
  |   |   | end
  |   | end
  | end
  |
  | Estimate empirical error covariance matrix Σ̂_δ using θ̂ and (3.12);
  | Determine covariance matrix of θ̂ in (3.15);
  | Draw sample θ̂^c from estimated coefficient distribution (3.15);
  | Draw error sample δ_T^c from N(0_H, Σ̂_δ);
  | Forecast return r_{T+1}^c from (3.12) based on θ̂^c, r_{T+1-q̃}, ..., r_T, F_{T+1-p̃}^c, ..., F_T^c and δ_T^c;
end
```

portfolios, whose standard deviation does not exceed our upper limit, e.g., 15 %, the one with the highest expected return. Here, we pursue the second approach and define it as follows:

Definition 3.1.15 (Mean-Variance Portfolio Optimization)
Assume a portfolio consisting of $H > 0$ assets with expected return $\mathbb{E}_\Theta\left[\boldsymbol{r}_{T+1}\right] \in \mathbb{R}^H$ and covariance matrix $\mathbb{V}ar_\Theta\left[\boldsymbol{r}_{T+1}\right] \in \mathbb{R}^{H\times H}$. The vector $w \in \mathbb{R}^H$ contains all asset weights, while $\sigma_p^2 \in \mathbb{R}_+$ is an upper limit of the admissible variance of the overall portfolio. If short selling ($0 > w_i$) and leverage ($w'\mathbb{1}_H > 1$) are excluded, a mean-variance optimal portfolio satisfies:

$$\begin{aligned} \max_w \quad & w'\,\mathbb{E}_\Theta\left[\boldsymbol{r}_{T+1}\right] \\ s.t. \quad & w'\,\mathbb{V}ar_\Theta\left[\boldsymbol{r}_{T+1}\right] w \le \sigma_p^2, \\ & w'\mathbb{1}_H = 1, 0 \le w \le 1. \end{aligned} \tag{3.18}$$

The relation $0 \le w \le 1$ holds for each single component of w.

Mean-variance optimization requires estimates of the expected returns $\mathbb{E}_\Theta\left[\boldsymbol{r}_{T+1}\right]$, which strongly affect the obtained portfolio. Small changes in the expected returns possibly provide different asset allocations, therefore, their estimation is crucial. E.g., risk-parity portfolio optimization waives return expectations. The idea behind risk-parity portfolios is that each asset contributes the same portion of risk to the overall risk. Thereby, risk often refers to the volatility of the single assets and the total portfolio. Let the vector $w \in \mathbb{R}^H$ and matrix $\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right] \in \mathbb{R}^{H\times H}$ denote the weights and expected covariance matrix of all assets a portfolio consists of. Then, the marginal risk vector is given by (Roncalli, 2013, p. 79, Section 2.1.2.1):

$$\frac{\partial\,\sigma_p}{\partial\,w} = \frac{\partial\sqrt{w'\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w}}{\partial\,w} = \frac{\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w}{\sqrt{w'\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w}} \in \mathbb{R}^H.$$

If $\left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_i$ with $1 \leq i \leq H$ stands for the i-th element of the vector $\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w$, we obtain for the risk contribution of asset i:

$$\frac{w_i\left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_i}{\sqrt{w'\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w}}. \tag{3.19}$$

In case of a risk-parity portfolio, all assets contribute to the total risk equally. This implies the subsequent non-linear optimization problem:

$$\min_w \sum_{i,j=1}^{H}\left(\frac{w_i\left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_i}{\sqrt{w'\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w}} - \frac{w_j\left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_j}{\sqrt{w'\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w}}\right)^2. \tag{3.20}$$

Because of the nonlinear objective function in (3.20), we pursue a slightly different approach in the sequel. Instead of the standard deviation, we consider the portfolio variance as risk measure. Thus, the marginal risk vector is given by $2\,\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w$. In addition, we aim at constructing a portfolio, whose assets have an equal marginal risk. In total, this results in the following unconstrained optimization problem:

$$\min_w \sum_{i,j=1}^{H}\Big(\left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_i - \left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_j\Big)^2.$$

By minimizing $\left(\left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_1 - \left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_2\right)^2$ and $\left(\left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_2 - \left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_3\right)^2$, we implicitly minimize $\left(\left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_1 - \left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_3\right)^2$. Furthermore, for any $i = j$ the difference is zero by definition. Therefore, we approach the objective function in (3.21) by:

$$\sum_{i=2}^{H}\left(\left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_{i-1} - \left(\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)_i\right)^2. \tag{3.21}$$

All in all, this forms the basis of our constrained marginal-risk-parity optimal portfolio.

Definition 3.1.16 (Marginal-Risk-Parity Portfolio Optimization)
For a portfolio of $H > 0$ entities with expected covariance matrix $\mathbb{V}ar_\Theta\left[\boldsymbol{r}_{T+1}\right] \in \mathbb{R}^{H\times H}$, the vector $w \in \mathbb{R}^H$ contains all asset weights. If short selling $(0 > w_i)$ and leverage $(w'\mathbb{1}_H > 1)$ are forbidden, the allocation of a marginal-risk-parity optimal portfolio obeys:

$$\begin{aligned}\min_w \quad & \left(\Delta\,\mathbb{V}ar_\Theta\left[\boldsymbol{r}_{T+1}\right]w\right)'\Delta\,\mathbb{V}ar_\Theta\left[\boldsymbol{r}_{T+1}\right]w \\ s.t. \quad & w'\mathbb{1}_H = 1, 0 \leq w \leq 1,\end{aligned} \tag{3.22}$$

with difference matrix $\Delta \in \mathbb{R}^{(H-1)\times H}$ defined as:

$$\Delta = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix}.$$

Hence, the objective function in Definition 3.1.16 is quadratic in w_i. For our empirical study in Section 3.2, the empirical mean $\hat{\boldsymbol{r}}_{T+1} \in \mathbb{R}^H$ and covariance matrix $\hat{\Sigma}_{\boldsymbol{r}_{T+1}} \in \mathbb{R}^{H \times H}$ of the predicted return sample of Algorithm 3.1.3 serve as estimates for $\mathbb{E}_\Theta\left[\boldsymbol{r}_{T+1}\right]$ and $\mathbb{V}\mathrm{ar}_\Theta\left[\Sigma_{\boldsymbol{r}_{T+1}}\right]$. For comparison purposes, we rerun the same portfolio optimizations based on historical return averages and covariance matrices.

At the end, we summarize some advantages of Algorithm 3.1.3: First, it admits the inclusion of incomplete panel data to model the dynamics of a multivariate return process. Second, we incorporate uncertainties caused by the estimated factors and VARX$(\tilde{q}, \tilde{p})$ parameters, since factor and parameter samples instead of their estimates enter our forecasts. Third, Algorithm 3.1.3 yields samples of predicted returns. On the one hand, those permit the construction of empirical prediction intervals for the monitoring of financial markets. On the other hand, they offer means and covariance matrices of expected returns and so, provide all inputs for mean-variance or risk-parity portfolio optimization. In this manner, its output itself or after some minor transformations perfectly fits in existing frameworks.

3.2 Empirical Application

Due to the size and importance of the United States, we apply the approach of Section 3.1 to US data. Thereby, the multivariate return process $\{\boldsymbol{r}_t\}$ in Definition 3.1.9 comprises weekly log-returns of price indices such as Gold, Nasdaq Composite (NASDAQ), Standard & Poor's 500 (S&P500) and Dow Jones Industrial Average (DJIA) over the period from February 1, 1985 until November 11, 2016. As observed panel data $\boldsymbol{X}^i_{\mathrm{obs}}$ in Corollary 3.1.8 serve 23 time series, which do not include the returns $\boldsymbol{r}_t$. On the one hand, those cover financial information, e.g., spreads between US Treasury Rates, spreads between London Interbank Offered Rates (LIBORs), corporate bond spreads as well as changes in Bank of America (BofA) Merrill Lynch Spreads and Chicago Board Options Exchange (CBOE) Volatility Indices. On the other hand, they include macroeconomic information, e.g., growth rates of US GDP and industrial production, unemployment and savings rates or personal consumption expenditures.

An overview of the returns $\boldsymbol{r}_t$ and observed panel data $\boldsymbol{X}^i_{\mathrm{obs}}$ is given in Section B.1, which describes in detail the total data. That is, we explain the preprocessing of the input data, the assumed variable types (stock, flow or change in flow), some time series characteristics (available period of time and frequency) and the data sources. To get a first impression, Figure 3.1 displays for each price index how an investment of 100 USD has evolved over the period from February 1, 1985 until November 11, 2016. Note, all price paths indicate the performances of simple Buy&Hold (B&H) strategies. With a view to Figure 3.1, we conclude: First, all stock indices outperformed the Gold strategy. Second, the similar behavior of the stock indices suggests that those are strongly correlated. Third, the correlation between Gold and stocks changed over time. For instance, from end of 2008 until 2009 stocks and Gold gained in value, whereas there were opposite trends in the years 2012-2016. Fourth, the NASDAQ shows the largest amplitude and so, appears to be the most volatile investment.

In the sequel, we successively shift a fixed rolling window of 208 returns until the end of the sample period is reached. In this manner, we ensure that our results remain comparable, when time goes by. Hence, the period from February 1, 1985 until January 27, 1989 serves as insample period to predict the returns for February 3, 1989. To be precise, the returns for February 3, 1989 denote the returns gained over the period from January 27, 1989 until February 3, 1989. In doing so, we run Algorithm 3.1.3 with $\zeta = 0.995$, $\xi = 10^{-4}$, $\bar{q} = \bar{p} = 2$, $V = 200{,}000$, $\nu = 0.95$ and $\sigma_p = 10\%$ p.a. As soon as we move forward, i.e., we forecast the returns for February 10, 1989, we run Algorithm 3.1.3 again. Although we keep the settings

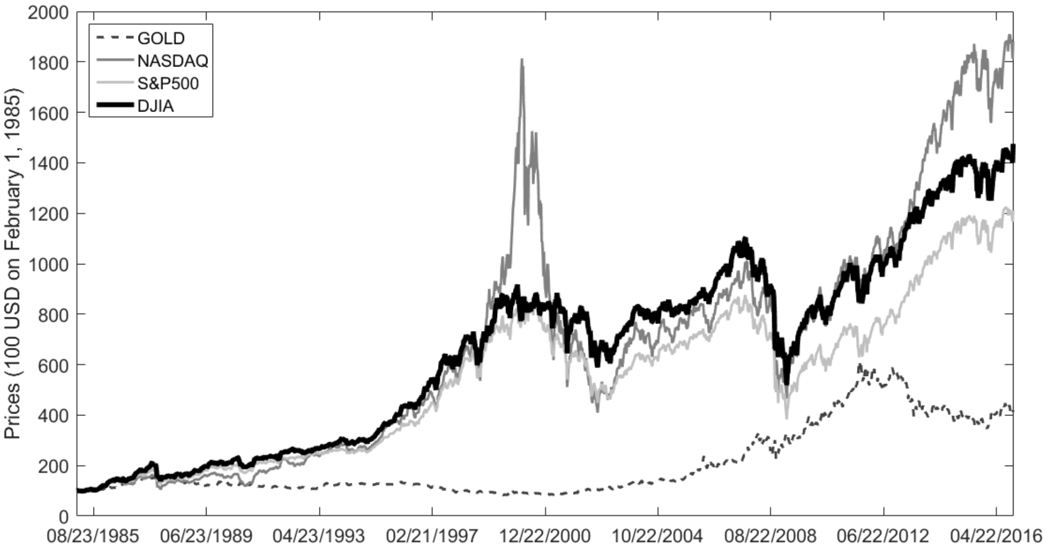

Figure 3.1: *Evolution of an initial investment of 100 USD in Gold, NASDAQ, S&P500 or DJIA over the period from February 1, 1985 until November 11, 2016 (Buy&Hold Strategies).*

for $\zeta, \xi, \tilde{q}, \tilde{p}, C, \nu$ and σ_p, our model selection procedure can provide new parameter dimensions.

A second fact that affects the parameter dimension is the dimension of the panel data. In total, we have 23 variables, however, some time series were first published after, e.g., January 27, 1989. That is, the respective panel data column of the insample period is missing and so, does not provide any information. To prevent us from empty columns, we check the panel data of each rolling window such that it comprises only time series with at least four observations. All in all, our out-of-sample period ranges from January 27, 1989 until November 11, 2016. For this purpose, Figure 3.2 shows the price paths of the single-index strategies and a $1/N$ Strategy, i.e., a Constant Mix Strategy, with weekly rebalancing. The idea behind Figure 3.2 is to highlight that all single-market strategies or the $1/N$ Strategy delivered less than 1400 USD for an initial investment of 100 USD on January 27, 1989. If we compare Figures 3.1 and 3.2, it follows that more than 400 USD of the final wealth of 1881.15 USD on November 11, 2016 in Figure 3.1 came from the NASDAQ performance over the insample period and its compound interest effects.

After the description of data and technical settings, we discuss our results. At the beginning, we claimed that the inclusion of mixed-frequency information may support asset allocation decisions. This is why we first focus on Figure 3.3, which illustrates the price paths of the $1/N$ Strategy and four mean-variance optimal portfolios for an initial investment of 100 USD on January 27, 1989. As mentioned before, we assume weekly rebalancing for the $1/N$ Strategy. The same holds for the mean-variance portfolios. That is, we use the information inherent in each rolling window to derive a mean-variance optimal asset allocation for the subsequent week. For simplicity reasons, all portfolios ignore transaction costs.

To support additional analyses, we distinguish the following four scenarios for estimating the expected returns and covariance matrices in case of mean-variance optimization: First, we add all available financial and macroeconomic time series to the panel data of the current rolling window, run Algorithm 3.1.3 and then, use the empirical means and covariance matrices of the predicted returns as estimates. Second, we pursue the same approach as in the first case, but restrict our panel data to financial time series. In this

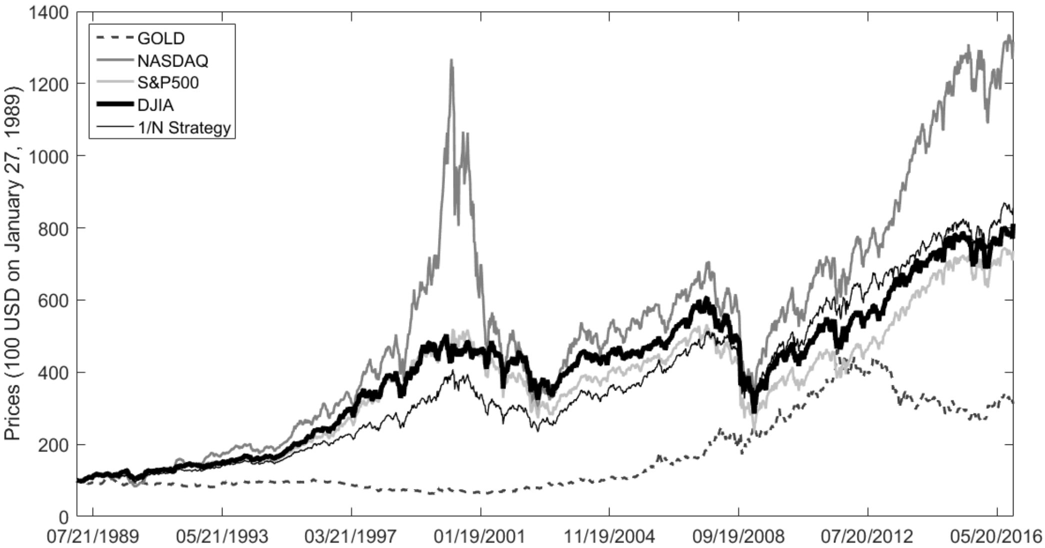

Figure 3.2: *Evolution of an initial investment of 100 USD in Gold, NASDAQ, S&P500, DJIA (Buy&Hold Strategies) or an 1/N Strategy with weekly rebalancing (constant mix) over the out-of-sample period from January 27, 1989 until November 11, 2016.*

way, we address the question whether macroeconomic variables shall be included. All financial time series are daily reported. However, when we update our scheme on Fridays after the close of trading and lack some observations, e.g., due to a public holiday, we keep this gap. Together with time series starting after February 1, 1985, this still causes incomplete panel data. Third, we assume a VAR(p), $1 \leq p \leq 2$, for the returns $\boldsymbol{r}_t$, modify Algorithm 3.1.3 accordingly and then, take the empirical means and covariance matrices of the forecasted VAR returns as estimates. Similar to the previous cases, we reestimate the lag order p for each new rolling window. Here, we have complete data, as there are no missing returns. The idea behind this scenario is to test, whether exogenous variables offer improvements. Fourth, for each rolling window the empirical means and covariance matrices of the historic returns serve as estimates. Hence, we skip all models to discuss whether the forecasting of return expectations and covariance matrices really pays off.

As shown in Figure 3.3, for mean-variance optimization the inclusion of exogenous variables, in particular, of finanicial and macroeconomic information offers an excess return. The $1/N$ Strategy has a final wealth of 855.80 USD, which is dominated by all mean-variance portfolios. In this context, the historic version provides 983.46 USD, the return VAR yields 1078.59 USD, the ESFM with incomplete financial panel data repays 1128.45 USD, while the ESFM with incomplete financial and macroeconomic panel data delivers 1223.37 USD.

In Figure 3.4, the same scenarios draw a different picture for marginal-risk-parity optimization, i.e., the $1/N$ Strategy with a final repayment of 855.80 USD outperforms all marginal-risk-parity portfolios. In addition, the historic approach exceeds with 514.11 USD the final value of the return VAR, which is given by 506.68 USD. The ESFM based versions still perform best, but the inclusion of macroeconomic variables could do more harm than good, since the ESFM with mixed-frequency financial and macroeconomic panel data finally provides 537.06 USD, whereas the restriction to incomplete financial data offers 537.79 USD.

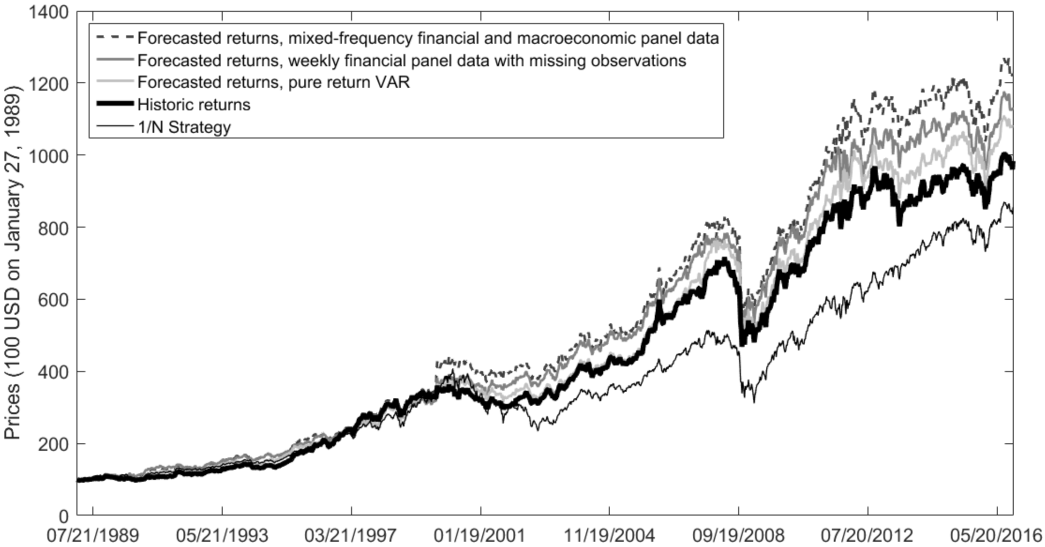

Figure 3.3: *Evolution of an initial investment of 100 USD in mean-variance optimal portfolios or an 1/N Strategy (constant mix) over the out-of-sample period from January 27, 1989 until November 11, 2016. All approaches involve weekly rebalancing of their asset allocations.*

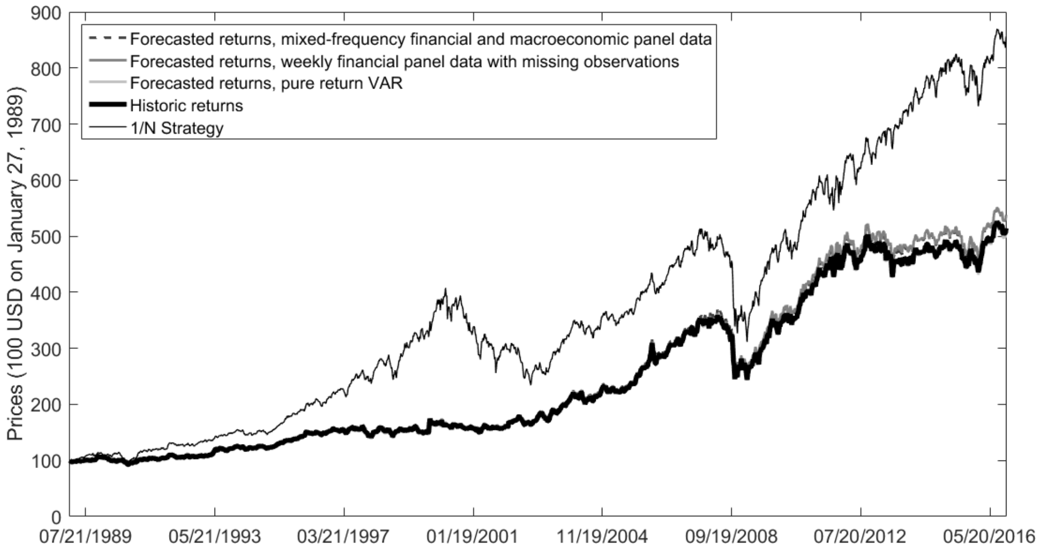

Figure 3.4: *Evolution of an initial investment of 100 USD in marginal-risk-parity optimal portfolios or an 1/N Strategy (constant mix) over the out-of-sample period from January 27, 1989 until November 11, 2016. All approaches involve weekly rebalancing of their asset allocations.*

Next, we dig deeper for the reasons behind the financial differences between mean-variance and marginal-risk-parity optimization. If we consider the asset allocations for mean-variance optimal portfolios in Figure 3.5, we conclude: First, all approches invest more or less the same ratio of the total wealth in Gold. Hence, the decomposition in Gold and stocks is the same for all. Second, each rolling window consists of 208 returns. Thus, migrating from one rolling window to the next replaces only the oldest of the 208 values, which enter the calculation of the empirical means and covariance matrices. In total, this explains why the asset allocations for historic means and covariance matrices (plot on the right bottom of Figure 3.5) are less volatile than the others. Third, the excess return of the mean-variance portfolios based on forecasted returns comes from changes in the stock exposure, which arise either from the return expectations or their covariance matrices. Note, there are no further inputs for mean-variance optimization.

In Figure 3.6, the portfolio weights of the marginal-risk-parity optimal portfolios are displayed. Irrespective of the considered scenario, we obtain almost the same asset allocation. On the one hand, this makes clear why the differences between these portfolios in Figure 3.4 were more or less negligible. On the other hand, it reveals that forecasted covariance matrices do not better describe the return dependencies than their historic counterparts. In case of mean-variance optimization, this implies that the means resulting from forecasted returns are the main driver of the excess performance.

To gain further insight into the financial properties of the presented portfolios, we have a look at Table 3.1, which lists some common performance and risk measures. As benchmarks for the mean-variance and marginal-risk-parity portfolios we added the single-market strategies and the $1/N$ Strategy. As before, all portfolios admit weekly rebalancing with zero transaction costs. At first glance, Table 3.1 confirms that NASDAQ offers the highest total log-return of 257.71% and the highest weekly volatility of 3.05%. The combination of both yields a Sharpe Ratio of 5.82%, which is the second smallest of all strategies. By contrast, the mean-variance optimal portfolio arising from the ESFM with macroeconomic and financial data has a total log-return of 250.42% and a standard deviation of 1.70%. Hence, it has a similar return, while its standard deviation is almost half in size. This is why its Sharpe Ratio of 10.18% is almost twice of the NASDAQ Sharpe Ratio of 5.82%. If we compare the four versions of mean-variance optimization, the inclusion of macroeconomic panel data improves the total log-return and so, the Sharpe Ratio. However, the restriction to financial data is still better than a pure return VAR. In our example, all forecasting-based mean-variance optimizations outperform the method with historic returns. Unfortunately, this conclusion does not hold for marginal-risk-parity optimization with Sharpe Ratios in the range of $[7.45, 7.69]$. Thus, the Sharpe Ratio of the historic approach of 7.52% is not always exceeded.

In the absence of a risk-free rate, the Omega Measure represents the ratio of the expected upside, i.e., the expectation of the postive part of the returns, and the expected downside, i.e., the expectation of their negative part. As shown in Table 3.1, the mean-variance portfolio based on financial and macroeconomic panel data offers an Omega Measure of 132.17, which is the overall highest. For mean-variance optimization, we get the same ranking for the Omega Measure as for the Sharpe Ratio. Again, there is no clear picture regarding marginal-risk-parity optimization.

By construction, marginal-risk-parity portfolios have the same marginal risk and so, behave rather conservatively. This assertion is supported by the fact that these portfolios have the lowest 95% VaR, 95% CVaR and Maximum Drowdown in Table 3.1. Unfortunately, there is no clear pattern in case of marginal-risk-parity optimization. For mean-variance optimization, we get 95% VaR, 95% CVaR and Maximum Drowdown figures, which are slightly worse than for the marginal-risk-parity portfolios, but a little bit better than for the $1/N$ Strategy. Because of diversification, it is natural that the 95% VaR, 95% CVaR and Maximum Drowdown of the single-market strategies exceed the ones of the portfolios by far. All in

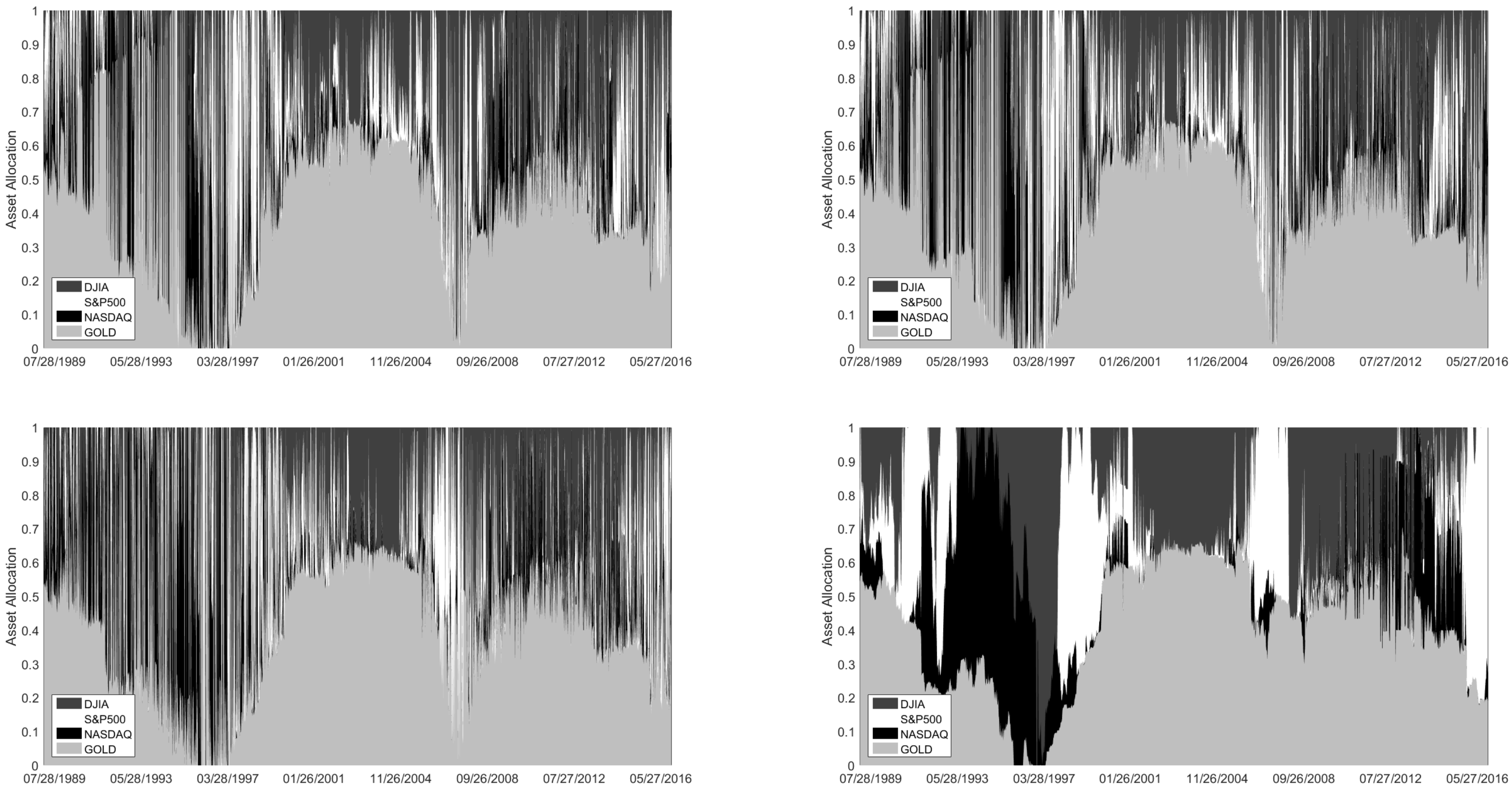

Figure 3.5: *Asset allocations for mean-variance optimal portfolios, where return expectations and covariance matrices are estimated by empirical means and covariance matrices based on: returns forecasted by an ESFM with incomplete panel data covering financial and macroeconomic time series (top, left), returns forecasted by an ESFM with incomplete financial panel data (top, right), returns forecasted by a pure return VAR(p) of dynamic order* $1 \leq p \leq 2$, *(bottom, left) and historic returns (bottom, right).*

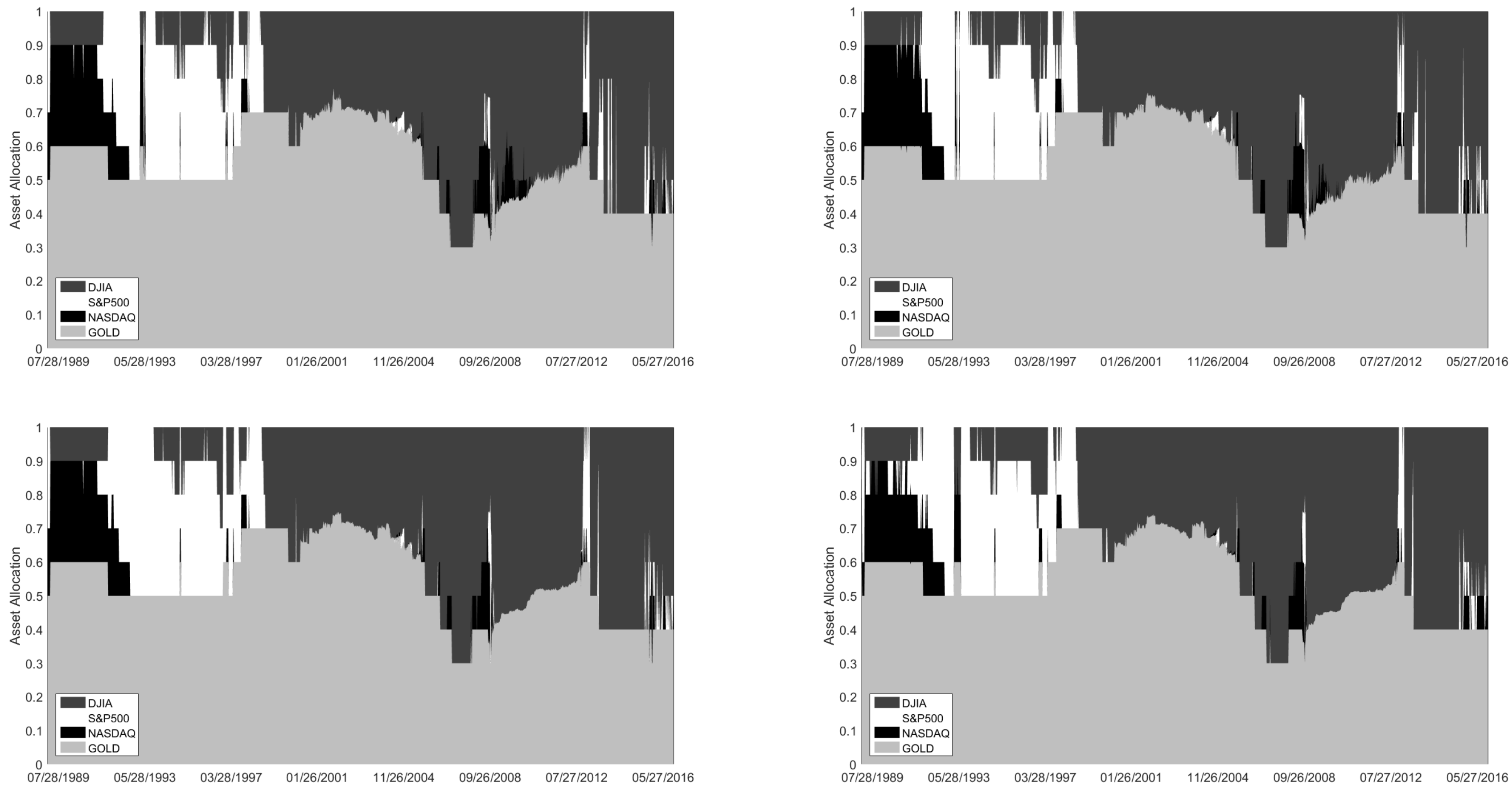

Figure 3.6: *Asset allocations for marginal-risk-parity optimal portfolios, where return expectations and covariance matrices are estimated by empirical means and covariance matrices based on: returns forecasted by an ESFM with incomplete panel data covering financial and macroeconomic time series (top, left), returns forecasted by an ESFM with incomplete financial panel data (top, right), returns forecasted by a pure return VAR(p) of dynamic order* $1 \leq p \leq 2$, *(bottom, left) and historic returns (bottom, right).*

Table 3.1: *Comparison of portfolio strategies for the out-of-sample period from January 27, 1989 until November 11, 2016*

	Indices				1/N	Mean-Variance Optimization				Marginal-Risk-Parity Optimization			
	Gold	NASDAQ	S&P500	DJIA		Forecasted			Hist.	Forecasted			Hist.
Panel Data						Fin.+Mac.	Fin.	Ret.		Fin.+Mac.	Fin.	Ret.	
Log-Return (Total, %)	113.88	**257.71**	199.70	209.36	214.69	250.42	242.34	237.82	228.59	168.09	168.23	162.27	163.73
Log-Return (Wkly., %)[a]	0.08	**0.18**	0.14	0.14	0.15	0.17	0.17	0.16	0.16	0.12	0.12	0.11	0.11
Std. Dev. (Wkly., %)[b]	2.21	3.05	2.28	2.24	1.84	1.70	1.68	1.66	1.68	1.51	1.51	**1.50**	**1.50**
Sharpe Ratio (Wkly., %)[c]	3.55	5.82	6.05	6.46	8.06	**10.18**	9.93	9.88	9.41	7.68	7.69	7.45	7.52
Omega (Wkly., %)[d]	110.36	117.86	118.32	119.44	124.82	**132.17**	131.06	130.48	128.40	123.55	123.56	122.67	122.87
95% VaR (Wkly., %)[e]	-3.37	-4.95	-3.55	-3.52	-2.90	-2.46	-2.36	-2.50	-2.56	**-2.18**	-2.20	**-2.18**	-2.23
95% CVaR (Wkly., %)[f]	-5.12	-7.46	-5.46	-5.22	-4.38	-3.79	-3.77	-3.79	-3.78	-3.43	-3.43	**-3.42**	**-3.42**
Max. Drawdown (%)[g]	-43.85	-77.42	-56.24	-52.98	-42.54	-35.23	-33.56	-36.07	-34.73	-32.86	**-31.74**	-32.38	-32.08

Bold figures highlight the best strategy of each category.

Note: The abbreviation “Fin.+Mac.” stands for mixed-frequency financial and macroeconomic panel data, while “Fin.” refers to weekly financial data with missing observations. By contrast, “Ret.” denotes a VAR restriced to complete, weekly returns of the four indices. In all three cases, we generate samples of predicted returns to derive return expectations and covariance matrices for our optimzation routines. With “Hist.” we highlight that historical returns, i.e., their empirical means and covariance matrices, serve as inputs of the given optimizations.

[a] The underlying period of time comprises 1450 weeks and hence, we have: Log-Return (Wkly.) = Log-Return (Total)/1450.

[b] As standard deviation the square root of the empirical variance, i.e., the squared deviation from the Log-Return (Wkly.) divided by 1449, is used.

[c] The Sharpe Ratio divides the expected excess return by its standard deviation. Here, our benchmark return is zero and so, we have : $\frac{\text{Log-Return (Wkly.)}}{\text{Stand. Deviation (Wkly.)}}$.

[d] The Omega Measure divides the upside by the downside of the expected excess returns, i.e., it is the ratio of the averaged positive and negative parts of Log-Return (Wkly.).

[e] For fixed time horizon and confidence level α, the Value at Risk reflects the maximal log-return that is not exceeded with probability $1-\alpha$. Mathematically, this means: 95% VaR (Wkly.) $= \sup\{r|\mathbb{P}(\text{Log-Return (Wkly.)} < r) \leq 0.05\}$.

[f] For fixed time horizon and confidence level α, the Conditional Value at Risk or Expected Shortfall is the expected return given that Log-Return (Wkly.) is below the α VaR, i.e.: 95% CVaR (Wkly.) $= \mathbb{E}[\text{Log-Return (Wkly.)}|\text{Log-Return (Wkly.)} < 95\% \text{ VaR (Wkly.)}]$.

[g] The Maximum Drawdown reveals the lowest discrete return, i.e., the highest loss, which could have been gained during the complete out-of-sample period.

all, the attractive upside with a tolerable downside let the mean-variance optimal portfolio derived from the ESFM with incomplete financial and macroeconomic panel data appear the best choice.

In Table 3.2, we take a closer look at the Sharpe Ratios in Table 3.1. That is, we use the test statistic from Jobson and Korkie (1981) to analyze whether the stated values are significantly different. Let $\tilde{\alpha}$ be the desired significance level. Then, we reject the null hypothesis for significance level $\tilde{\alpha}$, i.e., the Sharpe Ratios of two portfolios are equal, when it holds:

$$abs\left(\sqrt{T_o}\frac{\left(u_1\sqrt{u_4-u_3^2}-u_3\sqrt{u_2-u_1^2}\right)}{\hat{\sigma}_{\text{Sharpe}}}\right) > \Phi^{-1}\left(1-\tilde{\alpha}/2\right), \tag{3.23}$$

with $T_o > 0$ as the length of the out-of-sample period, i.e., $T_o = 1450$, and $\Phi^{-1}(\cdot)$ as the inverse of the cumulative distribution function of the standard normal distribution. Let $r_{t,1}$ and $r_{t,2}$ with $1 \leq t \leq T_o$ be the univariate returns of the two portfolios, whose Sharpe Ratios are compared. Then, the variables u_1, u_2, u_3 and u_4 are given by:

$$u_1 = \frac{1}{T_o}\sum_{t=1}^{T_o} r_{t,1}, \quad u_2 = \frac{1}{T_o}\sum_{t=1}^{T_o}\left(r_{t,1}\right)^2, \quad u_3 = \frac{1}{T_o}\sum_{t=1}^{T_o} r_{t,2}, \quad u_4 = \frac{1}{T_o}\sum_{t=1}^{T_o}\left(r_{t,2}\right)^2.$$

Let $\hat{\Sigma}_{\boldsymbol{u}} \in \mathbb{R}^{4\times 4}$ be the empirical covariance matrix of the vector $\left[r_{t,1}, r_{t,1}^2, r_{t,2}, r_{t,2}^2\right]' \in \mathbb{R}^4$. Furthermore, the vector $\boldsymbol{d}_{\text{Sharpe}} \in \mathbb{R}^4$ is defined as:

$$\boldsymbol{d}_{\text{Sharpe}} = \left[\sqrt{u_4-u_3^2}+\frac{u_3u_1}{\sqrt{u_2-u_1^2}}, -\frac{u_3}{2\sqrt{u_2-u_1^2}}, -\sqrt{u_2-u_1^2}-\frac{u_3u_1}{\sqrt{u_4-u_3^2}}, \frac{u_1}{2\sqrt{u_4-u_3^2}}\right]',$$

such that the standard deviation $\hat{\sigma}_{\text{Sharpe}}$ is given by:

$$\hat{\sigma}_{\text{Sharpe}} = \sqrt{\left(\boldsymbol{d}'_{\text{Sharpe}}\right)\hat{\Sigma}_{\boldsymbol{u}}\ \boldsymbol{d}_{\text{Sharpe}}}.$$

Using the test statistic in (3.23), Table 3.2 confirms that the Sharpe Ratios of the mean-variance portfolios and single-market strategies Gold, NASDAQ and S&P500 are significantly different for levels of 5% or 10%. Moreover, there are significant differences between the Sharpe Ratios of mean-variance and marginal-risk-parity portfolios. Unfortunately, there are no significant differences among the mean-variance portfolios or between the mean-variance portfolios and the $1/N$ Strategy.

For the significance of the Omega values in Table 3.1, we apply the test statistic in Schmid and Schmidt (2008). As before, let $\tilde{\alpha}$ be the desired significance level. Then, we reject the null hypothesis for significance level $\tilde{\alpha}$, i.e., the Omega Measures of two portfolios are equal, as soon as we receive:

$$abs\left(\sqrt{T_o}\frac{\left(v_1/v_2-v_3/v_4\right)}{\hat{\sigma}_{\text{Omega}}}\right) > \Phi^{-1}\left(1-\tilde{\alpha}/2\right), \tag{3.24}$$

where T_o is the length of the out-of-sample period. Let $r_{t,1}$ and $r_{t,2}, 1 \leq t \leq T_o$, be the portfolio returns to be compared. Then, for a risk-free rate of zero, the terms v_1, v_2, v_3 and v_4 are defined as follows:

$$v_1 = \frac{1}{T_o}\sum_{t=1}^{T_o}\max\left(r_{t,1},0\right), \quad v_2 = \frac{1}{T_o}\sum_{t=1}^{T_o}\max\left(-r_{t,1},0\right),$$
$$v_3 = \frac{1}{T_o}\sum_{t=1}^{T_o}\max\left(r_{t,2},0\right), \quad v_4 = \frac{1}{T_o}\sum_{t=1}^{T_o}\max\left(-r_{t,2},0\right).$$

For the vector $\left[\max\left(r_{t,1},0\right), \max\left(-r_{t,1},0\right), \max\left(r_{t,2},0\right), \max\left(-r_{t,2},0\right)\right]' \in \mathbb{R}^4, 1 \leq t \leq T_o$, let $\hat{\Sigma}_{\boldsymbol{v}} \in \mathbb{R}^{4\times 4}$ be its empirical covariance matrix and assume $\boldsymbol{d}_{\text{Omega}} = \left[1/v_2, -v_1/(v_2^2), 1/v_4, v_3/(v_4^2)\right]' \in \mathbb{R}^4$. Then, the

Table 3.2: *Test statistic of Jobson and Korkie (1981) for Sharpe Ratios*

	Indices				1/N	Mean-Variance Optimization				Marginal-Risk-Parity Optimization			
	Gold	NASDAQ	S&P500	DJIA		Forecasted			Hist.	Forecasted			Hist.
						Fin.+Mac.	Fin.	Ret.		Fin.+Mac.	Fin.	Ret.	
Gold	x	0.5969	0.6500	0.7426	1.3875	2.6615	2.5951	2.5284	2.3479	1.9645	1.9695	1.8836	1.9108
NASDAQ	x	x	0.1568	0.3272	1.7702	1.7681	1.6563	1.7192	1.5730	0.6947	0.6961	0.6053	0.6336
S&P500	x	x	x	0.4814	1.9921	1.9133	1.7754	1.8239	1.6539	0.6965	0.6977	0.5900	0.6242
DJIA	x	x	x	x	1.1816	1.6179	1.4977	1.5251	1.3486	0.5080	0.5095	0.4063	0.4396
$1/N$	x	x	x	x	x	1.3126	1.1461	1.2053	0.9645	0.2132	0.2120	0.3475	0.3070
Fin.+Mac.	x	x	x	x	x	x	0.4014	0.3356	0.7616	2.2558	2.2522	2.4420	2.3866
Fin.	x	x	x	x	x	x	x	0.0556	0.4927	2.0543	2.0617	2.2467	2.1862
Ret.	x	x	x	x	x	x	x	x	0.5438	1.9328	1.9310	2.1324	2.0655
Hist.	x	x	x	x	x	x	x	x	x	1.5019	1.5032	1.7169	1.6596
Fin.+Mac.	x	x	x	x	x	x	x	x	x	x	0.0243	1.4566	0.8499
Fin.	x	x	x	x	x	x	x	x	x	x	x	1.3603	0.8456
Ret.	x	x	x	x	x	x	x	x	x	x	x	x	0.5256
Hist.	x	x	x	x	x	x	x	x	x	x	x	x	x

Values marked in dark gray are significant for level 5% (test statistic: 1.96), while light gray ones are significant for level 10% (test statistic: 1.64).

standard deviation $\hat{\sigma}_{\text{Omega}}$ is given by:

$$\hat{\sigma}_{\text{Omega}} = \sqrt{\left(\boldsymbol{d}'_{\text{Omega}}\right) \hat{\Sigma}_{\boldsymbol{v}}\, \boldsymbol{d}_{\text{Omega}}}.$$

Based on the test statistic in (3.24), Table 3.3 shows that only Gold and the mean-variance portfolios are significantly different for levels of 5% or 10%.

Next, we consider some statistical properties of our ansatz. As illustrated in Figure 3.7, prediction intervals derived from an ESFM with incomplete financial panel data are wider than those of the return VAR. Similarly, the inclusion of macroeconomic information widens the prediction intervals once again. Besides the size of the prediction intervals, the underlying data also affects the factor dimension and lag lengths (see Figure 3.8). However, since the additional macroeconomic variables enlarge the panel data dimension, it is reasonable that more factors are required to map 99.5% of its variation. Except for a few outliers and the period after the financial crisis in 2008, the orders of return and factor lags are one. This indicates that the preprocessed panel data is little autoregressive. For the return histograms in Figure 3.9, there is neither a clear pattern for mean-variance portfolios nor for marginal-risk-parity portfolios.

Finally, we focus on Table 3.4, which lists some statistical figures. In case of the Root-Mean-Square Error (RMSE), which is specified in Definition A.3.2, the empirical means of the forecasted returns serve as point estimates. Despite the good financial characteristics, the inclusion of financial and macroeconomic data deteriorates the RMSE, which is surprising. Although exogenous information in the form of factors widens the prediction intervals, the Ratio of Interval Outlierss (RIOs) in Definition A.3.3 does not automatically decrease. The Mean Interval Score (MIS) in Definition A.3.4 takes the size of the prediction intervals and number of outliers into account, but still does not show any improvements for the inclusion of exogenous variables. So far, we considered each index separately. Therefore, let $\hat{\boldsymbol{r}}_t \in \mathbb{R}^H$ be the forecasted return of the afterwards realized return $\boldsymbol{r}_t \in \mathbb{R}^H$ for an out-of-sample period of length T_o. Then, we introduce the multivariate RMSE as follows:

$$\text{mRMSE} = \sqrt{\frac{1}{HT_o} \sum_{t=1}^{T_o} \left(\hat{\boldsymbol{r}}_t - \boldsymbol{r}_t\right)' \left(\hat{\boldsymbol{r}}_t - \boldsymbol{r}_t\right)}. \tag{3.25}$$

Similar to the RMSE, financial and macroeconomic information worsens the mRMSE in Table 3.4 and so, we are still missing an explanation why the proposed framework works from a statistical perspective. For this purpose, we define a new measure called Ranking Error (RE). For forecasted returns $\hat{\boldsymbol{r}}_t \in \mathbb{R}^H$ and afterwards realized returns $\boldsymbol{r}_t \in \mathbb{R}^H$, let the vectors $\hat{\boldsymbol{o}}_t \in \mathbb{R}^H$ and $\boldsymbol{o}_t \in \mathbb{R}^H$ contain the indices of the order statistics of $\hat{\boldsymbol{r}}_t$ and $\boldsymbol{r}_t$. For instance, for $\hat{\boldsymbol{r}}_t = [-0.01, 0.02, -0.05, 0.03]$ and $\boldsymbol{r}_t = [-0.02, 0.03, -0.03, 0]$ we have: $\hat{\boldsymbol{o}}_t = [3, 1, 2, 4]$ and $\boldsymbol{o}_t = [3, 1, 4, 2]$, since the smallest element of $\hat{\boldsymbol{r}}_t$ is the third, the second smallest one is the first, etc. Based on this, the RE is given by:

$$\text{RE} = \sqrt{\frac{1}{HT_o} \sum_{t=1}^{T_o} \left(\hat{\boldsymbol{o}}_t - \boldsymbol{o}_t\right)' \left(\hat{\boldsymbol{o}}_t - \boldsymbol{o}_t\right)}. \tag{3.26}$$

As the RE in Table 3.4 slightly improves, when financial and macroeconomic data is added, the outperformance of our mean-variance portfolios in Table 3.1 possibly comes from the ranking of the forecasted returns. That is, our approach more often predicts the correct return order. As mean-variance optimization maximizes the expected return subject to a variance constraint, for similar covariance matrices, the correct return order causes overweighted outperforming assets and so, yields an excess return. Note, if a vector comprises the same element twice, the indices of its order statistics are ambiguous. However, when working with real data, this occurs quite unlikely, if no rounding takes place.

Table 3.3: *Test statistic of Schmid and Schmidt (2008) for Omega Measures*

	Indices				1/N	Mean-Variance Optimization				Marginal-Risk-Parity Optimization			
	Gold	NASDAQ	S&P500	DJIA		Forecasted			Hist.	Forecasted			Hist.
						Fin.+Mac.	Fin.	Ret.		Fin.+Mac.	Fin.	Ret.	
Gold	x	0.7234	0.7683	0.8827	1.3987	2.1625	2.0629	2.0511	1.8713	1.3247	1.3237	1.2463	1.2701
NASDAQ	x	x	0.0490	0.1664	0.7404	1.4149	1.3094	1.2683	1.0823	0.5640	0.5651	0.4775	0.4976
S&P500	x	x	x	0.1219	0.6943	1.3990	1.2951	1.2434	1.0421	0.5316	0.5328	0.4419	0.4617
DJIA	x	x	x	x	0.5716	1.2846	1.1819	1.1280	0.9241	0.4194	0.4208	0.3293	0.3494
$1/N$	x	x	x	x	x	0.7333	0.6244	0.5746	0.3677	0.1261	0.1255	0.2142	0.1945
Fin.+Mac.	x	x	x	x	x	x	0.1089	0.1687	0.3768	0.8420	0.8409	0.9276	0.9100
Fin.	x	x	x	x	x	x	x	0.0578	0.2677	0.7395	0.7387	0.8259	0.8082
Ret.	x	x	x	x	x	x	x	x	0.2147	0.6939	0.6930	0.7825	0.7644
Hist.	x	x	x	x	x	x	x	x	x	0.4917	0.4908	0.5822	0.5630
Fin.+Mac.	x	x	x	x	x	x	x	x	x	x	0.0007	0.0891	0.0692
Fin.	x	x	x	x	x	x	x	x	x	x	x	0.0897	0.0699
Ret.	x	x	x	x	x	x	x	x	x	x	x	x	0.0202
Hist.	x	x	x	x	x	x	x	x	x	x	x	x	x

Values marked in dark gray are significant for level 5% (test statistic: 1.96), while light gray ones are significant for level 10% (test statistic: 1.64).

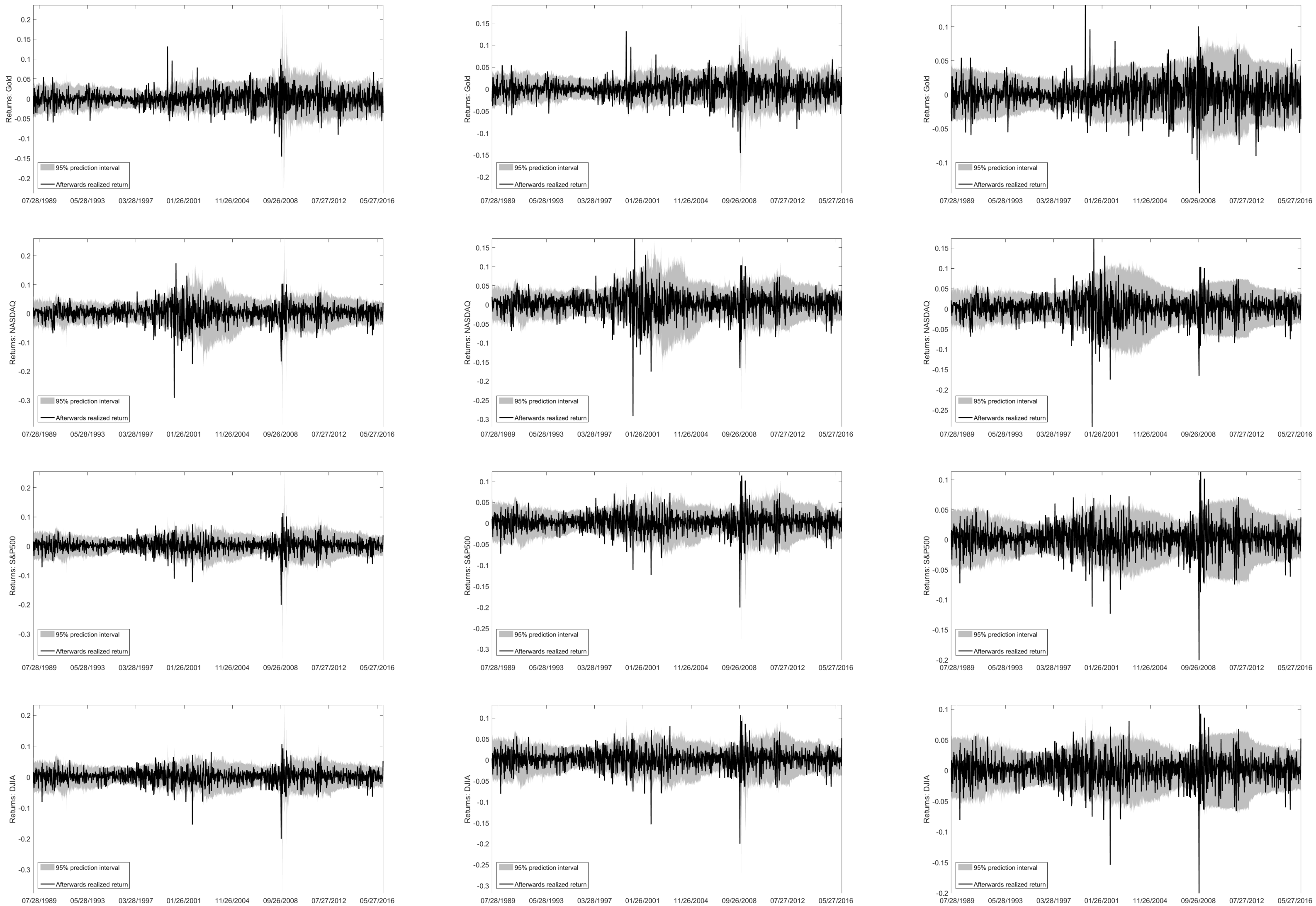

Figure 3.7: *95%-Prediction intervals (gray area) and afterwards realized returns (black line) for Gold, NASDAQ, S&P500 and DJIA. The first column displays returns forecasted by an ESFM with incomplete panel data covering financial and macroeconomic time series. The second column illustrates returns forecasted by an ESFM with incomplete financial data. The third column shows returns forecasted by a pure return VAR(p) of dynamic order* $1 \leq p \leq 2$.

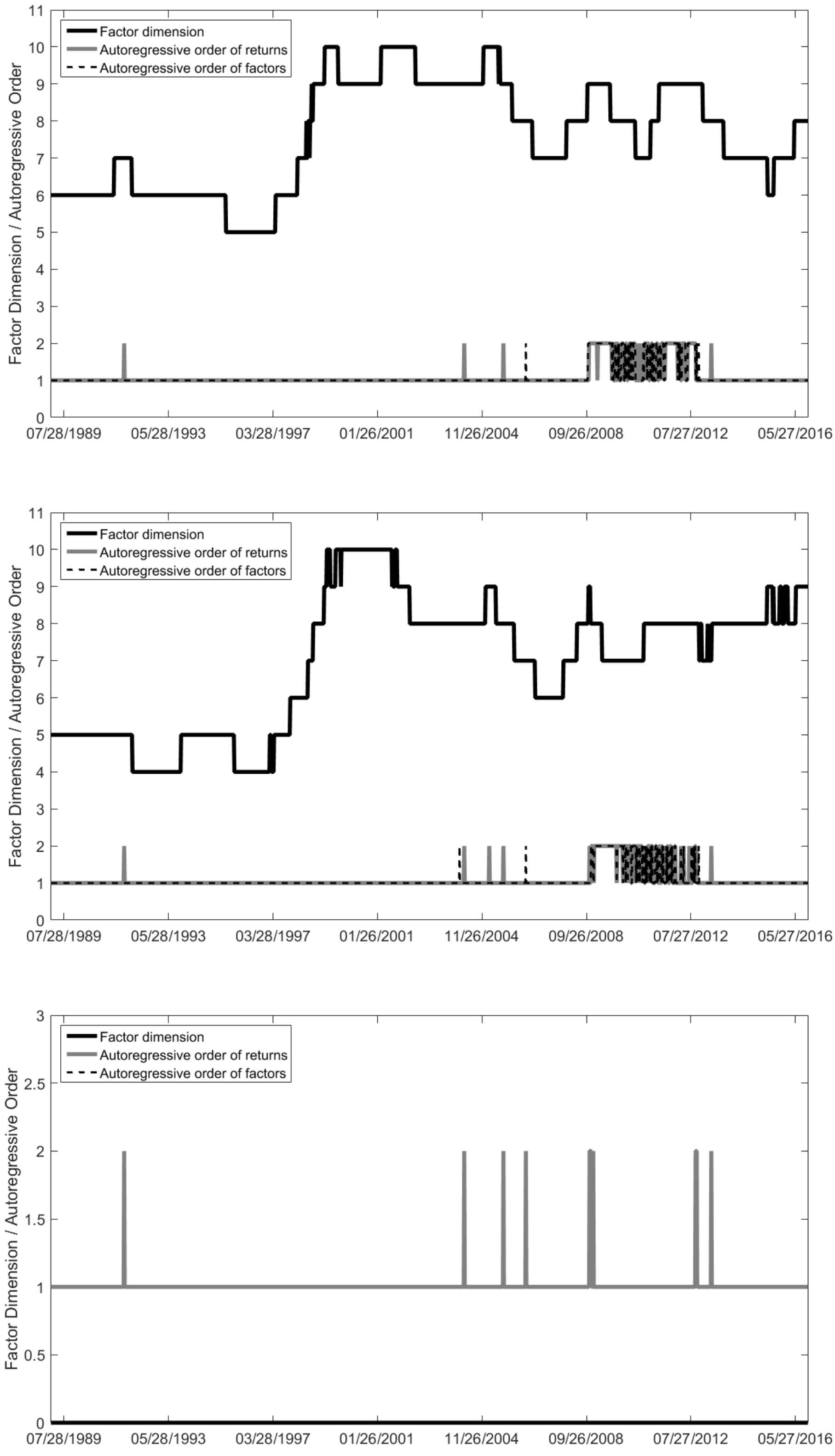

Figure 3.8: *Factor dimension and autoregressive orders for an ESFM with incomplete panel data covering financial and macroeconomic time series (top), for an ESFM with incomplete financial panel data (middle) and for a pure return VAR(p) of dynamic order* $1 \leq p \leq 2$.

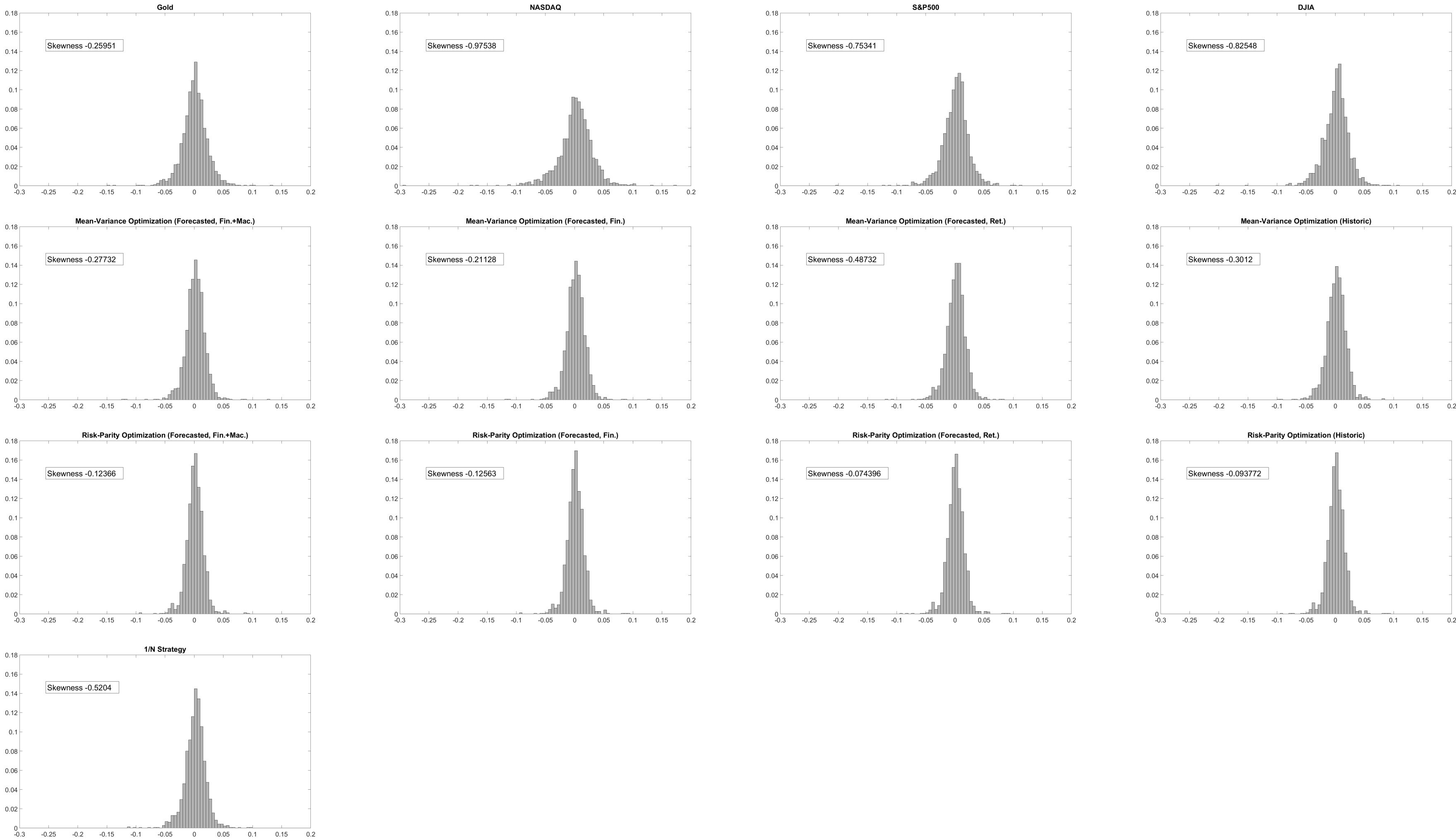

Figure 3.9: *Normalized histograms for returns of single-market strategies (first row) and portfolios (remaining rows). For mean-variance and marginal-risk-parity portfolios, expected returns and covariance matrices are estimated by the empirical means and covariance matrices of: returns forecasted by an ESFM with incomplete panel data covering financial and macroeconomic time series (first column), returns forecasted by an ESFM with incomplete financial panel data (second column), returns forecasted by a pure return VAR(p) of dynamic order* $1 \leq p \leq 2$ *(third column) and historic returns (last column).*

Table 3.4: *Estimation measures based on the out-of-sample period from January 27, 1989 until November 11, 2016*

		Panel Data		
		Financial and Macroeconomic	Financial	Returns
RMSE	Gold	0.0240	0.0239	**0.0227**
	NASDAQ	0.0339	0.0330	**0.0314**
	S&P500	0.0260	0.0249	**0.0234**
	DJIA	0.0255	0.0244	**0.0229**
RIO	Gold	0.0669	0.0655	**0.0607**
	NASDAQ	**0.0510**	0.0448	**0.0510**
	S&P500	0.0745	0.0724	**0.0628**
	DJIA	0.0690	**0.0648**	**0.0648**
MIS	Gold	0.1243	0.1244	**0.1211**
	NASDAQ	0.1792	0.1735	**0.1704**
	S&P500	0.1404	0.1342	**0.1288**
	DJIA	0.1384	0.1323	**0.1261**
mRMSE		0.0276	0.0268	**0.0254**
RE		**1.4376**	1.4389	1.4391

Bold figures highlight the best value of each category. That is, for each row the lowest RMSE, MIS, mRMSE and RE as well as RIO closest to 5% are marked in bold.

3.3 Conclusion and Future Research

In this chapter, we developed a comprehensive framework, which constructs prediction intervals for returns of future periods of time and supports portfolio optimization. Thereby, the alternating application of MLE and an EM estimates an ESFM with incomplete panel data. That is, in each loop the EM derives complete panel data from the observations and latest parameter estimates, before MLE reestimates the parameters. To ensure that parameter estimates and complete panel data fit together, this method runs until the change in the expected log-likelihood function is negligible. Besides the parameters, this ansatz estimates missing observations and provides timely indications for low-frequency variables (nowcasting). The underlying factor structure allows us to describe large-dimensional panel data by a few hidden factors with known distribution and thus, admits a reduction in dimension. Next, a VARX with the factors as exogenous variables maps the dynamics of a multivariate return process. At this stage, we use randomly drawn factor and coefficient samples instead of their estimates to generate samples of forecasted returns. In this way, we take into account parameter risks and uncertainties arising from Factor Analysis. Finally, such samples of predicted returns serve as starting point of our financial application. That means, they facilitate the construction of empirical prediction intervals. Furthermore, their empirical means and covariance matrices enter our mean-variance and marginal-risk-parity, respectively, portfolio optimizations.

Our empirical study considers an investment universe consisting of Gold and three US stock indices, i.e., Nasdaq Composite, Standard & Poor's 500 and Dow Jones Industrial Average. Besides financial figures, our panel data includes macroeconomic indicators. In total, a mixture of monthly, quarterly, discontinued and later starting time series entails incompleteness of the panel data. Among other things, the ability to work with ragged data is a main advantage of our approach. In particular, for applications in the area of asset and risk management this is an important feature. Based on our empirical example, we recommend our method for mean-variance portfolio optimization with incomplete financial and macroeconomic panel data, since it provided the best Sharpe Ratio and Omega Measure of all benchmark strategies.

This chapter contributes to the existing literature by showing how Factor Models for unbalanced panel data can support the monitoring of financial markets and portfolio decisions. The single building blocks of our framework were thoroughly discussed by the statistical community, but were primarily applied to non-financial data. Hence, the manner in which we combine them here and the chosen field of application, i.e., market monitoring and portfolio optimization, are the main novelties of this chapter.

Possible directions of the future research are as follows: First, the empirical analysis could be extended to other financial markets. For instance, non-US stock indices, Foreign Exchange rates, fixed income markets or commodity prices could be interesting. Second, instead of an ESFM an ADFM could be used, which we do in Chapter 4. For autoregressive factors, the preprocessing of the panel data is not obliged to remove their autoregressive nature. Third, the VARX can be replaced by the FAVAR of Bernanke et al. (2005) to investigate the impact of the factors on the returns as well as the reverse relation. Fourth, during crises the behavior of stocks is considerably different to normal times. For instance, they are strongly correlated and skewed. Similar to Hauptmann et al. (2014), regime switching in case of the return process possibly improves the performance of the overall framework.

Chapter 4

Estimation of Approximate Dynamic Factor Models

In this chapter, we estimate Approximate Dynamic Factor Models (ADFMs) with homoscedastic, cross-sectionally correlated shocks using incomplete panel data. Besides missing observations, data incompleteness systematically arises from the inclusion of mixed-frequency information. The factor dynamics obeys a Vector Autoregression Model of order $p \geq 0$. Unlike similar approaches such as in Bańbura and Modugno (2014), we alternately deploy two Expectation-Maximization Algorithms (EMs) for parameter estimation instead of a mix of an EM, the Kalman Filter and Smoother. In doing so, we trace estimated factor means back to the input data such that we know the contribution of each input signal to the expected result. On the one hand, this shows us which are the main drivers and so, the main sources of risk. On the other hand, it opens the door for further investigations and hence, indicates how reliable the received output is. For model selection, i.e., for determing the unknown factor dimension and autoregressive order, we derive a two-step model selection criterion, which we test within a comprehensive Monte Carlo simulation study. Finally, we apply the ADFM to real-economy data supporting investment decisions and risk management. In this context, an Autoregressive Model with estimated factors as exogenous variables maps the behavior of weekly S&P500 log-returns. As in Chapter 3, we construct prediction intervals for returns of future periods of time. In addition, we detect the drivers of our point forecasts and suggest a dynamic trading strategy to benefit from the gained information.

4.1 Mathematical Background

In Chapter 3, the factors were supposed to be identically and independently distributed and we assumed an isotropic error model for the idiosyncratic shocks. By contrast, we now allow for cross-sectionally and serially correlated factors. Furthermore, we admit cross-sectional correlation of the idiosyncratic errors. For this purpose, we work with the ADFM in Definition 2.1.4. As before, we first consider complete panel data and later on relax this restriction to treat unbalanced panel data, too.

4.1.1 Estimation of ADFMs with Complete Panel Data

Here, we assume for all points in time t the vector $\boldsymbol{X}_t \in \mathbb{R}^N$ as in Definition 2.1.1, that is, there are no missing entries. Let us recall the measurement and transition equations of the ADFM in Definition 2.1.4. If the joint vector $(\boldsymbol{F}_t', \boldsymbol{X}_t')' \in \mathbb{R}^{K+N}$ is observable, standard MLE can provide estimates for the ADFM parameters Θ. However, the hiddenness of the factors $\boldsymbol{F}_t \in \mathbb{R}^K$ calls for another approach. Namely, to treat the log-likelihood function $\mathcal{L}\left(\Theta|X, F\right)$ with latent factors $\boldsymbol{F}_t$, we have to estimate them. For this purpose, we replace the log-likelihood function by its expectation conditioned on the observed panel data $X \in \mathbb{R}^{T\times N}$. In this way, we intend to switch from the factors to their means and covariance matrices conditioned on the panel data, i.e., we deal with the moments of the random vectors $\boldsymbol{F}_t|\boldsymbol{X}_t$. Due to normality assumptions for the factors $\boldsymbol{F}_t$ and shocks $\boldsymbol{\epsilon}_t \in \mathbb{R}^N$ and $\boldsymbol{\delta}_t \in \mathbb{R}^K$, the vectors $\boldsymbol{F}_t|\boldsymbol{X}_t$ are still Gaussian and the switch from the factors to their conditional means and covariance matrices is possible. Furthermore, their moments arise from the Gaussian distribution of the vectors $\boldsymbol{X}_t|\boldsymbol{F}_t$ and the marginal distributions of the covariance-stationary vectors $\boldsymbol{F}_t$ by virtue of the Bayes' theorem. This is why the below Lemmata exactly pursue this roadmap. All in all, we adjust our estimation problem such that the EM in Definition 2.1.7 simultaneously estimates the hidden factors and unknown model parameters.

Lemma 4.1.1 (Conditional Distributions of ADFMs)
For the ADFMs in Definition 2.1.4, the conditional distributions $\boldsymbol{X}_t|\boldsymbol{F}_t$ and $\boldsymbol{F}_t|\boldsymbol{F}_{t-1},\dots,\boldsymbol{F}_{t-p}$ are multivariate Gaussian as follows:

$$\boldsymbol{X}_t|\boldsymbol{F}_t \sim \mathcal{N}\left(W\boldsymbol{F}_t + \boldsymbol{\mu}, \Sigma_{\boldsymbol{\epsilon}}\right), \tag{4.1}$$

$$\boldsymbol{F}_t|\boldsymbol{F}_{t-1},\dots,\boldsymbol{F}_{t-p} \sim \mathcal{N}\left(\sum_{i=1}^{p} A_i\boldsymbol{F}_{t-i}, \Sigma_{\boldsymbol{\delta}}\right). \tag{4.2}$$

Proof:
Because of the underlying normality assumptions, the random vectors $\boldsymbol{X}_t|\boldsymbol{F}_t$ and $\boldsymbol{F}_t|\boldsymbol{F}_{t-1},\dots,\boldsymbol{F}_{t-p}$ are multivariate Gaussian. Hence, we can restrict ourselves to determing their means and covariance matrices. Besides (2.4), the independence of $\boldsymbol{F}_t$ and $\boldsymbol{\epsilon}_t$ provides:

$$\begin{aligned}\mathbb{E}_\Theta\left[\boldsymbol{X}_t|\boldsymbol{F}_t\right] &= \mathbb{E}_\Theta\left[W\boldsymbol{F}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t|\boldsymbol{F}_t\right] = W\boldsymbol{F}_t + \boldsymbol{\mu} + \mathbb{E}_\Theta\left[\boldsymbol{\epsilon}_t|\boldsymbol{F}_t\right] = W\boldsymbol{F}_t + \boldsymbol{\mu},\\ \mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{X}_t|\boldsymbol{F}_t\right] &= \mathbb{V}\mathrm{ar}_\Theta\left[W\boldsymbol{F}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t|\boldsymbol{F}_t\right] = \mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{\epsilon}_t|\boldsymbol{F}_t\right] = \Sigma_{\boldsymbol{\epsilon}}.\end{aligned}$$

Similarly, (2.5) and the independence of $\boldsymbol{F}_{t-1},\dots,\boldsymbol{F}_{t-p}$ and $\boldsymbol{\delta}_t$ yield:

$$\begin{aligned}\mathbb{E}_\Theta\left[\boldsymbol{F}_t|\boldsymbol{F}_{t-1},\dots,\boldsymbol{F}_{t-p}\right] &= \mathbb{E}_\Theta\left[\sum_{i=1}^{p} A_i\boldsymbol{F}_{t-i} + \boldsymbol{\delta}_t \middle| \boldsymbol{F}_{t-1},\dots,\boldsymbol{F}_{t-p}\right]\\ &= \sum_{i=1}^{p} A_i\boldsymbol{F}_{t-i} + \mathbb{E}_\Theta\left[\boldsymbol{\delta}_t|\boldsymbol{F}_{t-1},\dots,\boldsymbol{F}_{t-p}\right] = \sum_{i=1}^{p} A_i\boldsymbol{F}_{t-i},\\ \mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{F}_t|\boldsymbol{F}_{t-1},\dots,\boldsymbol{F}_{t-p}\right] &= \mathbb{V}\mathrm{ar}_\Theta\left[\sum_{i=1}^{p} A_i\boldsymbol{F}_{t-i} + \boldsymbol{\delta}_t \middle| \boldsymbol{F}_{t-1},\dots,\boldsymbol{F}_{t-p}\right]\\ &= \mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{\delta}_t|\boldsymbol{F}_{t-1},\dots,\boldsymbol{F}_{t-p}\right] = \Sigma_{\boldsymbol{\delta}}.\end{aligned}$$

$\square$

Before we focus on the marginal distributions of the covariance-stationary vectors $\boldsymbol{X}_t$ and $\boldsymbol{F}_t$, we derive the covariance matrix of the factors $\boldsymbol{F}_t$. Thereby, the covariance-stationarity of the process $\{\boldsymbol{F}_t\}$ justifies the following infinite moving average, abbreviated by MA(∞), representation of $\boldsymbol{F}_t$ (Hamilton, 1994, p. 260, Eq. 10.1.15):

$$\boldsymbol{F}_t = \sum_{k=0}^{\infty} \mathbb{A}^k \boldsymbol{\delta}_{t-k}, \tag{4.3}$$

with matrices $\mathbb{A}^k \in \mathbb{R}^{K\times K}, k \geq 0$, as in (4.6). Note, the matrix sequence $\{\mathbb{A}^k\}$ is absolutely summable, that is, it holds for any $1 \leq i, j \leq K$ (Hamilton, 1994, p. 262-263):

$$\sum_{k=0}^{\infty} \left| \mathbb{A}^k_{(i,j)} \right| < \infty, \tag{4.4}$$

where $\mathbb{A}^k_{(i,j)}$ stands for the element in the i-th row and j-th column of matrix $\mathbb{A}^k$. With this in mind, we obtain the following covariance matrix of $\boldsymbol{F}_t$:

Lemma 4.1.2 (Covariance Matrix of Factors $\boldsymbol{F}_t$)
The covariance matrix of the covariance-stationary VAR(p) process $\{\boldsymbol{F}_t\}$ in (2.5) satisfies:

$$\Sigma_{\boldsymbol{F}} = \sum_{k=0}^{\infty} \left(\left(\mathbb{A}^k\right) \Sigma_{\boldsymbol{\delta}} \left(\mathbb{A}^k\right)' \right), \tag{4.5}$$

where for all $k \geq 1$ we define:

$$\mathbb{A}^k = \begin{bmatrix} A_1, \dots, A_p \end{bmatrix} \begin{bmatrix} \mathbb{A}^{k-1} \\ \vdots \\ \mathbb{A}^{k-p} \end{bmatrix} \quad \textit{with} \;\; \mathbb{A}^0 = I_K \;\; \textit{and} \;\; \mathbb{A}^{k-p} = O_K, \;\; \forall\,(k-p) < 0. \tag{4.6}$$

Proof:
For all points in time t, we have: $\boldsymbol{\delta}_t \sim \mathcal{N}\left(\mathbf{0}_K, \Sigma_{\boldsymbol{\delta}}\right)$ iid such that the dynamics of $\boldsymbol{F}_t$ in (2.5) coincides with the VAR(p) in Hamilton (1994, p. 258, Eq. 10.1.4-10.1.6). For the obtained MA(∞) representation, see Hamilton (1994, p. 260, Eq. 10.1.15) with matrices $\mathbb{A}^k$ as in Hamilton (1994, p. 260, Eq. 10.1.19). Finally, point *(a)* in Hamilton (1994, p. 263, Proposition 10.2) for $s = 0$ proves the claim. □

Next, we have for the marginal distributions of the covariance-stationary vectors $\boldsymbol{X}_t$ and $\boldsymbol{F}_t$.

Lemma 4.1.3 (Marginal Distributions of ADFMs)
Assume the ADFMs in Definition 2.1.4. Then, the covariance-stationary processes $\{\boldsymbol{F}_t\}$ and $\{\boldsymbol{X}_t\}$ are multivariate Gaussian. For any time t, their marginal distributions are given by:

$$\boldsymbol{F}_t \sim \mathcal{N}\left(\mathbf{0}_K, \Sigma_{\boldsymbol{F}}\right), \tag{4.7}$$

$$\boldsymbol{X}_t \sim \mathcal{N}\left(\boldsymbol{\mu}, W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}\right), \tag{4.8}$$

with the covariance matrix $\Sigma_{\boldsymbol{F}} \in \mathbb{R}^{K\times K}$ defined in Lemma 4.1.2.

Proof:
In this Gaussian setting, the independence of the factors $\boldsymbol{F}_t$ and shocks $\boldsymbol{\epsilon}_t$ ensures the normality of the marginal distributions. This is why we just determine the respective expectations and covariance matrices.

The claim $\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{F}_t\right] = \Sigma_{\boldsymbol{F}}$ is proven in Lemma 4.1.2. Similarly, the absolute summability of the MA(∞) representation and the properties of the white noise process $\{\boldsymbol{\delta}_t\}$ provide:

$$\mathbb{E}_\Theta\left[\boldsymbol{F}_t\right] = \mathbb{E}_\Theta\left[\sum_{k=0}^{\infty} \mathbb{A}^k \boldsymbol{\delta}_{t-k}\right] = \sum_{k=0}^{\infty} \mathbb{A}^k \mathbb{E}_\Theta\left[\boldsymbol{\delta}_{t-k}\right] = \mathbf{0}_K.$$

Using the factor distribution in (4.7) and the error distribution in (2.4), we get:

$$\begin{aligned}
\mathbb{E}_\Theta\left[\boldsymbol{X}_t\right] &= \mathbb{E}_\Theta\left[W\boldsymbol{F}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t\right] = W\mathbb{E}_\Theta\left[\boldsymbol{F}_t\right] + \boldsymbol{\mu} + \mathbb{E}_\Theta\left[\boldsymbol{\epsilon}_t\right] = \boldsymbol{\mu}, \\
\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{X}_t\right] &= \mathbb{V}\mathrm{ar}_\Theta\left[W\boldsymbol{F}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t\right] = W\,\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{F}_t\right] W' + \mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{\epsilon}_t\right] = W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}.
\end{aligned}$$

In case of $\mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{X}_t\right]$, the independence of $\boldsymbol{F}_t$ and $\boldsymbol{\epsilon}_t$ is essential. □

Finally, the Bayes' theorem yields the distribution of the factors $\boldsymbol{F}_t$ given the data $\boldsymbol{X}_t$, that is, $\boldsymbol{F}_t|\boldsymbol{X}_t$.

Theorem 4.1.4 (Conditional Factor Distribution)
Assume the ADFMs in Definition 2.1.4 with loadings matrix $W \in \mathbb{R}^{N\times K}$ of full column rank K. Then, the conditional factor distribution $\boldsymbol{F}_t|\boldsymbol{X}_t$ is given by:

$$\boldsymbol{F}_t|\boldsymbol{X}_t \sim \mathcal{N}\left(M^{-1}W'\Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right), M^{-1}\right) = \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}, \Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right), \tag{4.9}$$

where the matrix $M = W'\Sigma_{\boldsymbol{\epsilon}}^{-1}W + \Sigma_{\boldsymbol{F}}^{-1} \in \mathbb{R}^{K\times K}$ is symmetric and invertible for positive definite covariance matrices $\Sigma_{\boldsymbol{\epsilon}} \in \mathbb{R}^{N\times N}$ and $\Sigma_{\boldsymbol{F}} \in \mathbb{R}^{K\times K}$.

Proof:
By virtue of the Bayes' theorem, Lemmata 4.1.1 and 4.1.3, we have:

$$\begin{aligned}
f_\Theta\left(\boldsymbol{F}_t|\boldsymbol{X}_t\right) &= \frac{f_\Theta\left(\boldsymbol{X}_t|\boldsymbol{F}_t\right) f_\Theta\left(\boldsymbol{F}_t\right)}{f_\Theta\left(\boldsymbol{X}_t\right)} \propto f_\Theta\left(\boldsymbol{X}_t|\boldsymbol{F}_t\right) f_\Theta\left(\boldsymbol{F}_t\right) \\
&= (2\pi)^{-N/2}\left|\Sigma_{\boldsymbol{\epsilon}}\right|^{-1/2} exp\left(-\frac{1}{2}\left(\boldsymbol{X}_t - W\boldsymbol{F}_t - \boldsymbol{\mu}\right)'\Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t - W\boldsymbol{F}_t - \boldsymbol{\mu}\right)\right) \\
&\quad \cdot (2\pi)^{-K/2}\left|\Sigma_{\boldsymbol{F}}\right|^{-1/2} exp\left(-\frac{1}{2}\boldsymbol{F}_t'\Sigma_{\boldsymbol{F}}^{-1}\boldsymbol{F}_t\right) \\
&\propto exp\left(-\frac{1}{2}\left[\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)'\Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right) - \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)'\Sigma_{\boldsymbol{\epsilon}}^{-1}W\boldsymbol{F}_t\right]\right) \\
&\quad \cdot exp\left(-\frac{1}{2}\left[-\boldsymbol{F}_t'W'\Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right) + \boldsymbol{F}_t'W'\Sigma_{\boldsymbol{\epsilon}}^{-1}W\boldsymbol{F}_t + \boldsymbol{F}_t'\Sigma_{\boldsymbol{F}}^{-1}\boldsymbol{F}_t\right]\right) \\
&= exp\left(-\frac{1}{2}\left[\boldsymbol{F}_t'\left(W'\Sigma_{\boldsymbol{\epsilon}}^{-1}W + \Sigma_{\boldsymbol{F}}^{-1}\right)\boldsymbol{F}_t - \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)'\Sigma_{\boldsymbol{\epsilon}}^{-1}W\boldsymbol{F}_t\right]\right) \\
&\quad \cdot exp\left(-\frac{1}{2}\left[-\boldsymbol{F}_t'W'\Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right) + \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)'\Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)\right]\right).
\end{aligned}$$

Next, we set: $M = W'\Sigma_{\boldsymbol{\epsilon}}^{-1}W + \Sigma_{\boldsymbol{F}}^{-1} \in \mathbb{R}^{K\times K}$ such that it is symmetric by definition. In general, covariance matrices are at least positive semi-definite, but the definition of M requires positive-definite matrices $\Sigma_{\boldsymbol{\epsilon}}$ and $\Sigma_{\boldsymbol{F}}$ to ensure that their inverse matrices exist. The loadings matrix W is supposed to have rank K, otherwise, the collinearity of its columns supports the reduction of the factor dimension by 1. Thus, the transformation $W'\Sigma_{\boldsymbol{\epsilon}}^{-1}W$ preserves rank K. With this in mind, we proceed as follows:

$$f_\Theta\left(\boldsymbol{F}_t|\boldsymbol{X}_t\right) \propto exp\left(-\frac{1}{2}\left[\boldsymbol{F}_t'M\boldsymbol{F}_t - \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)'\Sigma_{\boldsymbol{\epsilon}}^{-1}WM^{-1}M\boldsymbol{F}_t\right]\right)$$

$$
\begin{aligned}
&\cdot\, exp\left(-\frac{1}{2}\left[-\boldsymbol{F}_t' M M^{-1} W' \Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)+\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)' \Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\right]\right)\\
&\cdot\, exp\left(-\frac{1}{2}\left[\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)' \Sigma_{\boldsymbol{\epsilon}}^{-1} W M^{-1} M M^{-1} W' \Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\right]\right)\\
&\cdot\, exp\left(+\frac{1}{2}\left[\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)' \Sigma_{\boldsymbol{\epsilon}}^{-1} W M^{-1} M M^{-1} W' \Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\right]\right)\\
\propto\, & exp\left(-\frac{1}{2}\left[\boldsymbol{F}_t' M \boldsymbol{F}_t-\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)' \Sigma_{\boldsymbol{\epsilon}}^{-1} W M^{-1} M \boldsymbol{F}_t\right]\right)\\
&\cdot\, exp\left(-\frac{1}{2}\left[-\boldsymbol{F}_t' M M^{-1} W' \Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\right]\right)\\
&\cdot\, exp\left(-\frac{1}{2}\left[\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)' \Sigma_{\boldsymbol{\epsilon}}^{-1} W M^{-1} M M^{-1} W' \Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\right]\right)\\
=\, & exp\left(-\frac{1}{2}\left[\boldsymbol{F}_t-M^{-1} W' \Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\right]' M\left[\boldsymbol{F}_t-M^{-1} W' \Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\right]\right),
\end{aligned}
$$

which is proportional to the probability density function of the claimed normal distribution. □

The restriction to positive-definite matrices $\Sigma_{\boldsymbol{\epsilon}}$ and $\Sigma_{\boldsymbol{F}}$ in Theorem 4.1.4 should not hit hard empirical studies, as empirical covariance matrices are usually positive definite. At least, we did not face this issue in our analyses. For an alternative proof of Theorem 4.1.4, see Appendix A.2. The shocks $\{\boldsymbol{\epsilon}_t\}$ in Definition 2.1.4 are iid such that the factors conditioned on the panel data, i.e., $\{\boldsymbol{F}_t|\boldsymbol{X}_t\}$, are uncorrelated. If we admit serial error correlation, the independence of the conditioned factors $\{\boldsymbol{F}_t|\boldsymbol{X}_t\}$ is lost.

Next, we estimate the ADFMs in Definition 2.1.4. To this, we derive the log-likelihood function $\mathcal{L}\left(\Theta|X,F\right)$ with model parameters $\Theta=\{W,\Sigma_{\boldsymbol{\epsilon}},A_1,\dots,A_p,\Sigma_{\boldsymbol{\delta}}\}$ for complete samples $X=[\boldsymbol{X}_1,\dots,\boldsymbol{X}_T]'\in\mathbb{R}^{T\times N}$ and $F=[\boldsymbol{F}_1,\dots,\boldsymbol{F}_T]'\in\mathbb{R}^{T\times K}$ of sufficient size $T>p$ given the first p factors. Note, the expectation $\boldsymbol{\mu}$ is not part of Θ, since we subtract from the vectors $\boldsymbol{X}_t$ their empirical means during the data preprocessing such that $\boldsymbol{\mu}=\boldsymbol{0}_N$ holds. In a next step, we integrate out the hidden factors F from $\mathcal{L}\left(\Theta|X,F\right)$, when we determine the expectation of the log-likelihood function conditioned on the panel data X and parameters Θ. Finally, we treat the conditional factor means and covariance matrices as constants and apply MLE for parameter estimation with known factor dimension $K\geq 1$ and VAR order $p\geq 1$. This procedure, which is similar to the EM in Definition 2.1.7, is also used in Bańbura and Modugno (2014). However, there are two exceptions. On the one hand, Bańbura and Modugno (2014) actually derived an estimation method for ESFMs and then, followed the argumentation in Doz et al. (2012) to extend it to ADFMs. On the other hand, we deploy closed-form solutions for the conditional factor expectations and covariance matrices as in Theorem 4.1.4, while Bańbura and Modugno (2014) relied on the standard KF and KS.

Lemma 4.1.5 (Conditional Log-Likelihood Function of ADFMs)
For the ADFMs in Definition 2.1.4, let $X\in\mathbb{R}^{T\times N}$ be complete panel data as in Definition 2.1.1 with $T>p$. Then, we have for the log-likelihood function $\mathcal{L}\left(\Theta|X,F\right)$ of the panel data X and factors F given the model parameters $\Theta=\left\{W,\Sigma_{\boldsymbol{\epsilon}},A_1,\dots,A_p,\Sigma_{\boldsymbol{\delta}}\right\}$ and first $p>1$ factors $[\boldsymbol{F}_1,\dots,\boldsymbol{F}_p]$:

$$
\begin{aligned}
\mathcal{L}\left(\Theta|X,F\right)=&-\frac{T}{2}\left[N\,ln\left(2\pi\right)+ln\left(|\Sigma_{\boldsymbol{\epsilon}}|\right)\right]-\frac{T-p}{2}\left[K\,ln\left(2\pi\right)+ln\left(|\Sigma_{\boldsymbol{\delta}}|\right)\right]\\
&-\frac{1}{2}\sum_{t=1}^{T}\left(\boldsymbol{X}_t-W\boldsymbol{F}_t-\boldsymbol{\mu}\right)'\Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t-W\boldsymbol{F}_t-\boldsymbol{\mu}\right)\\
&-\frac{1}{2}\sum_{t=p+1}^{T}\left(\boldsymbol{F}_t-\sum_{i=1}^{p}A_i\boldsymbol{F}_{t-i}\right)'\Sigma_{\boldsymbol{\delta}}^{-1}\left(\boldsymbol{F}_t-\sum_{i=1}^{p}A_i\boldsymbol{F}_{t-i}\right).
\end{aligned}
$$

Proof:
An iterative application of the Bayes' theorem provides for the joint likelihood given the first p factors:

$$\begin{aligned}
&f_\Theta\left(\boldsymbol{X}_T,\ldots,\boldsymbol{X}_1,\boldsymbol{F}_T,\ldots,\boldsymbol{F}_{p+1}|\boldsymbol{F}_p,\ldots,\boldsymbol{F}_1\right)\\
&\quad=\frac{f_\Theta\left(\boldsymbol{X}_T,\ldots,\boldsymbol{X}_1,\boldsymbol{F}_T,\ldots,\boldsymbol{F}_1\right)}{f_\Theta\left(\boldsymbol{F}_p,\ldots,\boldsymbol{F}_1\right)}\\
&\quad=\frac{f_\Theta\left(\boldsymbol{X}_T|\boldsymbol{X}_{T-1},\ldots,\boldsymbol{X}_1,\boldsymbol{F}_T,\ldots,\boldsymbol{F}_1\right)f_\Theta\left(\boldsymbol{X}_{T-1},\ldots,\boldsymbol{X}_1,\boldsymbol{F}_T,\ldots,\boldsymbol{F}_1\right)}{f_\Theta\left(\boldsymbol{F}_p,\ldots,\boldsymbol{F}_1\right)}\\
&\quad=\frac{f_\Theta\left(\boldsymbol{X}_T|\boldsymbol{F}_T\right)f_\Theta\left(\boldsymbol{F}_T|\boldsymbol{X}_{T-1},\ldots,\boldsymbol{X}_1,\boldsymbol{F}_{T-1},\ldots,\boldsymbol{F}_1\right)f_\Theta\left(\boldsymbol{X}_{T-1},\ldots,\boldsymbol{X}_1,\boldsymbol{F}_{T-1},\ldots,\boldsymbol{F}_1\right)}{f_\Theta\left(\boldsymbol{F}_p,\ldots,\boldsymbol{F}_1\right)}\\
&\quad=f_\Theta\left(\boldsymbol{X}_T|\boldsymbol{F}_T\right)f_\Theta\left(\boldsymbol{F}_T|\boldsymbol{F}_{T-1},\ldots,\boldsymbol{F}_{T-p}\right)\frac{f_\Theta\left(\boldsymbol{X}_{T-1},\ldots,\boldsymbol{X}_1,\boldsymbol{F}_{T-1},\ldots,\boldsymbol{F}_1\right)}{f_\Theta\left(\boldsymbol{F}_p,\ldots,\boldsymbol{F}_1\right)}\\
&\quad=f_\Theta\left(\boldsymbol{X}_T|\boldsymbol{F}_T\right)f_\Theta\left(\boldsymbol{F}_T|\boldsymbol{F}_{T-1},\ldots,\boldsymbol{F}_{T-p}\right)f_\Theta\left(\boldsymbol{X}_{T-1},\ldots,\boldsymbol{X}_1,\boldsymbol{F}_{T-1},\ldots,\boldsymbol{F}_{p+1}|\boldsymbol{F}_p,\ldots,\boldsymbol{F}_1\right)\\
&\quad=\ldots=\left(\prod_{t=1}^{T}f_\Theta\left(\boldsymbol{X}_t|\boldsymbol{F}_t\right)\right)\left(\prod_{t=p+1}^{T}f_\Theta\left(\boldsymbol{F}_t|\boldsymbol{F}_{t-1},\ldots,\boldsymbol{F}_{t-p}\right)\right). \qquad (4.10)
\end{aligned}$$

Thus, we obtain for the corresponding log-likelihood function $\mathcal{L}\left(\Theta|X,F\right)$:

$$\begin{aligned}
\mathcal{L}\left(\Theta|X,F\right)&=ln\left(f_\Theta\left(\boldsymbol{X}_T,\ldots,\boldsymbol{X}_1,\boldsymbol{F}_T,\ldots,\boldsymbol{F}_{p+1}|\boldsymbol{F}_p,\ldots,\boldsymbol{F}_1\right)\right)\\
&=\sum_{t=1}^{T}ln\left(f_\Theta\left(\boldsymbol{X}_t|\boldsymbol{F}_t\right)\right)+\sum_{t=p+1}^{T}ln\left(f_\Theta\left(\boldsymbol{F}_t|\boldsymbol{F}_{t-1},\ldots,\boldsymbol{F}_{t-p}\right)\right). \qquad (4.11)
\end{aligned}$$

Using the conditional distributions in Lemma 4.1.1, we continue with:

$$\begin{aligned}
\mathcal{L}\left(\Theta|X,F\right)=&\sum_{t=1}^{T}ln\left((2\pi)^{-N/2}\left|\Sigma_{\boldsymbol{\epsilon}}\right|^{-1/2}exp\left[\frac{-\left(\boldsymbol{X}_t-W\boldsymbol{F}_t-\boldsymbol{\mu}\right)'\Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t-W\boldsymbol{F}_t-\boldsymbol{\mu}\right)}{2}\right]\right)\\
&+\sum_{t=p+1}^{T}ln\left((2\pi)^{-K/2}\left|\Sigma_{\boldsymbol{\delta}}\right|^{-1/2}exp\left[\frac{-\left(\boldsymbol{F}_t-\sum_{i=1}^{p}A_i\boldsymbol{F}_{t-i}\right)'\Sigma_{\boldsymbol{\delta}}^{-1}\left(\boldsymbol{F}_t-\sum_{i=1}^{p}A_i\boldsymbol{F}_{t-i}\right)}{2}\right]\right).
\end{aligned}$$

Finally, the properties of the logarithm and rearranging terms confirm the claim. □

The log-likelihood function $\mathcal{L}\left(\Theta|X,F\right)$ in Lemma 4.1.5 depends on the hidden factors $\{\boldsymbol{F}_t\}_{1\leq t\leq T}$ and so, we cannot compute it directly. However, the EM of Dempster et al. (1977) in Definition 2.1.7 enables the estimation of Θ in a maximum likelihood framework, where Lemma 4.1.6 prepares the expectation step.

Lemma 4.1.6 (Conditional Expectation of Conditional Log-Likelihood)
For the ADFMs in Definition 2.1.4, let $\mathcal{L}\left(\Theta|X,F\right)$ be the conditional log-likelihood function in Lemma 4.1.5. Then, we have for the expectation of $\mathcal{L}\left(\Theta|X,F\right)$ given the panel data X and parameters Θ:

$$\begin{aligned}
\mathbb{E}_\Theta\left[\mathcal{L}\left(\Theta|X,F\right)|X\right]=&-\frac{T}{2}\left[N\,ln\left(2\pi\right)+ln\left(\left|\Sigma_{\boldsymbol{\epsilon}}\right|\right)\right]-\frac{T-p}{2}\left[K\,ln\left(2\pi\right)+ln\left(\left|\Sigma_{\boldsymbol{\delta}}\right|\right)\right]\\
&-\frac{1}{2}\sum_{t=1}^{T}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)'\Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)+\sum_{t=1}^{T}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)'\Sigma_{\boldsymbol{\epsilon}}^{-1}W\,\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\\
&-\frac{1}{2}\sum_{t=1}^{T}tr\left(W'\Sigma_{\boldsymbol{\epsilon}}^{-1}W\left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}+\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\right)\\
&+\sum_{t=p+1}^{T}\sum_{i=1}^{p}tr\left(\Sigma_{\boldsymbol{\delta}}^{-1}A_i\,\boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)
\end{aligned}$$

$$
\begin{aligned}
&- \sum_{t=p+1}^{T} \sum_{\substack{i,j=1 \\ i<j}}^{p} tr\left(A_i' \Sigma_{\boldsymbol{\delta}}^{-1} A_j\, \boldsymbol{\mu}_{\boldsymbol{F}_{t-j}|\boldsymbol{X}_{t-j}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\right) \\
&- \frac{1}{2} \sum_{t=p+1}^{T} \sum_{i=1}^{p} tr\left(A_i' \Sigma_{\boldsymbol{\delta}}^{-1} A_i \left(\Sigma_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}} + \boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\right)\right) \\
&- \frac{1}{2} \sum_{t=p+1}^{T} tr\left(\Sigma_{\boldsymbol{\delta}}^{-1} \left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} \boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\right),
\end{aligned}
\tag{4.12}
$$

with means and covariance matrices of the conditional factor distributions as in Theorem 4.1.4.

Proof:
The definition of $\mathcal{L}(\Theta|X,F)$ in Lemma 4.1.5 and the independence $\{\boldsymbol{F}_t|\boldsymbol{X}_t\}_{1\leq t\leq T}$ yield:

$$
\begin{aligned}
\mathbb{E}_\Theta\left[\mathcal{L}(\Theta|X,F)\,|X\right] &= -\frac{T}{2}\left[N\,ln\,(2\pi) + ln\,(|\Sigma_{\boldsymbol{\epsilon}}|)\right] - \frac{T-p}{2}\left[K\,ln\,(2\pi) + ln\left(|\Sigma_{\boldsymbol{\delta}}|\right)\right] \\
&\quad - \frac{1}{2}\sum_{t=1}^{T} \mathbb{E}_\Theta\left[(\boldsymbol{X}_t - W\boldsymbol{F}_t - \boldsymbol{\mu})' \Sigma_{\boldsymbol{\epsilon}}^{-1} (\boldsymbol{X}_t - W\boldsymbol{F}_t - \boldsymbol{\mu})\,|X\right] \\
&\quad - \frac{1}{2}\sum_{t=p+1}^{T} \mathbb{E}_\Theta\left[\left(\boldsymbol{F}_t - \sum_{i=1}^{p} A_i \boldsymbol{F}_{t-i}\right)' \Sigma_{\boldsymbol{\delta}}^{-1} \left(\boldsymbol{F}_t - \sum_{i=1}^{p} A_i \boldsymbol{F}_{t-i}\right)\middle| X\right] \\
&= -\frac{T}{2}\left[N\,ln\,(2\pi) + ln\,(|\Sigma_{\boldsymbol{\epsilon}}|)\right] - \frac{T-p}{2}\left[K\,ln\,(2\pi) + ln\left(|\Sigma_{\boldsymbol{\delta}}|\right)\right] \\
&\quad - \frac{1}{2}\sum_{t=1}^{T} (\boldsymbol{X}_t - \boldsymbol{\mu})' \Sigma_{\boldsymbol{\epsilon}}^{-1} (\boldsymbol{X}_t - \boldsymbol{\mu}) + \sum_{t=1}^{T} (\boldsymbol{X}_t - \boldsymbol{\mu})' \Sigma_{\boldsymbol{\epsilon}}^{-1} W\, \mathbb{E}_\Theta\left[\boldsymbol{F}_t|\boldsymbol{X}_t\right] \\
&\quad - \frac{1}{2}\sum_{t=1}^{T} \mathbb{E}_\Theta\left[\boldsymbol{F}_t' W' \Sigma_{\boldsymbol{\epsilon}}^{-1} W \boldsymbol{F}_t|\boldsymbol{X}_t\right] + \sum_{t=p+1}^{T}\sum_{i=1}^{p} \mathbb{E}_\Theta\left[\boldsymbol{F}_t' \Sigma_{\boldsymbol{\delta}}^{-1} A_i \boldsymbol{F}_{t-i}|\boldsymbol{X}_t, \boldsymbol{X}_{t-i}\right] \\
&\quad - \frac{1}{2}\sum_{t=p+1}^{T}\sum_{i,j=1}^{p} \mathbb{E}_\Theta\left[\boldsymbol{F}_{t-i}' A_i' \Sigma_{\boldsymbol{\delta}}^{-1} A_j \boldsymbol{F}_{t-j}|\boldsymbol{X}_{t-i}, \boldsymbol{X}_{t-j}\right] \\
&\quad - \frac{1}{2}\sum_{t=p+1}^{T} \mathbb{E}_\Theta\left[\boldsymbol{F}_t' \Sigma_{\boldsymbol{\delta}}^{-1} \boldsymbol{F}_t|\boldsymbol{X}_t\right].
\end{aligned}
$$

Using the trace properties in Lemma A.1.2 and Remark A.1.3, we get:

$$
\begin{aligned}
\mathbb{E}_\Theta\left[\mathcal{L}(\Theta|X,F)\,|X\right] &= -\frac{T}{2}\left[N\,ln\,(2\pi) + ln\,(|\Sigma_{\boldsymbol{\epsilon}}|)\right] - \frac{T-p}{2}\left[K\,ln\,(2\pi) + ln\left(|\Sigma_{\boldsymbol{\delta}}|\right)\right] \\
&\quad - \frac{1}{2}\sum_{t=1}^{T} (\boldsymbol{X}_t - \boldsymbol{\mu})' \Sigma_{\boldsymbol{\epsilon}}^{-1} (\boldsymbol{X}_t - \boldsymbol{\mu}) + \sum_{t=1}^{T} (\boldsymbol{X}_t - \boldsymbol{\mu})' \Sigma_{\boldsymbol{\epsilon}}^{-1} W\, \mathbb{E}_\Theta\left[\boldsymbol{F}_t|\boldsymbol{X}_t\right] \\
&\quad - \frac{1}{2}\sum_{t=1}^{T} tr\left(W' \Sigma_{\boldsymbol{\epsilon}}^{-1} W \mathbb{E}_\Theta\left[\boldsymbol{F}_t \boldsymbol{F}_t'|\boldsymbol{X}_t\right]\right) \\
&\quad + \sum_{t=p+1}^{T}\sum_{i=1}^{p} tr\left(\Sigma_{\boldsymbol{\delta}}^{-1} A_i \mathbb{E}_\Theta\left[\boldsymbol{F}_{t-i} \boldsymbol{F}_t'|\boldsymbol{X}_t, \boldsymbol{X}_{t-i}\right]\right) \\
&\quad - \frac{1}{2}\sum_{t=p+1}^{T}\sum_{i,j=1}^{p} tr\left(A_i' \Sigma_{\boldsymbol{\delta}}^{-1} A_j \mathbb{E}_\Theta\left[\boldsymbol{F}_{t-j} \boldsymbol{F}_{t-i}'|\boldsymbol{X}_{t-i}, \boldsymbol{X}_{t-j}\right]\right) \\
&\quad - \frac{1}{2}\sum_{t=p+1}^{T} tr\left(\Sigma_{\boldsymbol{\delta}}^{-1} \mathbb{E}_\Theta\left[\boldsymbol{F}_t \boldsymbol{F}_t'|\boldsymbol{X}_t\right]\right).
\end{aligned}
\tag{4.13}
$$

Finally, we benefit from the independence of $\{\boldsymbol{F}_t|\boldsymbol{X}_t\}_{1\leq t\leq T}$, as it holds:

$$\mathbb{E}_\Theta\left[\boldsymbol{F}_t\boldsymbol{F}_t'|\boldsymbol{X}_t\right] = \Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}',$$
$$\mathbb{E}_\Theta\left[\boldsymbol{F}_t\boldsymbol{F}_s'|\boldsymbol{X}_t,\boldsymbol{X}_s\right] = \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}_{\boldsymbol{F}_s|\boldsymbol{X}_s}', \text{ for all } t\neq s,$$

where the conditional factor means and covariance matrices are defined in Theorem 4.1.4. □

Usually, EMs for DFMs estimate the factor moments $\mathbb{E}_\Theta\left[\boldsymbol{F}_t|\boldsymbol{X}_t\right]$ and $\mathbb{E}_\Theta\left[\boldsymbol{F}_t\boldsymbol{F}_s'|\boldsymbol{X}_t,\boldsymbol{X}_s\right], 1\leq s,t\leq T$, in (4.13) by the KF or KS (Bork, 2009; Bork et al., 2010; Bańbura and Rünstler, 2011; Doz et al., 2011, 2012; Bańbura and Modugno, 2014). By contrast, we use the closed-form solutions in Theorem 4.1.4 such that Kalman Filtering is not needed. In principle, this is a minor advantage, since KF and KS are well-known concepts (see Section 2.1.4). Unfortunately, decomposing factor moments into the contributions of the input data is more complicated in case of the KF and KS than our closed-form ansatz. For further information about this, see the comparison between both procedures in Section 4.4. For simplicity reasons, we therefore proceed with the closed-form solutions for the factor moments.

In the maximization step of the EMs in Definition 2.1.7, we maximize the expected log-likelihood function with respect to Θ. Here, we do the same, when we search for the optimum of $\mathbb{E}_{\hat{\Theta}}[\mathcal{L}(\Theta|X,F)|X]$ in Lemma 4.1.6. However, the maximization is now done in a simplified way, as we ignore the dependence of $\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}$ and $\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}, 1\leq t\leq T$, on the parameters Θ and treat them as constants. Then, we obtain the partial derivatives of $\mathbb{E}_{\hat{\Theta}}[\mathcal{L}(\Theta|X,F)|X]$ regarding W, $\Sigma_{\boldsymbol{\epsilon}}$, $\Sigma_{\boldsymbol{\delta}}$ and $A_i, 1\leq i\leq p$, explicitly (see Lemma 4.1.7). To justify this simplification, be aware that $\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}$ and $\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}$ arise from the factors and hence, are interpreted as data or known parameters. The described maximization procedure follows the methods in Tipping and Bishop (1999) and Bańbura and Modugno (2014).

Lemma 4.1.7 (Maximum of the Expected Log-likelihood in ADFMs)
Let $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}(\Theta|X,F)|X\right]$ be the expected log-likelihood function in Lemma 4.1.6, where the factor means $\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}$ and covariance matrices $\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}$ are derived from the panel data X and latest parameter estimates $\hat{\Theta}_{(l)}$. Then, the maximum of $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}(\Theta|X,F)|X\right]$ is reached for the parameters $\hat{\Theta}$ given by:

$$\hat{W} = \left(\sum_{t=1}^T(\boldsymbol{X}_t-\boldsymbol{\mu})\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}'\right)\left(\sum_{t=1}^T\left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}+\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}'\right)\right)^{-1},$$

$$\hat{\Sigma}_{\boldsymbol{\epsilon}} = \frac{1}{T}\left(\sum_{t=1}^T(\boldsymbol{X}_t-\boldsymbol{\mu})(\boldsymbol{X}_t-\boldsymbol{\mu})' - \hat{W}\left(\sum_{t=1}^T\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}(\boldsymbol{X}_t-\boldsymbol{\mu})'\right)\right),$$

$$\begin{bmatrix}\hat{A}_1 & \cdots & \hat{A}_p\end{bmatrix} = \left(\sum_{t=p+1}^T\begin{bmatrix}\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}}' & \cdots & \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}}'\end{bmatrix}\right)\tilde{\Sigma}_{p,T}^{-1},$$

$$\begin{aligned}\hat{\Sigma}_{\boldsymbol{\delta}} = &-\frac{1}{T-p}\begin{bmatrix}\hat{A}_1 & \cdots & \hat{A}_p\end{bmatrix}\left(\sum_{t=p+1}^T\begin{bmatrix}\boldsymbol{\mu}_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}}\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}' \\ \vdots \\ \boldsymbol{\mu}_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}}\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}'\end{bmatrix}\right) \\ &+\frac{1}{T-p}\sum_{t=p+1}^T\left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}+\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}'\right),\end{aligned}$$

with invertible matrix $\tilde{\Sigma}_{p,T} \in \mathbb{R}^{pK \times pK}$ defined by:

$$\tilde{\Sigma}_{p,T} = \sum_{t=p+1}^{T} \begin{bmatrix} \Sigma_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} & O_K & \cdots & O_K \\ O_K & \Sigma_{\boldsymbol{F}_{t-2}|\boldsymbol{X}_{t-2}} & & \vdots \\ \vdots & & \ddots & O_K \\ O_K & \cdots & O_K & \Sigma_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \end{bmatrix}$$
$$+ \sum_{t=p+1}^{T} \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} & \cdots & \boldsymbol{\mu}_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\mu}_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} & \cdots & \boldsymbol{\mu}_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \end{bmatrix}$$

and the moments of the conditional factors as in Theorem 4.1.4.

Proof:
Here, we sequentially determine the partial derivatives of $\mathbb{E}_{\hat{\Theta}_{(l)}}[\mathcal{L}(\Theta|X,F)|X]$ with respect to Θ. At first, the trace properties in Lemma A.1.2, Remark A.1.3 and Lemma A.1.5 provide for W:

$$\begin{aligned}
\frac{\partial \mathbb{E}_{\hat{\Theta}_{(l)}}[\mathcal{L}(\Theta|X,F)|X]}{\partial W} &= \sum_{t=1}^{T} \frac{\partial}{\partial W} tr\left((\boldsymbol{X}_t - \boldsymbol{\mu})' \Sigma_{\boldsymbol{\epsilon}}^{-1} W \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right) \\
&\quad - \frac{1}{2} \sum_{t=1}^{T} \frac{\partial}{\partial W} tr\left(W' \Sigma_{\boldsymbol{\epsilon}}^{-1} W \left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} \boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\right) \\
&= \sum_{t=1}^{T} \frac{\partial}{\partial W} tr\left(\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} (\boldsymbol{X}_t - \boldsymbol{\mu})' \Sigma_{\boldsymbol{\epsilon}}^{-1} W\right) \\
&\quad - \sum_{t=1}^{T} \Sigma_{\boldsymbol{\epsilon}}^{-1} W \left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} \boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right) \\
&= \sum_{t=1}^{T} \left(\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} (\boldsymbol{X}_t - \boldsymbol{\mu})' \Sigma_{\boldsymbol{\epsilon}}^{-1}\right)' - \sum_{t=1}^{T} \Sigma_{\boldsymbol{\epsilon}}^{-1} W \left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} \boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right).
\end{aligned}$$

Thus, we have to solve the arising system of linear equations with respect to W, that is, we obtain:

$$\Sigma_{\boldsymbol{\epsilon}}^{-1} W \sum_{t=1}^{T} \left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} \boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right) = \Sigma_{\boldsymbol{\epsilon}}^{-1} \sum_{t=1}^{T} (\boldsymbol{X}_t - \boldsymbol{\mu}) \boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}.$$

This results in the estimated parameter $\hat{W}$ defined as:

$$\hat{W} = \left(\sum_{t=1}^{T} (\boldsymbol{X}_t - \boldsymbol{\mu}) \boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right) \left(\sum_{t=1}^{T} \left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} \boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\right)^{-1}.$$

Similarly, we obtain for $\Sigma_{\boldsymbol{\epsilon}}$ using the trace properties in Lemma A.1.2 and Remark A.1.3 together with its derivatives in Lemma A.1.5:

$$\begin{aligned}
\frac{\partial \mathbb{E}_{\hat{\Theta}_{(l)}}[\mathcal{L}(\Theta|X,F)|X]}{\partial \Sigma_{\boldsymbol{\epsilon}}} &= -\frac{T}{2} \frac{\partial}{\partial \Sigma_{\boldsymbol{\epsilon}}} ln\left(|\Sigma_{\boldsymbol{\epsilon}}|\right) - \frac{1}{2} \sum_{t=1}^{T} \frac{\partial}{\partial \Sigma_{\boldsymbol{\epsilon}}} tr\left((\boldsymbol{X}_t - \boldsymbol{\mu})(\boldsymbol{X}_t - \boldsymbol{\mu})' \Sigma_{\boldsymbol{\epsilon}}^{-1}\right) \\
&\quad - \frac{1}{2} \sum_{t=1}^{T} \frac{\partial}{\partial \Sigma_{\boldsymbol{\epsilon}}} tr\left(W \left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} \boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right) W' \Sigma_{\boldsymbol{\epsilon}}^{-1}\right) \\
&\quad + \frac{1}{2} \sum_{t=1}^{T} \frac{\partial}{\partial \Sigma_{\boldsymbol{\epsilon}}} tr\left(W \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} (\boldsymbol{X}_t - \boldsymbol{\mu})' \Sigma_{\boldsymbol{\epsilon}}^{-1}\right)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}\sum_{t=1}^{T}\frac{\partial}{\partial\Sigma_{\boldsymbol{\epsilon}}}tr\left(\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}W'\Sigma_{\boldsymbol{\epsilon}}^{-1}\right)\\
&=-\frac{T}{2}\Sigma_{\boldsymbol{\epsilon}}^{-1}+\frac{1}{2}\sum_{t=1}^{T}\left(\Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)'\Sigma_{\boldsymbol{\epsilon}}^{-1}\right)'\\
&+\frac{1}{2}\sum_{t=1}^{T}\left(\Sigma_{\boldsymbol{\epsilon}}^{-1}W\left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}+\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)W'\Sigma_{\boldsymbol{\epsilon}}^{-1}\right)'\\
&-\frac{1}{2}\sum_{t=1}^{T}\left(\Sigma_{\boldsymbol{\epsilon}}^{-1}W\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)'\Sigma_{\boldsymbol{\epsilon}}^{-1}\right)'\\
&-\frac{1}{2}\sum_{t=1}^{T}\left(\Sigma_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}W'\Sigma_{\boldsymbol{\epsilon}}^{-1}\right)'.
\end{aligned}
$$

Hence, we have the following system of linear equations:

$$
\begin{aligned}
T\,\Sigma_{\boldsymbol{\epsilon}}=&\sum_{t=1}^{T}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)'+W\left(\sum_{t=1}^{T}\left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}+\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\right)W'\\
&-\left(\sum_{t=1}^{T}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)W'-W\left(\sum_{t=1}^{T}\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)'\right).
\end{aligned}
$$

Note, the partial derivatives of the expected log-likelihood function $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,F\right)|X\right]$ with respect to W and $\Sigma_{\boldsymbol{\epsilon}}$ have to be zero at the same time. Therefore, we substitute the matrix W by its estimate $\hat{W}$. Moreover, to admit further simplifications, we insert the solution for $\hat{W}$ in the second summand on the right-hand side of the above equation. In total, this yields:

$$
\begin{aligned}
T\,\Sigma_{\boldsymbol{\epsilon}}&=\sum_{t=1}^{T}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)'+\left(\sum_{t=1}^{T}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\hat{W}'\\
&\quad-\left(\sum_{t=1}^{T}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\hat{W}'-\hat{W}\left(\sum_{t=1}^{T}\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)'\right)\\
&=\sum_{t=1}^{T}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)'-\hat{W}\left(\sum_{t=1}^{T}\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)'\right).
\end{aligned}
$$

Hence, the matrix $\Sigma_{\boldsymbol{\epsilon}}$ is estimated by $\hat{\Sigma}_{\boldsymbol{\epsilon}}$ defined as follows:

$$
\hat{\Sigma}_{\boldsymbol{\epsilon}}=\frac{1}{T}\left(\sum_{t=1}^{T}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)'-\hat{W}\left(\sum_{t=1}^{T}\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)'\right)\right).
$$

For any fixed $1\le k\le p$ the trace properties in Lemma A.1.2 and Remark A.1.3 as well as its derivatives in Lemma A.1.5 yield for the matrices A_k:

$$
\begin{aligned}
\frac{\partial\,\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,F\right)|X\right]}{\partial A_k}=&\sum_{t=p+1}^{T}\frac{\partial}{\partial A_k}tr\left(\boldsymbol{\mu}_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\Sigma_{\boldsymbol{\delta}}^{-1}A_k\right)\\
&-\frac{1}{2}\sum_{t=p+1}^{T}\sum_{\substack{i=1\\ i\neq k}}^{p}\frac{\partial}{\partial A_k}tr\left(A_i'\Sigma_{\boldsymbol{\delta}}^{-1}A_k\boldsymbol{\mu}_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\right)\\
&-\frac{1}{2}\sum_{t=p+1}^{T}\sum_{\substack{j=1\\ j\neq k}}^{p}\frac{\partial}{\partial A_k}tr\left(A_k'\Sigma_{\boldsymbol{\delta}}^{-1}A_j\boldsymbol{\mu}_{\boldsymbol{F}_{t-j}|\boldsymbol{X}_{t-j}}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}}\right)
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{1}{2} \sum_{t=p+1}^{T} \frac{\partial}{\partial A_k} tr \left(A_k' \Sigma_{\boldsymbol{\delta}}^{-1} A_k \left(\Sigma_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} + \boldsymbol{\mu}_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} \right) \right) \\
= &\sum_{t=p+1}^{T} \left(\boldsymbol{\mu}_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t}|\boldsymbol{X}_{t}} \Sigma_{\boldsymbol{\delta}}^{-1} \right)' \\
&- \frac{1}{2} \sum_{t=p+1}^{T} \sum_{\substack{i=1\\ i\neq k}}^{p} \left(\boldsymbol{\mu}_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}} A_i' \Sigma_{\boldsymbol{\delta}}^{-1} \right)' \\
&- \frac{1}{2} \sum_{t=p+1}^{T} \sum_{\substack{j=1\\ j\neq k}}^{p} \left(\Sigma_{\boldsymbol{\delta}}^{-1} A_j \, \boldsymbol{\mu}_{\boldsymbol{F}_{t-j}|\boldsymbol{X}_{t-j}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} \right) \\
&- \sum_{t=p+1}^{T} \left(\Sigma_{\boldsymbol{\delta}}^{-1} A_k \left(\Sigma_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} + \boldsymbol{\mu}_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} \right) \right) \\
= &\Sigma_{\boldsymbol{\delta}}^{-1} \sum_{t=p+1}^{T} \boldsymbol{\mu}_{\boldsymbol{F}_{t}|\boldsymbol{X}_{t}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} - \Sigma_{\boldsymbol{\delta}}^{-1} \sum_{t=p+1}^{T} \sum_{\substack{i=1\\ i\neq k}}^{p} \left(A_i \, \boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} \right) \\
&- \Sigma_{\boldsymbol{\delta}}^{-1} A_k \sum_{t=p+1}^{T} \left(\Sigma_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} + \boldsymbol{\mu}_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} \right).
\end{aligned}
$$

$\Sigma_{\boldsymbol{\delta}}$ has full rank, therefore, for all k we get the following system of linear equations with variables A_k:

$$
\begin{aligned}
\sum_{t=p+1}^{T} \boldsymbol{\mu}_{\boldsymbol{F}_{t}|\boldsymbol{X}_{t}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} = &\sum_{t=p+1}^{T} \sum_{\substack{i=1\\ i\neq k}}^{p} \left(A_i \, \boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} \right) \\
&+ A_k \sum_{t=p+1}^{T} \left(\Sigma_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} + \boldsymbol{\mu}_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-k}|\boldsymbol{X}_{t-k}} \right).
\end{aligned}
$$

Note, the previous equation system holds for all matrices $A_k, 1 \leq k \leq p$, at the same time. For clarity reasons, we define the matrix $\tilde{\Sigma}_{p,T} \in \mathbb{R}^{(pK)\times(pK)}$ as:

$$
\begin{aligned}
\tilde{\Sigma}_{p,T} = &\sum_{t=p+1}^{T} \begin{bmatrix} \Sigma_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} & O_K & \cdots & O_K \\ O_K & \Sigma_{\boldsymbol{F}_{t-2}|\boldsymbol{X}_{t-2}} & & \vdots \\ \vdots & & \ddots & O_K \\ O_K & \cdots & O_K & \Sigma_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \end{bmatrix} \\
&+ \sum_{t=p+1}^{T} \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} & \cdots & \boldsymbol{\mu}_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\mu}_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} & \cdots & \boldsymbol{\mu}_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \end{bmatrix},
\end{aligned}
$$

such that we can summarize the k systems of linear equations in the following manner:

$$
\begin{bmatrix} A_1 & \cdots & A_p \end{bmatrix} \tilde{\Sigma}_{p,T} = \sum_{t=p+1}^{T} \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{F}_{t}|\boldsymbol{X}_{t}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} & \cdots & \boldsymbol{\mu}_{\boldsymbol{F}_{t}|\boldsymbol{X}_{t}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \end{bmatrix},
$$

which is solved by $\begin{bmatrix} \hat{A}_1 & \cdots & \hat{A}_p \end{bmatrix}$ given by:

$$
\begin{bmatrix} \hat{A}_1 & \cdots & \hat{A}_p \end{bmatrix} = \left(\sum_{t=p+1}^{T} \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{F}_{t}|\boldsymbol{X}_{t}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} & \cdots & \boldsymbol{\mu}_{\boldsymbol{F}_{t}|\boldsymbol{X}_{t}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \end{bmatrix} \right) \tilde{\Sigma}_{p,T}^{-1}.
$$

Note, the covariance matrices $\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}$ in Theorem 4.1.4 are positive definite implying that the covariance matrix $\tilde{\Sigma}_{p,T}$ is positive definite, too.

Finally, we conclude for the covariance matrix $\Sigma_{\boldsymbol{\delta}}$ using the trace properties in Lemma A.1.2 and Remark A.1.3 together with its partial derivatives in Lemma A.1.5:

$$
\begin{aligned}
\frac{\partial \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,F\right)|X\right]}{\partial \Sigma_{\boldsymbol{\delta}}} &= -\frac{T-p}{2}\frac{\partial}{\partial \Sigma_{\boldsymbol{\delta}}} ln\left(|\Sigma_{\boldsymbol{\delta}}|\right) + \frac{1}{2}\sum_{t=p+1}^{T}\sum_{i=1}^{p}\frac{\partial}{\partial \Sigma_{\boldsymbol{\delta}}} tr\left(A_i\,\boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\Sigma_{\boldsymbol{\delta}}^{-1}\right) \\
&\quad + \frac{1}{2}\sum_{t=p+1}^{T}\sum_{i=1}^{p}\frac{\partial}{\partial \Sigma_{\boldsymbol{\delta}}} tr\left(\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}A_i'\,\Sigma_{\boldsymbol{\delta}}^{-1}\right) \\
&\quad - \sum_{t=p+1}^{T}\sum_{\substack{i,j=1\\ i<j}}^{p}\frac{\partial}{\partial \Sigma_{\boldsymbol{\delta}}} tr\left(A_j\,\boldsymbol{\mu}_{\boldsymbol{F}_{t-j}|\boldsymbol{X}_{t-j}}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}A_i'\,\Sigma_{\boldsymbol{\delta}}^{-1}\right) \\
&\quad - \frac{1}{2}\sum_{t=p+1}^{T}\sum_{i=1}^{p}\frac{\partial}{\partial \Sigma_{\boldsymbol{\delta}}} tr\left(A_i\left(\Sigma_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}} + \boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\right)A_i'\,\Sigma_{\boldsymbol{\delta}}^{-1}\right) \\
&\quad - \frac{1}{2}\sum_{t=p+1}^{T}\frac{\partial}{\partial \Sigma_{\boldsymbol{\delta}}} tr\left(\left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\Sigma_{\boldsymbol{\delta}}^{-1}\right) \\
&= -\frac{T-p}{2}\Sigma_{\boldsymbol{\delta}}^{-1} - \frac{1}{2}\sum_{t=p+1}^{T}\sum_{i=1}^{p}\left(\Sigma_{\boldsymbol{\delta}}^{-1}A_i\,\boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\Sigma_{\boldsymbol{\delta}}^{-1}\right)' \\
&\quad - \frac{1}{2}\sum_{t=p+1}^{T}\sum_{i=1}^{p}\left(\Sigma_{\boldsymbol{\delta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}A_i'\,\Sigma_{\boldsymbol{\delta}}^{-1}\right)' \\
&\quad + \sum_{t=p+1}^{T}\sum_{\substack{i,j=1\\ i<j}}^{p}\left(\Sigma_{\boldsymbol{\delta}}^{-1}A_j\,\boldsymbol{\mu}_{\boldsymbol{F}_{t-j}|\boldsymbol{X}_{t-j}}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}A_i'\,\Sigma_{\boldsymbol{\delta}}^{-1}\right)' \\
&\quad + \frac{1}{2}\sum_{t=p+1}^{T}\sum_{i=1}^{p}\left(\Sigma_{\boldsymbol{\delta}}^{-1}A_i\left(\Sigma_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}} + \boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\right)A_i'\,\Sigma_{\boldsymbol{\delta}}^{-1}\right)' \\
&\quad + \frac{1}{2}\sum_{t=p+1}^{T}\left(\Sigma_{\boldsymbol{\delta}}^{-1}\left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\Sigma_{\boldsymbol{\delta}}^{-1}\right)' \\
&= -\frac{T-p}{2}\Sigma_{\boldsymbol{\delta}}^{-1} - \frac{1}{2}\Sigma_{\boldsymbol{\delta}}^{-1}\sum_{t=p+1}^{T}\sum_{i=1}^{p}\left(\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}A_i'\right)\Sigma_{\boldsymbol{\delta}}^{-1} \\
&\quad - \frac{1}{2}\Sigma_{\boldsymbol{\delta}}^{-1}\sum_{t=p+1}^{T}\sum_{i=1}^{p}\left(A_i\boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\Sigma_{\boldsymbol{\delta}}^{-1} \\
&\quad + \Sigma_{\boldsymbol{\delta}}^{-1}\sum_{t=p+1}^{T}\sum_{\substack{i,j=1\\ i<j}}^{p}\left(A_i\,\boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-j}|\boldsymbol{X}_{t-j}}A_j'\right)\Sigma_{\boldsymbol{\delta}}^{-1} \\
&\quad + \frac{1}{2}\Sigma_{\boldsymbol{\delta}}^{-1}\sum_{t=p+1}^{T}\sum_{i=1}^{p}\left(A_i\left(\Sigma_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}} + \boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\right)A_i'\right)\Sigma_{\boldsymbol{\delta}}^{-1} \\
&\quad + \frac{1}{2}\Sigma_{\boldsymbol{\delta}}^{-1}\sum_{t=p+1}^{T}\left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\Sigma_{\boldsymbol{\delta}}^{-1}.
\end{aligned}
$$

Hence, the estimated matrix $\hat{\Sigma}_{\boldsymbol{\delta}}$ has to satisfy the following linear equation system:

$$
\begin{aligned}
(T-p)\,\Sigma_{\boldsymbol{\delta}} = &-\sum_{t=p+1}^{T}\sum_{i=1}^{p}\left(\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}A'_i\right) - \sum_{t=p+1}^{T}\sum_{i=1}^{p}\left(A_i\boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\\
&+\sum_{t=p+1}^{T}\sum_{\substack{i,j=1\\ i\neq j}}^{p}\left(A_i\,\boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-j}|\boldsymbol{X}_{t-j}}A'_j\right)\\
&+\sum_{t=p+1}^{T}\sum_{i=1}^{p}\left(A_i\left(\Sigma_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}+\boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}\right)A'_i\right)\\
&+\sum_{t=p+1}^{T}\left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}+\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\\
= &-\left(\sum_{t=p+1}^{T}\left[\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}}\ \cdots\ \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}}\right]\right)\begin{bmatrix}A'_1\\ \vdots\\ A'_p\end{bmatrix}\\
&-\begin{bmatrix}A_1 & \cdots & A_p\end{bmatrix}\left(\sum_{t=p+1}^{T}\begin{bmatrix}\boldsymbol{\mu}_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\\ \vdots\\ \boldsymbol{\mu}_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\end{bmatrix}\right)\\
&+\begin{bmatrix}A_1 & \cdots & A_p\end{bmatrix}\tilde{\Sigma}_{p,T}\begin{bmatrix}A'_1\\ \vdots\\ A'_p\end{bmatrix}+\sum_{t=p+1}^{T}\left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}+\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right).
\end{aligned}
$$

The partial derivatives of $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,F\right)|X\right]$ with respect to the matrices $\Sigma_{\boldsymbol{\delta}}$ and $A_k, 1\leq k\leq p$, have to be zero at the same time. This is why we replace the matrices A_k by their estimates $\hat{A}_k$ and receive for the estimated matrix $\hat{\Sigma}_{\boldsymbol{\delta}}$ the following solution:

$$
\begin{aligned}
\hat{\Sigma}_{\boldsymbol{\delta}} = &-\frac{1}{T-p}\left(\sum_{t=p+1}^{T}\left[\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}}\ \cdots\ \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}}\right]\right)\begin{bmatrix}\hat{A}'_1\\ \vdots\\ \hat{A}'_p\end{bmatrix}\\
&-\frac{1}{T-p}\begin{bmatrix}\hat{A}_1 & \cdots & \hat{A}_p\end{bmatrix}\left(\sum_{t=p+1}^{T}\begin{bmatrix}\boldsymbol{\mu}_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\\ \vdots\\ \boldsymbol{\mu}_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\end{bmatrix}\right)\\
&+\frac{1}{T-p}\begin{bmatrix}\hat{A}_1 & \cdots & \hat{A}_p\end{bmatrix}\tilde{\Sigma}_{p,T}\begin{bmatrix}\hat{A}'_1\\ \vdots\\ \hat{A}'_p\end{bmatrix}\\
&+\frac{1}{T-p}\sum_{t=p+1}^{T}\left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}+\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right),
\end{aligned}
$$

Finally, we insert the solution for $[\hat{A}_1\cdots\hat{A}_p]$ in the first matrix of the third row such that the expressions in the first and third row of the above representation cancel each other out and we get the stated version of the covariance matrix $\hat{\Sigma}_{\boldsymbol{\delta}}$.

So far, the estimates $\hat{\Theta}$ mark a saddle point of the expected log-likelihood function $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,F\right)|X\right]$, but the Gaussian framework causes that we have a maximum instead of a minimum or saddle point. □

If the conditional distributions in Theorem 4.1.4 and the cross-sectional error correlation enter the updates in Bańbura and Modugno (2014), we receive the formulas in Lemma 4.1.7 and thus, show the consistency of both results. Next, we define a general expression for the autocovariance matrix of the underlying panel data to shorten the updates in Lemma 4.1.7 and to support their implementation.

Definition 4.1.8 (Empirical Autocovariance Matrix of Panel Data)
For the ADFMs in Definition 2.1.4, let $X \in \mathbb{R}^{T\times N}$ be the complete panel data sample in Definition 2.1.1. Moreover, for any integers $1 \le b_1 < u_1 \le T$ and $1 \le b_2 < u_2 \le T$ with $u_1 - b_1 = u_2 - b_2$, let the matrices $[\boldsymbol{X}_{b_1}, \dots, \boldsymbol{X}_{u_1}]'$ and $[\boldsymbol{X}_{b_2}, \dots, \boldsymbol{X}_{u_2}]'$ be subsamples of X. Then, their empirical autocovariance matrix is defined as follows:

$$\left\{{}^{u_1}_{b_1}S^{u_2}_{b_2}\right\} = \frac{1}{u_1 - b_1}\left[\boldsymbol{X}_{b_1} - \boldsymbol{\mu}, \dots, \boldsymbol{X}_{u_1} - \boldsymbol{\mu}\right]\left[\boldsymbol{X}_{b_2} - \boldsymbol{\mu}, \dots, \boldsymbol{X}_{u_2} - \boldsymbol{\mu}\right]'.$$

In empirical studies, we estimate the mean $\boldsymbol{\mu}$ in Definition 4.1.8 by the empirical mean $\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}$ in Definition 2.1.2. The updates in Lemma 4.1.7 do not rely on the factors $\boldsymbol{F}_t$, but they are functions of the parameters in Theorem 4.1.4. If we insert them, we can simplify the expressions in Lemma 4.1.7 as follows.

Theorem 4.1.9 (EM for Complete ADFMs)
For the ADFMs in Definition 2.1.4, let (l) be the current loop of the EM resulting in factor means $\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}$ and covariance matrices $\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}$ based on the estimated parameters $\Theta_{(l)}$. For clarity reasons, we write $\Theta_{(l)}$ instead of $\hat{\Theta}_{(l)}$. Then, we obtain the parameters of the next loop $(l+1)$ in the following way:

$$\begin{aligned}
W_{(l+1)} &= \left\{{}^T_1S^T_1\right\}\Sigma^{-1}_{\boldsymbol{\epsilon}(l)}W_{(l)}\left(I_K + D_{(l)}\left\{{}^T_1S^T_1\right\}\Sigma^{-1}_{\boldsymbol{\epsilon}(l)}W_{(l)}\right)^{-1},\\
\Sigma_{\boldsymbol{\epsilon}(l+1)} &= \left(I_N - W_{(l+1)}D_{(l)}\right)\left\{{}^T_1S^T_1\right\},\\
\left[A_{1_{(l+1)}}, \dots, A_{p_{(l+1)}}\right] &= \left(\mathbb{1}_p' \otimes D_{(l)}\right)\tilde{S}'\left(I_p \otimes D'_{(l)}\right)\\
&\quad \cdot\left(\left(I_p \otimes M^{-1}_{(l)}\right) + \left(I_p \otimes D_{(l)}\right)\hat{S}'\left(I_p \otimes D'_{(l)}\right)\right)^{-1},\\
\Sigma_{\boldsymbol{\delta}(l+1)} &= M^{-1}_{(l)} + D_{(l)}\left\{{}^T_{p+1}S^T_{p+1}\right\}D'_{(l)}\\
&\quad - \left[A_{1_{(l+1)}}, \dots, A_{p_{(l+1)}}\right]\left(I_p \otimes D_{(l)}\right)\tilde{S}\left(\mathbb{1}_p \otimes D'_{(l)}\right),
\end{aligned}$$

where the matrices $\tilde{S} \in \mathbb{R}^{pN\times pN}$, $\hat{S} \in \mathbb{R}^{pN\times pN}$ and $D_{(l)} \in \mathbb{R}^{K\times N}$ are given by:

$$\begin{aligned}
\tilde{S} &= \begin{pmatrix} \left\{{}^{T-1}_{p}S^{T}_{p+1}\right\} & & O_N \\ & \ddots & \\ O_N & & \left\{{}^{T-p}_{1}S^{T}_{p+1}\right\} \end{pmatrix},\\
\hat{S} &= \begin{pmatrix} \left\{{}^{T-1}_{p}S^{T-1}_{p}\right\} & \cdots & \left\{{}^{T-1}_{p}S^{T-p}_{1}\right\} \\ \vdots & \ddots & \vdots \\ \left\{{}^{T-p}_{1}S^{T-1}_{p}\right\} & \cdots & \left\{{}^{T-p}_{1}S^{T-p}_{1}\right\} \end{pmatrix},\\
D_{(l)} &= M^{-1}_{(l)}W'_{(l)}\Sigma^{-1}_{\boldsymbol{\epsilon}(l)}.
\end{aligned}$$

Proof:
Lemma 4.1.7 provides for the updated matrix $W_{(l+1)}$:

$$W_{(l+1)} = \left(\sum_{t=1}^{T}\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\left(\sum_{t=1}^{T}\left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\right)^{-1}$$

$$
\begin{aligned}
&= \left(\sum_{t=1}^{T} \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right) \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} W_{(l)} M_{(l)}^{-1} \right) \\
&\quad \cdot \left(\sum_{t=1}^{T} \left(M_{(l)}^{-1} + M_{(l)}^{-1} W_{(l)}' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right) \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} W_{(l)} M_{(l)}^{-1} \right) \right)^{-1} \\
&= T \left\{ {}_{1}^{T}S_{1}^{T} \right\} \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} W_{(l)} M_{(l)}^{-1} \left(T\, M_{(l)}^{-1} + M_{(l)}^{-1} W_{(l)}' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} T \left\{ {}_{1}^{T}S_{1}^{T} \right\} \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} W_{(l)} M_{(l)}^{-1} \right)^{-1} \\
&= \left\{ {}_{1}^{T}S_{1}^{T} \right\} \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} W_{(l)} \left(I_K + M_{(l)}^{-1} W_{(l)}' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} \left\{ {}_{1}^{T}S_{1}^{T} \right\} \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} W_{(l)} \right)^{-1} .
\end{aligned}
$$

Replacing the term $M_{(l)}^{-1} W_{(l)}' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1}$ by the matrix $D_{(l)}$ proves the claim.

Similarly, we obtain for the matrix $\Sigma_{\boldsymbol{\epsilon}(l+1)}$, when starting with Lemma 4.1.7:

$$
\begin{aligned}
\Sigma_{\boldsymbol{\epsilon}(l+1)} &= \frac{1}{T} \left(\sum_{t=1}^{T} \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right) \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)' - W_{(l+1)} \left(\sum_{t=1}^{T} \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)' \right) \right) \\
&= \left\{ {}_{1}^{T}S_{1}^{T} \right\} - \frac{1}{T} W_{(l+1)} \sum_{t=1}^{T} \left(M_{(l)}^{-1} W_{(l)}' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right) \left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)' \right) \\
&= \left(I_N - W_{(l+1)} M_{(l)}^{-1} W_{(l)}' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} \right) \left\{ {}_{1}^{T}S_{1}^{T} \right\} .
\end{aligned}
$$

Again, the term $M_{(l)}^{-1} W_{(l)}' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1}$ is substituted by the matrix $D_{(l)}$, which results in the final expression.

Before we consider the matrices $A_{i_{(l+1)}}, 1 \leq i \leq p$, we simplify the matrix $\tilde{\Sigma}_{p,T}$:

$$
\begin{aligned}
\frac{1}{T-p} \tilde{\Sigma}_{p,T} &= \frac{1}{T-p} \sum_{t=p+1}^{T} \begin{bmatrix} \Sigma_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} & O_K & \cdots & O_K \\ O_K & \Sigma_{\boldsymbol{F}_{t-2}|\boldsymbol{X}_{t-2}} & & \vdots \\ \vdots & & \ddots & O_K \\ O_K & \cdots & O_K & \Sigma_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \end{bmatrix} \\
&\quad + \frac{1}{T-p} \sum_{t=p+1}^{T} \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} & \cdots & \boldsymbol{\mu}_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\mu}_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} & \cdots & \boldsymbol{\mu}_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \end{bmatrix} \\
&= \left(I_p \otimes M_{(l)}^{-1} \right) + \begin{bmatrix} M_{(l)}^{-1} W_{(l)}' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} \left\{ {}_{p}^{T-1}S_{p}^{T-1} \right\} \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} W_{(l)} M_{(l)}^{-1} & \cdots & M_{(l)}^{-1} W_{(l)}' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} \left\{ {}_{p}^{T-1}S_{1}^{T-p} \right\} \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} W_{(l)} M_{(l)}^{-1} \\ \vdots & \ddots & \vdots \\ M_{(l)}^{-1} W_{(l)}' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} \left\{ {}_{1}^{T-p}S_{p}^{T-1} \right\} \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} W_{(l)} M_{(l)}^{-1} & \cdots & M_{(l)}^{-1} W_{(l)}' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} \left\{ {}_{1}^{T-p}S_{1}^{T-p} \right\} \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} W_{(l)} M_{(l)}^{-1} \end{bmatrix} \\
&= \left(I_p \otimes M_{(l)}^{-1} \right) + \left(I_p \otimes \left(M_{(l)}^{-1} W_{(l)}' \Sigma_{\boldsymbol{\epsilon}(l)}^{-1} \right) \right) \hat{S} \left(I_p \otimes \left(\Sigma_{\boldsymbol{\epsilon}(l)}^{-1} W_{(l)} M_{(l)}^{-1} \right) \right) \\
&= \left(I_p \otimes M_{(l)}^{-1} \right) + \left(I_p \otimes D_{(l)} \right) \hat{S} \left(I_p \otimes D_{(l)}' \right) .
\end{aligned}
$$

Next, we have for matrices $A_{i_{(l+1)}}$ with $1 \leq i \leq p$:

$$
\begin{aligned}
\begin{bmatrix} A_{1_{(l+1)}} & \cdots & A_{p_{(l+1)}} \end{bmatrix} &= \left(\sum_{t=p+1}^{T} \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}} & \cdots & \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} \boldsymbol{\mu}'_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}} \end{bmatrix} \right) \tilde{\Sigma}_{p,T}^{-1} \\
&= (T-p) \left[D_{(l)} \left\{ {}_{p+1}^{T}S_{p}^{T-1} \right\} D_{(l)}' \cdots D_{(l)} \left\{ {}_{p+1}^{T}S_{1}^{T-p} \right\} D_{(l)}' \right] \tilde{\Sigma}_{p,T}^{-1} \\
&= \left(\mathbb{1}_p' \otimes D_{(l)} \right) \tilde{S}' \left(I_p \otimes D_{(l)}' \right) \left(\frac{1}{T-p} \tilde{\Sigma}_{p,T} \right)^{-1} \\
&= \left(\mathbb{1}_p' \otimes D_{(l)} \right) \tilde{S}' \left(I_p \otimes D_{(l)}' \right) \cdot \left(\left(I_p \otimes M_{(l)}^{-1} \right) + \left(I_p \otimes D_{(l)} \right) \hat{S} \left(I_p \otimes D_{(l)}' \right) \right)^{-1} .
\end{aligned}
$$

Finally, Lemma 4.1.7 provides for matrix $\Sigma_{\boldsymbol{\delta}(l+1)}$:

$$\begin{aligned}\Sigma_{\boldsymbol{\delta}(l+1)} &= -\frac{1}{T-p}\left[A_{1_{(l+1)}} \quad \cdots \quad A_{p_{(l+1)}}\right]\left(\sum_{t=p+1}^{T}\begin{bmatrix}\boldsymbol{\mu}_{\boldsymbol{F}_{t-1}|\boldsymbol{X}_{t-1}}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\\ \vdots \\ \boldsymbol{\mu}_{\boldsymbol{F}_{t-p}|\boldsymbol{X}_{t-p}}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\end{bmatrix}\right)\\ &\quad + \frac{1}{T-p}\sum_{t=p+1}^{T}\left(\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}\right)\\ &= -\left[A_{1_{(l+1)}}\cdots A_{p_{(l+1)}}\right]\left(\left(\mathbb{1}'_p\otimes D_{(l)}\right)\tilde{S}'\left(I_p\otimes D'_{(l)}\right)\right)'\\ &\quad + M^{-1} + D_{(l)}\left\{{}_{p+1}^{T}S_{p+1}^{T}\right\}D'_{(l)}.\end{aligned}$$

If we apply the matrix transpose to the inside of the brackets in the first row, the final version of matrix $\Sigma_{\boldsymbol{\delta}(l+1)}$ appears. □

As for iid factors, we must specify, when the updates in Theorem 4.1.9 stop. In (3.11), the absolute value of the relative change in the log-likelihood function was our termination criterion. Now, the hidden factors stay in the log-likelihood function in Lemma 4.1.5. This is why we modify (3.11) as follows:

$$\frac{abs\left(\mathbb{E}_{\hat{\Theta}_{(l+1)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l+1)}|X,F\right)|X\right] - \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l)}|X,F\right)|X\right]\right)}{\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l)}|X,F\right)|X\right]} < \eta, \tag{4.14}$$

with $\eta > 0$ as prespecified threshold. Doz et al. (2012) and Bańbura and Modugno (2014) only proceed, if the expected log-likelihood function improves and so, set $\eta = 10^{-4}$. We are less strict and use $\eta = 10^{-2}$, since we consider the absolute value of the change instead of the change itself.

Although we do not show it here, the EM in Theorem 4.1.9 can be adjusted such that the linear restrictions in Bork (2009) and Bork et al. (2010) are included. For this purpose, we have to replace the update formula of $W_{(l+1)}$ as in Bańbura and Modugno (2014) and reapply our subsequent steps.

The ADFM formulation in Definition 2.1.4, in particular, its estimation requires knowledge of the factor dimension and lag order. In empirical studies, both are a priori unknown and hence, must be determined. As a solution, we propose a two-step selection method. Bai and Ng (2002) investigated in detail how to choose the factor dimension for SFMs. As they considered diverse information and panel criteria, they offered a range of tools, which are common in empirical analyses, even though they may behave differently regarding small samples (Breitung and Eickmeier, 2006; Reis and Watson, 2007; Bai and Ng, 2008b; Bork, 2009; Stock and Watson, 2011). Thereafter, Amengual and Watson (2007) extended the approach of Bai and Ng (2002) to DFMs. Here, we adjust a panel criterion of Bai and Ng (2002) as follows:

Lemma 4.1.10 (Selection of Factor Dimension in SFMs)
For SFMs in Definition 2.1.3, the matrix $X \in \mathbb{R}^{T\times N}$ collects the complete panel data in Definition 2.1.1. Let assumptions A-D in Bai and Ng (2002, p. 196) be satisfied. As the true factor dimension is unknown, let $1 \leq \bar{K} \leq N$ be an upper limit to be tested. Then, we select the following K^*:

$$K^* = \underset{1\leq K\leq \bar{K}}{\arg\min}\left\{V(K) + K\hat{\sigma}^2\left(\frac{N+T}{NT}\right)ln\left[\min(N,T)\right]\right\}. \tag{4.15}$$

For multiplier $m \geq 0$ and for all $1 \leq K \leq \bar{K}$, we have:

$$V(K) = \frac{1}{NT}\sum_{t=1}^{T}\left(\boldsymbol{X}_t - \hat{W}\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} - \hat{\boldsymbol{\mu}}_{\boldsymbol{X}}\right)'\left(\boldsymbol{X}_t - \hat{W}\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} - \hat{\boldsymbol{\mu}}_{\boldsymbol{X}}\right). \tag{4.16}$$

$$\hat{\sigma}^2 = m\left(V_{PPCA}\left(1\right) - V_{PPCA}\left(N-1\right)\right), \tag{4.17}$$

with $V_{PPCA}\left(K\right)$ as the empirical variance of all residuals for a K-dimensional ESFM in Theorem 3.1.3, that means:

$$V_{PPCA}\left(K\right) = \frac{1}{NT}\sum_{t=1}^{T}\left(\boldsymbol{X}_t - \hat{W}\ \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} - \hat{\boldsymbol{\mu}}_{\boldsymbol{X}}\right)'\left(\boldsymbol{X}_t - \hat{W}\ \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} - \hat{\boldsymbol{\mu}}_{\boldsymbol{X}}\right).$$

Then, K^* consistently estimates the true factor dimension K.

Proof:

Bai and Ng (2002, p. 199, Theorem 2) showed that panel criteria defined by $PC(K) = V(K) + Kg\left(N,T\right)$ consistently estimate the true factor dimension, if assumptions A-D are satisfied, PCA is used for factor estimation and the penalty function obeys for $N, T \to \infty$:

$$g\left(N,T\right) \to 0 \quad \text{and} \quad \min\left(N,T\right) g\left(N,T\right) \to \infty.$$

Here, we have for the penalty function: $g\left(N,T\right) = \hat{\sigma}^2\left(\frac{N+T}{NT}\right) ln\left[\min\left(N,T\right)\right]$. Thus, the criterion for K^* in (4.15) coincides with the second panel criterion in Bai and Ng (2002, p. 201) except for $\hat{\sigma}^2$. Neither our version of $\hat{\sigma}^2$ nor the original proposal in Bai and Ng (2002), i.e., $\hat{\sigma}^2 = V\left(\bar{K}\right)$, affects the asymptotic behavior of the function $g\left(N,T\right)$ such that K^* in (4.15) consistently estimates the true dimension. □

Before we select the lag order p, we explain the ideas behind the above adjustments. For factor dimension K, the term $V\left(K\right)$ in (4.16) denotes the empirical variance of all residuals in Definition 2.1.3. Regarding empirical studies, Bai and Ng (2002) suggested $\hat{\sigma}^2 = V\left(\bar{K}\right)$ as proper scaling of the penalty in (4.15) with $V\left(\bar{K}\right)$ as minimum of (4.16) for fixed $\bar{K}$ with respect to $\hat{W}$, $\{\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\}_{1\le t\le T}$ and $\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}$. So, their penalty depends on the variance that remains, although the upper limit of the factor dimension was reached. If we use $\bar{K} = N$, the setting $\hat{W} = I_N, \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t} = \boldsymbol{X}_t - \hat{\boldsymbol{\mu}}_{\boldsymbol{X}}$ for all $1 \le t \le T$ provides a trivial solution for SFMs in Definition 2.1.3. Furthermore, it yields $\hat{\sigma}^2 = 0$ and thus, overrides the penalty. For any $\bar{K} < N$ the choice of $\bar{K}$ affects $\hat{\sigma}^2$ and hence, the penalty in (4.15). To avoid any undesirable degree of freedom arising from the choice of $\bar{K}$ we therefore propose the version in (4.17).

Irrespective of whether PCA or PPCA is deployed, the error variance decreases, when the factor dimension increases. Thus, $V_{PPCA}(1) - V_{PPCA}(N-1) \ge 0$ holds. The nonnegativity of m causes that $\hat{\sigma}^2$ in (4.17) and the penalty in (4.15) are nonnegative. This guarantees that large K are punished. Unlike $\hat{\sigma}^2 = V\left(\bar{K}\right)$, the strictness of $\hat{\sigma}^2$ depends on m instead of $\bar{K}$. Hence, the strictness of the penalty and the upper limit of the factor dimension are clearly separated from each other. The panel criteria of Bai and Ng (2002) are asymptotically equivalent as $N, T \to \infty$, but may differently behave for finite samples (Bai and Ng, 2002; Reis and Watson, 2007). For a better understanding of how m influences the penalty function, we exemplarily consider various multipliers $m \in [1/66, 1]$ in Section 4.4.

Finally, we answer why $\left(V_{PPCA}\left(1\right) - V_{PPCA}\left(N-1\right)\right)$ instead of $V_{PPCA}\left(1\right)$ is used. For $m = 1/\left(N-2\right)$ the term $\hat{\sigma}^2$ in (4.17) coincides with the negative slope of the straight line through the points $(1, V_{PPCA}(1))$ and $(N-1, V_{PPCA}(N-1))$. That is, we linearize the decay in $V_{PPCA}(K)$ over the interval $[1, N-1]$ and then, take its absolute value for penalty adjustment. In other words, for $m = 1/\left(N-2\right)$ the term $\hat{\sigma}^2$ in (4.17) describes the absolute value of the decay in $V_{PPCA}(K)$ per unit in dimension. In the empirical study of Section 4.5, we also use $m = 1/31 = 1/(N-2)$, since this provides a decent dimension reduction, but it is not such restrictive that changes in the economy are ignored. Regarding the DFMs in Definition

2.1.4, we neglect the factor dynamics and restrict ourselves to the measurement equation in (2.4). In this way, DFMs are treated as SFMs.

Next, we focus on the choice of the autoregressive order. As before, the true parameter $p \geq 0$ is unknown. In contrast to the factor dimension, we must address the VAR structure to define an optimal lag order $p^*(K) \geq 0$ from a fixed set of candidates $0 \leq p(K) \leq \bar{p}(K)$. Since the factors are unobserved, we adapt the usual AIC, when we replace the log-likelihood function by its expectation $\mathbb{E}_{\hat{\Theta}_{(l)}}[\mathcal{L}_F\left(\hat{\Theta}_{(l)}|F\right)|X]$.

Lemma 4.1.11 (Choice of Autoregressive Order)
For the ADFMs in Definition 2.1.4 with $X \in \mathbb{R}^{T\times N}$ as in Definition 2.1.1, let $1 \leq K \leq N$ be any fixed (estimated) factor dimension. Moreover, let $0 \leq \bar{p}(K) < T$ be an upper limit of the lag order to be tested. Then, based on the adjusted AIC we choose the following autoregressive order $p^*(K)$:

$$\begin{aligned} p^*(K) = \underset{0\leq p\leq \bar{p}(K)}{\arg\min} \Big\{ & tr\left(\Sigma_{\tilde{\boldsymbol{F}}}^{-1}(I_p\otimes D)\left(\tilde{\boldsymbol{X}}_p - \left(\mathbb{1}_p \otimes \hat{\boldsymbol{\mu}}_{\boldsymbol{X}}\right)\right)\left(\tilde{\boldsymbol{X}}_p - \left(\mathbb{1}_p \otimes \hat{\boldsymbol{\mu}}_{\boldsymbol{X}}\right)\right)'(I_p\otimes D')\right) \\ & + tr\left(\Sigma_{\tilde{\boldsymbol{F}}}^{-1}\left(I_p\otimes M^{-1}\right)\right) + 2pK^2 + K(K+1) + TK\ ln(2\pi) \\ & + (T-p)\ ln\left(\left|\Sigma_{\boldsymbol{\delta}}\right|\right) + ln\left(\left|\Sigma_{\tilde{\boldsymbol{F}}}\right|\right) + (T-p)K\Big\}. \end{aligned}$$

The vector $\tilde{\boldsymbol{X}}_p = [\boldsymbol{X}_p', \ldots, \boldsymbol{X}_1']' \in \mathbb{R}^{pN}$ comprises the first p rows of X. The matrices M, D and $\Sigma_{\tilde{\boldsymbol{F}}}$ result from Theorems 4.1.4 and 4.1.9 and Lemma A.2.6, after the EM in Theorem 4.1.9 stopped. Note, for clarity reasons, we skipped the loop index (l) in the above formulation. The estimated mean $\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}$ is given in Definition 2.1.2. For the static case $p = 0$, the argument of $\arg\min\{\cdot\}$ *is equal to:*

$$K(K+1) + TK\ ln(2\pi) + T\ ln\left(\left|\Sigma_{\boldsymbol{\delta}}\right|\right) + TK.$$

Proof:
If the factors $\boldsymbol{F}_t \in \mathbb{R}^K$ are observable, Akaike (1987, p. 323) yields for the AIC of the VAR in (2.5):

$$p^* = \underset{0\leq p\leq \bar{p}}{\arg\min} \left\{ 2\left(pK^2 + \frac{K(K+1)}{2} - \mathcal{L}_F\left(\hat{\Theta}_{(l)}|F\right)\right)\right\}, \tag{4.18}$$

where $\mathcal{L}_F\left(\hat{\Theta}_{(l)}|F\right)$ is the maximized log-likelihood function of the estimated factor VAR(p) given by:

$$\begin{aligned} \mathcal{L}_F\left(\hat{\Theta}_{(l)}|F\right) &= ln\left(f_{\hat{\Theta}_{(l)}}(\boldsymbol{F}_T, \ldots, \boldsymbol{F}_1)\right) = ln\left(f_{\hat{\Theta}_{(l)}}(\boldsymbol{F}_T|\boldsymbol{F}_{T-1}, \ldots, \boldsymbol{F}_1)\, f_{\hat{\Theta}_{(l)}}(\boldsymbol{F}_{T-1}, \ldots, \boldsymbol{F}_1)\right) \\ &= ln\left(f_{\hat{\Theta}_{(l)}}(\boldsymbol{F}_T|\boldsymbol{F}_{T-1}, \ldots, \boldsymbol{F}_{T-p})\, f_{\hat{\Theta}_{(l)}}(\boldsymbol{F}_{T-1}, \ldots, \boldsymbol{F}_1)\right) \\ &= ln\left(\left(\prod_{t=p+1}^{T} f_{\hat{\Theta}_{(l)}}(\boldsymbol{F}_t|\boldsymbol{F}_{t-1}, \ldots, \boldsymbol{F}_{t-p})\right) f_{\hat{\Theta}_{(l)}}(\boldsymbol{F}_p, \ldots, \boldsymbol{F}_1)\right) \\ &= \sum_{t=p+1}^{T} ln\left(f_{\hat{\Theta}_{(l)}}(\boldsymbol{F}_t|\boldsymbol{F}_{t-1}, \ldots, \boldsymbol{F}_{t-p})\right) + ln\left(f_{\hat{\Theta}_{(l)}}(\boldsymbol{F}_p, \ldots, \boldsymbol{F}_1)\right) \\ &= \sum_{t=p+1}^{T} ln\left(f_{\hat{\Theta}_{(l)}}(\boldsymbol{F}_t|\boldsymbol{F}_{t-1}, \ldots, \boldsymbol{F}_{t-p})\right) + ln\left(f_{\hat{\Theta}_{(l)}}\left(\tilde{\boldsymbol{F}}_p\right)\right), \end{aligned}$$

with $f_{\hat{\Theta}_{(l)}}(\boldsymbol{F}_t|\boldsymbol{F}_{t-1}, \ldots, \boldsymbol{F}_{t-p})$ defined in (4.2). For $\tilde{\boldsymbol{F}}_p = [\boldsymbol{F}_p', \ldots, \boldsymbol{F}_1']' \in \mathbb{R}^{pK}$, we convert the VAR($p$) in a VAR(1) as in Lemma A.2.2 such that we obtain $f_{\hat{\Theta}_{(l)}}\left(\tilde{\boldsymbol{F}}_p\right)$ as in Lemma A.2.6. For both distributions, we substitute the model parameters by $\hat{\Theta}_{(l)}$. In the subsequent derivation, we omit the loop index (l) and the hat symbol of $W, \Sigma_{\boldsymbol{\epsilon}}, A_i, 1 \leq i \leq p$, and $\Sigma_{\boldsymbol{\delta}}$ for better readability and proceed as follows:

$$
\begin{aligned}
&\mathcal{L}_F\left(\hat{\Theta}_{(l)}|F\right)\\
&=\sum_{t=p+1}^{T} ln\left((2\pi)^{-K/2}\left|\Sigma_{\boldsymbol{\delta}}\right|^{-1/2} exp\left(\frac{-\left(\boldsymbol{F}_t-\sum_{i=1}^{p}\left(A_i\boldsymbol{F}_{t-i}\right)\right)'\Sigma_{\boldsymbol{\delta}}^{-1}\left(\boldsymbol{F}_t-\sum_{i=1}^{p}\left(A_i\boldsymbol{F}_{t-i}\right)\right)}{2}\right)\right)\\
&\quad+ln\left((2\pi)^{-pK/2}\left|\Sigma_{\tilde{\boldsymbol{F}}}\right|^{-1/2} exp\left(\frac{-\tilde{\boldsymbol{F}}'_p\Sigma_{\tilde{\boldsymbol{F}}}^{-1}\tilde{\boldsymbol{F}}_p}{2}\right)\right)\\
&=-\frac{1}{2}\left(TK\ ln\left(2\pi\right)+\left(T-p\right)\ ln\left(\left|\Sigma_{\boldsymbol{\delta}}\right|\right)+\ ln\left(\left|\Sigma_{\tilde{\boldsymbol{F}}}\right|\right)+\tilde{\boldsymbol{F}}'_p\Sigma_{\tilde{\boldsymbol{F}}}^{-1}\tilde{\boldsymbol{F}}_p\right)\\
&\quad-\frac{1}{2}tr\left(\Sigma_{\boldsymbol{\delta}}^{-1}\sum_{t=p+1}^{T}\left(\boldsymbol{F}_t-\sum_{i=1}^{p}\left(A_i\boldsymbol{F}_{t-i}\right)\right)\left(\boldsymbol{F}_t-\sum_{i=1}^{p}\left(A_i\boldsymbol{F}_{t-i}\right)\right)'\right).
\end{aligned}
\tag{4.19}
$$

Unfortunately, the factors are hidden. To estimate them, we take the expectation of (4.18) given the panel data X and latest parameter estimates $\hat{\Theta}_{(l)}$. Thus, we integrate them out of the log-likelihood function and it follows for our modified AIC:

$$
p^*=\underset{0\leq p\leq\bar{p}}{\arg\min}\left\{2\left(pK^2+\frac{K\left(K+1\right)}{2}-\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}_F\left(\hat{\Theta}_{(l)}|F\right)|X\right]\right)\right\}. \tag{4.20}
$$

In case of (4.19), the expectation and the matrix trace properties in Lemma A.1.2 yield:

$$
\begin{aligned}
&\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}_F\left(\hat{\Theta}_{(l)}|F\right)|X\right]\\
&\quad=-\frac{1}{2}\left(TK\ ln\left(2\pi\right)+\left(T-p\right)\ ln\left(\left|\Sigma_{\boldsymbol{\delta}}\right|\right)+ln\left(\left|\Sigma_{\tilde{\boldsymbol{F}}}\right|\right)+\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\tilde{\boldsymbol{F}}'_p\Sigma_{\tilde{\boldsymbol{F}}}^{-1}\tilde{\boldsymbol{F}}_p|X\right]\right)\\
&\quad-\frac{1}{2}tr\left(\Sigma_{\boldsymbol{\delta}}^{-1}\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\left(\boldsymbol{F}_t-\sum_{i=1}^{p}\left(A_i\boldsymbol{F}_{t-i}\right)\right)\left(\boldsymbol{F}_t-\sum_{i=1}^{p}\left(A_i\boldsymbol{F}_{t-i}\right)\right)'\middle|X\right]\right)\\
&\quad=-\frac{1}{2}\left(TK\ ln\left(2\pi\right)+\left(T-p\right)\ ln\left(\left|\Sigma_{\boldsymbol{\delta}}\right|\right)+ln\left(\left|\Sigma_{\tilde{\boldsymbol{F}}}\right|\right)+tr\left(\Sigma_{\tilde{\boldsymbol{F}}}^{-1}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\tilde{\boldsymbol{F}}_p\tilde{\boldsymbol{F}}'_p|X\right]\right)\right)\\
&\quad-\frac{1}{2}\ tr\left(\Sigma_{\boldsymbol{\delta}}^{-1}\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\left(\boldsymbol{F}_t-\sum_{i=1}^{p}\left(A_i\boldsymbol{F}_{t-i}\right)\right)\left(\boldsymbol{F}_t-\sum_{i=1}^{p}\left(A_i\boldsymbol{F}_{t-i}\right)\right)'|X\right]\right).
\end{aligned}
$$

First, note that the solution of $\hat{\Sigma}_{\boldsymbol{\delta}}$ in Lemma 4.1.7 is equal to the following representation:

$$
\hat{\Sigma}_{\boldsymbol{\delta}}=\frac{1}{T-p}\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\left(\boldsymbol{F}_t-\sum_{i=1}^{p}\left(A_i\boldsymbol{F}_{t-i}\right)\right)\left(\boldsymbol{F}_t-\sum_{i=1}^{p}\left(A_i\boldsymbol{F}_{t-i}\right)\right)'|X\right].
$$

In addition, the independence of the factors conditioned on X (see Theorem 4.1.4) provides:

$$
\begin{aligned}
\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\tilde{\boldsymbol{F}}_p\tilde{\boldsymbol{F}}'_p|X\right]&=\mathbb{V}\mathrm{ar}_{\hat{\Theta}_{(l)}}\left[\tilde{\boldsymbol{F}}_p|X\right]+\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\tilde{\boldsymbol{F}}_p|X\right]\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\tilde{\boldsymbol{F}}'_p|X\right]\\
&=\begin{bmatrix}\Sigma_{\boldsymbol{F}_p|\boldsymbol{X}_p} & & O_K\\ & \ddots & \\ O_K & & \Sigma_{\boldsymbol{F}_1|\boldsymbol{X}_1}\end{bmatrix}+\begin{bmatrix}\boldsymbol{\mu}_{\boldsymbol{F}_p|\boldsymbol{X}_p}\\ \vdots\\ \boldsymbol{\mu}_{\boldsymbol{F}_1|\boldsymbol{X}_1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\mu}_{\boldsymbol{F}_p|\boldsymbol{X}_p} & \cdots & \boldsymbol{\mu}_{\boldsymbol{F}_1|\boldsymbol{X}_1}\end{bmatrix}\\
&=\left(I_p\otimes M^{-1}\right)+\begin{bmatrix}D\left(\boldsymbol{X}_p-\boldsymbol{\mu}\right)\\ \vdots\\ D\left(\boldsymbol{X}_1-\boldsymbol{\mu}\right)\end{bmatrix}\begin{bmatrix}\left(\boldsymbol{X}_p-\boldsymbol{\mu}\right)'D' & \cdots & \left(\boldsymbol{X}_1-\boldsymbol{\mu}\right)'D'\end{bmatrix}\\
&=\left(I_p\otimes M^{-1}\right)+\left(I_p\otimes D\right)\left(\tilde{\boldsymbol{X}}_p-\left(\mathbb{1}_p\otimes\boldsymbol{\mu}\right)\right)\left(\tilde{\boldsymbol{X}}_p-\left(\mathbb{1}_p\otimes\boldsymbol{\mu}\right)\right)'\left(I_p\otimes D'\right),
\end{aligned}
$$

with matrices $M = M_{(l)}$ and $D = D_{(l)}$ as in Theorems 4.1.4 and 4.1.9. The loop index (l) is skipped once again. By inserting both results and substituting $\boldsymbol{\mu}$ by $\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}$, we obtain:

$$\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}_F\left(\hat{\Theta}_{(l)}|F\right)|X\right] = -\frac{1}{2}\left(TK\ ln\left(2\pi\right) + \left(T-p\right)\ ln\left(\left|\Sigma_{\boldsymbol{\delta}}\right|\right) + ln\left(\left|\Sigma_{\tilde{\boldsymbol{F}}}\right|\right) + \left(T-p\right)K\right)$$
$$-\frac{1}{2}\ tr\left(\Sigma_{\tilde{\boldsymbol{F}}}^{-1}\left(\left(I_p\otimes M^{-1}\right) + \left(I_p\otimes D\right)\left(\tilde{\boldsymbol{X}}_p - \left(\mathbb{1}_p\otimes\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}\right)\right)\left(\tilde{\boldsymbol{X}}_p - \left(\mathbb{1}_p\otimes\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}\right)\right)'\left(I_p\otimes D'\right)\right)\right).$$

Finally, we insert the above expression for $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}_F\left(\hat{\Theta}_{(l)}|F\right)|X\right]$ in (4.20) and the claimed solution for $p>0$ follows. In the static case, i.e., $p=0$, we have $\boldsymbol{F}_t\sim\mathcal{N}\left(\boldsymbol{0}_K,\Sigma_{\boldsymbol{\delta}}\right)$ iid and get:

$$\begin{aligned}
\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}_F\left(\hat{\Theta}_{(l)}|F\right)|X\right] &= \sum_{t=1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[ln\left(f_{\hat{\Theta}_{(l)}}\left(\boldsymbol{F}_t\right)\right)|X\right]\\
&= \sum_{t=1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[ln\left(\left(2\pi\right)^{-K/2}\left|\Sigma_{\boldsymbol{\delta}}\right|^{-1/2}exp\left(-\frac{1}{2}\boldsymbol{F}_t'\Sigma_{\boldsymbol{\delta}}^{-1}\boldsymbol{F}_t\right)\right)|X\right]\\
&= -\frac{T}{2}\left(K\ ln\left(2\pi\right) + ln\left(\left|\Sigma_{\boldsymbol{\delta}}\right|\right)\right) - \frac{1}{2}\ tr\left(\Sigma_{\boldsymbol{\delta}}^{-1}\sum_{t=1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{F}_t\boldsymbol{F}_t'|X\right]\right)\\
&= -\frac{T}{2}\left(K\ ln\left(2\pi\right) + ln\left(\left|\Sigma_{\boldsymbol{\delta}}\right|\right) + K\right).
\end{aligned} \tag{4.21}$$

In the static case $K(K+1)/2$ parameters are estimated such that it holds for the argument of (4.20):

$$K\left(K+1\right) + TK\ ln\left(2\pi\right) + T\ ln\left(\left|\Sigma_{\boldsymbol{\delta}}\right|\right) + TK,$$

which proves the assertion for $p=0$. □

The matrix M in Theorem 4.1.4 requires the matrix $\Sigma_{\boldsymbol{F}}$. If the MA(∞) representation of $\Sigma_{\boldsymbol{F}}$ in Lemma 4.1.2 is used, we truncate the infinite series as soon as the contribution of the matrix, which was recently added, falls below the relative threshold $\eta_{\boldsymbol{F}}$. Similarly, the criterion $\eta_{\tilde{\boldsymbol{F}}}$ cuts the MA(∞) series of $\Sigma_{\tilde{\boldsymbol{F}}}$ in (A.4). Alternatively, the explicit solution in Lemma A.2.7 can be deployed.

As in Chapter 3, we summarize all steps in an algorithm and comment its properties. Thereby, Theorem 4.1.9, Lemmata 4.1.10 and 4.1.11 merge in Algorithm 4.1.1 estimating DFMs with complete panel data. Although we separate the choice of the factor dimension from the lag length, which simplifies the overall optimization, we might get stuck in a local maximum as any other EM. Let $R\in\mathbb{R}^{K\times K}$ be an invertible matrix. Then, the sets $\{WR^{-1},\Sigma_{\boldsymbol{\epsilon}},RA_1R^{-1},\dots,RA_pR^{-1},R\Sigma_{\boldsymbol{\delta}}R'\}$ and $\{W,\Sigma_{\boldsymbol{\epsilon}},A_1,\dots,A_p,\Sigma_{\boldsymbol{\delta}}\}$ define the same DFM in Definition 2.1.4. Thus, the output of Algorithm 4.1.1 is unique except for any invertible, linear transformation (Bańbura and Modugno, 2014, working paper). To prevent the runtime of Algorithm 4.1.1 from this parameter ambiguity the criterion in (4.14) controls, when the updates stop. As a solution of the left identification problem, the linear loadings contraints in Bork (2009) and Bork et al. (2010) may serve. However, Dempster et al. (1977) showed that EMs tend toward a particular parameter set and so, we are not ongoingly jumping between possbile results. This argument is repeated in the working paper version of Bańbura and Modugno (2014). For our purposes, i.e., the construction of prediction intervals for returns of future periods of time, the ambiguity of the factors up to a linear transformation R is not essential and will be addressed in Remark 4.2.3. For parameter initialization, Doz et al. (2012) and Bańbura et al. (2014) pursue a two-step approach: At first, they apply PCA for estimating the factors, loadings matrix and error covariance matrix $\Sigma_{\boldsymbol{\epsilon}}$. Thereafter, an OLS provides the VAR(p) parameters. By contrast, we use the PPCA of Tipping and Bishop (1999), which is a special case of the DFMs in Definition 2.1.4, such that our initial values for A_i, $i=1,\dots,p$ are zero matrices.

Algorithm 4.1.1: Estimate ADFMs based on complete panel data

Set relative termination criteria $\eta > 0, \eta_{\boldsymbol{F}} > 0$ and $\eta_{\tilde{\boldsymbol{F}}} > 0$;
Define upper limit of factor dimension $\bar{K}$ and lag length $\bar{p}$;
for $K = 1$ **to** $\bar{K}$ **do**
  for $p = 0$ **to** $\bar{p}$ **do**
    Initialize model parameters using PPCA in (3.4)-(3.5);
    Run EM in Theorem 4.1.9, store (K, p) and estimated parameters $\hat{\Theta}$;
  end
  Determine $p^*(K)$ using Lemma 4.1.11, store $(K, p^*(K))$ and estimated parameters $\hat{\Theta}$;
end
Determine K^* based on Lemma 4.1.10 and pairs $(K, p^*(K))$;

4.1.2 Model Estimation Based on Incomplete Panel Data

In this section, we extend the DFMs in Definition 2.1.4 to admit missing observations and mixed-frequency panel data. In doing so, we combine Definitions 2.1.4 and 2.2.1 in the following manner.

Definition 4.1.12 (Extended Dynamic Factor Model)
For $1 \leq i \leq N$ with $1 \leq T(i) \leq T$, let the vectors $\boldsymbol{X}^i_{obs} \in \mathbb{R}^{T(i)}$ and $\boldsymbol{X}^i \in \mathbb{R}^T$ and the matrix $Q_i \in \mathbb{R}^{T(i)\times T}$ be given as in Definition 2.2.1. Furthermore, the index $1 \leq t \leq T$ maps each point in time, when new data arrives. Then, we have for the extension of the DFMs in Definition 2.1.4:

$$\boldsymbol{X}^i_{obs} = Q_i \boldsymbol{X}^i, \ \ \forall \ \ 1 \leq i \leq N, \tag{4.22}$$

$$\boldsymbol{X}_t = W\boldsymbol{F}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \ \boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}_N, \Sigma_{\boldsymbol{\epsilon}}) \ iid, \ \ \forall \ \ 1 \leq t \leq T, \tag{4.23}$$

$$\boldsymbol{F}_t = \sum_{j=1}^{p} A_j \boldsymbol{F}_{t-j} + \boldsymbol{\delta}_t, \ \boldsymbol{\delta}_t \sim \mathcal{N}\left(\boldsymbol{0}_K, \Sigma_{\boldsymbol{\delta}}\right) \ iid, \ \ \forall \ \ 1 \leq t \leq T, \tag{4.24}$$

with constant parameters $\boldsymbol{\mu} \in \mathbb{R}^N, W \in \mathbb{R}^{N\times K}, \Sigma_{\boldsymbol{\epsilon}} \in \mathbb{R}^{N\times N}$, $A_j \in \mathbb{R}^{K\times K}, 1 \leq j \leq p$, and $\Sigma_{\boldsymbol{\delta}} \in \mathbb{R}^{K\times K}$. The errors in (4.23)-(4.24) are supposed to be independent, that is, $\boldsymbol{\epsilon}_t \perp \boldsymbol{\delta}_s \ \forall \, t, s$. As before, we assume the processes $\{\boldsymbol{X}_t\}$ and $\{\boldsymbol{F}_t\}$ to be covariance-stationary. In addition, the factor process $\{\boldsymbol{F}_t\}$ is supposed to be zero-mean and autoregressive. If the matrix $\Sigma_{\boldsymbol{\epsilon}}$ is diagonal, (4.22)-(4.24) describes an EDFM with incomplete panel data. Otherwise, it belongs to the ADFMs with incomplete panel data.

Assume complete, artificial panel data $X \in \mathbb{R}^{T\times N}$ as in Definition 2.2.1. Then, Equation (4.22) considers the columns of X separately such that data incompleteness can be taken into account, while (4.23)-(4.24) focus on the rows of X. That is, they reveal how all artificial signals $\boldsymbol{X}_t$ evolve over time. Thereby, cross-sectional correlations among the signals are covered. Similar to Lemma 3.1.7, for any signal $1 \leq i \leq N$ and loop $(l) \geq 0$ we receive the following updates for the artificial panel data:

$$\begin{aligned}\boldsymbol{X}^i_{(l+1)} = \mathbb{E}_{\Theta_{(l)}}\left[\boldsymbol{X}^i | F_{(l)}, \boldsymbol{X}^i_{\text{obs}}\right] = &\left(F_{(l)}\boldsymbol{W}'_{i(l)} + \mu_{i(l)}\mathbb{1}_T\right) \\ &+ Q'_i\left(Q_iQ'_i\right)^{-1}\left[\boldsymbol{X}^i_{\text{obs}} - Q_i\left(F_{(l)}\boldsymbol{W}'_{i(l)} + \mu_{i(l)}\mathbb{1}_T\right)\right].\end{aligned} \tag{4.25}$$

With this in mind, we adapt Algorithm 4.1.1 as shown in Algorithm 4.1.2. At first, complete panel data is initialized, if necessary, gaps are filled as explained earlier. Here, the univariate time series $\boldsymbol{X}^i_{(0)}$ are not needed to satisfy (4.22), since (4.25) ensures this until Algorithm 4.1.2 converges. As before, relative termination criteria diminish potential impact of the dimensions of the parameter space and data sample on the runtime of the algorithm. Furthermore, relative changes in $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l)}|X,F\right)|X\right]$ control, when Algorithm 4.1.2 stops such that neither changes in (K^*,p^*) nor the ambiguity of the parameters affect its convergence. After the initialization phase, Algorithm 4.1.2 alternately updates the complete panel data and reestimates the model parameters $\hat{\Theta}_{(l)}$ until a (local) maximum of the expected log-likelihood function $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l)}|X,F\right)|X\right]$ is reached. Besides the advantages of Algorithm 3.1.2, Algorithm 4.1.2 allows for autoregressive orders $p>0$ and hence, is not restricted to SFMs.

Algorithm 4.1.2: Estimate ADFMs based on incomplete panel data

```
### Initialization
Choose termination criterion ξ > 0;
Set loop index (l) = 0;
for i = 1 to N do
    Initialize X^i_(l) (if necessary, fill gaps);
    Specify matrix Q_i;
end
Estimate ADFM with X_(l) using Algorithm 4.1.1 and store parameters Θ̂_(l);
Determine expected log-likelihood E_Θ̂(l) [L(Θ̂_(l)|X_(l), F)|X_(l)] in (4.12);
for i = 1 to N do
    Derive updated panel data X^i_(l+1) from (4.25) and model parameters Θ̂_(l);
end
Estimate ADFM with X_(l+1) using Algorithm 4.1.1 and store parameters Θ̂_(l+1);
Determine expected log-likelihood E_Θ̂(l+1) [L(Θ̂_(l+1)|X_(l+1), F)|X_(l+1)] in (4.12);
### Alternating reconstruction and reestimation
while abs(E_Θ̂(l+1)[L(Θ̂_(l+1)|X_(l+1),F)|X_(l+1)] − E_Θ̂(l)[L(Θ̂_(l)|X_(l),F)|X_(l)]) / abs(E_Θ̂(l)[L(Θ̂_(l)|X_(l),F)|X_(l)]) > ξ do
    Set loop index (l) = (l + 1);
    for i = 1 to N do
        Derive updated panel data X^i_(l+1) from (4.25) and model parameters Θ̂_(l);
    end
    Estimate ADFM with X_(l+1) using Algorithm 4.1.1 and store parameters Θ̂_(l+1);
    Determine expected log-likelihood E_Θ̂(l+1) [L(Θ̂_(l+1)|X_(l+1), F)|X_(l+1)] in (4.12);
end
```

4.2 Modeling Index Returns

The preceding sections showed how to extract information from large panel data through a few factors with known distributions. In Section 3.1.3, the estimated factor distributions supported portfolio optimizations in a mean-variance and marginal-risk-parity framework, respectively. Now, we restrict ourselves to a single financial market and explain how these may improve asset allocation decisions and risk management. In doing so, we focus on the timely monitoring of financial market indices.

Like Bai and Ng (2006) and Luciani and Ricci (2014), we consider interval estimation. Here, we empirically construct prediction intervals, since the asymptotic ones in Bai and Ng (2006) rely on complete panel data. From our point of view, uncertainties arising from the estimation of the factors and model parameters shall affect the interval width. Therefore, we derive prediction intervals instead of confidence intervals as in Luciani and Ricci (2014). Additionally, we disclose the drivers of forecasted point estimates to open the door for further plausibility assessments. As any problems resulting from incomplete data were solved before, we assume that the updating frequencies of the factors and returns coincide. With this in mind, we introduce a univariate return process $\{r_t\}$ as follows:

Definition 4.2.1 (Returns of Financial Market Index)
Let $\{\boldsymbol{F}_t\}$ be the K-dimensional factor process in Definitions 2.1.3 and 2.1.4, respectively, and let $p \geq 0$ be its autoregressive order. The return process $\{r_t\}$ is supposed to be covariance-stationary and to satisfy a univariate ARX with the latent factors as exogenous variables. Let $0 \leq \tilde{q}$ and $0 \leq \tilde{p} \leq \max(1, p)$ be the lags of the returns and factors, respectively. Then, all assumptions result in the subsequent linear model:

$$r_t = \alpha + \sum_{i=1}^{\tilde{q}} (\beta_i r_{t-i}) + \sum_{i=1}^{\tilde{p}} (\boldsymbol{\gamma}_i' \boldsymbol{F}_{t-i}) + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right) iid,$$

where $\alpha, \beta_i, 1 \leq i \leq \tilde{q}, \sigma_\varepsilon^2 \in \mathbb{R}$ with $\sigma_\varepsilon^2 \geq 0$ and $\boldsymbol{\gamma}_i \in \mathbb{R}^K, 1 \leq i \leq \tilde{p}$, are constants. In addition, we assume that the factors and errors are independent, i.e., $\boldsymbol{F}_t \perp \varepsilon_s$ for all s, t and that neither the return process $\{r_t\}$ nor any of its transformations enters the panel data of the factors.

For $\tilde{p} = 0$, the returns in Definition 4.2.1 obey an Autoregressive Model (AR) of order $\tilde{q}$. In case of DFMs, the VAR(p) in (2.5) requires the contraint $0 \leq \tilde{p} \leq p$, otherwise, the ARX parameters in Definition 4.2.1 are not identifiable. For SFMs, $p = 0$ holds, but this may not mean that static factors cannot have any impact on the returns. So, we adapt the original restriction towards $0 \leq \tilde{p} \leq \max(1, p)$. Although $\tilde{p} > 1$ would be possible for static factors, we restrict ourselves to $\tilde{p} = 1$ for simplicity reasons. The exclusion of the returns as well as any transformation of $\{r_t\}$ from the panel data X is another important assumption ensuring the uniqueness of the ARX parameters. For instance, consider the case, when $\{r_t\}$ is a column of X such that it coincides with the first factor and $\tilde{p} = \tilde{q}$ holds.

Similar to (3.13), the vector $\boldsymbol{\theta} = \left[\alpha, \beta_1, \ldots, \beta_{\tilde{q}}, \boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_{\tilde{p}}'\right]' \in \mathbb{R}^{1+\tilde{q}+\tilde{p}K}$ collects the parameters of the ARX($\tilde{q}, \tilde{p}$) in Definition 4.2.1. If $\boldsymbol{r} = [r_1, \ldots, r_T]' \in \mathbb{R}^T$ and $F = [\boldsymbol{F}_1, \ldots, \boldsymbol{F}_T]' \in \mathbb{R}^{T \times K}$ are return and factor samples of the same length and time horizon, for $\tilde{m} = \max\{\tilde{q}, \tilde{p}\}$ the model in Definition 4.2.1 can be rewritten in matrix form as follows:

$$\underbrace{\begin{bmatrix} r_{\tilde{m}+1} \\ \vdots \\ r_T \end{bmatrix}}_{\boldsymbol{r}_T^{\tilde{m}+1}} = \underbrace{\begin{bmatrix} 1 & r_{\tilde{m}} & \cdots & r_{\tilde{m}-\tilde{q}+1} & \boldsymbol{F}_{\tilde{m}}' & \cdots & \boldsymbol{F}_{\tilde{m}-\tilde{p}+1}' \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & r_{T-1} & \cdots & r_{T-\tilde{q}} & \boldsymbol{F}_{T-1}' & \cdots & \boldsymbol{F}_{T-\tilde{p}}' \end{bmatrix}}_{G} \boldsymbol{\theta} + \underbrace{\begin{bmatrix} \varepsilon_{\tilde{m}+1} \\ \vdots \\ \varepsilon_T \end{bmatrix}}_{\boldsymbol{\varepsilon}_T^{\tilde{m}+1}}, \tag{4.26}$$

with $\boldsymbol{r}_T^{\tilde{m}+1} \in \mathbb{R}^{T-\tilde{m}}, G \in \mathbb{R}^{(T-\tilde{m}) \times (1+\tilde{q}+\tilde{p}K)}$ and $\boldsymbol{\varepsilon}_T^{\tilde{m}+1} \in \mathbb{R}^{T-\tilde{m}}$.

Lemma 4.2.2 (Estimation of ARX for Returns)
Assume the matrix form (4.26) of the return ARX in Definition 4.2.1. Then, the OLS estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is

given by $\hat{\boldsymbol{\theta}} = (G'G)^{-1} G'\boldsymbol{r}_T^{\tilde{m}+1}$, *which is asymptotically normal with mean* $\boldsymbol{\theta}$ *and covariance matrix* $\Sigma_{\boldsymbol{\theta}}$. *The parameters of the asymptotic distribution* $\mathcal{N}(\boldsymbol{\theta}, \Sigma_{\boldsymbol{\theta}})$ *are consistently estimated by:*

$$\hat{\boldsymbol{\theta}} = (G'G)^{-1} G'\boldsymbol{r}_T^{\tilde{m}+1} \tag{4.27}$$

$$\text{and} \quad \widehat{\Sigma}_{\boldsymbol{\theta}} = \hat{\sigma}_\varepsilon^2 (G'G)^{-1}, \tag{4.28}$$

$$\text{where} \quad \hat{\sigma}_\varepsilon^2 = \frac{1}{T-\tilde{m}} \left(\boldsymbol{r}_T^{\tilde{m}+1} - G (G'G)^{-1} G'\boldsymbol{r}_T^{\tilde{m}+1}\right)' \left(\boldsymbol{r}_T^{\tilde{m}+1} - G (G'G)^{-1} G'\boldsymbol{r}_T^{\tilde{m}+1}\right). \tag{4.29}$$

Proof:
Hamilton (1994, pp. 215-216, Case 4) shows this statement for autoregressions of order $\tilde{q}$. Using similar steps and the properties of the factor process $\{\boldsymbol{F}_t\}$, the proof for the ARX in Definition 4.2.1 follows. □

Besides uncertainties arising from the estimation of the factors, uncertainties caused by the estimation of the ARX parameters $\boldsymbol{\theta}$ shall drive the width of the prediction intervals. For this purpose, the asymptotic distribution in Lemma 4.2.2 is essential, since our algorithm for the construction of the prediction intervals randomly draws unknown parameter vectors $\hat{\boldsymbol{\theta}}^c$ from it.

Remark 4.2.3 (Impact of Factor Ambiguity on Return ARX)
As mentioned in the scope of Algorithm 4.1.1, factors are unique except for an invertible, linear transformation. For a non-singular matrix $R \in \mathbb{R}^{K\times K}$, *the return dynamics are equivalently represented by:*

$$r_t = \alpha + \sum_{i=1}^{\tilde{q}} (\beta_i r_{t-i}) + \sum_{i=1}^{\tilde{p}} \left(\left(\boldsymbol{\gamma}_i' R^{-1}\right)(R\boldsymbol{F}_{t-i})\right) + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right) \text{ iid},$$

which clearly shows that for all $1 \leq i \leq \tilde{p}$ *the respective vector* $\boldsymbol{\gamma}_i$ *is accordingly adjusted such that* R *does not affect the returns* $\{r_t\}$.

In preparation for Remark 4.2.9, which traces point forecasts of future returns back to the original panel data and estimation risks, we show in Remark 4.2.4 how to substitute the factors $\boldsymbol{F}_t$ in the return ARX by their means and some multivariate Gaussian random variables. To ensure that our prediction intervals capture uncertainties caused by the estimation of the hidden factors, we use random samples drawn from the conditional factor distribution in (4.9) instead of the factor estimates.

Remark 4.2.4 (Decomposition of Factor Impact)
Irrespective of whether SFMs in Definition 2.1.3 or DFMs in Definition 2.1.4 are considered, we receive Gaussian factors for given panel data, i.e., $\boldsymbol{F}_t|\boldsymbol{X}_t \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}, \Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t})$. *Let the matrix* $\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}^{1/2}$ *be the square root matrix of* $\Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t}$ *and assume* $\boldsymbol{Z}_t \sim \mathcal{N}(\boldsymbol{0}_K, I_K)$ *iid with* $\boldsymbol{Z}_t \perp \varepsilon_s$, *for all* t, s. *Then, the return ARX in Definition 4.2.1 can be rewritten as follows:*

$$r_t = \alpha + \sum_{i=1}^{\tilde{q}} (\beta_i r_{t-i}) + \sum_{i=1}^{\tilde{p}} \left(\boldsymbol{\gamma}_i' \left(\boldsymbol{\mu}_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}} + \Sigma_{\boldsymbol{F}_{t-i}|\boldsymbol{X}_{t-i}}^{1/2} \boldsymbol{Z}_{t-i}\right)\right) + \varepsilon_t.$$

When we empirically construct prediction intervals for r_{T+1}, we use MC simulations. Let V denote the number of simulated r_{T+1}. After the conditional distributions in Theorem 4.1.4 have been determined from the parameter estimates of Algorithm 4.1.2, for each trajectory $1 \leq c \leq V$ a randomly drawn sample $\boldsymbol{F}_1^c, \ldots \boldsymbol{F}_T^c$ enters the OLS in Lemma 4.2.2 implying that the distribution of the estimates $\hat{\boldsymbol{\theta}}$ depends on path c. To highlight this, we write $\hat{\boldsymbol{\theta}}^c$, if applicable. In total, this procedure covers both estimation risks despite their nonlinear relation.

If we combine Definitions 2.1.3 or 2.1.4 with the ARX in Definition 4.2.1, the total model has the VAR(1) representation in Lemma 4.2.5 for the process $\{\tilde{\boldsymbol{B}}_t\}$. As shown in Lemma A.2.9, the covariance-stationary process $\{\tilde{\boldsymbol{B}}_t\}$ has a MA(∞) respresentation, which we use in Lemmata A.2.10 and A.2.11 to calculate its mean and covariance matrix. Based on those, we derive in Lemma A.2.12 the mean μ_r and variance σ_r^2 of the returns r_t in Definition 4.2.1. For clarity reasons, we moved the technical Lemmata A.2.9-A.2.12 to the appendix, since we are at this stage interested in mean μ_r and variance σ_r^2 as inputs of the log-likelihood function in Lemma 4.2.6. Therefore, we have:

Lemma 4.2.5 (VAR(1) Representation of ARX)
Let $\{\boldsymbol{F}_t\}$ be the factor process of the SFM in Definition 2.1.3 with $p = 0$ or the DFM in Definition 2.1.4 with $p \geq 1$. For lag orders $0 \leq \tilde{q}$ and $0 \leq \tilde{p} \leq \max(1, p)$ and any point in time t, we define:

$$\tilde{\boldsymbol{B}}_t' = \left(r_t, \ldots, r_{t+1-\max(1,\tilde{q})}, \tilde{\boldsymbol{F}}_t'\right) \in \mathbb{R}^d, \tag{4.30}$$

with $d = \max(1, \tilde{q}) + \max(1, p)\, K$. For $p \geq 1$, the vector $\tilde{\boldsymbol{F}}_t \in \mathbb{R}^{pK}$ is given by Lemma A.2.2. For $p = 0$, we have: $\tilde{\boldsymbol{F}}_t = \boldsymbol{F}_t \in \mathbb{R}^K$. Then, the ARX in Definition 4.2.1 has the following VAR(1) representation:

$$\tilde{\boldsymbol{B}}_t = \underbrace{\begin{bmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\boldsymbol{a}} + \underbrace{\begin{bmatrix} \beta_1 & \cdots & \beta_{\tilde{q}-1} & \beta_{\max(1,\tilde{q})} & \boldsymbol{\gamma}_1' & \cdots & \boldsymbol{\gamma}_{\tilde{p}}' & \mathbf{0}_K' & \cdots & \mathbf{0}_K' \\ 1 & & 0 & 0 & \mathbf{0}_K' & \cdots & \mathbf{0}_K' & \mathbf{0}_K' & \cdots & \mathbf{0}_K' \\ & \ddots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & & 1 & 0 & \mathbf{0}_K' & \cdots & \mathbf{0}_K' & \mathbf{0}_K' & \cdots & \mathbf{0}_K' \\ \mathbf{0}_K & \cdots & \cdots & \mathbf{0}_K & A_1 & \cdots & \cdots & \cdots & A_{p-1} & A_{\max(1,p)} \\ \vdots & & & \vdots & I_K & O_K & \cdots & \cdots & O_K & O_K \\ \vdots & & & \vdots & O_K & \ddots & \ddots & & \vdots & \vdots \\ \vdots & & & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & & \vdots & \vdots & & \ddots & \ddots & O_K & \vdots \\ \mathbf{0}_K & \cdots & \cdots & \mathbf{0}_K & O_K & \cdots & \cdots & O_K & I_K & O_K \end{bmatrix}}_{\mathbb{H}} \tilde{\boldsymbol{B}}_{t-1} + \underbrace{\begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \\ \boldsymbol{\delta}_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\boldsymbol{e}_t},$$

with vector $\boldsymbol{a} \in \mathbb{R}^d$ and matrix $\mathbb{H} \in \mathbb{R}^{d \times d}$ as constants. For the errors $\boldsymbol{e}_t \in \mathbb{R}^d$, it holds: $\boldsymbol{e}_t \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{e}}, \Sigma_{\boldsymbol{e}}\right)$ iid for all points in time t with mean and covariance matrix given by:

$$\boldsymbol{\mu}_{\boldsymbol{e}} = \mathbf{0}_d \quad \text{and} \quad \Sigma_{\boldsymbol{e}} = \begin{bmatrix} \sigma_\varepsilon^2 & 0 & \cdots & 0 & \mathbf{0}_K' & \mathbf{0}_K' & \cdots & \mathbf{0}_K' \\ 0 & 0 & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & \cdots & 0 & \mathbf{0}_K' & \mathbf{0}_K' & \cdots & \mathbf{0}_K' \\ \mathbf{0}_K & \cdots & \cdots & \mathbf{0}_K & \Sigma_{\boldsymbol{\delta}} & O_K & \cdots & O_K \\ \vdots & & & \vdots & O_K & O_K & & \vdots \\ \vdots & & & \vdots & \vdots & & \ddots & \vdots \\ \mathbf{0}_K & \cdots & \cdots & \mathbf{0}_K & O_K & O_K & \cdots & O_K \end{bmatrix}.$$

If $\tilde{p} = p$ holds, there are no zeros in the first row of $\mathbb{H}$. For $p = 0$, we can simplify the matrix $\mathbb{H}$ and *obtain the following formulation:*

$$\mathbb{H} = \begin{bmatrix} \beta_1 & \cdots & \beta_{\tilde{q}-1} & \beta_{\max(1,\tilde{q})} & \gamma_1' \\ 1 & & 0 & 0 & \mathbf{0}_K' \\ & \ddots & & \vdots & \vdots \\ 0 & & 1 & 0 & \mathbf{0}_K' \\ \mathbf{0}_K & \cdots & \cdots & \mathbf{0}_K & O_K \end{bmatrix}.$$

For $\tilde{q} = 0$, we have $\beta_1 = 0$.

Proof:
Follows directly from Definitions 2.1.3, 2.1.4 and 4.2.1 and Lemma A.2.2. □

As shown in Lemmata A.2.9-A.2.11, the process $\{\tilde{\boldsymbol{B}}_t\}$ is covariance-stationary, has a MA(∞) representation, satisfies the absolute summability condition and is normally distributed. Next, we select the lag orders $\tilde{q}$ and $\tilde{p}$. A simple approach applies AIC based on the estimated factor means. We aim to include the estimated factor variance using the distortion in the form of $\boldsymbol{F}_1^c, \ldots \boldsymbol{F}_{\tilde{T}}^c$ to allow for the factors' hiddenness. Therefore, we replace the log-likelihood function in the usual AIC by the log-likelihood function of $\boldsymbol{r}$ conditioned on the factor sample F^c as follows.

Lemma 4.2.6 (Conditional Log-Likelihood Function of ARX)
Let $\boldsymbol{r} = [r_1, \ldots, r_T] \in \mathbb{R}^T$ and $(F^c)' = [\boldsymbol{F}_1^c, \ldots, \boldsymbol{F}_T^c] \in \mathbb{R}^{K\times T}$ be return and factor samples, respectively. For $\tilde{m} = \max(\tilde{q}, \tilde{p})$, the returns are supposed to obey the ARX in Definition 4.2.1 based on F^c, i.e.:

$$r_t = \alpha + \sum_{i=1}^{\tilde{q}} (\beta_i r_{t-i}) + \sum_{i=1}^{\tilde{p}} \left(\boldsymbol{\gamma}_i' \boldsymbol{F}_{t-i}^c\right) + \varepsilon_t, \ \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right) \ iid. \tag{4.31}$$

Then, we have for the log-likelihood function of (4.31) given $\boldsymbol{r}$ and F^c, i.e., $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{r}, F^c)$:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{r}, F^c) = &-\frac{T\ ln(2\pi)}{2} - \frac{1}{2}\sum_{t=(\tilde{m}+1)}^{T}\left(ln\left(\sigma^2_{r_t|Full}\right) + \frac{\left(r_t - \mu_{r_t|Full}\right)^2}{\sigma^2_{r_t|Full}}\right) \\ &- \frac{1}{2}\sum_{t=2}^{\tilde{m}}\left(ln\left(\sigma^2_{r_t|Part}\right) + \frac{\left(r_t - \mu_{r_t|Part}\right)^2}{\sigma^2_{r_t|Part}}\right) - \frac{1}{2}\left(ln\left(\sigma_r^2\right) + \frac{\left(r_t - \mu_r\right)^2}{\sigma_r^2}\right), \end{aligned}$$

where $\mu_r, \mu_{r_t|Part}, \mu_{r_t|Full}, \sigma_r^2, \sigma^2_{r_t|Part}$ and $\sigma^2_{r_t|Full}$ are defined in Lemmata A.2.12-A.2.14.

Proof:
By virtue of the Bayes' theorem we get for the likelihood function of $\boldsymbol{r}$ conditioned on F^c in (4.31):

$$\begin{aligned} f_{\boldsymbol{\theta}}(\boldsymbol{r}|F^c) &= f_{\boldsymbol{\theta}}(r_T, \ldots, r_1|F^c) = \frac{f_{\boldsymbol{\theta}}(r_T, \ldots, r_1, F^c)}{f_{\boldsymbol{\theta}}(F^c)} = \frac{f_{\boldsymbol{\theta}}(r_T|r_{T-1}, \ldots, r_1, F^c)\, f_{\boldsymbol{\theta}}(r_{T-1}, \ldots, r_1, F^c)}{f_{\boldsymbol{\theta}}(F^c)} \\ &= f_{\boldsymbol{\theta}}(r_T|r_{T-1}, \ldots, r_{T-\tilde{q}}, F^c)\, f_{\boldsymbol{\theta}}(r_{T-1}, \ldots, r_1|F^c) \\ &= \left(\prod_{t=(\tilde{m}+1)}^{T} f_{\boldsymbol{\theta}}(r_t|r_{t-1}, \ldots, r_{t-\tilde{q}}, F^c)\right)\left(\prod_{t=2}^{\tilde{m}} f_{\boldsymbol{\theta}}\left(r_t|r_{t-1}, \ldots, r_{\max(1,t-\tilde{q})}, F^c\right)\right) f_{\boldsymbol{\theta}}(r_1|F^c). \end{aligned}$$

Hence, we obtain for the corresponding log-likelihood function $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{r}, F^c)$:

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{r}, F^c) = \sum_{t=(\tilde{m}+1)}^{T} ln\left(f_{\boldsymbol{\theta}}(r_t|r_{t-1}, \ldots, r_{t-\tilde{q}}, F^c)\right) + \sum_{t=2}^{\tilde{m}} ln\left(f_{\boldsymbol{\theta}}\left(r_t|r_{t-1}, \ldots, r_{\max(1,t-\tilde{q})}, F^c\right)\right)$$

$$+\, ln \left(f_{\boldsymbol{\theta}} \left(r_1 | F^c \right) \right) .$$

Next, we insert the distributions in Lemmata A.2.12-A.2.14. In this manner, receive:

$$\begin{aligned}
\mathcal{L}\left(\boldsymbol{\theta}|\boldsymbol{r},F^c\right) = & \sum_{t=(\tilde{m}+1)}^{T} ln \left(\left(2\pi\sigma^2_{r_t|Full}\right)^{-1/2} exp\left(-\frac{\left(r_t - \mu_{r_t|Full}\right)^2}{2\,\sigma^2_{r_t|Full}} \right) \right) \\
& + \sum_{t=2}^{\tilde{m}} ln \left(\left(2\pi\sigma^2_{r_t|Part}\right)^{-1/2} exp\left(-\frac{\left(r_t - \mu_{r_t|Part}\right)^2}{2\,\sigma^2_{r_t|Part}} \right) \right) \\
& + ln \left(\left(2\pi\sigma^2_{r}\right)^{-1/2} exp\left(-\frac{\left(r_t - \mu_{r}\right)^2}{2\,\sigma^2_{r}} \right) \right) .
\end{aligned}$$

Finally, summarizing equal expressions leads to the stated formulation. □

The likelihood $f_{\boldsymbol{\theta}}\left(\boldsymbol{r}|F^c\right)$ in the above proof consists of three constituents. The first group ($\tilde{m}+1 \leq t \leq T$) comprises the mean and variance of r_t, when all required lags are observed and thus, the *Full* lag history is available. Since samples are of finite size, the second group ($2 \leq t \leq \tilde{m}$) covers the mean and variance, when only a few of the necessary lags are given. Thereby, we talk about *Partial* lag information. For the last term $f_{\boldsymbol{\theta}}\left(r_1|F^c\right)$ we do not have any lagged returns or factors such that we have to restrict ourselves to its stationary behavior. Because of its definition, $\tilde{m}$ is the minimal number of lags, which the full lag history calls for. Hence, it determines the upper limit of the second group.

In Lemma 4.2.6, the samples F^c and $\boldsymbol{r} = [r_1, \ldots, r_T]$ have same length. In particular, there is no lead time with information in the form of $\boldsymbol{F}^c_t, t \leq 0$. As this assumption is made for convenience, it can easily be relaxed. Assume there is a run-up period with additional factors $\boldsymbol{F}^c_t, 1-\tilde{p} \leq t \leq 0$. Then, the classification of $f_{\boldsymbol{\theta}}\left(\boldsymbol{r}|F^c\right)$ in three groups remains valid, but the upper and lower limits of the middle group may change. For each former member of the middle group we have to check, whether it still belongs to it because of the additional factors. Similarly, we have to verify, whether there is partial information for the last multiplier $f_{\boldsymbol{\theta}}\left(r_1|F^c\right)$ such that it is treated like the middle group. As soon as the new classification is known, the respective return moments in Lemmata A.2.12-A.2.14 can be applied. By similar reasoning, we can include additional returns $r_t, 1-\tilde{q} \leq t \leq 0$.

Based on the usual AIC in Akaike (1987, p. 323), we now proceed with the following modification:

Remark 4.2.7 (Selection of ARX Lag Orders)
For the setting in Lemma 4.2.6, let $0 \leq K$ and $0 \leq p$ be the dimension and lag order, respectively, of the factor sample F^c. Furthermore, the vector $\boldsymbol{r} \in \mathbb{R}^T$ collects the returns and let $\bar{q} \geq 0$ be the upper limit of the autoregressive order, which we test for the returns. Then, our modified selection criterion chooses the pair $(\tilde{q}^*, \tilde{p}^*)$ as follows:

$$\left(\tilde{q}^*, \tilde{p}^*\right) = \underset{0 \leq \tilde{q} \leq \bar{q}, 0 \leq \tilde{p} \leq p}{\arg\min} \left\{ 2\left(2 + \tilde{q} + \tilde{p}K - \mathcal{L}\left(\hat{\boldsymbol{\theta}}\left(\tilde{q},\tilde{p}\right)|\boldsymbol{r},F^c\right)\right)\right\}, \tag{4.32}$$

where $\mathcal{L}\left(\hat{\boldsymbol{\theta}}\left(\tilde{q},\tilde{p}\right)|\boldsymbol{r},F^c\right)$ denotes the log-likelihood function in Lemma 4.2.6 with the estimated parameters $\hat{\boldsymbol{\theta}}$ in Lemma 4.2.2 for autoregressive orders $(\tilde{q},\tilde{p})$.

Proof:
The ARX in (4.31) involves $2+\tilde{q}+\tilde{p}K$ parameters and $\mathcal{L}\left(\hat{\boldsymbol{\theta}}\left(\tilde{q},\tilde{p}\right)|\boldsymbol{r},F^c\right)$ is the log-likelihood function, which we want to use instead of the usual one. Therefore, the general definition of AIC in Akaike (1987,

p. 323) results in the stated version. □

Remark 4.2.8 (Simplified Conditional Log-Likelihood Function $\mathcal{L}(\hat{\boldsymbol{\theta}}\,(\tilde{q},\tilde{p})\,|\boldsymbol{r},F^c))$
The definition of $\hat{\sigma}^2_\varepsilon$ in (4.29), the equality $\sigma^2_{r_t|Full}=\sigma^2_\varepsilon$ and the definition of $\mu_{r_t|Full}$ for all $\tilde{m}+1\le t\le T$ in Lemma A.2.13 enable us to shorten the expression for $\mathcal{L}\left(\boldsymbol{\theta}|\boldsymbol{r},F^c\right)$ in Lemma 4.2.6 as follows:

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{\theta}|\boldsymbol{r},F^c\right)=&-\frac{1}{2}\left[T\ ln\left(2\pi\right)+\left(T-\tilde{m}\right)\left(1+ln\left(\sigma^2_\varepsilon\right)\right)+ln\left(\sigma^2_r\right)+\frac{\left(r_t-\mu_r\right)^2}{\sigma^2_r}\right]\\&-\frac{1}{2}\sum_{t=2}^{\tilde{m}}\left(ln\left(\sigma^2_{r_t|Part}\right)+\frac{\left(r_t-\mu_{r_t|Part}\right)^2}{\sigma^2_{r_t|Part}}\right).\end{aligned}$$

Similar to $\eta_{\boldsymbol{F}}$ and $\eta_{\tilde{\boldsymbol{F}}}$, the termination criterion $\eta_{\tilde{\boldsymbol{B}}}>0$ truncates the mean and covariance matrix series of $\tilde{\boldsymbol{B}}_t$ in Lemma A.2.10. For reasons of clarity, the panel data $X=[\boldsymbol{X}_1,\ldots,\boldsymbol{X}_T]'$ in Definition 2.1.1 and return sample $\boldsymbol{r}=[r_1,\ldots,r_T]$ have the same length. However, this assumption is not crucial, since the intersection of both samples is otherwise taken. Again, we summarize all steps in an algorithm. As before, Algorithm 4.2.1 adds as much uncertainty as possible to the constructed prediction intervals.

Remark 4.2.9 (Drivers of the 1-Step Ahead Returns)
The mean and covariance matrix of the OLS estimate $\hat{\boldsymbol{\theta}}$ in (4.27)-(4.29) are functions of the factors such that the asymptotic distribution of $\hat{\boldsymbol{\theta}}^c$ in Algorithm 4.2.1 depends on F^c. If we neglect the impact of F^c on the mean and covariance matrix of $\hat{\boldsymbol{\theta}}^c$ for a second, e.g., in case of a sufficiently long sample and little varying factors, we may decompose the forecasted returns as follows:

$$r^c_{T+1}=\underbrace{\bar{\alpha}^c+\sum_{i=1}^{\tilde{q}}\left(\bar{\beta}^c_i r_{T+1-i}\right)}_{AR\ Nature}+\underbrace{\sum_{i=1}^{\tilde{p}}\left(\boldsymbol{w}'_i\left(\boldsymbol{X}_{T+1-i}-\boldsymbol{\mu}\right)\right)}_{Factor\ Impact}+\underbrace{\sum_{i=1}^{\tilde{p}}\left(\left(\bar{\gamma}^c_i\right)'\boldsymbol{Z}^c_{T+1-i}\right)}_{Factor\ Risk}+\underbrace{\hat{\sigma}^c_\varepsilon\,Z^c}_{AR\ Risk}, \tag{4.33}$$

with $\boldsymbol{w}'_i=\left(\bar{\gamma}^c_i\right)'M^{-1}W'\Sigma^{-1}_{\boldsymbol{\epsilon}}\in\mathbb{R}^N$ and $\boldsymbol{Z}^c_{T+1-i}=\boldsymbol{F}^c_{T+1-i}-\boldsymbol{\mu}_{\boldsymbol{F}_{T+1-i}|\boldsymbol{X}_{T+1-i}}\in\mathbb{R}^K$ for all $1\le i\le\tilde{p}$. If neither the returns $\boldsymbol{r}$ nor any transformation of $\boldsymbol{r}$ are part of the panel data X, the distinction between the four pillars in (4.33) is more precise. For the SFMs in Definition 2.1.3, the classification in (4.33) is the same, but the means and covariance matrices of $\boldsymbol{F}_t|\boldsymbol{X}_t$ must be replaced by the ones in Lemma 3.1.5.

In Remark 4.2.9, the pillar *AR Nature* covers the autoregressive return behavior, whereas *Factor impact* maps the information extracted from the panel data X. So, both affect the direction of r^c_{T+1}. By contrast, the latter treat estimation uncertainties. In this context, *Factor Risk* reveals the distortion caused by F^c and hence, indicates the variation inherent in the estimated factors. This is of particular importance for data sets of small size or with many gaps. Finally, *AR Risk* incorporates deviations from the expected trend, since it adds the standard deviation of the ARX residuals.

The four pillars in (4.33) support the detection of model inadequacies and the construction of extensions, since each driver can be treated separately or as part of a group. For instance, a comparison of the pillars *AR Nature* and *Factor Impact* shows, whether a market has an own behavior like a trend and seasonalities or is triggered by exogenous events. Next, we trace the total contribution of *Factor Impact* to its single constituents such that the influence of a single signal may be analyzed. For this purpose, we additionally store the single constituents of *Factor Risk* in the outer for loop of Algorithm 4.2.1, sort all time series in line with the ascendingly ordered returns and then, derive prediction intervals for both, i.e., the returns

Algorithm 4.2.1: Empirical prediction intervals of ARX($\tilde{q}, \tilde{p}$) returns for next period of time

Initialization

Define number $V > 0$ of returns to be predicted;

Choose prediction level $\nu > 0$ and termination criterion $\eta_{\tilde{\boldsymbol{B}}} > 0$;

Set upper limits of autoregressive orders $\bar{q} \geq 0$ and $\bar{p} \geq 1$;

Determine factor distribution (4.9) for $1 \leq t \leq T$ using Algorithm 4.1.2;

Generation of Return Sample

for $c = 1$ **to** V **do**

  Draw factor sample $(F^c)' = [\boldsymbol{F}_1^c, \ldots, \boldsymbol{F}_T^c]$ from (4.9);

  Initialize coefficient vector $\hat{\boldsymbol{\theta}}_{\text{ov}} = \emptyset$;

  Reset overall AIC value $\text{AIC}_{\text{ov}} = \infty$ (or any sufficiently large number);

  # Model Selection for Return ARX

  for $q = 0$ **to** $\bar{q}$ **do**

    for $p = 1$ **to** $\bar{p}$ **do**

      Estimate mean of $\hat{\boldsymbol{\theta}}$ in (4.27) with $\boldsymbol{r} = [r_1, \ldots, r_T]$, F^c, q and p;

      Determine temporary AIC value AIC_{tmp} for $\hat{\boldsymbol{\theta}}$ using q, p and (4.32);

      if $AIC_{tmp} < AIC_{ov}$ **then**

        Renew overall AIC value by $\text{AIC}_{\text{ov}} = \text{AIC}_{\text{tmp}}$;

        Update overall coefficient vector by $\hat{\boldsymbol{\theta}}_{\text{ov}} = \hat{\boldsymbol{\theta}}$;

      end

    end

  end

  Determine asymptotic distribution of $\hat{\boldsymbol{\theta}}_{\text{ov}}$ in (4.27)-(4.29) for chosen orders $(\tilde{q}, \tilde{p})$;

  Draw sample $\boldsymbol{\theta}^c$ from asymptotic distribution of $\hat{\boldsymbol{\theta}}_{\text{ov}}$ in (4.27)-(4.29);

  Draw random variable Z^c from $\mathcal{N}(0, 1)$;

  Set $r_{T+1}^c = \bar{\alpha}^c + \sum_{i=1}^{\tilde{q}} \left(\bar{\beta}_i^c r_{T+1-i}\right) + \sum_{i=1}^{\tilde{p}} \left((\bar{\gamma}_i^c)' \boldsymbol{F}_{T+1-i}^c\right) + \hat{\sigma}_\varepsilon^c \, Z^c$;

end

Sort returns in ascending order $r_{T+1}^{(1)} \leq \ldots \leq r_{T+1}^{(V)}$;

Prediction interval is given by $\left[r_{T+1}^{(\lfloor(1-\nu)V/2\rfloor)}, r_{T+1}^{(\lceil(1+\nu)V/2\rceil)}\right]$;

and their single drivers. This procedure prevents us from discrepancies caused by data aggregation and thus, ensures the matching between the expectations of r_{T+1}^c and its drivers.

All in all, the presented approach for modeling the 1-step ahead returns of a financial index offers several advantages for asset and risk management applications: First, it admits the inclusion of mixed-frequency information and hence, supports the treatment of incomplete data. Especially, when, e.g., macroeconomic data, flows, technical findings and valuation results are included, data and calendar irregularities cannot be neglected. Second, for each low-frequency signal a high-frequency counterpart is constructed (nowcasting) such that, e.g., structural changes in the real economy may be identified at an early stage. Third, the ARX in Definition 4.2.1 links the empirical behavior of an asset class with exogenous information to provide interval and point estimations. Besides the direction of the future returns, the derived prediction intervals measure estimation uncertainties. In addition to risk-return characteristics, investors take a great interest in the drivers of a market movement, as those indicate its sustainability. For instance, if increased inflows caused by an extremely loose monetary policy triggered a stock market ralley and an asset manager is aware of this, he would be more afraid of an unexpected change in monetary policy than poor macroeconomic figures. As soon as the drivers have been identified, alternative hedging strategies can be developed. In our example, derivatives based on fixed income products might serve for hedging purposes instead of derivatives with stocks as underlying, if the first are, e.g., more liquid or cheaper.

So far, we considered prediction intervals for the next period of time. Within a small excursion, we focus on the general case, that is, prediction intervals covering s-step ahead returns with $s \geq 1$.

Definition 4.2.10 (Shifted Returns of Financial Market Index)
Assume the setting in Definition 4.2.1, but let $s \geq 1$ denote a shift in time. Then, we have:

$$r_t = \alpha + \sum_{i=1}^{\tilde{q}} (\beta_i r_{t-s+1-i}) + \sum_{i=1}^{\tilde{p}} (\gamma_i' \boldsymbol{F}_{t-s+1-i}) + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right) \textit{ iid},$$

where $\alpha, \beta_i, 1 \leq i \leq \tilde{q}, \sigma_\varepsilon^2 \in \mathbb{R}$ with $\sigma_\varepsilon^2 \geq 0$ and $\boldsymbol{\gamma}_j \in \mathbb{R}^K, 1 \leq j \leq \tilde{p}$, are constants. As before, we assume that factors and errors are independent, i.e., $\boldsymbol{F}_t \perp \varepsilon_u$ for all u, t and that neither the return process $\{r_t\}$ nor any of its transformations enters the panel data of the factors.

Let the vector $\boldsymbol{\theta} = \left[\alpha, \beta_1, \ldots, \beta_{\tilde{q}}, \boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_{\tilde{p}}'\right]' \in \mathbb{R}^{1+\tilde{q}+\tilde{p}K}$ collect the ARX$(\tilde{q}, \tilde{p})$ parameters in Definition 4.2.10, while the vector $\boldsymbol{r} = [r_1, \ldots, r_T]' \in \mathbb{R}^T$ and matrix $F = [\boldsymbol{F}_1, \ldots, \boldsymbol{F}_T]' \in \mathbb{R}^{T \times K}$ are return and factor samples, respectively, of same length and time horizon. Then, for $\tilde{m} = \max\{\tilde{q}, \tilde{p}\} + s - 1$ the ARX in Definition 4.2.10 can be rewritten in matrix form as follows:

$$\underbrace{\begin{bmatrix} r_{\tilde{m}+1} \\ \vdots \\ r_T \end{bmatrix}}_{\boldsymbol{r}_T^{\tilde{m}+1}} = \underbrace{\begin{bmatrix} 1 & r_{\tilde{m}+1-s} & \cdots & r_{\tilde{m}+2-s-\tilde{q}} & \boldsymbol{F}'_{\tilde{m}+1-s} & \cdots & \boldsymbol{F}'_{\tilde{m}+2-s-\tilde{p}} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & r_{T-s} & \cdots & r_{T+1-s-\tilde{q}} & \boldsymbol{F}'_{T-s} & \cdots & \boldsymbol{F}'_{T+1-s-\tilde{p}} \end{bmatrix}}_{G_s} \boldsymbol{\theta} + \underbrace{\begin{bmatrix} \varepsilon_{\tilde{m}+1} \\ \vdots \\ \varepsilon_T \end{bmatrix}}_{\boldsymbol{\varepsilon}_T^{\tilde{m}+1}}, \tag{4.34}$$

with $\boldsymbol{r}_T^{\tilde{m}+1} \in \mathbb{R}^{T-\tilde{m}}, G_s \in \mathbb{R}^{(T-\tilde{m})\times(1+\tilde{q}+\tilde{p}K)}$ and $\boldsymbol{\varepsilon}_T^{\tilde{m}+1} \in \mathbb{R}^{T-\tilde{m}}$.

If we replace the matrix G in Lemma 4.2.2 by the matrix G_s in (4.34), the asymptotic distribution of the OLS estimate $\hat{\boldsymbol{\theta}}$ for the general case follows. Furthermore, the findings in Remarks 4.2.3 and 4.2.4 remain valid for $s \geq 1$. In case of its VAR(1) representation the following changes are required.

Lemma 4.2.11 (VAR(1) Representation of Shifted ARX)
Let $\{\boldsymbol{F}_t\}$ be the factor process of the SFM in Definition 2.1.3 with $p = 0$ or the DFM in Definition 2.1.4 with $p \geq 1$. For lag lengths $0 \leq \tilde{q}$ and $0 \leq \tilde{p} \leq \max(1, p)$, shift $s \geq 1$ and any point in time t, we define:

$$\tilde{\boldsymbol{S}}_t = \left(r_t, \ldots, r_{t+2-s-\max(1,\tilde{q})}, \boldsymbol{F}'_t, \ldots, \boldsymbol{F}'_{t+2-s-\max(1,p)}\right)' \in \mathbb{R}^d,$$

with vector dimension $d = \max(1, \tilde{q}) + \max(1, p) K + (s-1)(K+1)$. Then, the ARX in Definition 4.2.10 *has the following VAR(1) representation:*

$$\tilde{\boldsymbol{S}}_t = \underbrace{\begin{bmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\boldsymbol{a}} + \underbrace{\begin{bmatrix}
0 & \cdots & 0 & \beta_1 & \cdots & \beta_{\tilde{q}-1} & \beta_{\max(1,\tilde{q})} & \mathbf{0}'_K & \cdots & \mathbf{0}'_K & \boldsymbol{\gamma}'_1 & \cdots & \boldsymbol{\gamma}'_{\tilde{p}} & \mathbf{0}'_K & \cdots & \mathbf{0}'_K \\
1 & & 0 & 0 & \cdots & \cdots & 0 & \mathbf{0}'_K & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0}'_K \\
 & \ddots & & \vdots & & & \vdots & \vdots & & & & & & & & \vdots \\
0 & & 1 & 0 & \cdots & \cdots & 0 & \vdots & & & & & & & & \vdots \\
0 & \cdots & 0 & 1 & & 0 & 0 & \vdots & & & & & & & & \vdots \\
\vdots & & \vdots & & \ddots & & \vdots & \vdots & & & & & & & & \vdots \\
0 & \cdots & 0 & 0 & & 1 & 0 & \mathbf{0}'_K & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0}'_K \\
\mathbf{0}_K & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_K & A_1 & \cdots & \cdots & \cdots & A_{\max(1,p)} & O_K & \cdots & \cdots & O_K \\
\vdots & & & & & & \vdots & I_K & O_K & \cdots & \cdots & O_K & \vdots & & & \vdots \\
\vdots & & & & & & \vdots & O_K & \ddots & \ddots & & \vdots & \vdots & & & \vdots \\
\vdots & & & & & & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & & & \vdots \\
\vdots & & & & & & \vdots & \vdots & & \ddots & \ddots & O_K & \vdots & & & \vdots \\
\vdots & & & & & & \vdots & O_K & \cdots & \cdots & O_K & I_K & O_K & \cdots & \cdots & O_K \\
\vdots & & & & & & \vdots & \vdots & & & \vdots & O_K & I_K & & O_K & O_K \\
\vdots & & & & & & \vdots & \vdots & & & \vdots & \vdots & & \ddots & & \vdots \\
\mathbf{0}_K & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_K & O_K & \cdots & \cdots & O_K & O_K & O_K & & I_K & O_K
\end{bmatrix}}_{\mathbb{H}_s} \tilde{\boldsymbol{S}}_{t-1} + \underbrace{\begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \\ \boldsymbol{\delta}_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\boldsymbol{e}_t},$$

with constant vector $\boldsymbol{a} \in \mathbb{R}^d$ and matrix $\mathbb{H}_s \in \mathbb{R}^{d \times d}$. For the shocks $\boldsymbol{e}_t \in \mathbb{R}^d$, it holds: $\boldsymbol{e}_t \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{e}}, \Sigma_{\boldsymbol{e}}\right)$ iid for all points in time t, whose mean and covariance matrix are defined by:

$$\boldsymbol{\mu}_{\boldsymbol{e}} = \mathbf{0}_d \quad \text{and} \quad \Sigma_{\boldsymbol{e}} = \begin{bmatrix}
\sigma_\varepsilon^2 & 0 & \cdots & 0 & \mathbf{0}'_K & \mathbf{0}'_K & \cdots & \mathbf{0}'_K \\
0 & 0 & & \vdots & \vdots & \vdots & & \vdots \\
\vdots & & \ddots & \vdots & \vdots & \vdots & & \vdots \\
0 & \cdots & \cdots & 0 & \mathbf{0}'_K & \mathbf{0}'_K & \cdots & \mathbf{0}'_K \\
\mathbf{0}_K & \cdots & \cdots & \mathbf{0}_K & \Sigma_{\boldsymbol{\delta}} & O_K & \cdots & O_K \\
\vdots & & & \vdots & O_K & O_K & & \vdots \\
\vdots & & & \vdots & \vdots & & \ddots & \vdots \\
\mathbf{0}_K & \cdots & \cdots & \mathbf{0}_K & O_K & O_K & \cdots & O_K
\end{bmatrix}.$$

In case of $\tilde{p} = p$, there are no zeros at the end of the first row in matrix $\mathbb{H}_s$. For $s = 1$, there are no zeros in the first row in front of β_1 and $\boldsymbol{\gamma}_1$. In addition, some of the last rows are removed. If $p = 0$ holds, we obtain $A_1 = O_K$. Similarly, $\tilde{q} = 0$ results in $\beta_1 = 0$.

Proof:
Follows directly from Definitions 2.1.3, 2.1.4 and 4.2.10. □

Besides the VAR(1) representation, the log-likelihood function of the shifted ARX given the sample F^c calls for adjustments. Thereby, return means and variances must be recalculated. For clarity reasons, we moved rather technical Lemmata to the appendix. That means, in Lemma A.2.15 we obtain the MA(∞) representation of $\tilde{\boldsymbol{S}}_t$ to determine in Lemmata A.2.16 or A.2.17 the mean and covariance matrix of $\tilde{\boldsymbol{S}}_t$. Eventually, we receive for shift $s \geq 1$ in Lemma A.2.18 the mean $\boldsymbol{\mu}_{r,s}$ and covariance matrix $\Sigma_{r,s}$ of the vectors $\boldsymbol{r}_{t,s} = [r_t, \ldots, r_{t-s+1}]' \in \mathbb{R}^s$, which enter the following log-likelihood function.

Lemma 4.2.12 (Conditional Log-Likelihood of Shifted ARX)
For a time shift $s \geq 1$, let $\boldsymbol{r} = [r_1, \ldots, r_T] \in \mathbb{R}^T$ and $(F^c)' = [\boldsymbol{F}_1^c, \ldots, \boldsymbol{F}_T^c] \in \mathbb{R}^{K \times T}$ be return and factor samples, respectively. Moreover, we set: $\tilde{m} = \max(\tilde{q}, \tilde{p}) + s - 1$ and assume that the returns obey the ARX in Definition 4.2.10 based on F^c, i.e.:

$$r_t = \alpha + \sum_{i=1}^{\tilde{q}} \left(\beta_i r_{t-s+1-i}\right) + \sum_{i=1}^{\tilde{p}} \left(\boldsymbol{\gamma}'_i \boldsymbol{F}^c_{t-s+1-i}\right) + \varepsilon_t, \ \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right) \text{ iid.} \tag{4.35}$$

Then, we have for its log-likelihood function given $\boldsymbol{r}, F^c$ and shift s, i.e., $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{r}, F^c, s)$:

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{r}, F^c, s) = -\frac{T\ ln(2\pi)}{2} - \frac{1}{2} ln(|\Sigma_{r,s}|) - \frac{1}{2}\sum_{t=(\tilde{m}+1)}^{T}\left(ln\left(\sigma^2_{r_{t,s}|Full}\right) + \frac{\left(r_t - \mu_{r_{t,s}|Full}\right)^2}{\sigma^2_{r_{t,s}|Full}}\right)$$
$$-\frac{1}{2}\sum_{t=(s+1)}^{\tilde{m}}\left(ln\left(\sigma^2_{r_{t,s}|Part}\right) + \frac{\left(r_t - \mu_{r_{t,s}|Part}\right)^2}{\sigma^2_{r_{t,s}|Part}}\right) - \frac{1}{2}\left(\boldsymbol{r}_{s,s} - \boldsymbol{\mu}_{r,s}\right)' \Sigma^{-1}_{r,s}\left(\boldsymbol{r}_{s,s} - \boldsymbol{\mu}_{r,s}\right),$$

with $\boldsymbol{r}_{s,s}, \boldsymbol{\mu}_{r,s}, \mu_{r_{t,s}|Part}, \mu_{r_{t,s}|Full}, \Sigma_{r,s}, \sigma^2_{r_{t,s}|Part}$ and $\sigma^2_{r_{t,s}|Full}$ as in Lemmata A.2.18-A.2.20.

Proof:
By virtue of the Bayes' theorem we receive for the likelihood function of $\boldsymbol{r}$ conditioned on F^c in (4.35):

$$f_{\boldsymbol{\theta}}(\boldsymbol{r}|F^c) = f_{\boldsymbol{\theta}}(r_T, \ldots, r_1|F^c) = \frac{f_{\boldsymbol{\theta}}(r_T, \ldots, r_1, F^c)}{f_{\boldsymbol{\theta}}(F^c)} = \frac{f_{\boldsymbol{\theta}}(r_T|r_{T-1}, \ldots, r_1, F^c)\, f_{\boldsymbol{\theta}}(r_{T-1}, \ldots, r_1, F^c)}{f_{\boldsymbol{\theta}}(F^c)}$$
$$= f_{\boldsymbol{\theta}}(r_T|r_{T-s}, \ldots, r_{T-s+1-\tilde{q}}, F^c)\, f_{\boldsymbol{\theta}}(r_{T-1}, \ldots, r_1|F^c)$$
$$= \left(\prod_{t=(\tilde{m}+1)}^{T} f_{\boldsymbol{\theta}}(r_t|r_{t-s}, \ldots, r_{t-s+1-\tilde{q}}, F^c)\right)\left(\prod_{t=(s+1)}^{\tilde{m}} f_{\boldsymbol{\theta}}\left(r_t|r_{t-s}, \ldots, r_{\max(1,t-s+1-\tilde{q})}, F^c\right)\right) f_{\boldsymbol{\theta}}(\boldsymbol{r}_{s,s}).$$

Hence, we obtain for its log-likelihood function $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{r}, F^c, s)$:

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{r}, F^c, s) = \sum_{t=(\tilde{m}+1)}^{T} ln\left(f_{\boldsymbol{\theta}}(r_t|r_{t-s}, \ldots, r_{t-s+1-\tilde{q}}, F^c)\right) + ln\left(f_{\boldsymbol{\theta}}(\boldsymbol{r}_{s,s})\right)$$
$$+ \sum_{t=(s+1)}^{\tilde{m}} ln\left(f_{\boldsymbol{\theta}}\left(r_t|r_{t-s}, \ldots, r_{\max(1,t-s+1-\tilde{q})}, F^c\right)\right).$$

Next, the distributions in Lemmata A.2.18-A.2.20 provide:

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{r}, F^c, s) = \sum_{t=(\tilde{m}+1)}^{T} ln\left(\left(2\pi\sigma^2_{r_{t,s}|Full}\right)^{-1/2} exp\left(-\frac{\left(r_t - \mu_{r_{t,s}|Full}\right)^2}{2\,\sigma^2_{r_{t,s}|Full}}\right)\right)$$
$$+ \sum_{t=(s+1)}^{\tilde{m}} ln\left(\left(2\pi\sigma^2_{r_{t,s}|Part}\right)^{-1/2} exp\left(-\frac{\left(r_{t,s} - \mu_{r_{t,s}|Part}\right)^2}{2\,\sigma^2_{r_{t,s}|Part}}\right)\right)$$
$$+ ln\left((2\pi)^{-s/2}|\Sigma_{r,s}|^{-1/2} exp\left(-\frac{1}{2}\left(\boldsymbol{r}_{s,s} - \boldsymbol{\mu}_{r,s}\right)' \Sigma^{-1}_{r,s}\left(\boldsymbol{r}_{s,s} - \boldsymbol{\mu}_{r,s}\right)\right)\right).$$

Finally, summarizing equal expressions leads to the stated formulation. □

For any $s \geq 1$, the shifted ARX in Definition 4.2.10 has the same number of parameters as the special case with $s = 1$ in Definition 4.2.1. Therefore, Remark 4.2.7 can be kept for selecting the autoregressive orders, if $\mathcal{L}\left(\hat{\boldsymbol{\theta}}(\tilde{q}, \tilde{p})|\boldsymbol{r}, F^c\right)$ is replaced by $\mathcal{L}\left(\hat{\boldsymbol{\theta}}(\tilde{q}, \tilde{p})|\boldsymbol{r}, F^c, s\right)$. Due to $\sigma^2_{r_{t,s}|Full} = \sigma^2_{\varepsilon}$ for all $\tilde{m}+1 \leq t \leq T$ in Lemma A.2.19 and the definition of the OLS estimate $\hat{\boldsymbol{\theta}}$ in Lemma 4.2.2, we can simplify the log-likelihood $\mathcal{L}\left(\hat{\boldsymbol{\theta}}(\tilde{q}, \tilde{p})|\boldsymbol{r}, F^c, s\right)$ as in Remark 4.2.8. Finally, the structure of Algorithm 4.2.1 is kept, since only minor changes are required. Furthermore, the classification in Remark 4.2.9 remains valid. In this case, the return equation has to take the shift s into account, but the meaning of the pillars stays the same.

4.3 Interval-Based Trading Strategies for Single Markets

The location and width of the prediction intervals from Section 4.2 cover the trend and uncertainty of the forcasted returns. Therefore, we now propose some simple and risk-adjusted dynamic trading strategies, which incoporate those. Unlike the portfolio perspective in Section 3.1.3, the current investment strategies are restricted to a single financial market and a bank account.

Definition 4.3.1 (Basic Single-Market Trading Strategy)
Let $\hat{l}_t, \hat{u}_t \in \mathbb{R}$ with $\hat{l}_t \leq \hat{u}_t$ be the lower and upper limits of the ν-prediction interval provided by Algorithm 4.2.1 for the period $(t-1, t]$. Moreover, let $\pi_t \in [0, 1]$ be the percentage of the total wealth invested in the financial market over the period $(t-1, t]$. The remaining proportion $1 - \pi_t$ is deposited on a bank account for a risk-free rate $\tilde{r}_t$. Then, we have for the market exposure of the basic trading strategy:

$$\pi_t = \begin{cases} 1 & \text{if } \hat{l}_t \geq 0 \text{ and } \hat{u}_t \geq 0, \\ \frac{\hat{u}_t}{\hat{u}_t - \hat{l}_t} & \text{if } \hat{l}_t < 0 \text{ and } \hat{u}_t > 0, \\ 0 & \text{if } \hat{l}_t \leq 0 \text{ and } \hat{u}_t \leq 0. \end{cases} \tag{4.36}$$

Whenever the prediction interval is centered around zero, except for a possible lateral movement, no clear trend is detected. Irrespective of the interval width, the basic strategy in (4.36) recommends the neutral allocation of a 50% market exposure and a 50% bank account deposit for this case. If a prediction interval is shifted to the positive (negative) half-plane, the market exposure increases up to 100% (decreases down to 0%). Depending on the interval width, the same shift size may result in different proportions π_t. That is, for large intervals indicating a high degree of uncertainty, a shift to the positive (negative) half-plane causes a smaller increase (decrease) in π_t compared to tight intervals with low uncertainties. Besides temporal uncertainties, the prediction level ν affects the interval size and thus, the market exposure π_t. Thereby, we have: The higher the prediction level ν is the lower and rarer deviations from the neutral allocations are. All in all, the market exposure is on average 50% with temporary over- and underweightings. Since the basic strategy in Defintion 4.3.1 is not always appropriate for applications in practice, we provide an extension that allows for the risk-return profile of an investor.

Definition 4.3.2 (Risk-Adjusted Single-Market Trading Strategy)
Assume π_t in (4.36) and let $\pi^L, \pi^U \in \mathbb{R}$ with $\pi^L \leq \pi^U$ denote the lower and upper limits, respectively, of the overall market exposure, which may never be exceeded. In addition, let $\alpha^A \geq 0$ cover the risk appetite of the investor. Then, the market exposure $\hat{\pi}_t$ of the risk-adjusted trading strategy is defined by:

$$\hat{\pi}_t = \max\left[\min\left[\alpha^A \pi_t, \pi^U - \pi^L\right], 0\right] + \pi^L. \tag{4.37}$$

The max-min-construction in (4.37) defines a piecewise linear function bounded below by π^L and bounded above by π^U. Within the limits the product $\alpha^A \pi_t$ drives the market exposure $\hat{\pi}_t$. For $\alpha^A > 1$ changes in π_t are scaled-up. Consequently, the fluctuation margin of $\alpha^A \pi_t$ exceeds the one of π_t. Furthermore, the bounds are reached more likely and changes in $\hat{\pi}_t$ are of bigger size. Because of this, we call an investor with $\alpha^A > 1$ risk affine. By contrast, $0 \leq \alpha^A \leq 1$ reduces the fluctuation margin of $\alpha^A \pi_t$ and thus, of $\hat{\pi}_t$. Therefore, $0 \leq \alpha^A \leq 1$ covers a risk-averse attitude. As an example, we set $\pi^L = -1, \pi^U = 1$ and $\alpha^A = 2$, which implies: $\hat{\pi}_t \in [-1, 1]$ such that short sales are possible.

4.4 Simulation Study

Now, we analyze the performance of the two-step estimation method for ADFMs in Sections 4.1.1 and 4.1.2 within a comprehensive MC simulation study. Among other things, we address the following questions: *(i)* does the size of the data sample, i.e., its length and number of time series, affect the estimation quality? *(ii)* to what extent does data incompleteness deteriorate our estimation results? *(iii)* does the choice between stock, flow and change in flow variables for the underlying panel data matter? *(iv)* does our model selection procedure detect the true factor dimension and lag order, even for $K > 1$ and $p > 1$? *(v)* how does our two-step approach perform compared to the estimation method of Bańbura and Modugno (2014)? *(vi)* are factor means and covariance matrices more accurate, if we use our closed-form solutions in Theorem 4.1.4 instead of the standard KF and KS?

Before we answer the previous questions, we explain how our random samples are generated. For $a, b \in \mathbb{R}$ with $a < b$, let $\mathcal{U}(a, b)$ stand for the uniform distribution on the interval $[a, b]$ and let $\operatorname{diag}(\boldsymbol{z}) \in \mathbb{R}^{K \times K}$ be a diagonal matrix with elements $\boldsymbol{z} = [z_1, \ldots, z_K] \in \mathbb{R}^K$. For fixed data and factor dimensions (T, N, K, p), let $V_i \in \mathbb{R}^{K \times K}, 1 \leq i \leq p, V_{\boldsymbol{\delta}} \in \mathbb{R}^{K \times K}$ and $V_{\boldsymbol{\epsilon}} \in \mathbb{R}^{N \times N}$ represent arbitrary orthonormal matrices. Then, we receive the parameters of the ADFMs in Definition 2.1.4 in the following manner:

$$\begin{aligned}
A_i &= V_i \operatorname{diag}\left(\frac{z_{i,1}}{p}, \ldots, \frac{z_{i,K}}{p}\right)(V_i'), && z_{i,j} \sim \mathcal{U}(0.25, 0.75) \,\text{iid}, 1 \leq i \leq p, 1 \leq j \leq K, \\
\Sigma_{\boldsymbol{\delta}} &= V_{\boldsymbol{\delta}} \operatorname{diag}\left(z_{\boldsymbol{\delta},1}, \ldots, z_{\boldsymbol{\delta},K}\right)\left(V_{\boldsymbol{\delta}}'\right), && z_{\boldsymbol{\delta},j} \sim \mathcal{U}(0.25, 0.50) \,\text{iid}, 1 \leq j \leq K, \\
W &= (w_{n,j})_{n,j}, && w_{n,j} \sim \mathcal{N}(0,1) \,\text{iid}, 1 \leq n \leq N, 1 \leq j \leq K, \\
\mu &= (\mu_n)_{n,1}, && \mu_n \sim \mathcal{N}(0,1) \,\text{iid}, 1 \leq n \leq N, \\
\Sigma_{\boldsymbol{\epsilon}} &= V_{\boldsymbol{\epsilon}} \operatorname{diag}(z_{\boldsymbol{\epsilon},1}, \ldots, z_{\boldsymbol{\epsilon},N})\left(V_{\boldsymbol{\epsilon}}'\right), && z_{\boldsymbol{\epsilon},n} \sim \mathcal{U}(0.05, 0.25) \,\text{iid}, 1 \leq n \leq N.
\end{aligned} \tag{4.38}$$

In total, the parameters in (4.38) specify ADFMs with cross-sectionally, but not serially correlated shocks. To prevent us from the implicit construction of SFMs with eigenvalues of A_i close to zero, the eigenvalues of $A_i, 1 \leq i \leq p$, lie within the range $[0.25/p, 0.75/p]$. Here, the division by p balances the total sum of all eigenvalues with respect to the autoregressive order p. For simplicity reasons, we consider matrices A_i with positive eigenvalues. However, this assumption, the restriction to eigenvalues in the range $[0.25/p, 0.75/p]$ and the division by p can be skipped. If all matrices $A_i, 1 \leq i \leq p$, meet the covariance-stationarity conditions in Lemma A.2.3, we simulate factor samples $F \in \mathbb{R}^{T \times K}$ and panel data $X \in \mathbb{R}^{T \times N}$ through the transition equation (2.5) and observation equation (2.4). Otherwise, all matrices A_i are drawn again until covariance-stationarity is reached. Similarly, we choose only matrices W of full column rank K.

So far, we have complete panel data. Let $\rho_m \in [0, 1]$ be the ratio of gaps arising from missing observations and low-frequency time series, respectively. To achieve incomplete data we remove $\lceil \rho_m T \rceil$ elements from each time series. For stock variables, we randomly delete $\lceil \rho_m T \rceil$ values to end up with irregularly scattered gaps. At this stage, flow and change in flow variables serve as low-frequency information, which is supposed to have an ordered pattern of gaps. Therefore, an observation is made at time $t = \lceil 1 + s/(1 - \rho_m) \rceil$ with $0 \leq s \leq \lfloor (T-1)(1-\rho_m) \rfloor$ and $s \in \mathbb{N}_0$. In line with Definition 2.2.2, an observed flow variable represents the sum or average of the corresponding high-frequency analogs. As in Lemma 2.2.3, an observed change in flow variable is a linear combination of high-frequency changes. In Tables 4.1-4.13, the same ρ_m holds for all univariate time series in X, i.e., for each of its columns, such that gaps of flow or change in flow variables occur at the same time. If the panel data contains a single point in time without any observation, neither our closed-form solution nor the standard KF and KS provide factor estimates. To avoid such scenarios, i.e., empty rows of the observed panel data X_{obs}, each panel data in the second (third) column of Tables 4.1-4.6 comprises $\lceil N/2 \rceil$ times series modeled as stock variable and $\lfloor N/2 \rfloor$ time series serving as flow (change in flow) variable. To ensure that each row of X_{obs} contains at least a single observation, we check each panel data sample, before we proceed. If there is an empty row of matrix X_{obs}, we reapply our routine for preparing missing data to the complete panel data X.

As mentioned in the scope of Algorithm 4.1.1, the estimated factors are unique except for an invertible, linear transformation. Since we aim at assessing the quality of diverse estimation methods for the hidden factors F, we have to take this ambiguity into account. In doing so, we follow Stock and Watson (2002a), Boivin and Ng (2006), Schumacher and Breitung (2008), Doz et al. (2012) and Bańbura and Modugno (2014). That is, we use the trace R^2 in Definition A.3.1 as appropriate performance measure.

Eventually, we provide some general settings. First, we choose for the termination criteria: $\eta = \xi = 10^{-2}$ and $\eta_{\boldsymbol{F}} = 10^{-6}$. That is, we have the same η, ξ and $\eta_{\boldsymbol{F}}$ as in the empirical study of Section 4.5. Second, we use constant interpolation in case of incomplete panel data, when we initialize the set $X_{(0)}$.

In Tables 4.1-4.3, we consider for known factor dimension K and lag order p, whether the standard KF and KS in Lemmata 2.1.8-2.1.9 with lag-one autocovariance smoother in Lemma 2.1.10 should be used for estimating the factor means $\mathbb{E}_\Theta\left[\boldsymbol{F}_t|X\right]$ and covariance matrices $\mathbb{C}\mathrm{ov}_\Theta\left[\boldsymbol{F}_t, \boldsymbol{F}_s|X\right], 1 \leq t, s \leq T$, instead of the closed-form solutions in Theorem 4.1.4. To be more precise, Table 4.2 shows trace R^2 means, each based on 500 MC simulations, when we estimate the parameters Θ as in Theorem 4.1.9. For the same MC paths, Table 4.1 provides trace R^2 means, when we combine the standard KF and KS with the following parameter updates:

$$\hat{W}_{(l+1)} = \left(\sum_{t=1}^{T} (\boldsymbol{X}_t - \boldsymbol{\mu})\, \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{F}_t'|X\right]\right)\left(\sum_{t=1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{F}_t\boldsymbol{F}_t'|X\right]\right)^{-1},$$

$$\hat{\Sigma}_{\boldsymbol{\epsilon}(l+1)} = \frac{1}{T}\left(\sum_{t=1}^{T} (\boldsymbol{X}_t - \boldsymbol{\mu})(\boldsymbol{X}_t - \boldsymbol{\mu})' - \hat{W}_{(l+1)}\left(\sum_{t=1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{F}_t|X\right](\boldsymbol{X}_t - \boldsymbol{\mu})'\right)\right),$$

$$\left[\hat{A}_{1_{(l+1)}} \quad \cdots \quad \hat{A}_{p_{(l+1)}}\right] = \left(\sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{F}_t\tilde{\boldsymbol{F}}_{t-1}'|X\right]\right)\left(\sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\tilde{\boldsymbol{F}}_{t-1}\tilde{\boldsymbol{F}}_{t-1}'|X\right]\right)^{-1},$$

$$\hat{\Sigma}_{\boldsymbol{\delta}(l+1)} = \frac{1}{T-p}\sum_{t=p+1}^{T}\left(\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{F}_t\boldsymbol{F}_t'|X\right] - \left[\hat{A}_{1_{(l+1)}} \quad \cdots \quad \hat{A}_{p_{(l+1)}}\right]\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\tilde{\boldsymbol{F}}_{t-1}\boldsymbol{F}_t'|X\right]\right),$$

with stacked factors $\tilde{\boldsymbol{F}}_t = [\boldsymbol{F}_t', \ldots, \boldsymbol{F}_{t-p+1}']' \in \mathbb{R}^{pK}$ as in Lemma A.2.2. By comparing Tables 4.1 and 4.2 we conclude: First, both estimation methods offer high values for the trace R^2 regardless the underlying data type. Hence, the choice between stock, flow and change in flow variables does not considerably affect the trace R^2. Second, the higher the percentage of missing data the worse the trace R^2. Third, for larger samples, i.e., more or longer time series, the trace R^2 increases, which sounds reasonable. Fourth, for more complex structures in the form of larger K and p the trace R^2, ceteris paribus, deteriorates. Fifth, our estimation method based on closed-form factor moments appears more robust than the Kalman approach. For instance, in Table 4.1 for $K = 3, p = 1, N = 75, T = 100$ and 40% of missing data the trace R^2 is NaN, which is an abbreviation for *Not a Number*. That is, there was at least one MC path the Kalman approach could not estimate. By contrast, the respective trace R^2 in Table 4.2 is 0.95 and so, all 500 MC paths were estimated without any problems. The means in Tables 4.1 and 4.2 are pretty close, this is why Table 4.3 divides the means in Table 4.2 by their counterparts in Table 4.1. Hence, ratios larger than one indicate that our estimation method outperforms the Kalman approach, while ratios less than one do the opposite. Since all ratios on Table 4.3 are at least one, our method performs better.

In Tables 4.4-4.6, we compare again for known factor dimension K and lag order p, whether the standard KF and KS in Lemmata 2.1.8-2.1.9 with lag-one autocovariance smoother in Lemma 2.1.10 should be used for estimating the factor moments $\mathbb{E}_\Theta\left[\boldsymbol{F}_t|X\right]$ and $\mathbb{E}_\Theta\left[\boldsymbol{F}_t\boldsymbol{F}_s'|X\right], 1 \leq t, s \leq T$, instead of the closed-form solutions in Theorem 4.1.4. In both cases, we estimate the model parameters Θ as in Lemma 4.1.7. The only difference between the two estimation methods comes from the estimation of the factor moments.

Table 4.1: *Means of trace R^2 for random ADFMs using standard KF and KS*

		stock[a]				**stock/flow (average)**[b]				**stock/change in flow (average)**[c]			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
						$K = 1, p = 1$							
25	100	0.94	0.93	0.93	0.91	0.94	0.93	0.92	0.86	0.94	0.93	0.92	0.87
25	250	0.97	0.96	0.96	0.95	0.97	0.96	0.96	0.93	0.97	0.96	0.96	0.92
25	500	0.98	0.98	0.97	0.96	0.98	0.97	0.97	0.94	0.98	0.97	0.97	0.95
50	100	0.95	0.94	0.94	0.91	0.94	0.94	0.92	0.84	0.95	0.94	0.93	0.86
50	250	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.95	0.97	0.97	0.97	0.94
50	500	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.96	0.98	0.98	0.98	0.96
75	100	0.95	0.95	0.92	0.83	0.95	0.94	0.89	0.89	0.95	0.94	0.88	NaN
75	250	0.98	0.98	0.98	0.97	0.98	0.97	0.97	0.95	0.98	0.97	0.97	0.95
75	500	0.99	0.99	0.98	0.98	0.99	0.98	0.98	0.96	0.99	0.98	0.98	0.97
						$K = 3, p = 1$							
25	100	0.92	0.92	0.91	0.88	0.92	0.92	0.90	0.84	0.92	0.92	0.89	0.85
25	250	0.96	0.96	0.95	0.93	0.96	0.96	0.94	0.91	0.96	0.95	0.94	0.90
25	500	0.98	0.97	0.97	0.95	0.98	0.97	0.96	0.93	0.98	0.97	0.96	0.93
50	100	0.94	0.94	0.92	0.90	0.94	0.93	0.91	0.85	0.94	0.93	0.91	0.86
50	250	0.97	0.97	0.97	0.96	0.97	0.97	0.96	0.93	0.97	0.96	0.96	0.93
50	500	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.95	0.98	0.97	0.98	0.96
75	100	0.94	0.93	0.92	0.89	0.94	0.92	0.90	NaN	0.94	0.93	0.90	NaN
75	250	0.97	0.97	0.97	0.96	0.97	0.97	0.97	0.94	0.97	0.97	0.97	0.94
75	500	0.99	0.99	0.98	0.98	0.99	0.98	0.98	0.96	0.99	0.98	0.98	0.97
						$K = 3, p = 2$							
25	100	0.88	0.87	0.85	0.80	0.88	0.87	0.84	0.76	0.88	0.87	0.83	0.75
25	250	0.95	0.94	0.93	0.90	0.95	0.94	0.91	0.83	0.95	0.93	0.90	0.82
25	500	0.97	0.96	0.96	0.94	0.97	0.96	0.94	0.87	0.97	0.95	0.94	0.88
50	100	0.90	0.90	0.88	0.85	0.90	0.89	0.86	0.77	0.90	0.89	0.85	0.77
50	250	0.96	0.96	0.95	0.94	0.96	0.95	0.94	0.87	0.96	0.95	0.94	0.86
50	500	0.98	0.98	0.97	0.96	0.98	0.97	0.97	0.91	0.98	0.97	0.96	0.92
75	100	0.91	0.90	0.88	0.84	0.90	0.89	0.85	NaN	0.91	0.89	0.85	0.75
75	250	0.96	0.96	0.96	0.95	0.96	0.96	0.95	0.89	0.96	0.96	0.95	0.88
75	500	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.93	0.98	0.97	0.97	0.94
						$K = 5, p = 1$							
25	100	0.92	0.91	0.89	0.85	0.92	0.91	0.88	0.83	0.92	0.91	0.88	0.83
25	250	0.96	0.96	0.94	0.92	0.96	0.95	0.93	0.89	0.96	0.95	0.93	0.88
25	500	0.97	0.97	0.96	0.94	0.97	0.97	0.95	0.91	0.97	0.96	0.95	0.92
50	100	0.93	0.93	0.91	0.89	0.93	0.92	0.90	0.85	0.93	0.92	0.90	0.85
50	250	0.97	0.97	0.96	0.95	0.97	0.97	0.95	0.92	0.97	0.96	0.95	0.92
50	500	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.94	0.98	0.97	0.97	0.95
75	100	0.93	0.93	0.92	0.89	0.93	0.92	0.90	0.85	0.93	0.92	0.90	0.85
75	250	0.97	0.97	0.97	0.96	0.97	0.97	0.96	0.93	0.97	0.97	0.96	0.93
75	500	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.98	0.96
						$K = 5, p = 2$							
25	100	0.88	0.87	0.84	0.79	0.88	0.87	0.82	0.75	0.88	0.86	0.82	0.75
25	250	0.95	0.94	0.92	0.88	0.95	0.93	0.90	0.81	0.95	0.92	0.89	0.81
25	500	0.97	0.96	0.95	0.92	0.97	0.96	0.93	0.85	0.97	0.95	0.93	0.85
50	100	0.89	0.89	0.87	0.84	0.89	0.88	0.85	0.78	0.89	0.88	0.84	0.77
50	250	0.95	0.95	0.94	0.92	0.95	0.95	0.92	0.85	0.95	0.94	0.92	0.84
50	500	0.97	0.97	0.97	0.96	0.97	0.97	0.96	0.89	0.97	0.96	0.95	0.90
75	100	0.90	0.89	0.87	0.85	0.89	0.88	0.85	0.78	0.90	0.88	0.85	0.78
75	250	0.96	0.96	0.95	0.94	0.96	0.95	0.94	0.88	0.96	0.95	0.93	0.86
75	500	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.91	0.98	0.97	0.97	0.92

The displayed means are derived from 500 MC simulations for known dimensions K and p.

[a] For incomplete time series a stock variable is assumed.
[b] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series are stock and flow (average formulation) variables, respectively.
[c] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series serve as stock or change in flow (average formulaton) variables.

Table 4.2: *Means of trace R^2 for random ADFMs using closed-form factor moments*

		stock[a]				**stock/flow (average)**[b]				**stock/change in flow (average)**[c]			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
						$K = 1, p = 1$							
25	100	0.96	0.95	0.95	0.95	0.96	0.95	0.95	0.92	0.96	0.95	0.94	0.92
25	250	0.97	0.97	0.97	0.96	0.97	0.97	0.96	0.95	0.97	0.96	0.97	0.94
25	500	0.98	0.98	0.97	0.97	0.98	0.98	0.97	0.95	0.98	0.97	0.97	0.95
50	100	0.96	0.96	0.96	0.95	0.96	0.96	0.95	0.94	0.96	0.95	0.96	0.94
50	250	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.96	0.98	0.97	0.98	0.95
50	500	0.99	0.99	0.98	0.98	0.99	0.98	0.98	0.97	0.99	0.98	0.98	0.97
75	100	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.94	0.97	0.96	0.96	0.94
75	250	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.96
75	500	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	0.98	0.99	0.98
						$K = 3, p = 1$							
25	100	0.95	0.95	0.95	0.94	0.95	0.95	0.94	0.92	0.95	0.94	0.94	0.91
25	250	0.97	0.97	0.97	0.96	0.97	0.97	0.96	0.94	0.97	0.96	0.96	0.93
25	500	0.98	0.98	0.97	0.96	0.98	0.97	0.97	0.94	0.98	0.97	0.97	0.95
50	100	0.96	0.96	0.96	0.95	0.96	0.95	0.95	0.94	0.96	0.95	0.95	0.93
50	250	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.96	0.98	0.97	0.98	0.95
50	500	0.99	0.99	0.98	0.98	0.99	0.98	0.98	0.96	0.99	0.98	0.98	0.97
75	100	0.96	0.96	0.96	0.96	0.96	0.95	0.96	0.95	0.96	0.96	0.96	0.94
75	250	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.96
75	500	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	0.98	0.99	0.98
						$K = 3, p = 2$							
25	100	0.95	0.95	0.94	0.93	0.95	0.94	0.94	0.90	0.95	0.94	0.94	0.90
25	250	0.97	0.97	0.96	0.95	0.97	0.96	0.96	0.93	0.97	0.96	0.96	0.92
25	500	0.98	0.97	0.97	0.96	0.98	0.97	0.96	0.93	0.98	0.96	0.96	0.94
50	100	0.96	0.95	0.95	0.95	0.96	0.95	0.95	0.93	0.96	0.95	0.95	0.93
50	250	0.98	0.98	0.97	0.97	0.98	0.97	0.97	0.95	0.98	0.97	0.97	0.95
50	500	0.99	0.98	0.98	0.98	0.99	0.98	0.98	0.96	0.99	0.98	0.98	0.96
75	100	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.94	0.96	0.95	0.96	0.94
75	250	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.96	0.98	0.97	0.98	0.96
75	500	0.99	0.99	0.99	0.98	0.99	0.99	0.98	0.96	0.99	0.98	0.98	0.97
						$K = 5, p = 1$							
25	100	0.95	0.95	0.94	0.93	0.95	0.94	0.94	0.90	0.95	0.94	0.93	0.90
25	250	0.97	0.97	0.96	0.95	0.97	0.96	0.96	0.93	0.97	0.96	0.95	0.92
25	500	0.98	0.98	0.97	0.96	0.98	0.97	0.96	0.93	0.98	0.97	0.96	0.93
50	100	0.96	0.96	0.96	0.95	0.96	0.95	0.95	0.93	0.96	0.95	0.95	0.93
50	250	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.96	0.98	0.97	0.97	0.95
50	500	0.99	0.99	0.98	0.98	0.99	0.98	0.98	0.96	0.99	0.98	0.98	0.97
75	100	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.94	0.96	0.96	0.96	0.94
75	250	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.96
75	500	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	0.98	0.99	0.97
						$K = 5, p = 2$							
25	100	0.95	0.94	0.94	0.93	0.95	0.94	0.93	0.89	0.95	0.93	0.93	0.88
25	250	0.97	0.97	0.96	0.95	0.97	0.96	0.95	0.92	0.97	0.95	0.95	0.91
25	500	0.98	0.97	0.97	0.95	0.98	0.97	0.96	0.92	0.98	0.96	0.96	0.92
50	100	0.96	0.96	0.95	0.95	0.96	0.95	0.95	0.93	0.96	0.95	0.95	0.93
50	250	0.98	0.98	0.97	0.97	0.98	0.97	0.97	0.95	0.98	0.97	0.97	0.95
50	500	0.99	0.98	0.98	0.98	0.99	0.98	0.98	0.95	0.99	0.98	0.98	0.96
75	100	0.96	0.96	0.96	0.96	0.96	0.95	0.96	0.94	0.96	0.95	0.96	0.94
75	250	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.96	0.98	0.97	0.98	0.96
75	500	0.99	0.99	0.99	0.98	0.99	0.98	0.98	0.96	0.99	0.98	0.98	0.97

The displayed means are derived from 500 MC simulations for known dimensions K and p.

[a] For incomplete time series a stock variable is assumed.
[b] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series are stock and flow (average formulation) variables, respectively.
[c] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series serve as stock or change in flow (average formulaton) variables.

Table 4.3: *Ratios of trace R^2 means for random ADFMs using both approaches*

		stock[a]				**stock/flow (average)**[b]				**stock/change in flow (average)**[c]			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
						$K=1, p=1$							
25	100	1.02	1.02	1.02	1.04	1.02	1.02	1.03	1.08	1.02	1.02	1.03	1.06
25	250	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.01	1.01	1.01	1.02
25	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.01
50	100	1.02	1.02	1.02	1.05	1.02	1.02	1.03	1.13	1.02	1.02	1.03	1.09
50	250	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.01	1.01	1.01	1.02
50	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.01
75	100	1.01	1.02	1.04	1.15	1.01	1.02	1.08	1.07	1.01	1.02	1.09	NaN
75	250	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.01	1.01	1.01	1.01
75	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00
						$K=3, p=1$							
25	100	1.03	1.04	1.04	1.07	1.03	1.03	1.05	1.09	1.03	1.03	1.05	1.08
25	250	1.01	1.01	1.02	1.03	1.01	1.01	1.02	1.04	1.01	1.01	1.02	1.03
25	500	1.00	1.00	1.01	1.01	1.00	1.00	1.01	1.02	1.00	1.00	1.01	1.01
50	100	1.03	1.03	1.04	1.06	1.03	1.03	1.05	1.10	1.03	1.03	1.05	1.09
50	250	1.01	1.01	1.01	1.02	1.01	1.01	1.01	1.04	1.01	1.01	1.01	1.03
50	500	1.00	1.00	1.00	1.01	1.00	1.00	1.01	1.02	1.00	1.00	1.01	1.01
75	100	1.02	1.03	1.05	1.08	1.02	1.03	1.06	NaN	1.02	1.03	1.07	NaN
75	250	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.03	1.01	1.01	1.01	1.03
75	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.01
						$K=3, p=2$							
25	100	1.08	1.08	1.11	1.16	1.08	1.08	1.12	1.19	1.08	1.08	1.13	1.20
25	250	1.02	1.03	1.03	1.06	1.02	1.03	1.05	1.12	1.02	1.03	1.06	1.12
25	500	1.01	1.01	1.02	1.02	1.01	1.01	1.02	1.07	1.01	1.01	1.02	1.07
50	100	1.06	1.07	1.08	1.13	1.06	1.07	1.11	1.21	1.06	1.07	1.12	1.21
50	250	1.02	1.02	1.02	1.04	1.02	1.02	1.04	1.10	1.02	1.02	1.04	1.11
50	500	1.01	1.01	1.01	1.02	1.01	1.01	1.02	1.05	1.01	1.01	1.02	1.05
75	100	1.06	1.07	1.09	1.14	1.06	1.07	1.12	NaN	1.06	1.07	1.13	1.24
75	250	1.02	1.02	1.02	1.03	1.02	1.02	1.03	1.09	1.02	1.02	1.03	1.09
75	500	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.04	1.01	1.01	1.01	1.03
						$K=5, p=1$							
25	100	1.04	1.04	1.06	1.09	1.04	1.04	1.06	1.09	1.04	1.04	1.06	1.08
25	250	1.01	1.01	1.02	1.04	1.01	1.01	1.02	1.04	1.01	1.01	1.02	1.04
25	500	1.00	1.01	1.01	1.02	1.00	1.00	1.01	1.02	1.00	1.00	1.01	1.02
50	100	1.03	1.03	1.04	1.07	1.03	1.03	1.06	1.10	1.03	1.03	1.06	1.09
50	250	1.01	1.01	1.01	1.02	1.01	1.01	1.02	1.04	1.01	1.01	1.02	1.04
50	500	1.00	1.00	1.01	1.01	1.00	1.00	1.01	1.02	1.00	1.00	1.01	1.02
75	100	1.03	1.03	1.05	1.07	1.03	1.04	1.06	1.11	1.03	1.04	1.06	1.10
75	250	1.01	1.01	1.01	1.02	1.01	1.01	1.02	1.04	1.01	1.01	1.02	1.03
75	500	1.00	1.00	1.00	1.01	1.00	1.00	1.01	1.02	1.00	1.00	1.01	1.01
						$K=5, p=2$							
25	100	1.08	1.08	1.12	1.18	1.08	1.08	1.13	1.18	1.08	1.08	1.14	1.18
25	250	1.02	1.03	1.04	1.07	1.02	1.03	1.06	1.13	1.02	1.03	1.07	1.12
25	500	1.01	1.01	1.02	1.03	1.01	1.01	1.03	1.08	1.01	1.01	1.03	1.09
50	100	1.08	1.08	1.09	1.13	1.08	1.08	1.12	1.19	1.08	1.08	1.13	1.20
50	250	1.03	1.03	1.03	1.05	1.03	1.02	1.05	1.12	1.03	1.03	1.06	1.12
50	500	1.01	1.01	1.01	1.02	1.01	1.01	1.02	1.07	1.01	1.01	1.03	1.07
75	100	1.07	1.08	1.10	1.13	1.07	1.08	1.12	1.20	1.07	1.08	1.13	1.21
75	250	1.02	1.02	1.03	1.04	1.02	1.02	1.04	1.10	1.02	1.02	1.05	1.11
75	500	1.01	1.01	1.01	1.02	1.01	1.01	1.02	1.05	1.01	1.01	1.02	1.05

The displayed ratios are derived from 500 MC simulations for known dimensions K and p. In doing so, each figure represents the mean of the trace R^2 in Table 4.2 divided by its counterpart in Table 4.1.

[a] For incomplete time series a stock variable is assumed.
[b] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series are stock and flow (average formulation) variables, respectively.
[c] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series serve as stock or change in flow (average formulaton) variables.

In opposition to Tables 4.1-4.3, the two approaches in Tables 4.4-4.6 take the independence of the factors $\boldsymbol{F}_t$ given the observed data $\boldsymbol{X}_t$ into account, i.e., it holds:

$$\mathbb{E}_\Theta\left[\boldsymbol{F}_t\boldsymbol{F}_t'|\boldsymbol{X}_t\right] = \Sigma_{\boldsymbol{F}_t|\boldsymbol{X}_t} + \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_t|\boldsymbol{X}_t}, \tag{4.39}$$

$$\mathbb{E}_\Theta\left[\boldsymbol{F}_t\boldsymbol{F}_s'|\boldsymbol{X}_t, \boldsymbol{X}_s\right] = \boldsymbol{\mu}_{\boldsymbol{F}_t|\boldsymbol{X}_t}\boldsymbol{\mu}'_{\boldsymbol{F}_s|\boldsymbol{X}_s}, \text{ for all } t \neq s. \tag{4.40}$$

With a view to the trace R^2 means in Tables 4.4-4.6, we can confirm the previous findings. Furthermore, this shows that the simplification in the form of (4.39)-(4.40) does not harm.

Tables 4.1-4.6 do not indicate any distinctive differences in the trace R^2 means caused by data type. For the sake of simplicity, we therefore proceed with stock variables only. That is, in the sequel we treat all incomplete time series as stock variables. This restriction is convenient for Tables 4.7 and 4.8, which compare the single-step estimation method in Bańbura and Modugno (2014) with our two-step approach. At first glance, one step less speaks in favor of the single-step ansatz. However, one step less comes with a price, i.e., its state-space representation. Whenever there is a switch from one data type to another, the state-space representation of the overall model in Bańbura and Modugno (2014) calls for adjustments resulting in further effort. In addition the inclusion of mixed-frequency information should rely on a regular scheme, e.g., as for months and quarters. Otherwise, e.g., as for weeks and months, the state-space representation in Bańbura and Modugno (2014) becomes tremendous or requires a recursive implementation of the temporal aggregation in (2.9) as in Bańbura and Rünstler (2011). By contrast, our two-step approach permits any data type and calendar structure through the linear relation in (2.9) and leaves the overall model untouched. This is easy and reduces the risk of mistakes. As already mentioned, Bańbura and Modugno (2014) first derived their estimation method for EDFMs. Thereafter, they followed the argumentation in Doz et al. (2012) to admit weakly cross-sectionally correlated shocks $\boldsymbol{\epsilon}_t$. Since Doz et al. (2012) provided asymptotic results, we would like to assess how the method of Bańbura and Modugno (2014) performs for finite samples with cross-sectionally correlated shocks.

With a view to Tables 4.7 and 4.8, which summarize the results of our MC study, we conclude: First, the general facts remain valid. That is, the higher the ratio of missing data, the worse are the trace R^2 means. Similarly, for more complex dynamics, i.e., larger K and p, the trace R^2 means, ceteris paribus, deteriorate. By contrast, for larger panel data the trace R^2 means improve. Second, for simple factor dynamics, i.e., small K and p, or sufficiently large panel data, cross-sectional correlation of the idiosyncratic shocks does not matter, if the ratio of missing data is low. This is in line with the argumentation in Doz et al. (2012) and Bańbura and Modugno (2014). However, for small panel data, e.g., $N = 25$ and $T = 100$, with 40% missing observations and factor dimensions $K \geq 5$ cross-sectional error correlation matters. This is why our two-step estimation method outperforms the one-step approach of Bańbura and Modugno (2014) in such scenarios.

Next, we focus on our two-step model selection procedure. Thereby, we address the impact of the multiplier m in Lemma 4.1.10 on the estimated factor dimension. For Tables 4.9-4.13, we set $\eta_{\tilde{\boldsymbol{F}}} = 10^{-6}$ in Algorithm 4.1.1. Since Tables 4.9 and 4.10 treat ADFMs with $K \leq 5$ and $p \leq 2$, the upper limits in Lemmata 4.1.10 and 4.1.11 are $\bar{K} = 7$ and $\bar{p} = \bar{p}(K) = 4$. In Tables 4.11-4.13, we have trace R^2 means, estimated factor dimensions and lag orders of ADFMs with $16 \leq K \leq 18$ and $p \leq 2$. Therefore, we specify $\bar{K} = 22$ and $\bar{p} = \bar{p}(K) = 4$ in these cases. For efficiency reasons, the criterion in Lemma 4.1.10 tests factor dimensions in the range $[12, 22]$ instead of the overall range $[1, 22]$.

A comparison of Tables 4.9 and 4.10 shows: First, except for $K = 1$ and $p = 1$, the multipliers $m = 1$ and $m = \frac{1}{2}$ detect the true factor dimension and hence, support that the true lag order is identified. In doing so, larger panel data increases the estimation quality. That is, trace R^2 means increase, while estimated

Table 4.4: *Means of trace R^2 for random ADFMs using the standard KF and KS for the conditional factor moments in Lemma 4.1.7*

		stock[a]				**stock/flow (average)**[b]				**stock/change in flow (average)**[c]			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
						$K=1, p=1$							
25	100	0.94	0.93	0.93	0.90	0.94	0.93	0.92	0.85	0.94	0.93	0.91	0.87
25	250	0.97	0.97	0.96	0.95	0.97	0.96	0.96	0.93	0.97	0.96	0.96	0.92
25	500	0.98	0.98	0.97	0.96	0.98	0.97	0.97	0.94	0.98	0.97	0.97	0.95
50	100	0.95	0.95	0.94	0.91	0.95	0.94	0.93	0.83	0.95	0.94	0.93	0.85
50	250	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.95	0.98	0.96	0.97	0.94
50	500	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.96	0.99	0.98	0.98	0.96
75	100	0.95	0.95	0.92	0.84	0.95	0.94	0.89	0.89	0.95	0.94	0.89	0.86
75	250	0.98	0.98	0.98	0.97	0.98	0.97	0.97	0.95	0.98	0.97	0.97	0.95
75	500	0.99	0.99	0.98	0.98	0.99	0.98	0.98	0.96	0.99	0.98	0.98	0.97
						$K=3, p=1$							
25	100	0.92	0.92	0.90	0.87	0.92	0.92	0.90	0.84	0.92	0.92	0.90	0.85
25	250	0.96	0.96	0.95	0.93	0.96	0.96	0.94	0.91	0.96	0.95	0.94	0.90
25	500	0.98	0.97	0.97	0.96	0.98	0.97	0.96	0.92	0.98	0.97	0.96	0.93
50	100	0.94	0.93	0.93	0.90	0.93	0.93	0.91	0.85	0.94	0.93	0.91	0.86
50	250	0.97	0.97	0.97	0.96	0.97	0.97	0.96	0.93	0.97	0.96	0.96	0.93
50	500	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.98	0.96
75	100	0.94	0.93	0.91	0.89	0.94	0.93	0.90	NaN	0.94	0.92	0.90	0.83
75	250	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.94	0.97	0.97	0.97	0.94
75	500	0.99	0.98	0.98	0.98	0.99	0.98	0.98	0.96	0.99	0.98	0.98	0.97
						$K=3, p=2$							
25	100	0.88	0.88	0.85	0.80	0.88	0.87	0.83	0.75	0.88	0.87	0.83	0.75
25	250	0.95	0.94	0.93	0.90	0.95	0.94	0.91	0.83	0.95	0.93	0.91	0.82
25	500	0.97	0.96	0.96	0.94	0.97	0.96	0.94	0.87	0.97	0.95	0.94	0.88
50	100	0.90	0.90	0.88	0.85	0.90	0.89	0.85	0.77	0.90	0.89	0.85	0.76
50	250	0.96	0.96	0.95	0.94	0.96	0.95	0.94	0.87	0.96	0.95	0.94	0.86
50	500	0.98	0.98	0.97	0.96	0.98	0.97	0.96	0.91	0.98	0.97	0.96	0.92
75	100	0.90	0.90	0.87	0.84	0.91	0.89	0.85	0.74	0.90	0.89	0.84	0.76
75	250	0.96	0.96	0.96	0.95	0.96	0.96	0.95	0.89	0.96	0.96	0.95	0.88
75	500	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.93	0.98	0.97	0.97	0.94
						$K=5, p=1$							
25	100	0.92	0.91	0.89	0.85	0.92	0.91	0.89	0.83	0.92	0.91	0.89	0.83
25	250	0.96	0.96	0.94	0.92	0.96	0.95	0.93	0.89	0.96	0.95	0.94	0.89
25	500	0.97	0.97	0.96	0.94	0.97	0.97	0.95	0.91	0.97	0.96	0.95	0.92
50	100	0.93	0.93	0.92	0.89	0.93	0.92	0.90	0.85	0.93	0.92	0.90	0.86
50	250	0.97	0.97	0.96	0.95	0.97	0.97	0.96	0.92	0.97	0.96	0.95	0.92
50	500	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.94	0.98	0.97	0.97	0.95
75	100	0.93	0.93	0.91	0.89	0.93	0.92	0.90	0.85	0.93	0.92	0.90	0.85
75	250	0.97	0.97	0.97	0.96	0.97	0.97	0.96	0.93	0.97	0.97	0.96	0.93
75	500	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.98	0.96
						$K=5, p=2$							
25	100	0.87	0.87	0.83	0.78	0.87	0.86	0.81	0.74	0.88	0.86	0.81	0.74
25	250	0.95	0.94	0.92	0.88	0.95	0.93	0.90	0.81	0.95	0.92	0.88	0.80
25	500	0.97	0.96	0.95	0.92	0.97	0.96	0.93	0.85	0.97	0.95	0.92	0.85
50	100	0.89	0.89	0.87	0.83	0.89	0.88	0.85	0.77	0.89	0.88	0.84	0.77
50	250	0.95	0.95	0.94	0.92	0.95	0.95	0.92	0.85	0.95	0.94	0.92	0.84
50	500	0.97	0.97	0.97	0.96	0.97	0.97	0.96	0.89	0.97	0.96	0.95	0.90
75	100	0.90	0.89	0.87	0.85	0.90	0.88	0.85	0.78	0.90	0.88	0.84	0.77
75	250	0.96	0.96	0.95	0.94	0.96	0.96	0.94	0.87	0.96	0.95	0.93	0.86
75	500	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.91	0.98	0.97	0.97	0.92

The displayed means are derived from 500 MC simulations for known dimensions K and p.

[a] For incomplete time series a stock variable is assumed.
[b] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series are stock and flow (average formulation) variables, respectively.
[c] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series serve as stock or change in flow (average formulaton) variables.

Table 4.5: *Means of trace R^2 for random ADFMs using closed-form factor moments for the conditional factor moments in Lemma 4.1.7*

		stock[a]				**stock/flow (average)**[b]				**stock/change in flow (average)**[c]			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
$K = 1, p = 1$													
25	100	0.96	0.95	0.95	0.94	0.96	0.94	0.95	0.92	0.96	0.95	0.94	0.92
25	250	0.97	0.97	0.97	0.96	0.97	0.97	0.97	0.95	0.97	0.96	0.96	0.94
25	500	0.98	0.98	0.97	0.97	0.98	0.98	0.97	0.95	0.98	0.97	0.97	0.96
50	100	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.94	0.97	0.95	0.96	0.93
50	250	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.96	0.98	0.97	0.98	0.96
50	500	0.99	0.99	0.99	0.98	0.99	0.98	0.98	0.97	0.99	0.98	0.98	0.97
75	100	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.95	0.96	0.96	0.96	0.94
75	250	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.96
75	500	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	0.98	0.99	0.98
$K = 3, p = 1$													
25	100	0.95	0.95	0.95	0.94	0.95	0.94	0.94	0.92	0.95	0.94	0.94	0.91
25	250	0.97	0.97	0.97	0.96	0.97	0.97	0.96	0.94	0.97	0.96	0.96	0.93
25	500	0.98	0.98	0.97	0.97	0.98	0.97	0.97	0.94	0.98	0.97	0.97	0.95
50	100	0.96	0.96	0.96	0.95	0.96	0.95	0.96	0.94	0.96	0.95	0.96	0.93
50	250	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.96	0.98	0.97	0.98	0.95
50	500	0.99	0.99	0.98	0.98	0.99	0.98	0.98	0.96	0.99	0.98	0.98	0.97
75	100	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.95	0.96	0.95	0.96	0.94
75	250	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.96
75	500	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	0.98	0.99	0.98
$K = 3, p = 2$													
25	100	0.95	0.95	0.94	0.93	0.95	0.94	0.94	0.91	0.95	0.94	0.94	0.90
25	250	0.97	0.97	0.96	0.95	0.97	0.96	0.96	0.93	0.97	0.96	0.96	0.92
25	500	0.98	0.98	0.97	0.96	0.98	0.97	0.96	0.93	0.98	0.96	0.96	0.94
50	100	0.96	0.96	0.96	0.95	0.96	0.95	0.95	0.93	0.96	0.95	0.95	0.93
50	250	0.98	0.98	0.98	0.97	0.98	0.97	0.97	0.96	0.98	0.97	0.97	0.95
50	500	0.99	0.98	0.98	0.98	0.99	0.98	0.98	0.96	0.99	0.98	0.98	0.96
75	100	0.96	0.96	0.96	0.96	0.96	0.95	0.95	0.94	0.96	0.95	0.95	0.94
75	250	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.96
75	500	0.99	0.99	0.99	0.98	0.99	0.99	0.98	0.96	0.99	0.98	0.98	0.97
$K = 5, p = 1$													
25	100	0.95	0.95	0.94	0.93	0.95	0.94	0.94	0.90	0.95	0.94	0.94	0.90
25	250	0.97	0.97	0.96	0.95	0.97	0.96	0.96	0.93	0.97	0.96	0.96	0.92
25	500	0.98	0.98	0.97	0.96	0.98	0.97	0.96	0.93	0.98	0.97	0.96	0.93
50	100	0.96	0.96	0.96	0.95	0.96	0.95	0.95	0.94	0.96	0.95	0.95	0.93
50	250	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.96	0.98	0.97	0.97	0.95
50	500	0.99	0.99	0.98	0.98	0.99	0.98	0.98	0.96	0.99	0.98	0.98	0.97
75	100	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.94	0.96	0.95	0.96	0.94
75	250	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.96
75	500	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	0.98	0.99	0.97
$K = 5, p = 2$													
25	100	0.95	0.95	0.94	0.93	0.95	0.94	0.93	0.89	0.95	0.93	0.93	0.88
25	250	0.97	0.97	0.96	0.95	0.97	0.96	0.95	0.92	0.97	0.95	0.95	0.91
25	500	0.98	0.97	0.97	0.95	0.98	0.97	0.96	0.92	0.98	0.96	0.96	0.93
50	100	0.96	0.96	0.96	0.95	0.96	0.95	0.95	0.93	0.96	0.95	0.95	0.93
50	250	0.98	0.98	0.97	0.97	0.98	0.97	0.97	0.95	0.98	0.97	0.97	0.95
50	500	0.99	0.98	0.98	0.98	0.99	0.98	0.98	0.95	0.99	0.97	0.98	0.96
75	100	0.96	0.96	0.96	0.96	0.96	0.95	0.95	0.94	0.96	0.95	0.95	0.94
75	250	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.96	0.98	0.97	0.98	0.96
75	500	0.99	0.99	0.99	0.98	0.99	0.98	0.98	0.96	0.99	0.98	0.98	0.97

The displayed means are derived from 500 MC simulations for known dimensions K and p.

[a] For incomplete time series a stock variable is assumed.

[b] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series are stock and flow (average formulation) variables, respectively.

[c] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series serve as stock or change in flow (average formulaton) variables.

Table 4.6: *Ratios of trace R^2 means for random ADFMs using both approaches for the conditional factor moments in Lemma 4.1.7*

		stock[a]				**stock/flow (average)**[b]				**stock/change in flow (average)**[c]			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
						$K = 1, p = 1$							
25	100	1.02	1.02	1.02	1.05	1.02	1.02	1.03	1.08	1.02	1.02	1.03	1.06
25	250	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.01	1.01	1.01	1.02
25	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.01
50	100	1.02	1.02	1.02	1.05	1.02	1.02	1.03	1.13	1.02	1.02	1.03	1.10
50	250	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.01	1.01	1.01	1.02
50	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.01
75	100	1.01	1.02	1.05	1.15	1.01	1.02	1.08	1.06	1.01	1.02	1.08	1.10
75	250	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.01	1.01	1.01	1.01
75	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00
						$K = 3, p = 1$							
25	100	1.03	1.04	1.05	1.07	1.03	1.03	1.05	1.09	1.03	1.03	1.05	1.08
25	250	1.01	1.01	1.02	1.03	1.01	1.01	1.02	1.04	1.01	1.01	1.02	1.03
25	500	1.00	1.00	1.01	1.01	1.00	1.00	1.01	1.02	1.00	1.00	1.01	1.02
50	100	1.02	1.03	1.04	1.06	1.03	1.03	1.05	1.10	1.02	1.03	1.05	1.09
50	250	1.01	1.01	1.01	1.02	1.01	1.01	1.01	1.04	1.01	1.01	1.01	1.03
50	500	1.00	1.00	1.00	1.01	1.00	1.00	1.01	1.02	1.00	1.00	1.01	1.01
75	100	1.02	1.03	1.05	1.08	1.03	1.03	1.06	NaN	1.02	1.03	1.07	1.13
75	250	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.03	1.01	1.01	1.01	1.03
75	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.01
						$K = 3, p = 2$							
25	100	1.08	1.08	1.11	1.17	1.08	1.08	1.13	1.21	1.08	1.08	1.14	1.21
25	250	1.02	1.03	1.03	1.06	1.03	1.03	1.05	1.12	1.02	1.03	1.06	1.12
25	500	1.01	1.01	1.01	1.02	1.01	1.01	1.02	1.07	1.01	1.01	1.03	1.07
50	100	1.06	1.06	1.09	1.13	1.06	1.07	1.11	1.21	1.06	1.07	1.12	1.22
50	250	1.02	1.02	1.02	1.04	1.02	1.02	1.04	1.10	1.02	1.02	1.04	1.11
50	500	1.01	1.01	1.01	1.02	1.01	1.01	1.02	1.05	1.01	1.01	1.02	1.05
75	100	1.06	1.07	1.10	1.14	1.06	1.07	1.13	1.27	1.06	1.08	1.13	1.24
75	250	1.02	1.02	1.02	1.03	1.02	1.02	1.03	1.09	1.02	1.02	1.03	1.09
75	500	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.04	1.01	1.01	1.01	1.03
						$K = 5, p = 1$							
25	100	1.04	1.04	1.06	1.10	1.04	1.04	1.06	1.09	1.04	1.04	1.06	1.08
25	250	1.01	1.01	1.02	1.04	1.01	1.01	1.02	1.05	1.01	1.01	1.02	1.04
25	500	1.00	1.01	1.01	1.02	1.00	1.00	1.01	1.02	1.00	1.00	1.01	1.02
50	100	1.03	1.03	1.04	1.07	1.03	1.03	1.06	1.10	1.03	1.03	1.06	1.09
50	250	1.01	1.01	1.01	1.02	1.01	1.01	1.02	1.05	1.01	1.01	1.02	1.04
50	500	1.00	1.00	1.01	1.01	1.00	1.00	1.01	1.02	1.00	1.00	1.01	1.02
75	100	1.03	1.03	1.05	1.07	1.03	1.04	1.06	1.11	1.03	1.04	1.06	1.10
75	250	1.01	1.01	1.01	1.02	1.01	1.01	1.02	1.04	1.01	1.01	1.02	1.03
75	500	1.00	1.00	1.00	1.01	1.00	1.00	1.01	1.02	1.00	1.00	1.01	1.01
						$K = 5, p = 2$							
25	100	1.08	1.09	1.12	1.19	1.09	1.09	1.14	1.20	1.08	1.09	1.15	1.20
25	250	1.02	1.03	1.04	1.08	1.02	1.03	1.06	1.13	1.02	1.03	1.07	1.13
25	500	1.01	1.01	1.01	1.03	1.01	1.01	1.03	1.09	1.01	1.01	1.03	1.09
50	100	1.08	1.08	1.10	1.14	1.08	1.08	1.12	1.20	1.08	1.08	1.13	1.21
50	250	1.03	1.03	1.03	1.05	1.03	1.03	1.05	1.12	1.03	1.03	1.06	1.12
50	500	1.01	1.01	1.01	1.02	1.01	1.01	1.02	1.07	1.01	1.01	1.03	1.07
75	100	1.07	1.08	1.10	1.13	1.07	1.08	1.13	1.21	1.07	1.08	1.13	1.21
75	250	1.02	1.02	1.03	1.04	1.02	1.02	1.04	1.10	1.02	1.02	1.05	1.11
75	500	1.01	1.01	1.01	1.02	1.01	1.01	1.02	1.05	1.01	1.01	1.02	1.05

The displayed ratios are derived from 500 MC simulations for known dimensions K and p. In doing so, each figure represents the mean of the trace R^2 in Table 4.5 divided by its counterpart in Table 4.4.

[a] For incomplete time series a stock variable is assumed.

[b] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series are stock and flow (average formulation) variables, respectively.

[c] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series serve as stock or change in flow (average formulaton) variables.

factor dimensions and lag orders converge to the true ones. By contrast, more gaps deteriorate the results. Second, for $K = 1$ and $p = 1$ the multiplier $m = \frac{1}{2}$ struggles with finding the true factor dimension. This is because of the definition of $\hat{\sigma}^2$ in (4.17). As the true factor dimension is one, the empirical variance of all residuals, i.e., $V_{PPCA}(K)$ in Lemma 4.1.10, is already quite small for $K = 1$ and of course, for $K = N - 1$. In total, the difference $V_{PPCA}(1) - V_{PPCA}(K)$ is also very small, which $m = \frac{1}{2}$ further diminishes. In a nutshell, ADFMs with $K = 1$ and $p = 1$ are possibly too clean for small m.

For a better understanding of the meaning of m, we have a look at Tables 4.11-4.13 and conclude: First, for ADFMs with $16 \leq K \leq 18$, the multiplier $m = \frac{1}{2}$ is too strict, since it provides 12 for the estimated factor dimension, which is the lower limit of our tests, for all cases in Table 4.11. Fortunately, the criterion in Lemma 4.1.11 for estimating the autoregressive order tends to the true one, even though the estimated factor dimension is too small. Second, for $N = 35$ the slope argumentation in the sequel of Lemma 4.1.10 yields $m = \frac{1}{33}$, which properly estimates the true factor dimension for all scenarios in Table 4.12. As a consequence, the trace R^2 means in Table 4.12 clearly dominate their analogs in Table 4.11. Third, we consider $m = \frac{1}{2 \cdot 33} = \frac{1}{66}$ in Table 4.13 for some additional sensitivity analyses. If 40% of the panel data is missing, $m = \frac{1}{66}$ overshoots the true factor dimension, which is reflected in slightly smaller trace R^2 means than in Table 4.12. For lower ratios of missing observations, our two-step estimation method with $m = \frac{1}{66}$ also works well, i.e., it delivers high trace R^2 means and the estimated factor dimensions and lag orders tend towards the true values. With Tables 4.11-4.13 in mind, we recommend for empirical studies to have m rather too small than too big.

Table 4.7: *Comparison of trace R^2 means for random ADFMs using the approach of Bańbura and Modugno (2014) and our two-step estimation method in Section 4.1*

		BM[a]				**CFM**[b]				**CFM/BM**			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
						$K=1,p=1$							
25	100	0.95	0.94	0.94	0.93	0.96	0.95	0.95	0.94	1.01	1.01	1.01	1.01
25	250	0.97	0.97	0.96	0.96	0.97	0.97	0.97	0.96	1.00	1.00	1.00	1.00
25	500	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.97	1.00	1.00	1.00	1.00
50	100	0.95	0.95	0.95	0.95	0.96	0.96	0.96	0.96	1.01	1.01	1.01	1.01
50	250	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.98	1.00	1.00	1.00	1.00
50	500	0.99	0.99	0.98	0.98	0.99	0.99	0.98	0.98	1.00	1.00	1.00	1.00
75	100	0.96	0.95	0.95	0.95	0.97	0.96	0.96	0.96	1.01	1.01	1.01	1.01
75	250	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	1.00	1.00	1.00	1.00
75	500	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00
						$K=3,p=1$							
25	100	0.94	0.94	0.93	0.93	0.95	0.95	0.94	0.94	1.01	1.01	1.01	1.01
25	250	0.97	0.97	0.96	0.96	0.97	0.97	0.97	0.96	1.00	1.00	1.00	1.00
25	500	0.98	0.98	0.97	0.96	0.98	0.98	0.97	0.97	1.00	1.00	1.00	1.00
50	100	0.95	0.95	0.95	0.94	0.96	0.96	0.96	0.95	1.01	1.01	1.01	1.01
50	250	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.97	1.00	1.00	1.00	1.00
50	500	0.98	0.98	0.98	0.98	0.99	0.99	0.98	0.98	1.00	1.00	1.00	1.00
75	100	0.96	0.95	0.95	0.95	0.97	0.96	0.96	0.96	1.01	1.01	1.01	1.01
75	250	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	1.00	1.00	1.00	1.00
75	500	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00
						$K=3,p=2$							
25	100	0.93	0.93	0.92	0.91	0.95	0.95	0.94	0.93	1.02	1.02	1.02	1.02
25	250	0.96	0.96	0.96	0.95	0.97	0.97	0.96	0.95	1.01	1.01	1.01	1.01
25	500	0.97	0.97	0.97	0.96	0.98	0.97	0.97	0.96	1.00	1.00	1.00	1.00
50	100	0.94	0.94	0.93	0.93	0.96	0.96	0.95	0.95	1.02	1.02	1.02	1.02
50	250	0.97	0.97	0.97	0.96	0.98	0.98	0.98	0.97	1.01	1.01	1.01	1.01
50	500	0.98	0.98	0.98	0.98	0.99	0.98	0.98	0.98	1.00	1.00	1.00	1.00
75	100	0.94	0.94	0.94	0.94	0.96	0.96	0.96	0.96	1.02	1.02	1.02	1.02
75	250	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98	1.01	1.01	1.01	1.01
75	500	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.98	1.00	1.00	1.00	1.00
						$K=5,p=1$							
25	100	0.94	0.94	0.93	0.92	0.95	0.95	0.94	0.93	1.01	1.01	1.01	1.01
25	250	0.97	0.97	0.96	0.95	0.97	0.97	0.96	0.95	1.00	1.00	1.00	1.00
25	500	0.98	0.97	0.97	0.96	0.98	0.98	0.97	0.96	1.00	1.00	1.00	1.00
50	100	0.95	0.95	0.95	0.94	0.96	0.96	0.96	0.95	1.01	1.01	1.01	1.01
50	250	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.97	1.00	1.00	1.00	1.00
50	500	0.98	0.98	0.98	0.98	0.99	0.99	0.98	0.98	1.00	1.00	1.00	1.00
75	100	0.95	0.95	0.95	0.95	0.96	0.96	0.96	0.96	1.01	1.01	1.01	1.01
75	250	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.98	1.00	1.00	1.00	1.00
75	500	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00
						$K=5,p=2$							
25	100	0.93	0.93	0.92	0.90	0.95	0.95	0.94	0.93	1.02	1.02	1.02	1.03
25	250	0.96	0.96	0.95	0.94	0.97	0.97	0.96	0.95	1.01	1.01	1.01	1.01
25	500	0.97	0.97	0.96	0.95	0.98	0.97	0.97	0.95	1.00	1.00	1.00	1.00
50	100	0.94	0.94	0.93	0.93	0.96	0.96	0.95	0.95	1.02	1.02	1.02	1.03
50	250	0.97	0.97	0.97	0.96	0.98	0.98	0.98	0.97	1.01	1.01	1.01	1.01
50	500	0.98	0.98	0.98	0.97	0.99	0.98	0.98	0.98	1.00	1.00	1.00	1.00
75	100	0.94	0.94	0.94	0.94	0.96	0.96	0.96	0.96	1.02	1.02	1.02	1.02
75	250	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98	1.01	1.01	1.01	1.01
75	500	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.98	1.00	1.00	1.00	1.00

The means in columns **BM** and **CFM** are derived from 500 MC simulations for known dimensions K and p. The ratios in column **CFM/BM** denote the means in column **CFM** divided by their counterparts in column **BM**. In case of incomplete data, all time series are supposed to be stock variables.

[a] Abbreviation for the estimation method in Bańbura and Modugno (2014).
[b] Abbreviation for closed-form factor moments, that is, the estimation method proposed in Section 4.1.

Table 4.8: *Comparison of trace R^2 means for random ADFMs of higher dimension with many lags using the approach of Bańbura and Modugno (2014) and our two-step estimation method in Section 4.1*

		BM[a]				**CFM**[b]				**CFM/BM**			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
						$K = 5, p = 3$							
25	100	0.91	0.91	0.90	0.88	0.95	0.94	0.94	0.92	1.03	1.04	1.04	1.05
25	250	0.96	0.95	0.94	0.93	0.97	0.96	0.96	0.94	1.01	1.01	1.01	1.01
25	500	0.97	0.97	0.96	0.95	0.98	0.97	0.96	0.95	1.01	1.01	1.01	1.01
50	100	0.92	0.93	0.92	0.92	0.95	0.96	0.95	0.95	1.03	1.03	1.03	1.04
50	250	0.97	0.96	0.96	0.96	0.98	0.98	0.97	0.97	1.01	1.01	1.01	1.01
50	500	0.98	0.98	0.98	0.97	0.99	0.98	0.98	0.98	1.01	1.01	1.01	1.01
75	100	0.93	0.93	0.92	0.92	0.96	0.96	0.96	0.96	1.03	1.03	1.03	1.03
75	250	0.97	0.97	0.97	0.96	0.98	0.98	0.98	0.98	1.01	1.01	1.01	1.01
75	500	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.98	1.01	1.01	1.01	1.01
						$K = 5, p = 4$							
25	100	0.90	0.90	0.88	0.85	0.94	0.94	0.93	0.92	1.05	1.05	1.06	1.08
25	250	0.95	0.95	0.94	0.92	0.97	0.96	0.96	0.94	1.02	1.02	1.02	1.02
25	500	0.97	0.96	0.96	0.94	0.97	0.97	0.96	0.95	1.01	1.01	1.01	1.01
50	100	0.91	0.91	0.91	0.90	0.96	0.95	0.95	0.95	1.05	1.05	1.05	1.05
50	250	0.96	0.96	0.96	0.95	0.98	0.98	0.97	0.97	1.02	1.02	1.02	1.02
50	500	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.98	1.01	1.01	1.01	1.01
75	100	0.92	0.92	0.91	0.91	0.96	0.96	0.96	0.95	1.05	1.05	1.05	1.05
75	250	0.96	0.96	0.96	0.96	0.98	0.98	0.98	0.97	1.02	1.02	1.02	1.02
75	500	0.98	0.98	0.98	0.97	0.99	0.99	0.99	0.98	1.01	1.01	1.01	1.01
						$K = 6, p = 2$							
25	100	0.93	0.92	0.91	0.89	0.95	0.94	0.94	0.92	1.02	1.02	1.03	1.04
25	250	0.96	0.96	0.95	0.93	0.97	0.96	0.96	0.94	1.01	1.01	1.01	1.01
25	500	0.97	0.97	0.96	0.95	0.97	0.97	0.96	0.95	1.00	1.00	1.00	1.00
50	100	0.94	0.93	0.93	0.93	0.96	0.96	0.95	0.95	1.02	1.02	1.02	1.03
50	250	0.97	0.97	0.97	0.96	0.98	0.98	0.98	0.97	1.01	1.01	1.01	1.01
50	500	0.98	0.98	0.98	0.97	0.99	0.98	0.98	0.98	1.00	1.00	1.00	1.00
75	100	0.94	0.94	0.94	0.93	0.96	0.96	0.96	0.95	1.02	1.02	1.02	1.02
75	250	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98	1.01	1.01	1.01	1.01
75	500	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.98	1.00	1.00	1.00	1.00
						$K = 6, p = 3$							
25	100	0.91	0.91	0.89	0.86	0.94	0.94	0.93	0.92	1.04	1.04	1.05	1.07
25	250	0.95	0.95	0.94	0.92	0.97	0.96	0.95	0.94	1.01	1.01	1.01	1.02
25	500	0.97	0.96	0.96	0.94	0.97	0.97	0.96	0.95	1.01	1.01	1.01	1.01
50	100	0.92	0.93	0.92	0.91	0.96	0.96	0.95	0.95	1.03	1.03	1.04	1.04
50	250	0.97	0.96	0.96	0.96	0.98	0.98	0.97	0.97	1.01	1.01	1.01	1.01
50	500	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.98	1.01	1.01	1.01	1.01
75	100	0.93	0.93	0.93	0.92	0.96	0.96	0.96	0.96	1.03	1.03	1.03	1.04
75	250	0.97	0.97	0.97	0.96	0.98	0.98	0.98	0.98	1.01	1.01	1.01	1.01
75	500	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.98	1.01	1.01	1.01	1.01
						$K = 7, p = 3$							
25	100	0.91	0.91	0.88	0.83	0.95	0.94	0.93	0.91	1.04	1.04	1.05	1.09
25	250	0.95	0.95	0.94	0.91	0.97	0.96	0.95	0.93	1.01	1.01	1.02	1.02
25	500	0.97	0.96	0.95	0.94	0.97	0.97	0.96	0.94	1.01	1.01	1.01	1.01
50	100	0.92	0.92	0.92	0.91	0.96	0.95	0.95	0.95	1.03	1.03	1.04	1.04
50	250	0.96	0.96	0.96	0.96	0.98	0.98	0.97	0.97	1.01	1.01	1.01	1.01
50	500	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.98	1.01	1.01	1.01	1.01
75	100	0.93	0.93	0.93	0.92	0.96	0.96	0.96	0.95	1.03	1.03	1.03	1.04
75	250	0.97	0.97	0.97	0.96	0.98	0.98	0.98	0.98	1.01	1.01	1.01	1.01
75	500	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.98	1.01	1.01	1.01	1.01

The means in columns **BM** and **CFM** are derived from 500 MC simulations for known dimensions K and p. The ratios in column **CFM/BM** denote the means in column **CFM** divided by their counterparts in column **BM**. In case of incomplete data, all time series are supposed to be stock variables.

[a] Abbreviation for the estimation method in Bańbura and Modugno (2014).

[b] Abbreviation for closed-form factor moments, that is, the estimation method proposed in Section 4.1.

Table 4.9: *Means of trace R^2 for random ADFMs of low dimensions using our two-step estimation method with $m = 1$*

		trace R^2				**estimated K**				**estimated p**			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
						$K = 1, p = 1$							
25	100	0.96	0.95	0.95	0.94	1.00	1.00	1.08	1.46	1.37	1.34	1.35	1.18
25	250	0.97	0.97	0.97	0.96	1.00	1.00	1.13	1.42	1.37	1.42	1.31	1.25
25	500	0.98	0.98	0.97	0.97	1.00	1.00	1.15	1.44	1.45	1.45	1.31	1.31
50	100	0.96	0.96	0.96	0.96	1.00	1.00	1.07	1.28	1.42	1.38	1.34	1.36
50	250	0.98	0.98	0.98	0.98	1.00	1.00	1.12	1.34	1.42	1.40	1.30	1.37
50	500	0.99	0.99	0.98	0.98	1.00	1.00	1.16	1.45	1.39	1.34	1.35	1.34
75	100	0.96	0.96	0.96	0.96	1.00	1.00	1.01	1.19	1.35	1.40	1.40	1.32
75	250	0.98	0.98	0.98	0.98	1.00	1.00	1.11	1.26	1.39	1.35	1.32	1.32
75	500	0.99	0.99	0.99	0.99	1.00	1.00	1.13	1.42	1.43	1.47	1.38	1.38
						$K = 3, p = 1$							
25	100	0.95	0.95	0.93	0.89	2.99	2.98	2.91	2.77	1.04	1.08	1.07	1.09
25	250	0.97	0.97	0.96	0.94	3.00	3.00	2.97	2.92	1.05	1.04	1.04	1.06
25	500	0.98	0.98	0.97	0.96	3.00	3.00	2.99	2.96	1.05	1.04	1.02	1.06
50	100	0.96	0.96	0.96	0.95	3.00	3.00	3.00	2.97	1.05	1.07	1.05	1.08
50	250	0.98	0.98	0.98	0.97	3.00	3.00	3.00	3.00	1.04	1.04	1.04	1.06
50	500	0.99	0.99	0.98	0.98	3.00	3.00	3.00	3.00	1.03	1.05	1.04	1.04
75	100	0.96	0.96	0.96	0.96	3.00	3.00	3.00	3.00	1.05	1.06	1.09	1.05
75	250	0.98	0.98	0.98	0.98	3.00	3.00	3.00	3.00	1.05	1.04	1.04	1.04
75	500	0.99	0.99	0.99	0.99	3.00	3.00	3.00	3.00	1.04	1.05	1.05	1.05
						$K = 3, p = 2$							
25	100	0.95	0.94	0.92	0.87	2.99	2.98	2.89	2.72	1.65	1.74	1.69	1.76
25	250	0.97	0.97	0.96	0.93	3.00	3.00	2.98	2.90	2.03	2.02	2.04	2.02
25	500	0.98	0.97	0.97	0.94	3.00	3.00	2.99	2.93	2.04	2.04	2.05	2.04
50	100	0.96	0.96	0.96	0.94	3.00	3.00	3.00	2.95	1.72	1.78	1.72	1.70
50	250	0.98	0.98	0.98	0.97	3.00	3.00	3.00	3.00	2.03	2.03	2.04	2.03
50	500	0.99	0.98	0.98	0.98	3.00	3.00	3.00	3.00	2.03	2.03	2.06	2.06
75	100	0.96	0.96	0.96	0.96	3.00	3.00	3.00	3.00	1.68	1.71	1.76	1.78
75	250	0.98	0.98	0.98	0.98	3.00	3.00	3.00	3.00	2.05	2.04	2.02	2.03
75	500	0.99	0.99	0.99	0.98	3.00	3.00	3.00	3.00	2.05	2.05	2.05	2.06
						$K = 5, p = 1$							
25	100	0.78	0.74	0.68	0.60	3.74	3.50	3.16	2.74	1.01	1.03	1.05	1.04
25	250	0.83	0.80	0.74	0.65	4.08	3.91	3.57	3.10	1.01	1.02	1.03	1.05
25	500	0.85	0.83	0.77	0.68	4.17	4.08	3.75	3.32	1.00	1.01	1.02	1.06
50	100	0.92	0.89	0.83	0.74	4.62	4.39	3.95	3.44	1.01	1.01	1.02	1.04
50	250	0.97	0.96	0.93	0.86	4.96	4.89	4.65	4.18	1.00	1.00	1.01	1.02
50	500	0.98	0.98	0.96	0.91	4.99	4.97	4.85	4.52	1.00	1.00	1.01	1.01
75	100	0.96	0.94	0.89	0.82	4.93	4.75	4.36	3.87	1.01	1.01	1.01	1.02
75	250	0.98	0.98	0.98	0.94	5.00	5.00	4.96	4.74	1.00	1.01	1.00	1.01
75	500	0.99	0.99	0.99	0.98	5.00	5.00	5.00	4.96	1.01	1.00	1.01	1.00
						$K = 5, p = 2$							
25	100	0.77	0.72	0.64	0.56	3.78	3.45	3.05	2.64	1.60	1.62	1.72	1.80
25	250	0.83	0.79	0.71	0.61	4.13	3.90	3.49	3.05	1.99	1.99	2.01	1.99
25	500	0.85	0.81	0.74	0.64	4.23	4.06	3.67	3.21	2.00	2.01	2.01	2.01
50	100	0.92	0.88	0.79	0.70	4.70	4.36	3.82	3.29	1.47	1.56	1.66	1.79
50	250	0.98	0.97	0.93	0.83	4.99	4.93	4.68	4.10	1.98	1.99	1.99	2.00
50	500	0.98	0.98	0.96	0.89	4.99	4.98	4.83	4.45	2.01	2.00	2.00	2.00
75	100	0.95	0.94	0.87	0.79	4.94	4.83	4.28	3.73	1.44	1.48	1.59	1.71
75	250	0.98	0.98	0.98	0.94	5.00	5.00	4.97	4.72	1.99	1.99	2.00	1.99
75	500	0.99	0.99	0.99	0.98	5.00	5.00	5.00	4.97	2.01	2.00	2.00	2.00

The means in column **trace** R^2 are derived from 500 MC simulations for unknown dimensions K and p. Therefore, columns two and three show the corresponding means of the estimted factor dimension K and lag length p. In case of incomplete data, all time series are supposed to be stock variables.

Table 4.10: *Means of trace R^2 for random ADFMs of low dimensions using our two-step estimation method with $m = \frac{1}{2}$*

		trace R^2				**estimated** K				**estimated** p			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
						$K = 1, p = 1$							
25	100	0.96	0.95	0.95	0.94	2.70	2.90	3.52	4.06	0.81	0.77	0.66	0.59
25	250	0.97	0.97	0.97	0.96	2.13	2.58	3.64	4.33	1.09	1.03	0.89	0.86
25	500	0.98	0.98	0.98	0.97	1.76	2.23	3.48	4.47	1.19	1.12	0.99	0.96
50	100	0.96	0.96	0.96	0.96	1.61	2.11	3.58	4.41	1.13	0.95	0.68	0.63
50	250	0.98	0.98	0.98	0.97	1.20	1.63	3.78	5.28	1.27	1.21	0.90	0.75
50	500	0.99	0.99	0.98	0.98	1.04	1.24	3.35	5.41	1.41	1.33	1.01	0.91
75	100	0.96	0.96	0.96	0.96	1.10	1.62	3.58	4.79	1.34	1.09	0.74	0.53
75	250	0.98	0.98	0.98	0.98	1.02	1.26	3.66	5.73	1.44	1.32	0.94	0.73
75	500	0.99	0.99	0.99	0.99	1.00	1.06	3.27	6.12	1.43	1.41	1.00	0.91
						$K = 3, p = 1$							
25	100	0.95	0.95	0.94	0.94	3.00	3.00	3.00	3.01	1.04	1.04	1.06	1.06
25	250	0.97	0.97	0.97	0.96	3.00	3.00	3.00	3.01	1.03	1.02	1.06	1.04
25	500	0.98	0.98	0.97	0.97	3.00	3.00	3.00	3.01	1.04	1.05	1.03	1.03
50	100	0.96	0.96	0.96	0.95	3.00	3.00	3.00	3.01	1.06	1.07	1.06	1.07
50	250	0.98	0.98	0.98	0.97	3.00	3.00	3.00	3.03	1.03	1.06	1.05	1.06
50	500	0.99	0.99	0.98	0.98	3.00	3.00	3.00	3.01	1.06	1.05	1.04	1.03
75	100	0.96	0.96	0.96	0.96	3.00	3.00	3.00	3.00	1.05	1.06	1.05	1.10
75	250	0.98	0.98	0.98	0.98	3.00	3.00	3.00	3.01	1.04	1.06	1.04	1.05
75	500	0.99	0.99	0.99	0.99	3.00	3.00	3.00	3.01	1.05	1.03	1.03	1.04
						$K = 3, p = 2$							
25	100	0.95	0.95	0.94	0.93	3.00	3.00	3.00	3.01	1.67	1.67	1.63	1.65
25	250	0.97	0.97	0.96	0.95	3.00	3.00	3.00	3.00	2.04	2.01	2.03	2.01
25	500	0.98	0.98	0.97	0.96	3.00	3.00	3.00	3.02	2.05	2.03	2.03	2.04
50	100	0.96	0.95	0.96	0.95	3.00	3.00	3.00	3.01	1.72	1.75	1.77	1.76
50	250	0.98	0.98	0.98	0.97	3.00	3.00	3.00	3.02	2.03	2.05	2.03	2.05
50	500	0.99	0.98	0.98	0.98	3.00	3.00	3.00	3.02	2.03	2.04	2.06	2.04
75	100	0.96	0.96	0.96	0.96	3.00	3.00	3.00	3.00	1.70	1.72	1.76	1.73
75	250	0.98	0.98	0.98	0.98	3.00	3.00	3.00	3.01	2.04	2.03	2.04	2.04
75	500	0.99	0.99	0.99	0.98	3.00	3.00	3.00	3.02	2.03	2.04	2.04	2.05
						$K = 5, p = 1$							
25	100	0.94	0.92	0.90	0.87	4.87	4.75	4.65	4.50	1.01	1.01	1.01	1.01
25	250	0.97	0.96	0.95	0.92	4.95	4.95	4.90	4.79	1.00	1.00	1.00	1.01
25	500	0.97	0.97	0.96	0.94	4.96	4.97	4.92	4.85	1.00	1.00	1.00	1.01
50	100	0.96	0.96	0.95	0.95	5.00	5.00	4.98	4.91	1.00	1.01	1.01	1.01
50	250	0.98	0.98	0.98	0.97	5.00	5.00	5.00	5.00	1.01	1.01	1.00	1.00
50	500	0.99	0.99	0.98	0.98	5.00	5.00	5.00	5.00	1.00	1.00	1.01	1.00
75	100	0.96	0.96	0.96	0.96	5.00	5.00	5.00	4.99	1.00	1.00	1.01	1.01
75	250	0.98	0.98	0.98	0.98	5.00	5.00	5.00	5.00	1.00	1.00	1.00	1.01
75	500	0.99	0.99	0.99	0.99	5.00	5.00	5.00	5.00	1.00	1.00	1.00	1.00
						$K = 5, p = 2$							
25	100	0.94	0.92	0.89	0.85	4.88	4.78	4.60	4.47	1.35	1.42	1.43	1.47
25	250	0.96	0.96	0.94	0.91	4.96	4.95	4.88	4.78	1.98	1.99	1.99	1.97
25	500	0.97	0.97	0.95	0.93	4.98	4.96	4.92	4.85	2.00	2.00	2.00	2.00
50	100	0.96	0.96	0.95	0.94	5.00	5.00	4.98	4.92	1.45	1.46	1.43	1.47
50	250	0.98	0.98	0.98	0.97	5.00	5.00	5.00	5.00	1.99	1.99	1.99	2.00
50	500	0.99	0.98	0.98	0.98	5.00	5.00	5.00	5.00	2.00	2.01	2.00	2.00
75	100	0.96	0.96	0.96	0.96	5.00	5.00	5.00	4.99	1.46	1.42	1.42	1.41
75	250	0.98	0.98	0.98	0.98	5.00	5.00	5.00	5.00	1.99	1.99	2.00	1.98
75	500	0.99	0.99	0.99	0.98	5.00	5.00	5.00	5.00	2.00	2.00	2.00	2.01

The means in column **trace** R^2 are derived from 500 MC simulations for unknown dimensions K and p. Therefore, columns two and three show the corresponding means of the estimted factor dimension K and lag length p. In case of incomplete data, all time series are supposed to be stock variables.

Table 4.11: *Means of trace R^2 for random ADFMs of large dimensions using our two-step estimation method with $m = \frac{1}{2}$*

		trace R^2				**estimated K**				**estimated p**			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
$K = 16, p = 1$													
30	300	0.78	0.78	0.75	0.71	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
30	350	0.78	0.77	0.75	0.71	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
30	400	0.78	0.77	0.75	0.71	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
35	300	0.79	0.78	0.76	0.73	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
35	350	0.79	0.78	0.76	0.73	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
35	400	0.78	0.78	0.76	0.73	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
40	300	0.79	0.79	0.77	0.75	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
40	350	0.79	0.78	0.77	0.75	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
40	400	0.79	0.78	0.77	0.74	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
$K = 16, p = 2$													
30	300	0.77	0.76	0.74	0.69	12.00	12.00	12.00	12.00	1.82	1.84	1.83	1.79
30	350	0.77	0.76	0.74	0.69	12.00	12.00	12.00	12.00	1.97	1.97	1.96	1.95
30	400	0.77	0.76	0.74	0.69	12.00	12.00	12.00	12.00	1.99	1.99	1.99	1.99
35	300	0.78	0.77	0.75	0.72	12.00	12.00	12.00	12.00	1.87	1.87	1.87	1.82
35	350	0.78	0.77	0.75	0.71	12.00	12.00	12.00	12.00	1.97	1.97	1.95	1.96
35	400	0.77	0.77	0.75	0.71	12.00	12.00	12.00	12.00	1.99	1.99	2.00	1.99
40	300	0.78	0.78	0.76	0.73	12.00	12.00	12.00	12.00	1.92	1.86	1.88	1.89
40	350	0.78	0.77	0.76	0.73	12.00	12.00	12.00	12.00	1.97	1.98	1.98	1.98
40	400	0.78	0.77	0.76	0.73	12.00	12.00	12.00	12.00	1.99	1.99	2.00	2.00
$K = 17, p = 1$													
30	300	0.75	0.74	0.72	0.67	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
30	350	0.75	0.74	0.71	0.67	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
30	400	0.74	0.73	0.71	0.67	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
35	300	0.75	0.75	0.73	0.69	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
35	350	0.75	0.74	0.73	0.69	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
35	400	0.75	0.74	0.72	0.69	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
40	300	0.76	0.75	0.74	0.71	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
40	350	0.76	0.75	0.73	0.70	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
40	400	0.75	0.75	0.73	0.70	12.00	12.00	12.00	12.00	1.00	1.00	1.00	1.00
$K = 17, p = 2$													
30	300	0.74	0.73	0.70	0.66	12.00	12.00	12.00	12.00	1.80	1.78	1.81	1.77
30	350	0.73	0.72	0.70	0.65	12.00	12.00	12.00	12.00	1.95	1.96	1.94	1.92
30	400	0.73	0.72	0.70	0.65	12.00	12.00	12.00	12.00	1.98	1.98	1.99	1.98
35	300	0.74	0.73	0.71	0.68	12.00	12.00	12.00	12.00	1.88	1.83	1.84	1.83
35	350	0.74	0.73	0.71	0.67	12.00	12.00	12.00	12.00	1.97	1.95	1.97	1.95
35	400	0.74	0.73	0.71	0.67	12.00	12.00	12.00	12.00	2.00	2.00	2.00	1.99
40	300	0.75	0.74	0.72	0.69	12.00	12.00	12.00	12.00	1.89	1.87	1.88	1.89
40	350	0.74	0.74	0.72	0.69	12.00	12.00	12.00	12.00	1.97	1.98	1.98	1.98
40	400	0.74	0.73	0.72	0.69	12.00	12.00	12.00	12.00	2.00	2.00	1.99	1.99
$K = 18, p = 2$													
30	300	0.70	0.69	0.67	0.62	12.00	12.00	12.00	12.00	1.76	1.78	1.78	1.73
30	350	0.70	0.69	0.66	0.62	12.00	12.00	12.00	12.00	1.93	1.94	1.94	1.93
30	400	0.70	0.69	0.66	0.62	12.00	12.00	12.00	12.00	1.98	1.98	2.00	1.98
35	300	0.71	0.70	0.68	0.64	12.00	12.00	12.00	12.00	1.80	1.81	1.84	1.81
35	350	0.70	0.70	0.68	0.64	12.00	12.00	12.00	12.00	1.97	1.97	1.95	1.95
35	400	0.70	0.69	0.67	0.64	12.00	12.00	12.00	12.00	1.98	1.99	1.98	1.98
40	300	0.71	0.71	0.69	0.66	12.00	12.00	12.00	12.00	1.87	1.87	1.87	1.86
40	350	0.71	0.70	0.68	0.65	12.00	12.00	12.00	12.00	1.97	1.96	1.97	1.96
40	400	0.71	0.70	0.68	0.65	12.00	12.00	12.00	12.00	1.99	1.99	2.00	1.99

The means in column **trace** R^2 are derived from 500 MC simulations for unknown dimensions K and p. Therefore, columns two and three show the corresponding means of the estimted factor dimension K and lag length p. In case of incomplete data, all time series are supposed to be stock variables.

Table 4.12: *Means of trace R^2 for random ADFMs of large dimensions using our two-step estimation method with $m = \frac{1}{33}$*

		trace R^2				**estimated K**				**estimated p**			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
						$K = 16, p = 1$							
30	300	0.97	0.96	0.94	0.84	15.98	15.96	15.99	17.79	1.00	1.00	1.00	1.00
30	350	0.97	0.96	0.94	0.85	15.96	15.99	15.99	17.38	1.00	1.00	1.00	1.00
30	400	0.97	0.96	0.94	0.85	15.98	15.98	15.99	17.06	1.00	1.00	1.00	1.00
35	300	0.97	0.97	0.96	0.90	16.00	16.00	16.01	17.04	1.00	1.00	1.00	1.00
35	350	0.98	0.97	0.96	0.90	16.00	16.00	16.00	17.03	1.00	1.00	1.00	1.00
35	400	0.98	0.97	0.96	0.90	16.00	16.00	16.01	16.95	1.00	1.00	1.00	1.00
40	300	0.98	0.97	0.97	0.92	16.00	16.00	16.01	17.27	1.00	1.00	1.00	1.00
40	350	0.98	0.98	0.97	0.92	16.00	16.00	16.01	17.08	1.00	1.00	1.00	1.00
40	400	0.98	0.98	0.97	0.93	16.00	16.00	16.01	16.91	1.00	1.00	1.00	1.00
						$K = 16, p = 2$							
30	300	0.96	0.96	0.93	0.81	15.98	15.98	16.03	19.36	1.53	1.56	1.50	1.06
30	350	0.97	0.96	0.93	0.82	15.98	15.99	16.03	18.89	1.84	1.83	1.74	1.17
30	400	0.97	0.96	0.93	0.83	15.99	15.99	16.00	18.32	1.94	1.96	1.92	1.41
35	300	0.97	0.97	0.95	0.88	16.00	16.00	16.02	17.55	1.66	1.60	1.57	1.31
35	350	0.97	0.97	0.95	0.89	16.00	16.00	16.02	17.40	1.84	1.85	1.83	1.49
35	400	0.97	0.97	0.95	0.88	16.00	16.00	16.01	17.48	1.95	1.97	1.96	1.59
40	300	0.97	0.97	0.96	0.92	16.00	16.00	16.01	17.34	1.66	1.66	1.65	1.41
40	350	0.98	0.97	0.96	0.91	16.00	16.00	16.02	17.49	1.89	1.89	1.88	1.53
40	400	0.98	0.97	0.96	0.91	16.00	16.00	16.03	17.43	1.97	1.97	1.95	1.65
						$K = 17, p = 1$							
30	300	0.96	0.95	0.93	0.82	16.92	16.88	16.92	18.74	1.00	1.00	1.00	1.00
30	350	0.97	0.96	0.93	0.83	16.91	16.88	16.90	18.33	1.00	1.00	1.00	1.00
30	400	0.97	0.96	0.93	0.83	16.94	16.91	16.92	18.19	1.00	1.00	1.00	1.00
35	300	0.97	0.97	0.95	0.89	16.99	16.99	16.99	17.80	1.00	1.00	1.00	1.00
35	350	0.97	0.97	0.95	0.89	17.00	17.00	17.00	17.72	1.00	1.00	1.00	1.00
35	400	0.98	0.97	0.95	0.89	16.99	16.99	17.00	17.68	1.00	1.00	1.00	1.00
40	300	0.98	0.97	0.96	0.91	17.00	17.00	17.00	18.00	1.00	1.00	1.00	1.00
40	350	0.98	0.97	0.96	0.92	17.00	17.00	17.00	17.78	1.00	1.00	1.00	1.00
40	400	0.98	0.98	0.97	0.92	17.00	17.00	17.00	17.62	1.00	1.00	1.00	1.00
						$K = 17, p = 2$							
30	300	0.96	0.95	0.91	0.79	16.92	16.92	16.95	19.85	1.40	1.41	1.31	1.03
30	350	0.96	0.95	0.92	0.80	16.95	16.93	16.95	19.58	1.76	1.75	1.65	1.13
30	400	0.96	0.95	0.92	0.80	16.95	16.95	16.94	19.19	1.91	1.89	1.87	1.32
35	300	0.97	0.96	0.95	0.87	16.99	17.00	17.00	18.37	1.45	1.46	1.46	1.21
35	350	0.97	0.97	0.95	0.88	16.99	17.00	17.00	18.07	1.79	1.79	1.74	1.46
35	400	0.97	0.97	0.95	0.88	17.00	17.00	17.01	17.99	1.95	1.93	1.91	1.64
40	300	0.97	0.97	0.96	0.91	17.00	17.00	17.01	17.95	1.53	1.55	1.51	1.37
40	350	0.98	0.97	0.96	0.91	17.00	17.00	17.01	18.16	1.81	1.81	1.81	1.51
40	400	0.98	0.97	0.96	0.91	17.00	17.00	17.00	18.08	1.94	1.95	1.94	1.65
						$K = 18, p = 2$							
30	300	0.96	0.94	0.90	0.77	17.82	17.81	17.82	20.46	1.28	1.32	1.26	1.01
30	350	0.96	0.95	0.90	0.77	17.78	17.83	17.82	20.19	1.65	1.61	1.52	1.07
30	400	0.96	0.95	0.91	0.78	17.84	17.81	17.83	19.89	1.88	1.85	1.80	1.20
35	300	0.97	0.96	0.94	0.86	17.99	17.98	17.99	19.43	1.39	1.35	1.35	1.11
35	350	0.97	0.96	0.94	0.86	17.99	17.99	17.99	19.00	1.69	1.70	1.65	1.40
35	400	0.97	0.97	0.94	0.86	17.99	17.99	17.98	18.81	1.89	1.90	1.88	1.56
40	300	0.97	0.97	0.95	0.90	18.00	18.00	18.01	18.61	1.44	1.37	1.40	1.24
40	350	0.98	0.97	0.96	0.91	18.00	18.00	18.01	18.65	1.75	1.79	1.72	1.53
40	400	0.98	0.97	0.96	0.91	18.00	18.00	18.00	18.64	1.91	1.91	1.93	1.71

The means in column **trace** R^2 are derived from 500 MC simulations for unknown dimensions K and p. Therefore, columns two and three show the corresponding means of the estimted factor dimension K and lag length p. In case of incomplete data, all time series are supposed to be stock variables.

Table 4.13: *Means of trace R^2 for random ADFMs of large dimensions using our two-step estimation method with $m = \frac{1}{66}$*

		trace R^2				**estimated K**				**estimated p**			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	10%	25%	40%	0%	10%	25%	40%	0%	10%	25%	40%
						$K = 16, p = 1$							
30	300	0.97	0.96	0.93	0.81	16.00	16.00	16.80	21.87	1.00	1.00	1.00	1.00
30	350	0.97	0.96	0.93	0.82	16.00	16.00	16.73	21.82	1.00	1.00	1.00	1.00
30	400	0.97	0.96	0.93	0.83	16.00	16.00	16.59	21.77	1.00	1.00	1.00	1.00
35	300	0.97	0.97	0.95	0.88	16.00	16.00	16.40	21.89	1.00	1.00	1.00	1.00
35	350	0.97	0.97	0.95	0.89	16.00	16.00	16.35	21.81	1.00	1.00	1.00	1.00
35	400	0.98	0.97	0.96	0.89	16.00	16.00	16.18	21.62	1.00	1.00	1.00	1.00
40	300	0.98	0.97	0.96	0.93	16.00	16.01	16.18	21.03	1.00	1.00	1.00	1.00
40	350	0.98	0.98	0.97	0.93	16.00	16.00	16.18	20.11	1.00	1.00	1.00	1.00
40	400	0.98	0.98	0.97	0.94	16.00	16.01	16.14	19.48	1.00	1.00	1.00	1.00
						$K = 16, p = 2$							
30	300	0.97	0.96	0.91	0.79	16.00	16.01	17.62	21.97	1.56	1.58	1.23	1.00
30	350	0.97	0.96	0.91	0.80	16.00	16.00	17.49	21.95	1.79	1.84	1.42	1.00
30	400	0.97	0.96	0.92	0.80	16.00	16.00	17.33	21.93	1.92	1.95	1.61	1.01
35	300	0.97	0.97	0.94	0.86	16.00	16.01	17.10	21.99	1.60	1.64	1.35	0.97
35	350	0.97	0.97	0.95	0.87	16.00	16.01	16.77	21.98	1.88	1.89	1.64	1.00
35	400	0.97	0.97	0.95	0.88	16.00	16.00	16.45	21.96	1.97	1.94	1.86	1.03
40	300	0.97	0.97	0.96	0.91	16.00	16.02	16.50	21.84	1.68	1.67	1.51	0.92
40	350	0.98	0.97	0.96	0.92	16.00	16.01	16.32	21.63	1.89	1.89	1.83	1.03
40	400	0.98	0.97	0.96	0.92	16.00	16.01	16.23	21.30	1.96	1.96	1.93	1.19
						$K = 17, p = 1$							
30	300	0.97	0.96	0.92	0.80	17.00	17.00	17.66	21.90	1.00	1.00	1.00	1.00
30	350	0.97	0.96	0.92	0.80	17.00	17.00	17.56	21.87	1.00	1.00	1.00	1.00
30	400	0.97	0.96	0.92	0.81	17.00	17.00	17.48	21.84	1.00	1.00	1.00	1.00
35	300	0.97	0.97	0.95	0.87	17.00	17.00	17.39	21.95	1.00	1.00	1.00	1.00
35	350	0.97	0.97	0.95	0.87	17.00	17.00	17.31	21.92	1.00	1.00	1.00	1.00
35	400	0.98	0.97	0.95	0.88	17.00	17.00	17.20	21.81	1.00	1.00	1.00	1.00
40	300	0.98	0.97	0.96	0.92	17.00	17.00	17.20	21.39	1.00	1.00	1.00	1.00
40	350	0.98	0.97	0.96	0.92	17.00	17.00	17.15	21.03	1.00	1.00	1.00	1.00
40	400	0.98	0.98	0.96	0.93	17.00	17.00	17.11	20.18	1.00	1.00	1.00	1.00
						$K = 17, p = 2$							
30	300	0.96	0.95	0.90	0.78	17.00	17.00	18.32	21.97	1.41	1.39	1.18	1.00
30	350	0.96	0.96	0.91	0.78	17.00	17.00	18.15	21.97	1.76	1.71	1.35	1.00
30	400	0.97	0.96	0.91	0.79	17.00	17.00	17.95	21.97	1.92	1.88	1.59	1.02
35	300	0.97	0.96	0.94	0.85	17.00	17.00	17.98	22.00	1.49	1.48	1.29	0.99
35	350	0.97	0.97	0.94	0.86	17.00	17.01	17.70	21.98	1.79	1.78	1.53	1.01
35	400	0.97	0.97	0.95	0.86	17.00	17.00	17.38	21.99	1.92	1.93	1.85	1.06
40	300	0.97	0.97	0.96	0.91	17.00	17.01	17.47	21.95	1.54	1.54	1.44	0.95
40	350	0.98	0.97	0.96	0.91	17.00	17.00	17.28	21.77	1.82	1.81	1.71	1.06
40	400	0.98	0.97	0.96	0.92	17.00	17.01	17.21	21.66	1.96	1.96	1.91	1.16
						$K = 18, p = 2$							
30	300	0.96	0.95	0.90	0.76	17.99	18.00	18.79	21.99	1.28	1.29	1.13	1.00
30	350	0.96	0.95	0.90	0.76	18.00	18.00	18.80	21.98	1.61	1.57	1.33	1.00
30	400	0.96	0.95	0.90	0.77	18.00	18.00	18.71	21.97	1.85	1.82	1.56	1.02
35	300	0.97	0.96	0.93	0.84	18.00	18.00	18.74	22.00	1.32	1.40	1.22	1.00
35	350	0.97	0.96	0.93	0.85	18.00	18.00	18.52	22.00	1.71	1.69	1.48	1.02
35	400	0.97	0.97	0.94	0.85	18.00	18.00	18.42	21.99	1.91	1.88	1.71	1.10
40	300	0.97	0.97	0.95	0.90	18.00	18.01	18.36	21.98	1.45	1.44	1.33	0.98
40	350	0.98	0.97	0.95	0.90	18.00	18.01	18.31	21.95	1.74	1.71	1.62	1.07
40	400	0.98	0.97	0.96	0.91	18.00	18.00	18.15	21.85	1.90	1.92	1.86	1.21

The means in column **trace** R^2 are derived from 500 MC simulations for unknown dimensions K and p. Therefore, columns two and three show the corresponding means of the estimted factor dimension K and lag length p. In case of incomplete data, all time series are supposed to be stock variables.

4.5 Empirical Application

This section applies the developed framework to real-world data. Because of its importance, liquidity and size, we focus on the US stock market represented by the S&P500. Diverse publication conventions and delays require us to declare, when we run our updates. From a business perspective the period between the end of trading on Friday and its restart on Monday is reasonable. On the one hand, there is plenty of time after the day-to-day business is done. On the other hand, there is enough time left to prepare changes in existing asset allocations triggered by the gained information, e.g., the weekly prediction intervals, until the stock exchange reopens. In this example, we have a weekly horizon such that the obtained prediction intervals cover the expected S&P500 log-return until the next Friday.

At the beginning, we describe our panel data. For the convenience of the reader, we summarize the panel data including its sources in Appendix B.2. Here, we mention some characteristics of the raw information, explain the preprocessing of the inputs and state the data types (stock, flow or change in flow variable) of the transformed time series. Some inputs are related with each other, therefore, we group them, before we analyze the drivers of the predicted log-returns. This improves the clarity of our results, in particular, when we illustrate them.

US Treasuries rank among the most important assets for investors. They offer a steady income (coupons, accrued interest and profits, if bought below par) and are deemed as comparably less risky. The Treasuries' market size and liquidity support their trading. Their risk-return profile affects investors in their asset allocation decisions and hence, in their willingness to buy or sell stocks, which finally moves the S&P500. This is why our panel data comprises constant maturity rates of US Treasuries for maturities of three months up to ten years. For stationarity reasons, we use changes, i.e., first differences, in the weekly rates. Those are regarded as stock variables, as there are no data gaps.

Corporate Bonds and London Interbank Offered Rates (LIBOR) provide further investment opportunities and thus, also influence portfolio decisions. To take US Corporate Bonds into account we add proper yields and spreads published by Moody's and Merrill Lynch to our panel data. In terms of LIBOR, we use rates for maturities of one month until twelve months. As before, we determine differences between weekly rates to bypass potential stationarity violations and treat them as stock variables.

Foreign Exchange (FX) Rates and Gold are investment possibilities, which enable investors to leave the USD. For clarity reasons, we therefore form the joint driver FX&Gold. As their quotes are always positive, we take their weekly log-returns to ensure stationarity. Again, there are no missing observations and so, those serve as stock variables.

So far, our panel data is complete and only covers financial information. The intention behind and area of application affect how real-economy data is interpreted. This is why our subsequent classification serves as an example. The first macroeconomic pillar collects all drivers of the Demand for goods and services. The lower the Unemployment Rate, the more people spend and invest money. Similarly, the higher the Personal Income, the more money for spending and investing is available. The higher the Personal Savings Rate, the less money is spent for consumption, but the demand for stocks may increase. As measure of real consumption, we add the Personal Consumption Expenditures. In addition, the Governmental Total Expenditures affect the demand side. In case of Unemployment and Personal Savings Rate, the difference between two subsequent months serves as average formulation of a change in flow variable. For Personal Income, Personal Consumption Expenditures and Governmental Total Expenditures the annualized log-returns take the role of average versions of flow variables.

GDP, Industrial Production Index, Real Exports of Goods & Services and Real Imports of Goods & Services build the second macroeconomic pillar measuring Supply. For stationarity reasons, their annualized log-returns are regarded as flow variables (average formulation).

GDP aggregates the worth of all goods and services an economy produces within a certain period of time, e.g., one year. For the US this value is expressed in USD. If an increase in GDP occurs, the figure itself does not provide any information whether the output was boosted, prices raised or a mixture of both was causal. To partially overcome this problem the third pillar, which comprises the Consumer Price Index for All Urban Consumers and the Producer Price Index for All Commodities, reveals how Inflation evolves. In this context, the annualized log-returns of both enter our investigations and serve as average versions of flow variables.

Next, we summarize some general facts. The overall sample ranges from January 15, 1999 until February 5, 2016 and is weekly updated. We specify a rolling window of 364 weeks, i.e., seven years, such that the period from January 15, 1999 until December 30, 2005 constitutes our insample period. Based on this we construct the first prediction interval for the S&P500 log-return from December 30, 2005 until January 6, 2006. Then, we shift the rolling window by one week and repeat all steps (including model selection and parameter estimation) to derive the second prediction interval. In the sequel, we proceed in this way until the sample end is reached. As the length of the rolling window is kept, the estimated contributions remain comparable, when time goes by. Furthermore, our prediction intervals react on structural changes, e.g., crises, more quickly compared to an increasing insample period. As upper limits of the factor dimension, factor lags and return lags we choose $\bar{K} = 22, \bar{p} = 4$ and $\bar{q} = 5$, respectively. For the termination criteria, we have: $\xi = 10^{-2}, \eta = 10^{-2}, \eta_{\boldsymbol{F}} = 10^{-6}, \eta_{\tilde{\boldsymbol{F}}} = 10^{-6}$ and $\eta_{\tilde{\boldsymbol{B}}} = 10^{-3}$. Note, the latter criteria control, when the infinite series of the covariance matrices $\Sigma_{\tilde{\boldsymbol{F}}}$ in (A.4) or $\Sigma_{\tilde{\boldsymbol{B}}}$ in (A.8) are truncated. Hence, we work with their MA(∞) versions instead of the vectorized solutions in (A.6) or (A.10). To avoid any bias caused by simulation each prediction interval relies on $C = 500$ trajectories.

Based on the above settings, Algorithm 4.2.1 provides the prediction intervals in Figure 4.1, which cover S&P500 log-returns for the subsequent week. To be precise, the light gray area reveals the 50%-prediction intervals, while the black areas specify the 90%-prediction intervals. Here, each new, slightly darker area increases the prediction level by 10%. In addition, the red line shows the afterwards realized S&P500 log-returns and so, displays, when the afterwards realized returns exceed the prediction intervals. At first glance, the prediction intervals cover the S&P500 log-returns quite well, as there is a moderate number of interval outliers. However, during the financial crisis in 2008/2009 we have a cluster of interval outliers, which calls for further analyses. Perhaps, the additional inclusion of regime-switching concepts remedies this circumstance. Now, we just describe our first findings. Later on, we discuss our results from a financial point of view, before we do the same from a statistical perspective. That is, we behave as in Section 3.2.

As supplement to Figure 4.1, Figure 4.2 breaks the means of the predicted S&P500 log-returns down into the contributions of our panel data groups. In contrast to Figure 4.1, where Factor and AR Risks widened the prediction intervals, both do not matter in Figure 4.2. This makes sense, as we average the predicted returns, whose Factor and AR Risks are assumed to have zero mean. Dark and light blue areas detect how financial data affected our return predictions. In particular, during the financial crisis in 2008/2009 as well as in the years 2010-2012, when the US Federal Reserve intervened on captial markets in the form of its quantitative easing programs, financial aspects mainly drove our return predictions. Since the year 2012, the decomposition is more scattered and changes quite often. That is, macroeconomic and financial events matter. Figure 4.3 also supports the hypothesis that exogenous information increasingly affected the S&P500 log-returns in recent years. Although the factor dimension stayed within the range [15, 16]

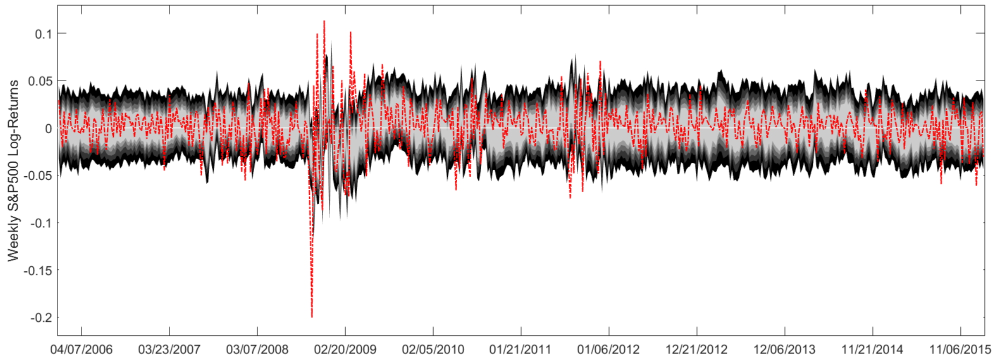

Figure 4.1: *Prediction intervals for S&P500 log-returns of the subsequent week (gray and black areas) and afterwards realized S&P500 log-returns (red line). The light gray area reveals the 50%-prediction intervals, whereas the black areas define the 90%-prediction intervals. Here, the prediction level gradually increases by 10% for each new, slightly darker area.*

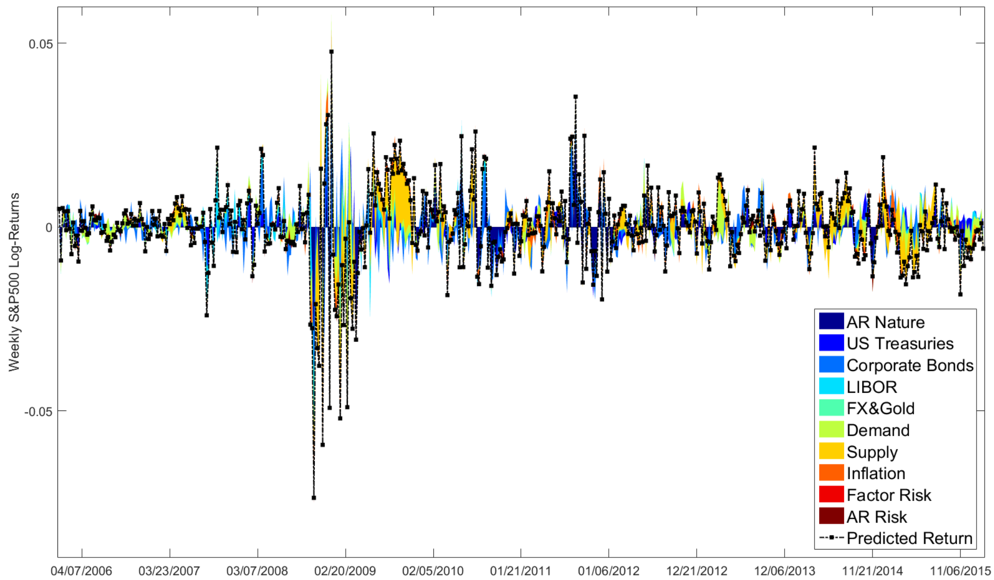

Figure 4.2: *Decomposition of the S&P500 log-returns predicted for the next week.*

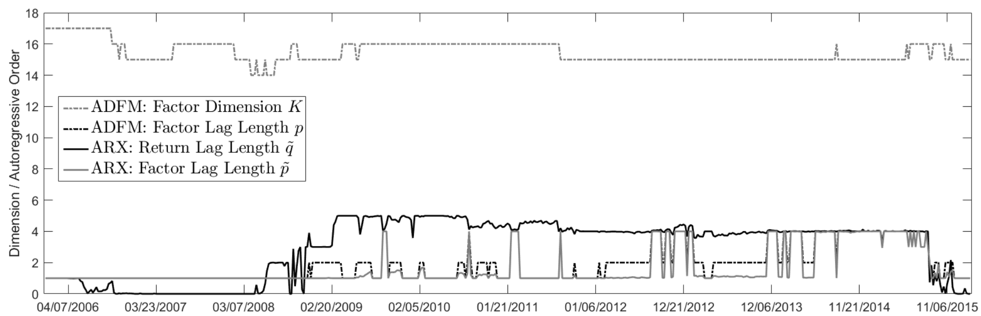

Figure 4.3: *Dimensions and lag orders of factors or returns of the predicted S&P500 log-returns.*

and we had for the autoregressive return order $\tilde{q} = 4$, from mid-2013 until mid-2015 the factor lags p and $\tilde{p}$ increased. Hence, our estimation method indicated a more complex ADFM and ARX modeling.

After these first impressions, we now turn our attention to some financial characteristics of the presented approach. So far, the prediction intervals in Figure 4.1 seem to properly map the behavior of the S&P500 log-returns. Therefore, we investigate whether the trading strategies in (4.36) and (4.37) benefit from this. Here, we abbreviate the trading strategy from (4.36) based on the 50%-prediction intervals by Prediction Level (PL) 50, while PL 60 is its analog using the 60%-prediction intervals, etc. For simplicity, our cash account does not offer any interest rate, i.e., $\tilde{r}_t \equiv 0$ for all times $t \geq 0$ and transaction costs are neglected. In total, Figure 4.4 illustrates how an initial investment of 100 USD on December 30, 2005 in the trading strategies PL 50, PL 60, PL 70, PL 80 and PL 90 with weekly reblancing would have evolved. Hence, it shows a classical backtest. Thereby, one of our intentions is to disclose the impact of the chosen prediction interval on the performance of the trading strategy in (4.36).

Next, we investigate how Leverage & Short Sales (L&S) change the risk-return profile of the basic trading strategy. That is, how deviates the risk-return profile of the trading strategy in (4.37) from the one in (4.36) and to what extend depends it on the parameters α^A, π^U and π^L. In Figure 4.4, L&S 2/1/0 stands for the trading strategy in (4.37) with weekly rebalancing based on PL 50 with parameters $\alpha^A = 2, \pi^U = 1$ and $\pi^L = 0$. The trading strategy L&S 2/1/-1 is also based on PL 50, but has the parameters $\alpha^A = 2, \pi^U = 1$ and $\pi^L = -1$. Again, we have for the cash account: $\tilde{r}_t \equiv 0$ and there are zero transaction costs.

In Figure 4.4, the strategy S&P500 reveals how a pure investment in the S&P500 would have performed. Moreover, Figure 4.4 illustrates the price evolution of two Buy&Hold (B&H) and two Constant Proportion Portfolio Insurance (CPPI)[1] strategies with weekly rebalancing. Hence, the Buy&Hold strategies may be interpreted as Constant Mix strategies. Here, the notation B&H 50 denotes a Buy&Hold strategy, whose rebalanced S&P500 exposure is the averaged S&P500 exposure of PL 50, which is 51%. Similarly, B&H 90 invests the averaged S&P500 exposure of PL 90, i.e., 50%, in the S&P500. In Figure 4.4, CPPI 2/60 stands for a CPPI strategy with multiplier 2 and floor 60%.

In addition to Figure 4.4, Table 4.14 lists the values of some common performance and risk measures for all trading strategies. Using both we conclude: First, the higher the prediction level, the lower the Log-Return (Total, %) of its PL strategy. E.g., compare PL 50 and PL 90. By definition, a high prediction level widens the intervals such that shifts in their location have less impact on the stock exposure π_t in (4.36). As shown in Figure 4.5, all PL strategies are centered around a level of 50%, but PL 50 adjusts its stock exposure more often and to a bigger extent than PL 90. Second, all PL strategies have periods of time with a lasting stock exposure $\geq 50\%$ or $\leq 50\%$. Over our out-of-sample period, PL 50 invested on average 51% of its wealth in the S&P 500, but it outperformed B&H 50, which is weekly rebalanced to 51%, by far. Hence, changing our asset allocation by π_t in (4.36) really paid off.

Except for the L&S strategies, PL 50 has the highest Log-Return (Total, %) and therefore, appears very attractive. However, the upside usually comes with a price. This is why we next focus on the volatilities of our trading strategies. In this regard, CPPI 2/80 offers with 0.93% the lowest weekly standard deviation.

[1] The floor of a CPPI strategy denotes the minimum repayment at maturity. For any point in time before maturity, the cushion represents the difference between the current portfolio value and the discounted floor. Here, discounting does not matter, since $\tilde{r}_t \equiv 0 \ \forall t \geq 0$ holds. The multiplier of a CPPI strategy constitutes to what extent the positive cushion is leveraged. As long as the cuhsion is positive, the cushion times the multiplier, which is called exposure, is invested in the risky assets. Because of $\tilde{r}_t \equiv 0 \ \forall t \geq 0$, there is no penalty, if the exposure exceeds the current portfolio value. To avoid borrowing money, the portfolio value at a given rebalancing date caps the risky exposure in this section. As soon as the cushion is zero or becomes negative, the total wealth is deposited on the bank account with $\tilde{r}_t \equiv 0$ for the remaining time to maturity. Further information about CPPI strategies is stated in, e.g., Black and Perold (1992).

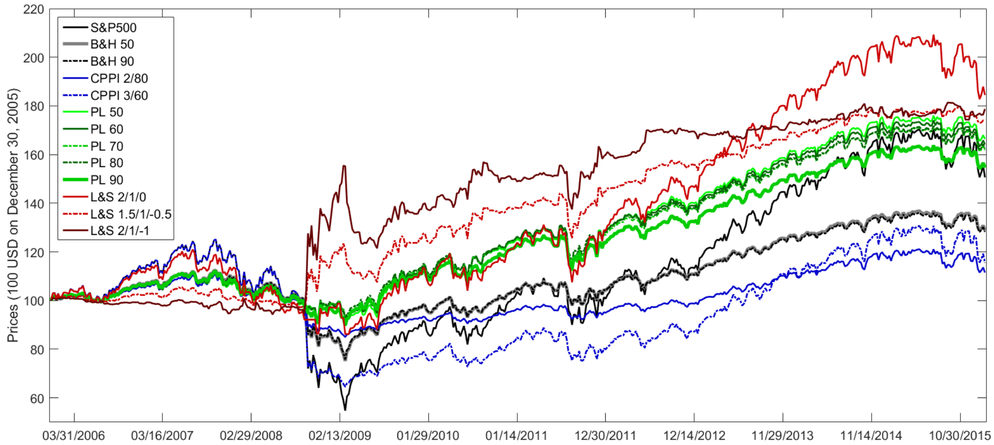

Figure 4.4: *Evolution of an initial investment of 100 USD in diverse single-market strategies (S&P500, B&H, CPPI, PL and L&S) over the out-of-sample period from December 30, 2005 until February 5, 2016. All strategies are weekly rebalanced and have zero transaction costs.*

With its allocation in Figure 4.5 in mind, this makes sense, as CPPI 2/80 was much less exposed to the S&P500 than all others. Note, Figure 4.5 also shows how CPPI 3/60 was hit by the finanical crisis in 2007/2008, when its S&P500 exposure dramatically dropped from 100% on October 3, 2008 to 21% on March 13, 2009. In case of the PL strategies, we get for the volatility an opposite picture compared to the Log-Return (Total, %). That is, the higher the prediction level, the lower the weekly standard deviation is. This sounds reasonable, as PL 90 makes smaller bets than PL 50. For the L&S strategies, Table 4.14 confirms that leveraging works as usual. Both, i.e., return and volatility, increased at the same time.

The Sharpe Ratio links the return and volatility of a trading strategy. Except for L&S 1.5/1/-0.5, the PL strategies offer the largest Sharpe Ratios. Thereby, PL 80 has a weekly Sharpe Ratio of 7.39%, which is biggest. As supplement, Table 4.15 reveals that the Sharpe Ratios of PL 80 and PL 90 are significantly different to those of S&P500, CPPI 2/80 and CPPI 3/60. Unfortunately, the differences within or between the PL and L&S strategies are not significant.

The Omega Measure compares the upside and downside of a strategy. Based on Table 4.14, L&S 1.5/1/-0.5 and L&S 2/1/-1 have the largest Omega Measures given by 134.92% and 132.39%. The Omega Measures of the PL strategies lie in the range of [121.34%, 124.94%], which are larger than those of the benchmark strategies in the range of [103.86%, 111.16%]. As shown in Table 4.16, the differences between all Omega Measures are not significant.

Similar to the volatility, CPPI 2/80 has the smallest 95% Value at Risk and 95% Conditional Value at Risk. This comes again from its reduced S&P500 exposure in Figure 4.5. The PL strategies have more or less the same weekly 95% VaR, since all lie in the range $[-1.99\%, -1.90\%]$. However, their 95% CVaR ranges from -3.19% to -2.78% and so, reflects that PL 50 makes bigger bets than PL 90. In case of the L&S strategies, there is no clear picture how leveraging and short selling affects the 95% VaR and CVaR.

Finally, we consider the Maximum Drawdown based on the complete out-of-sample period from December 30, 2005 until February 5, 2016. Note, Figures 4.4-4.5 and Table 4.14 confirm that CPPI 3/60 behaved like the S&P500, until it was knocked out by the financial crises in 2007/2008. This is why its Maximum Drawdown of -48.43% is close to the -56.24% of the S&P500. By contrast, the Maximum Drawdowns of

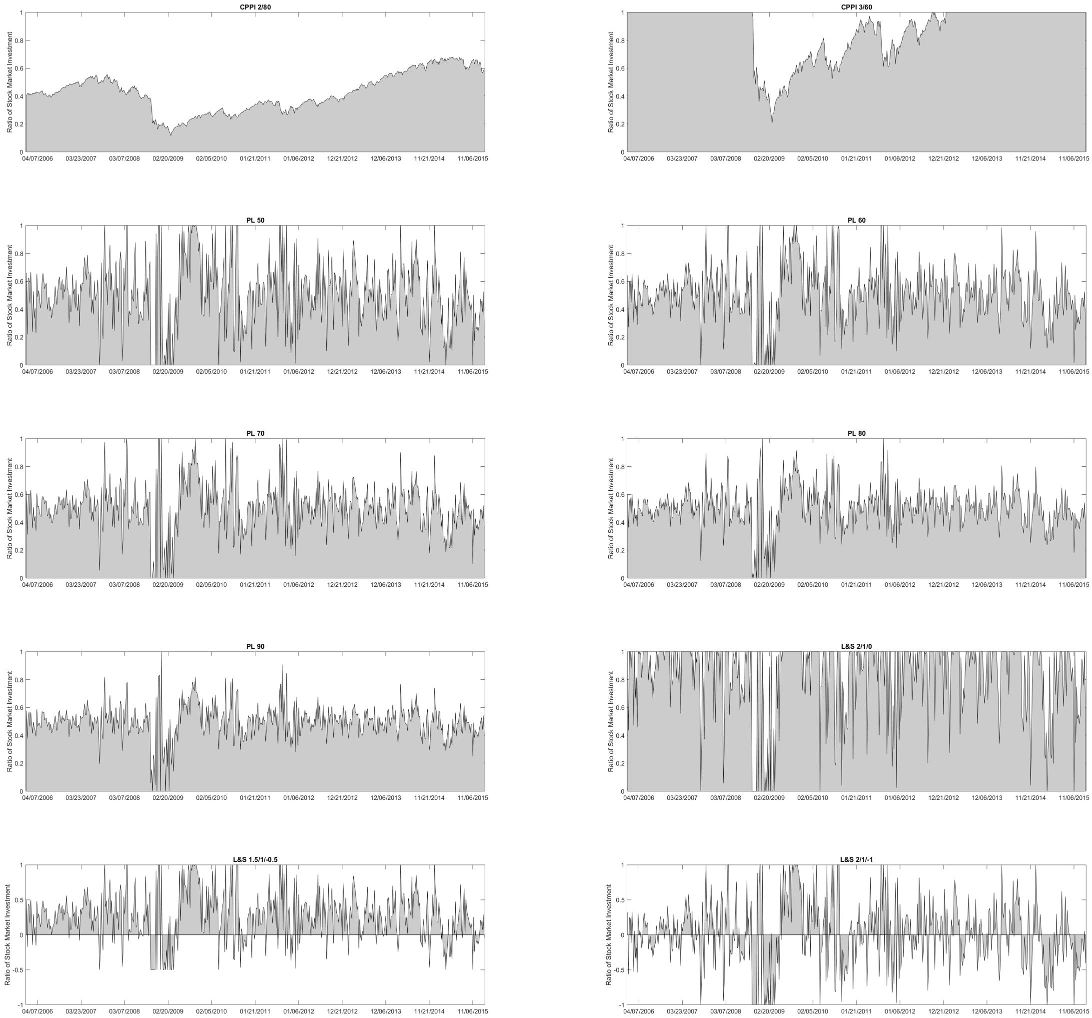

Figure 4.5: *Percentage of total wealth invested in the S&P500 for diverse single-market trading strategies (CPPI, PL and L&S) over the out-of-sample period from December 30, 2005 until February 5, 2016.*

the PL strategies lie in the range of [-19.91%, -17.37%], which is less than half. They are even smaller than the Maximum Drawdown of CPPI 2/80, which is -23.18%. For the L&S strategies, we have on the one hand that short sales admit us to gain from a drop on the stock market. On the other hand, leveraging boosts profits and losses. In total, this yields a scattered picture for their Maximum Drawdowns.

With the financial figures in mind, we recommend PL 50 for several reasons: First, it provides a decent return, which is steadily gained over the total period. Second, it has an acceptable volatiliy and a moderate downside. For reasons of completeness, Figure 4.6 illustrates the normalized histograms of the log-returns for all trading strategies. Note, all PL strategies, L&S 1.5/1/-0.5 and L&S 2/1/-1 are positively skewed, which indicates a capped downside.

Table 4.14: *Comparison of trading strategies for the out-of-sample period from December 30, 2005 until February 5, 2016*

	S&P500	B&H		CPPI		Prediction Level					Leverage & Short Sales		
		50	90	2/80	3/60	50	60	70	80	90	2/1/0	1.5/1/-0.5	2/1/-1
Log-Return (Total, %)	40.95	25.30	24.84	10.87	14.45	**<u>50.93</u>**	49.55	48.12	48.47	43.57	**61.23**	55.59	58.08
Log-Return (Wkly., %)[a]	0.08	0.05	0.05	0.02	0.03	**<u>0.10</u>**	0.09	0.09	0.09	0.08	**0.12**	0.11	0.11
Std. Dev. (Wkly., %)[b]	2.59	1.31	1.29	**0.93**	2.04	1.37	1.32	1.28	1.25	1.22	1.97	1.32	1.76
Sharpe Ratio (Wkly., %)[c]	3.00	3.65	3.67	2.22	1.34	7.05	7.12	7.11	**<u>7.39</u>**	6.78	5.89	**7.97**	6.26
Omega Measure (Wkly., %)[d]	109.08	111.12	111.16	106.24	103.86	**<u>124.94</u>**	124.37	123.70	124.03	121.34	118.51	**134.92**	132.39
95% VaR (Wkly., %)[e]	-4.39	-2.22	-2.17	**-1.53**	-3.37	-1.99	-1.92	-1.90	-1.95	-1.92	-3.41	-1.59	-1.67
95% CVaR (Wkly., %)[f]	-6.50	-3.25	-3.18	**-2.45**	-5.20	-3.19	-3.07	-2.97	-2.84	-2.78	-4.87	-2.85	-3.46
Max. Drawdown (in %) [g]	-56.24	-33.12	-32.50	-23.18	-48.43	-18.61	-17.63	**<u>-17.37</u>**	-17.84	-19.91	-28.71	**-12.90**	-21.87

*For each criterion, the **bold value** highlights the overall best strategy, whereas the **<u>bold and underlined value</u>** emphasizes the best strategy without Leverage & Short Sales (L&S).*

[a] The considered period of time consists of 527 weeks. Therefore, it holds: Log-Return (Wkly.) = Log-Return (Total)/527.
[b] As standard deviation the square root of the empirical variance, i.e., the squared deviation from the Log-Return (Wkly.) divided by 526, is used.
[c] The Sharpe Ratio divides the expected excess return by its standard deviation. As the cash account does not provide any yield, the return of the benchmark is zero.
[d] The Omega Measure divides the upside by the downside of the expected excess returns, i.e., it is the ratio of the averaged positive and negative parts of Log-Return (Wkly.).
[e] For a fixed time horizon and confidence level α, the Value at Risk reflects the maximal Log-Return (Wkly.) that is not exceeded with probability $1-\alpha$. Mathematically, this means: 95% VaR (Wkly.) $= \sup\{r \mid \mathbb{P}(\text{Log-Return (Wkly.)} < r) \leq 0.05\}$.
[f] For a fixed time horizon and confidence level α, the Conditional Value at Risk or Expected Shortfall is the expected return given that Log-Return (Wkly.) is below the α VaR, i.e.: 95% CVaR (Wkly.) $= \mathbb{E}[\text{Log-Return (Wkly.)} \mid \text{Log-Return (Wkly.)} < \text{95\% VaR (Wkly.)}]$.
[g] The Maximum Drawdown reveals the lowest discrete return, i.e., the highest loss, which could have been gained during the complete out-of-sample period.

Table 4.15: *Test statistic of Jobson and Korkie (1981) for Sharpe Ratios*

	S&P500	B&H		CPPI		Prediction Level					Leverage & Short Sales		
		50	90	2/80	3/60	50	60	70	80	90	2/1/0	1.5/1/-0.5	2/1/-1
S&P500	x	4.6611	4.6634	0.4329	1.1120	1.3056	1.3796	1.4393	1.6756	1.7459	1.0333	0.8908	0.4720
B&H 50	x	x	4.7699	0.7931	1.4806	1.1193	1.1882	1.2419	1.4708	1.4989	0.8171	0.7854	0.3820
B&H 90	x	x	x	0.8012	1.4885	1.1149	1.1837	1.2372	1.4659	1.4930	0.8120	0.7829	0.3799
CPPI 2/80	x	x	x	x	0.8032	1.5177	1.6087	1.6703	1.8772	1.8728	1.3677	1.0658	0.6031
CPPI 3/60	x	x	x	x	x	1.6094	1.6792	1.7237	1.8974	1.9583	1.4495	1.1254	0.6900
PL 50	x	x	x	x	x	x	0.2315	0.1025	0.3835	0.2205	1.0267	0.3433	0.1782
PL 60	x	x	x	x	x	x	x	0.0441	0.4052	0.3256	1.2036	0.2979	0.1878
PL 70	x	x	x	x	x	x	x	x	0.6822	0.3846	1.2559	0.2830	0.1784
PL 80	x	x	x	x	x	x	x	x	x	1.1568	1.5309	0.1754	0.2242
PL 90	x	x	x	x	x	x	x	x	x	x	0.7785	0.3188	0.0976
L&S 2/1/0	x	x	x	x	x	x	x	x	x	x	x	0.6283	0.0751
L&S 1.5/1/-0.5	x	x	x	x	x	x	x	x	x	x	x	x	0.8438
L&S 2/1/-1	x	x	x	x	x	x	x	x	x	x	x	x	x

Values marked in dark gray are significant for level 5% (test statistic: 1.96), while light gray ones are significant for level 10% (test statistic: 1.64).

Table 4.16: *Test statistic of Schmid and Schmidt (2008) for Omega Measures*

	S&P500	B&H		CPPI		Prediction Level					Leverage&Short Sales		
		50	90	2/80	3/60	50	60	70	80	90	2/1/0	1.5/1/-0.5	2/1/-1
S&P500	x	0.1363	0.1394	0.2034	0.3813	0.8530	0.8467	0.8279	0.8667	0.7430	0.5644	1.2257	0.9483
B&H 50	x	x	0.0031	0.3441	0.5212	0.7367	0.7269	0.7055	0.7406	0.6120	0.4377	1.1177	0.8573
B&H 90	x	x	x	0.3473	0.5243	0.7340	0.7241	0.7027	0.7377	0.6090	0.4348	1.1153	0.8553
CPPI 2/80	x	x	x	x	0.1861	1.0860	1.0894	1.0778	1.1301	1.0049	0.8046	1.4157	1.0901
CPPI 3/60	x	x	x	x	x	1.2026	1.2089	1.2005	1.2583	1.1531	0.9346	1.5804	1.2432
PL 50	x	x	x	x	x	x	0.0283	0.0624	0.0465	0.1860	0.3379	0.4113	0.2631
PL 60	x	x	x	x	x	x	x	0.0346	0.0179	0.1613	0.3178	0.4417	0.2859
PL 70	x	x	x	x	x	x	x	x	0.0175	0.1286	0.2881	0.4743	0.3116
PL 80	x	x	x	x	x	x	x	x	x	0.1488	0.3114	0.4633	0.3012
PL 90	x	x	x	x	x	x	x	x	x	x	0.1640	0.5930	0.4087
L&S 2/1/0	x	x	x	x	x	x	x	x	x	x	x	0.7180	0.5101
L&S 1.5/1/-0.5	x	x	x	x	x	x	x	x	x	x	x	x	0.0779
L&S 2/1/-1	x	x	x	x	x	x	x	x	x	x	x	x	x

Values marked in dark gray are significant for level 5% (test statistic: 1.96), while light gray ones are significant for level 10% (test statistic: 1.64).

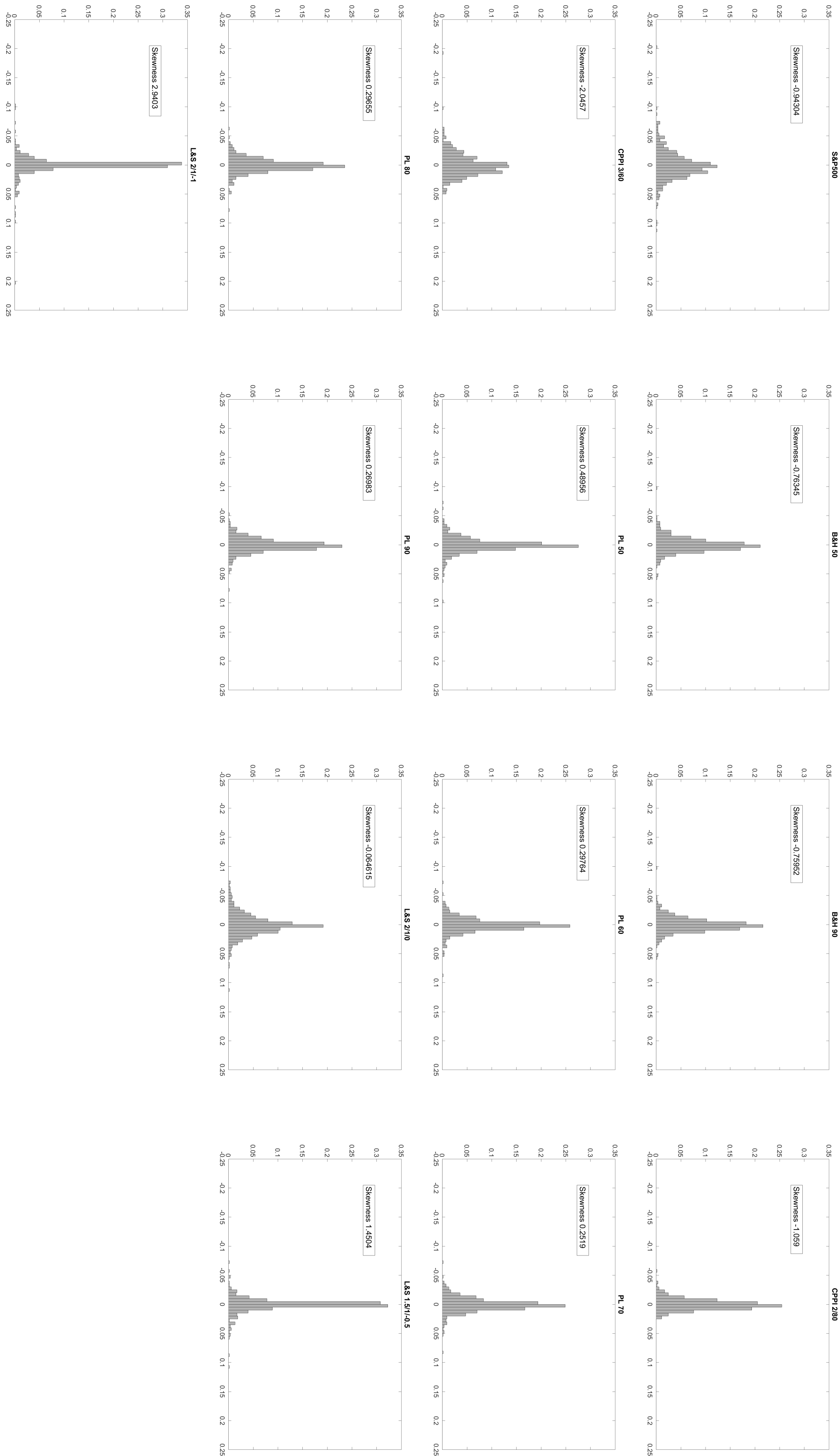

Figure 4.6: *Normalized histograms of weekly log-returns of single-market trading strategies (S&P500, B&H, CPPI, PL and L&S) over the out-of-sample period from December 30, 2005 until February 5, 2016.*

Before we discuss some statistical properties of the predicted log-returns, we repeat the previous analysis for the case of complete panel data. Instead of all 33 time series in Appendix B.2, we restrict ourselves to the groups US Treasuries, Corporate Bonds, LIBOR and FX&Gold. So, we have 22 time series without any missing observations. Our idea behind this ansatz is to check whether the inclusion of mixed-frequency information pays off. To this, we keep our rolling window of 364 weeks and gradually shift it over time, until we reach the end of the overall sample period from January 15, 1999 until February 5, 2016. Because of the total 22 time series, we replace the upper limit of the factor dimension by $\bar{K} = 21$. All other settings remain untouched. Then, Algorithm 4.2.1 generates the prediction intervals in Figure 4.7. At this stage, there are no obvious differences between the prediction intervals in Figures 4.1 and 4.7.

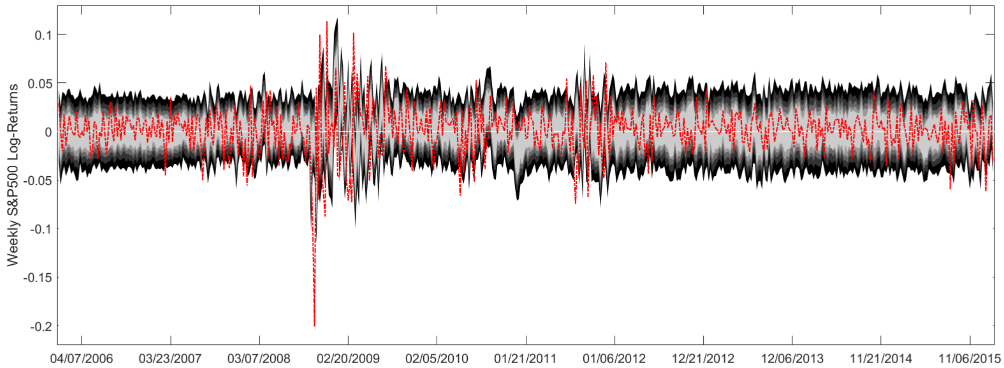

Figure 4.7: *Prediction intervals for S&P500 log-returns of next week (gray and black areas) and afterwards realized S&P500 log-returns (red line) based on complete panel data. The light gray area reveals the 50%-prediction intervals, whereas the black areas define the 90%-prediction intervals. Here, the prediction level gradually increases by 10% for each new, slightly darker area.*

In a next step, we break the means of the predicted log-returns in Figure 4.7 down into the contributions of the respective groups as shown in Figure 4.8. Here, we have a different pattern than in Figure 4.2. For instance, Figure 4.2 identified Supply as main driver at the turn of the year 2009/2010, whereas Figure 4.8 suggests a scattered pattern of US Treasuries and Corporate Bonds. However, in the years 2010-2012 US Treasuries gained in importance in Figure 4.8, which also indicates the interventions of the US Federal Reserve through its quantitative easing programs.

For the estimated dimensions and autoregressive orders of factors and returns, respectively, Figures 4.3 and 4.9 also draw different pictures. Of course, the estimated factor dimension in Figure 4.9 is smaller due to less panel data. However, in Figure 4.9 the factor dimension increases at the end, whereas it keeps its level in Figure 4.3. In Figure 4.9, the autoregressive order p of the ADFM mainly stayed between one and two up to 2013, while it holds $p = 4$ for most of the time in the years 2009-2015 in Figure 4.9. In Figure 4.9, the return lag order $\tilde{q}$ is much smaller than in Figure 4.9. That is, in the absence of macroeconomic data much of the autoregressive return behavior is mapped by the exogenous factors. Similarly, there are differences between the factor lag orders in Figures 4.3 and 4.9.

As before, we analyze the impact of the prediction intervals on the trading strategies in (4.36) and (4.37). For this purpose, Figure 4.10 displays the PL and L&S strategies of Figure 4.4 based on panel data with 33 variables once again and their analogs arising from the 22 complete time series. Note, the expression PL 50 (no) in Figure 4.10 is an abbreviation for PL 50 using panel data having no gaps. The same holds for L&S 2/1/0, etc. Besides the prices in Figure 4.10, Figure 4.11 illustrates the S&P500 exposures of the

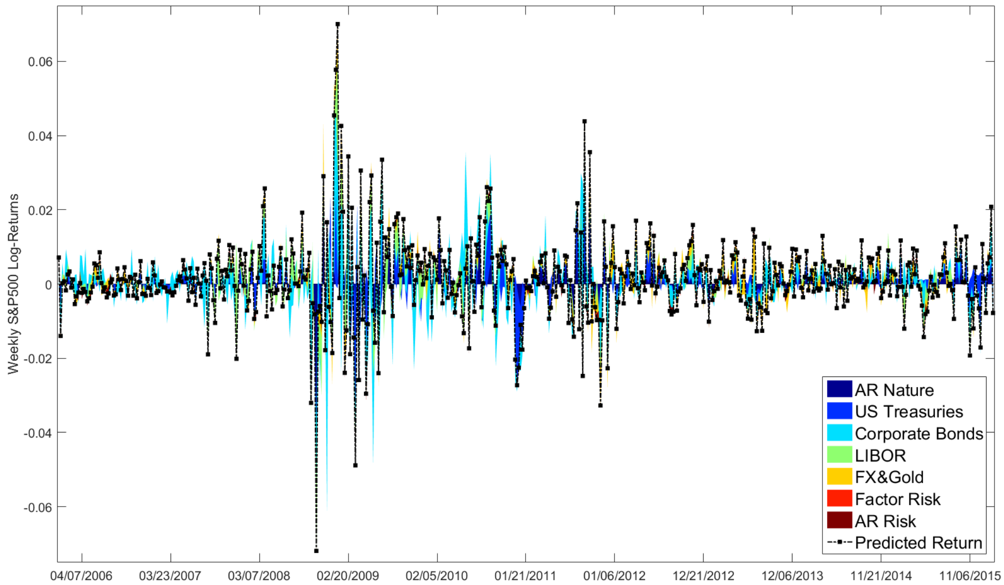

Figure 4.8: *Decomposition of S&P500 log-returns predicted for next week, when the panel data is restricted to complete time series.*

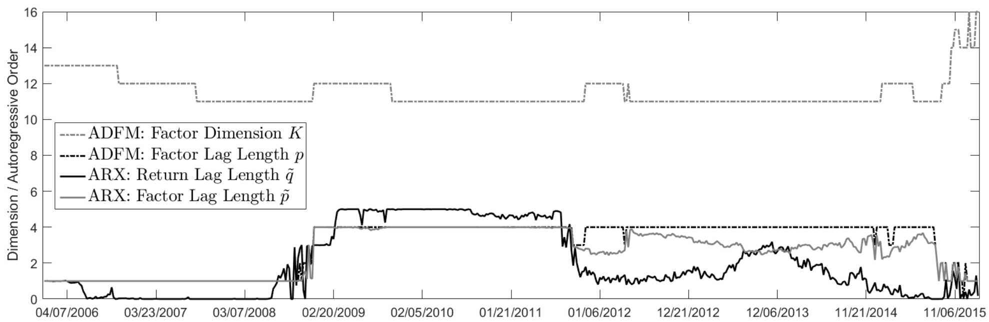

Figure 4.9: *Dimensions and lag orders of factors and returns, respectively, of predicted S&P500 log-returns, when the panel data is restricted to complete time series.*

new PL and L&S strategies and Table 4.17 lists their performance and risk measures. Then, we conclude: First, PL 50 (no) has a total log-return of 30.22%, which exceeds all other PL (no) strategies, but is much less than the 50.93% of PL 50. Similarly, the L&S (no) strategies have a much lower log-return than their L&S counterparts. Second, PL 50 (no) changes its S&P500 exposure more often and to a larger extent than PL 90 (no), which is in line with the PL strategies. Third, the standard deviations of the PL (no) strategies exceed their PL analogs such that their Sharpe Ratios are about half of the PL Sharpe Ratios. As shown in Table 4.18, the Sharpe Ratios of PL and PL (no) strategies are significantly different.

Fourth, PL (no) strategies are dominated by their PL versions regarding the Omega Measure, even though such differences are not significant as shown in Table 4.19. Fifth, the 95% VaR and CVaR of the PL (no) strategies are slightly worse than of the PL alternatives, but their Maximum Drawdowns almost doubled in the absence of macroeconomic signals. As stated in Figure 4.12, the returns of all PL (no) strategies,

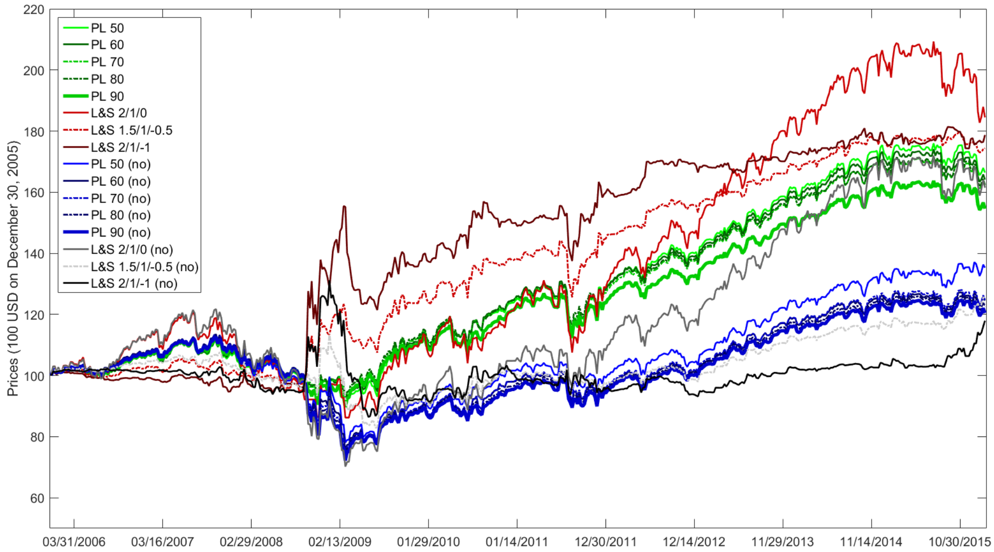

Figure 4.10: *Evolution of an initial investment of 100 USD in PL and L&S strategies based on complete and incomplete panel data over the out-of-sample period from December 30, 2005 until February 5, 2016. All strategies are weekly rebalanced and have zero transaction costs.*

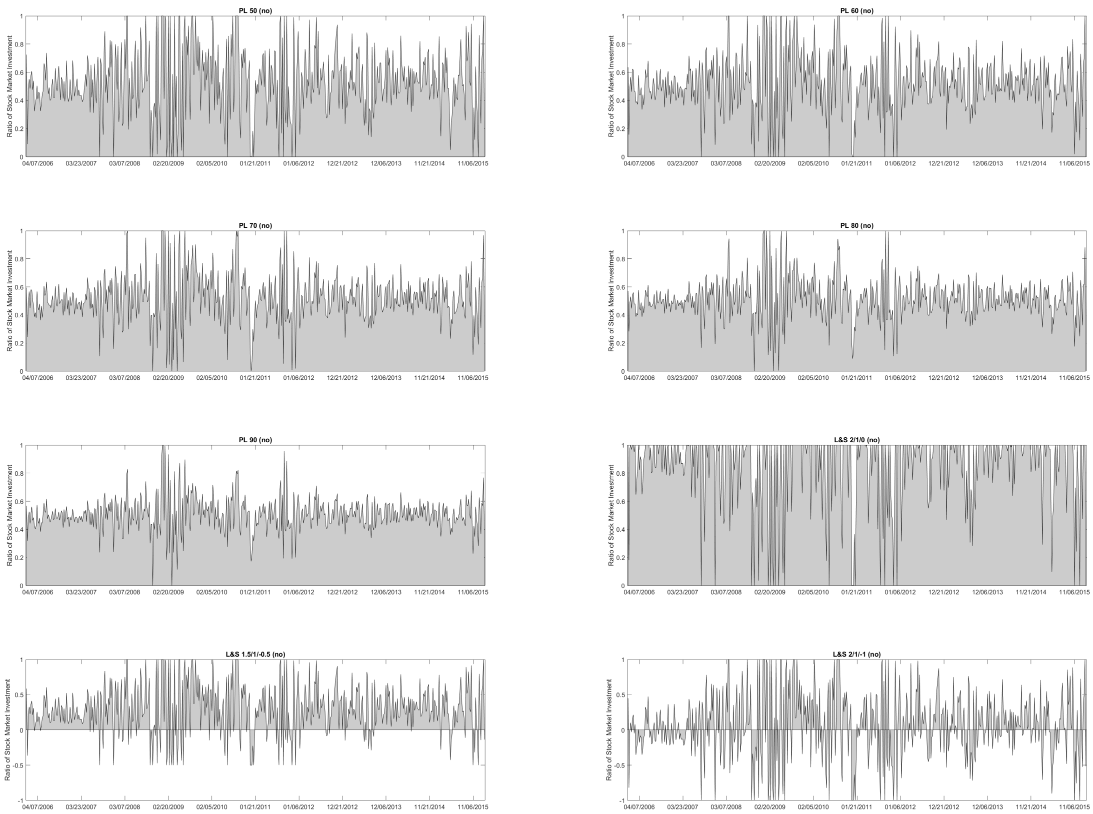

Figure 4.11: *Ratio of total wealth invested in the S&P500 for PL and L&S strategies based on complete panel data over the out-of-sample period from December 30, 2005 until February 5, 2016.*

Table 4.17: *Comparison of trading strategies for the out-of-sample period from December 30, 2005 until February 5, 2016*

	Prediction Level (no)					Leverage & Short Sales (no)		
	50	60	70	80	90	2/1/0	1.5/1/-0.5	2/1/-1
Log-Return (Total, %)	<u>**30.22**</u>	21.92	22.12	20.35	18.62	**47.92**	23.11	16.50
Log-Return (Wkly., %)[a]	<u>**0.06**</u>	0.04	0.04	0.04	0.04	**0.09**	0.04	0.0300
Std. Dev. (Wkly., %)[b]	1.53	1.49	1.44	1.41	<u>**1.38**</u>	2.10	**1.35**	1.57
Sharpe Ratio (Wkly., %)[c]	<u>**3.75**</u>	2.80	2.91	2.74	2.56	**4.33**	3.24	2.00
Omega Measure (Wkly., %)[d]	<u>**112.79**</u>	109.18	109.34	108.66	107.97	**113.33**	112.67	108.28
95% VaR (Wkly., %)[e]	-2.18	-2.06	<u>**-1.99**</u>	-2.03	-2.15	-3.21	**-1.73**	-2.05
95% CVaR (Wkly., %)[f]	-3.88	-3.85	-3.70	-3.56	**-3.47**	-5.17	-3.52	-3.97
Max. Drawdown (in %) [g]	<u>**-30.75**</u>	-33.67	-34.56	-35.13	-35.82	-42.34	**-27.24**	-34.12

*For each criterion, the **bold value** highlights the overall best strategy, whereas the <u>**bold and underlined value**</u> emphasizes the best strategy without Leverage & Short Sales (L&S).*

[a] The considered period of time consists of 527 weeks. Therefore, it holds: Log-Return (Wkly.) = Log-Return (Total)/527.
[b] As standard deviation the square root of the empirical variance, i.e., the squared deviation from the Log-Return (Wkly.) divided by 526, is used.
[c] The Sharpe Ratio divides the expected excess return by its standard deviation. As the cash account does not provide any yield, the return of the benchmark is zero.
[d] The Omega Measure divides the upside by the downside of the expected excess returns, i.e., it is the ratio of the averaged positive and negative parts of Log-Return (Wkly.).
[e] For a fixed time horizon and confidence level α, the Value at Risk reflects the maximal Log-Return (Wkly.) that is not exceeded with probability $1-\alpha$. Mathematically, this means: 95% VaR (Wkly.) $= \sup\{r|\mathbb{P}(\text{Log-Return (Wkly.)} < r) \leq 0.05\}$.
[f] For a fixed time horizon and confidence level α, the Conditional Value at Risk or Expected Shortfall is the expected return given that Log-Return (Wkly.) is below the α VaR, i.e.: 95% CVaR (Wkly.) $= \mathbb{E}[\text{Log-Return (Wkly.)}|\text{Log-Return (Wkly.)} < 95\% \text{ VaR (Wkly.)}]$.
[g] The Maximum Drawdown reveals the lowest discrete return, i.e, the highest loss, which could have been gained during the complete out-of-sample period.

Table 4.18: *Test statistic of Jobson and Korkie (1981) for Sharpe Ratios*

	Prediction Level					Leverage & Short Sales			Prediction Level (no)					Leverage & Short Sales (no)		
	50	60	70	80	90	2/1/0	1.5/1/-0.5	2/1/-1	50	60	70	80	90	2/1/0	1.5/1/-0.5	2/1/-1
PL 50	x	0.2315	0.1025	0.3835	0.2205	1.0267	0.3433	0.1782	1.5261	1.8144	1.7131	1.7162	1.7233	1.1995	1.2113	1.1017
PL 60	x	x	0.0441	0.4052	0.3256	1.2036	0.2979	0.1878	1.6177	1.9247	1.8228	1.8264	1.8322	1.2891	1.2235	1.1001
PL 70	x	x	x	0.6822	0.3846	1.2559	0.2830	0.1784	1.6705	2.0073	1.9112	1.9209	1.9286	1.3535	1.2035	1.0779
PL 80	x	x	x	x	1.1568	1.5309	0.1754	0.2242	1.9030	2.2925	2.2170	2.2419	2.2509	1.6464	1.2526	1.1023
PL 90	x	x	x	x	x	0.7785	0.3188	0.0976	1.6910	2.2467	2.2325	2.3126	2.3633	1.5731	1.0252	0.9401
L&S 2/1/0	x	x	x	x	x	x	0.6283	0.0751	1.0206	1.3953	1.3294	1.3703	1.4097	0.7876	0.7989	0.8041
L&S 1.5/1/-0.5	x	x	x	x	x	x	x	0.8438	1.0064	1.1360	1.0758	1.0748	1.0796	0.7995	1.3368	1.6229
L&S 2/1/-1	x	x	x	x	x	x	x	x	0.4439	0.5771	0.5461	0.5590	0.5753	0.3216	0.6635	1.1484
PL (no) 50	x	x	x	x	x	x	x	x	x	1.5643	0.9987	0.9268	0.9027	0.4772	0.2334	0.4231
PL (no) 60	x	x	x	x	x	x	x	x	x	x	0.3328	0.0933	0.2763	1.3698	0.1665	0.1762
PL (no) 70	x	x	x	x	x	x	x	x	x	x	x	0.5149	0.5916	1.3364	0.1152	0.1903
PL (no) 80	x	x	x	x	x	x	x	x	x	x	x	x	0.6066	1.4478	0.1578	0.1495
PL (no) 90	x	x	x	x	x	x	x	x	x	x	x	x	x	1.5453	0.2008	0.1087
L&S (no) 2/1/0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	0.3485	0.4753
L&S (no) 1.5/1/-0.5	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	0.5895
L&S (no) 2/1/-1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Values marked in dark gray are significant for level 5% (test statistic: 1.96), while light gray ones are significant for level 10% (test statistic: 1.64).

Table 4.19: *Test statistic of Schmid and Schmidt (2008) for Omega Measures*

	Prediction Level					Leverage & Short Sales			Prediction Level (no)					Leverage & Short Sales (no)		
	50	60	70	80	90	2/1/0	1.5/1/-0.5	2/1/-1	50	60	70	80	90	2/1/0	1.5/1/-0.5	2/1/-1
PL 50	x	0.0283	0.0624	0.0465	0.1860	0.3379	0.4113	0.2631	0.5879	0.7799	0.7797	0.8205	0.8581	0.5945	0.5480	0.7257
PL 60	x	x	0.0346	0.0179	0.1613	0.3178	0.4417	0.2859	0.5767	0.7739	0.7734	0.8152	0.8538	0.5829	0.5347	0.7148
PL 70	x	x	x	0.0175	0.1286	0.2881	0.4743	0.3116	0.5550	0.7557	0.7553	0.7978	0.8375	0.5611	0.5112	0.6935
PL 80	x	x	x	x	0.1488	0.3114	0.4633	0.3012	0.5817	0.7874	0.7880	0.8316	0.8728	0.5909	0.5322	0.7169
PL 90	x	x	x	x	x	0.1640	0.5930	0.4087	0.4563	0.6689	0.6697	0.7148	0.7583	0.4604	0.4146	0.6132
L&S 2/1/0	x	x	x	x	x	x	0.7180	0.5101	0.3080	0.5163	0.5139	0.5573	0.5989	0.2987	0.2853	0.4833
L&S 1.5/1/-0.5	x	x	x	x	x	x	x	0.0779	0.8716	1.0198	1.0158	1.0465	1.0715	0.8747	0.8455	0.9595
L&S 2/1/-1	x	x	x	x	x	x	x	x	0.6704	0.7951	0.7901	0.8157	0.8369	0.6645	0.6621	0.7640
PL (no) 50	x	x	x	x	x	x	x	x	x	0.1974	0.1913	0.2307	0.2697	0.0303	0.0058	0.2135
PL (no) 60	x	x	x	x	x	x	x	x	x	x	0.0089	0.0308	0.0712	0.2452	0.1727	0.0446
PL (no) 70	x	x	x	x	x	x	x	x	x	x	x	0.0405	0.0818	0.2401	0.1666	0.0529
PL (no) 80	x	x	x	x	x	x	x	x	x	x	x	x	0.0417	0.2848	0.2020	0.0192
PL (no) 90	x	x	x	x	x	x	x	x	x	x	x	x	x	0.3302	0.2368	0.0154
L&S (no) 2/1/0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	0.0329	0.2513
L&S (no) 1.5/1/-0.5	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	0.1859
L&S (no) 2/1/-1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Values marked in dark gray are significant for level 5% (test statistic: 1.96), while light gray ones are significant for level 10% (test statistic: 1.64).

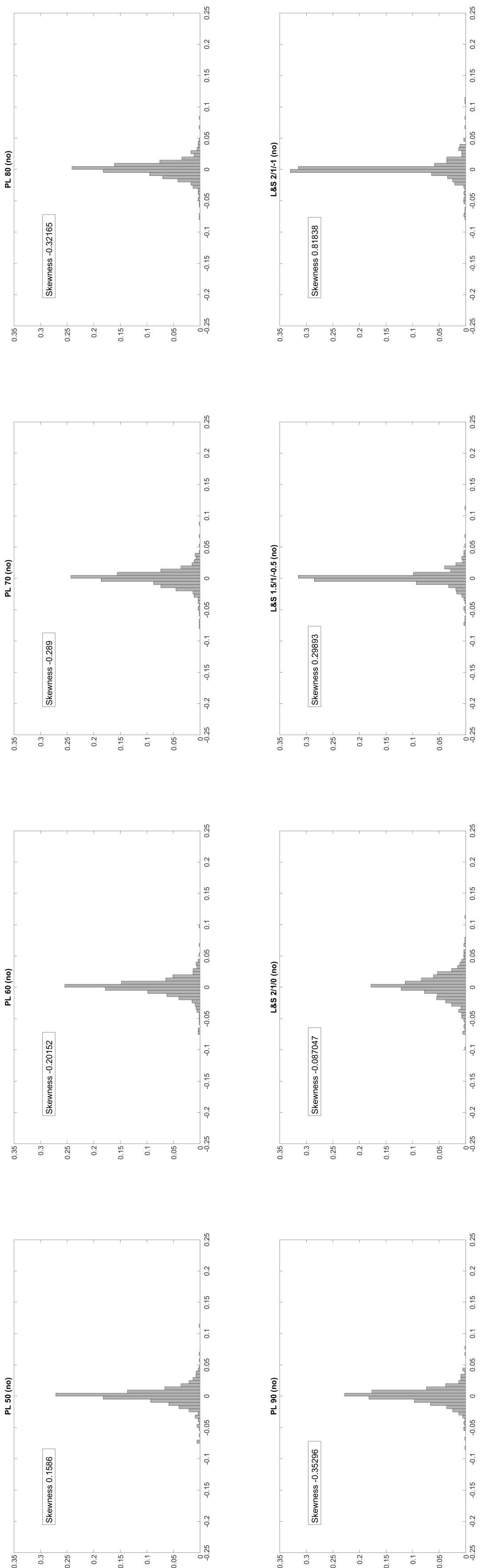

Figure 4.12: *Normalized histograms for log-returns of PL and L&S strategies, which are derived from complete panel data, over the out-of-sample period from December 30, 2005 until February 5, 2016.*

except for PL 50 (no), are negatively skewed. This indicates that large profits were removed or big losses were added. All in all, we therefore suggest the inclusion of macroeconomic variables, although this results in mixed-frequency panel data.

Next, we focus on some statistical properties. In doing so, we start with the RMSE in Definition A.3.2 for point forecasts of the S&P500 log-returns. That is, we replace the sampled factors and ARX coefficients by their estimates to predict the log-return of next week. Then, the ARX based on mixed-frequency panel data has a RMSE of 0.0272, whereas the ARX restricted to 22 variables provides a RMSE of 0.0292. For comparison reasons, the constant forecast $\hat{r}_t \equiv 0$ yields a RMSE of 0.0259, the RMSEs of ARs with orders from 1-12 lie in the range of $[0.0260, 0.0266]$ and the RMSEs of Random Walks with and without drift are 0.0380 and 0.0379, respectively. So, our model is rather mediocre in terms of RMSE. Since the RMSE controls the size, but not the direction of the deviations, Figure 4.13 illustrates the deviations $\hat{r}_t - r_t$ of our ARX based on all panel data and the AR(3), which was best regarding RMSE. As Figure 4.13 shows, the orange histogram has 4 data points with $\hat{r}_t - r_t \leq -0.10$. That is, our ARX predictions for 10/17/2008, 10/31/2008, 11/28/2008 and 03/13/2009 were too conservative, which deteriorated its RMSE. To verify this we determine the RMSE once again, but exclude these four dates. Then, the mixed-frequency ARX has a RMSE of 0.0251, which beats all other models.

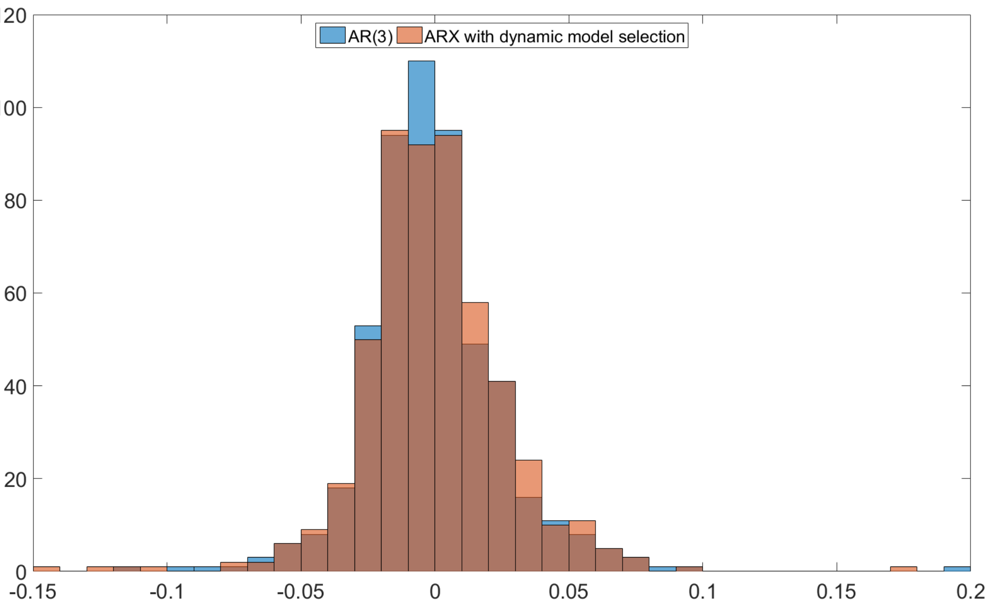

Figure 4.13: *Differences between point forecasts and realizations of weekly S&P500 log-returns. The blue histogram shows such differences, when the return predictions arise from an AR(3). The orange histogram uses forecasts of our ARX based on mixed-frequency panel data.*

Finally, we investigate the quality of our interval forecasts. For this purpose, Table 4.20 collects the Ratio of Interval Outliers from Definition A.3.3 and the Mean Interval Score from Definition A.3.4 for prediction intervals based on the incomplete panel data covering all 33 variables or the 22 complete time series. In this context, the inclusion of mixed-frequency information can provide some statistical improvements. As shown in Table 4.20, except for the 50%-prediction intervals, we have more outliers, when the ARX relies on 22 complete time series than all 33 variables. Thus, the included macroeconomic indicators make our model more cautious. Except for the 90%-prediction intervals based on complete panel data, all Ratios of

Interval Outliers are below the aimed threshold. In contrast to RIO, which counts the number of interval outliers, the Mean Interval Score takes into account by how much the prediction intervals are exceeded. In this regard, the ARX using incomplete panel data dominates the ARX restricted to the 22 time series. All in all, this confirms once again that the inclusion of macroeconomic information makes sense.

Table 4.20: *Comparison of RIO and MIS for weekly S&P500 log-returns based on the out-of-sample period from December 30, 2005 until February 5, 2016*

Measure	Panel Data	Prediction Level				
		50%	60%	70%	80%	90%
RIO	incomplete	**0.4402**	0.3397	0.2524	0.1594	**0.0930**
RIO	complete	0.4269	**0.3548**	**0.2638**	**0.1765**	0.1139
MIS	incomplete	**0.0635**	**0.0713**	**0.0816**	**0.0963**	**0.1240**
MIS	complete	0.0663	0.0749	0.0854	0.1010	0.1303

Bold figures highlight the best value of each category. That is, for the ν-prediction interval in rows 1-2 the RIO closest to $(1-\nu)$ and in rows 3-4 the lowest MIS are marked in bold.

4.6 Conclusion and Future Research

In this chapter, we considered the estimation of ADFMs with homoscedastic, cross-sectionally correlated errors using incomplete panel data. The proposed approach alternately applies two EMs until convergence is reached, that is, the relative change in the expected log-likelihood function falls below a prespecified threshold. Besides parameter estimation, our framework automatically selects the dimension of the hidden factors as well as the autoregressive orders of the factor and return dynamics. This is an important feature for empirical studies. When we empirically construct prediction intervals for returns of subsequent periods of time, randomly drawn factor samples instead of estimated factor means ensure that uncertainties in the estimated factors are taken into account. To be precise, such factor samples serve as exogenous variables in an Autoregressive Extended Model, which shall describe the dynamics of the considered returns. Similar to the factors, we draw samples from the asymptotic distribution of the estimated ARX coefficients and do not stick to their estimates. Consequently, our prediction intervals also cover uncertainties inherent in the estimated ARX parameters. Note, our OLS of the ARX parameters relies on sampled factors such that both risks are jointly treated despite their nonlinear relation. At the end, our empirical example applies the presented procedure to US data. Thereby, we determine prediction intervals for weekly log-returns of the S&P500 price index, detect the contributions of the panel data to our point forecasts and define two dynamic trading strategies based on the obtained prediction intervals.

With a view to the existing literature, this chapter makes several contributions: First, recent publications on ADFMs mainly focused on serially correlated errors, but excluded cross-sectional ones. Doz et al. (2012) showed that EDFMs may be regarded as misspecified ADFMs and allow for consistent factor estimation. Therefore, cross-sectionality of shocks is usually ignored. For instance, in case of mixed-frequency panel data the derivation of the conditional log-likelihood function in Bańbura and Modugno (2014) essentially requires a diagonal structure of the error covariance matrix, i.e., it actually relies on uncorrelated shocks. However, Doz et al. (2012) provide asymptotic results, which not necessarily remain valid for small, ragged samples as in our MC simulation study. Since we explicitly allow for cross-sectionally correlated errors, our approach outperforms the benchmark estimation of ADFMs in Bańbura and Modugno (2014).

Second, our MLE does not link an EM and the Kalman Filter or Smoother. We instead obtain the means and covariance matrices of the latent factors in closed form. Our MC simulation study shows that MLE based on closed-form factor moments dominates MLE with the Kalman Filter and Smoother. In addition, we propose the alternating usage of two EMs for incomplete data. The first EM makes sure that the relation between low-frequency observations and their artificial analogs of higher frequency holds (Stock and Watson, 1999a, 2002b). Furthermore, it reconstructs complete panel data from the latest parameter estimates and observations. The second EM performs the actual MLE based on the augmented data.

Third, we treat the stochastic factor dynamics in its general form. On the one hand, we do not convert the original VAR(p) into a VAR(1) during the estimation of ADFMs. Therefore, our approach does not need additional loadings constraints as discussed in Bork (2009). On the other hand, we also address the selection of the factor dimension and autoregressive order.

Fourth, the processing of the estimated factors is novel. Instead of point estimates, we construct empirical prediction intervals for a return time series. Besides exogenous information and autoregressive return characteristics, the prediction intervals incorporate uncertainties arising from the estimation of the factors and model parameters. The underlying time horizon is weekly, which is rather rare than common in the literature.

Fifth, we trace our point forecasts of the returns back to the original panel data and their high-frequency counterparts, respectively. This is an important feature for practitioners, as it offers them the possibility to compare our model based output with their expectations. Furthermore, if the drivers of the expected market behavior are detected at an early stage, professionals can take suitable measures to reduce the inherent risks.

Finally, we propose two dynamic trading strategies. Thereby, the first determines how much of the total wealth should be invested in the financial index depending on the width and location of the prediction intervals. In a next step, the second strategy shows how the risk-return characteristics of the first can be adapted to the needs of an investor. E.g., how an upper limit for the investment in the financial index can be implemented. The idea behind both trading strategies is to show how one can profit from the gained information about the future index behavior.

Unfortunately, our approach does not cover serially correlated errors. As the case of cross-sectionally and serially correlated shocks would exceed the scope of this work, a possible direction for future reseach is to combine both in the form of homoscedastic, serially and cross-sectionally correlated idiosyncratic errors. In a next step, an extension to heteroscedasticity or the incorporation of regime-switching concepts would be worthwhile. Finally, several ADFMs could be coupled by copulas to capture inter-market dependencies.

Chapter 5

FAVARs for Incomplete Panel Data

We extend the Factor-Augmented Vector Autoregression Model in Bernanke et al. (2005) to ragged panel data. Within the scope of a fully parametric two-step approach, the alternating use of two Expectation-Maximization Algorithms simultaneously estimates model parameters and missing observations. Furthermore, it addresses the selection of the factor dimension and lag order. In opposition to non-parametric two-step estimation methods linking PCA and linear regressions, we apply MLE. Thereby, we adapt the standard Kalman Filter and Smoother to explicitly take into account that factors are partially observed. To eliminate any identification issues of the model parameters we linearly constrain the loadings matrix or VAR(p) coefficients of the factor dynamics. Furthermore, in the scope of an intense MC simulation study, we compare our estimation method with several alternatives existing in the literature. In our empirical application, the presented framework analyzes US data for measuring the effects of the monetary policy on the real economy and financial markets. Here, the consequences for the quarterly GDP growth rates are of particular importance.

5.1 Mathematical Background

We start with the definition of FAVARs. At this stage, we show why certain assumptions with respect to the covariance matrices of the idiosyncratic shocks are less restrictive than it may appear at first glance. Then, we derive an EM for parameter estimation with complete data as in Bork (2009). Finally, we add the EM of Stock and Watson (1999a, 2002b) to the estimation procedure such that incomplete time series can be treated and estimates for missing values are obtained.

5.1.1 Rotations and Identification Restrictions

The idea behind FAVARs is to gain from the advantages of FMs and VARs. Although VARs are well-known in the literature and offer many methods for measuring the impact of certain variables on the whole system, they are still restricted to a few time series and so, can only incorporate a limited number of variables. By contrast, FA enables a significant reduction in dimension, when the panel data is described by a few latent factors. Similar to DFMs, FAVARs have a transition equation and a measurement equation. However, the factors part of FAVARs are partially observed. In this section, we stick to FAVARs with complete time series, that is, neither the panel data nor the observed factor variables have any data gaps.

Definition 5.1.1 (Factor-Augmented Vector Autoregression Model)
For any point in time t, the vectors $\boldsymbol{X}_t \in \mathbb{R}^N$ and $\boldsymbol{Y}_t \in \mathbb{R}^M$ consist of panel data and other variables. Here, $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$ are not necessarily disjoint, i.e., they can share some variables. Furthermore, all univariate times series part of the processes $\{\boldsymbol{X}_t\}$ and $\{\boldsymbol{Y}_t\}$ are supposed to be complete, of the same frequency and standardized with zero mean and standard deviation of one. If the vector $\boldsymbol{F}_t \in \mathbb{R}^K$ denotes the unobserved factors at time t, the measurement and transition equations of a FAVAR are defined as follows:

$$\boldsymbol{X}_t = \begin{bmatrix}\Lambda^f & \Lambda^y\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_t \\ \boldsymbol{Y}_t\end{bmatrix} + \boldsymbol{e}_t,\ \boldsymbol{e}_t \sim \mathcal{N}\left(\boldsymbol{0}_N, \Sigma_{\boldsymbol{e}}\right) iid, \tag{5.1}$$

$$\begin{bmatrix}\boldsymbol{F}_t \\ \boldsymbol{Y}_t\end{bmatrix} = \Phi(L)\begin{bmatrix}\boldsymbol{F}_{t-1} \\ \boldsymbol{Y}_{t-1}\end{bmatrix} + \boldsymbol{v}_t = \begin{bmatrix}\Phi^{ff}(L) & \Phi^{fy}(L) \\ \Phi^{yf}(L) & \Phi^{yy}(L)\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_{t-1} \\ \boldsymbol{Y}_{t-1}\end{bmatrix} + \boldsymbol{v}_t,\ \boldsymbol{v}_t \sim \mathcal{N}\left(\boldsymbol{0}_{K+M}, \Sigma_{\boldsymbol{v}}\right) iid, \tag{5.2}$$

with constant loadings matrices $\Lambda^f \in \mathbb{R}^{N\times K}$ and $\Lambda^y \in \mathbb{R}^{N\times M}$. The operator $\Phi(L)$ in (5.2) represents a conformable lag polynomial of order $p \geq 1$ given by $\Phi(L) = \Phi_1 + \Phi_2 L^1 + ... + \Phi_p L^{p-1}$ and constant coefficient matrices $\Phi_j \in \mathbb{R}^{(K+M)\times(K+M)}$ for $1 \leq j \leq p$. Moreover, we assume the idiosyncratic shocks $\boldsymbol{e}_t \in \mathbb{R}^N$ and $\boldsymbol{v}_t \in \mathbb{R}^{K+M}$ to be iid Gaussian and independent of each other. That is, $\boldsymbol{e}_t \perp \boldsymbol{v}_s \,\forall\, t, s$.

The above FAVAR definition complies with the FAVAR definition in Bernanke et al. (2005). They combine PCA and OLS regressions for parameter estimation. To this, they assume the shocks $\boldsymbol{v}_t$ to be zero mean with covariance matrix $\Sigma_{\boldsymbol{v}}$, but do not pinpoint a concrete distribution. For the errors $\boldsymbol{e}_t$, they assume zero mean, but admit two correlation cases: the errors $\boldsymbol{e}_t$ are either uncorrelated or weakly cross-sectionally correlated. In the sequel, we will discuss restrictions to avoid parameter ambiguity and so, identification issues. Therefore, we do not further comment the shock distributions in Definition 5.1.1 at this point in time. In particular, we do not assume the covariance matrix $\Sigma_{\boldsymbol{e}}$ as diagonal and follow the argumentation in Doz et al. (2012) to justify that weakly cross-sectionally correlated errors can be ignored as in Bańbura and Modugno (2014). Note, Bernanke et al. (2005) also presented a Bayesian estimation method, which exceeds the scope of this section.

In the sequel, we assume the VAR(p) process in (5.2) to be covariance-stationary as in Definition A.2.1. Equation (5.2) is a standard VAR(p) in the observed variables $\boldsymbol{Y}_t$, if all matrix elements of $\Phi(L)$ covering the impact of $\boldsymbol{F}_{t-1}$ on $\boldsymbol{Y}_t$ are zero (Bernanke et al., 2005). Otherwise, Bernanke et al. (2005) call (5.2) the transition equation a FAVAR. Moreover, they note: On the one hand, the FAVAR in (5.2) nests a VAR(p) supporting comparisons with general VAR(p) results and assessments of the marginal contribution of the factors $\boldsymbol{F}_t$. Second, if the true system is a FAVAR, neglecting the hidden factors $\boldsymbol{F}_t$ and sticking to the simple VAR in $\boldsymbol{Y}_t$ will cause biased estimation results such that the interpretation of Impulse Response Functions (IRFs) and Forecast Error Variance Decompositions (FEVDs) may be faulty.

For $M = 0$, i.e., in the absence of observed factor elements, FAVARs coincide with ADFMs by definition. However, for $M > 0$, Bork (2009) and Marcellino and Sivec (2016) showed that special loadings constraints and properly sorted panel data, where the observed variables $\boldsymbol{Y}_t$ are a subset of $\boldsymbol{X}_t$, result in the common state-space representation of ADFMs. Since the variables $\boldsymbol{Y}_t$ part of $\boldsymbol{X}_t$ on the left-hand side in (5.1) are identically mapped to the vector $\boldsymbol{Y}_t$ on the right-hand side in (5.1), some identification problems of the model parameters are implicitly solved. But this mapping also provides that the respective entries of the idiosyncratic shocks $\boldsymbol{e}_t$ should be zero and thus, the covariance matrix $\Sigma_{\boldsymbol{e}}$ does not have full rank. From a theoretical perspective the reduced rank of the covariance matrix $\Sigma_{\boldsymbol{e}}$ causes that its inverse matrix $\Sigma_{\boldsymbol{e}}^{-1}$ part of the log-likelihood function is not well-defined. In addition, the reduced rank of $\Sigma_{\boldsymbol{e}}$ and the fact that the variables $\boldsymbol{Y}_t$ are observed such that the covariance matrix of $\boldsymbol{Y}_t$ conditioned on the information up to time t is a zero matrix have to be addressed, when the standard KF and KS in Section 2.1.4 are applied. In such cases, the inverse matrices in the Kalman gains of Lemmata 2.1.8 and 2.1.9, perhaps, do not always exist. If this reduced rank issue is neglected on purpose, as the standard KF and KS still work, one implicitly ignores the observability of $\boldsymbol{Y}_t$ on the right-hand side in (5.2). Regardless the performed adjustments, rather specific ADFMs than real FAVARs are estimated.

Because of (5.1), the vector $\left[\boldsymbol{F}_t', \boldsymbol{Y}_t'\right]'$ drives the dynamics of $\boldsymbol{X}_t$. This explains why Bernanke et al. (2005) regard all $\boldsymbol{X}_t$ as "noisy measures of the underlying unobserved factors $\boldsymbol{F}_t$". For monetary policy analysis, the variables $\boldsymbol{Y}_t$ often cover policy instruments, e.g., the US Effective Federal Funds Rate (FEDFUNDS) or Monetary Base. By contrast, in conventional VARs, the vector $\boldsymbol{Y}_t$ contains all data and thus, lacks the ability to include additional information in the form of the factors $\boldsymbol{F}_t$. In case of FAVARs, the size of the panel data can be large such that we often receive: $K + M \ll N$. For reasons of clarity, we proceed with the assumption $N < T$ in the sequel.

Lemma 5.1.2 (Ambiguity of FAVAR Parameters)
Let $R \in \mathbb{R}^{(K+M)\times(K+M)}$ be a non-singular matrix defined as follows:

$$R = \begin{pmatrix} R_1 & R_2 \\ O_{M\times K} & I_M \end{pmatrix},$$

with $O_{M\times K} \in \mathbb{R}^{M\times K}$ as zero matrix. The matrices $R_1 \in \mathbb{R}^{K\times K}$ and $R_2 \in \mathbb{R}^{K\times M}$ are arbitrary as long as the non-singularity of the matrix R is kept. Then, the transformed vector $\left[\breve{\boldsymbol{F}}_t', \boldsymbol{Y}_t'\right]' = R\left[\boldsymbol{F}_t', \boldsymbol{Y}_t'\right]' \in \mathbb{R}^{K+M}$ can be equivalently rewritten in the form of (5.1)-(5.2)*.*

Proof:
In case of (5.1), we receive with the non-singular matrix R:

$$\boldsymbol{X}_t = \begin{bmatrix} \Lambda^f & \Lambda^y \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_t \\ \boldsymbol{Y}_t \end{bmatrix} + \boldsymbol{e}_t = \begin{bmatrix} \Lambda^f & \Lambda^y \end{bmatrix} R^{-1} R \begin{bmatrix} \boldsymbol{F}_t \\ \boldsymbol{Y}_t \end{bmatrix} + \boldsymbol{e}_t. \tag{5.3}$$

As mentioned in Bai et al. (2015), the matrix R has to map the observed vector $\boldsymbol{Y}_t$ to itself, i.e., we get:

$$\boldsymbol{Y}_t = \begin{bmatrix} R_3 & R_4 \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_t \\ \boldsymbol{Y}_t \end{bmatrix} \Leftrightarrow \begin{bmatrix} R_3 & (R_4 - I_M) \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_t \\ \boldsymbol{Y}_t \end{bmatrix} = \boldsymbol{0}_M.$$

Since this must hold for all point in times t, we obtain: $R_3 = O_{M\times K}$ and $R_4 = I_M$. For the inverse of the resulting matrix R we have:

$$R^{-1} = \begin{pmatrix} R_1^{-1} & -R_1^{-1}R_2 \\ O_{M\times K} & I_M \end{pmatrix}, \tag{5.4}$$

such that $K\,(K+M)$ degrees of freedom are still left. □

Lemma 5.1.2 confirms that the model in (5.1)-(5.2) is econometrically unidentified. For instance, $R_1 = I_K$ and $R_2 = O_{K\times M}$ or $R_1 = I_K$ and $R_2 = \mathbb{1}_K \mathbb{1}_M'$ are possible choices. Moreover, Lemma 5.1.2 emphasizes a feature of FAVARs, that is, the observability of the vector $\boldsymbol{Y}_t$ constrains the matrix R to map $\boldsymbol{Y}_t$ on itself. Next, we follow the ideas in Bai et al. (2015) once again to further simplify our FAVAR representation.

Lemma 5.1.3 (FAVAR Formulation with Partially Uncorrelated VAR Shocks)
For the non-singular matrix $R \in \mathbb{R}^{(K+M)\times(K+M)}$ in Lemma 5.1.2, the vector $\breve{\boldsymbol{v}}_t = R\boldsymbol{v}_t \in \mathbb{R}^{K+M}$ denotes

the transformed errors of the VAR(p) in (5.2) *and is iid Gaussian as follows:*

$$\breve{\boldsymbol{v}}_t \sim \mathcal{N}\left(\mathbf{0}_{K+M}, \begin{bmatrix} \Sigma_{\breve{\boldsymbol{v}}}^{ff} & \Sigma_{\breve{\boldsymbol{v}}}^{fy} \\ \Sigma_{\breve{\boldsymbol{v}}}^{yf} & \Sigma_{\breve{\boldsymbol{v}}}^{yy} \end{bmatrix}\right). \tag{5.5}$$

By defining the matrix $H \in \mathbb{R}^{(K+M)\times(K+M)}$ as

$$H = \begin{bmatrix} \left(\Sigma_{\breve{\boldsymbol{v}}|y}^{ff}\right)^{-\frac{1}{2}} & -\left(\Sigma_{\breve{\boldsymbol{v}}|y}^{ff}\right)^{-\frac{1}{2}} \Sigma_{\breve{\boldsymbol{v}}}^{fy} \left(\Sigma_{\breve{\boldsymbol{v}}}^{yy}\right)^{-1} \\ O_{M\times K} & I_M \end{bmatrix} \quad \text{with} \quad \Sigma_{\breve{\boldsymbol{v}}|y}^{ff} = \Sigma_{\breve{\boldsymbol{v}}}^{ff} - \Sigma_{\breve{\boldsymbol{v}}}^{fy} \left(\Sigma_{\breve{\boldsymbol{v}}}^{yy}\right)^{-1} \Sigma_{\breve{\boldsymbol{v}}}^{yf},$$

the FAVAR in (5.1)-(5.2) *with matrices $\Sigma_{\breve{\boldsymbol{v}}|y}^{ff} \in \mathbb{R}^{K\times K}$ and $\Sigma_{\breve{\boldsymbol{v}}}^{yy} \in \mathbb{R}^{M\times M}$ of full rank can be rewritten as:*

$$\boldsymbol{X}_t = \bar{\Lambda} \begin{bmatrix} \bar{\boldsymbol{F}}_t \\ \boldsymbol{Y}_t \end{bmatrix} + \boldsymbol{e}_t = \begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{F}}_t \\ \boldsymbol{Y}_t \end{bmatrix} + \boldsymbol{e}_t,\ \boldsymbol{e}_t \sim \mathcal{N}\left(\mathbf{0}_N, \Sigma_{\boldsymbol{e}}\right)\ \textit{iid}, \tag{5.6}$$

$$\begin{bmatrix} \bar{\boldsymbol{F}}_t \\ \boldsymbol{Y}_t \end{bmatrix} = \sum_{i=1}^{p} \bar{\Phi}_i \begin{bmatrix} \bar{\boldsymbol{F}}_{t-i} \\ \boldsymbol{Y}_{t-i} \end{bmatrix} + \bar{\boldsymbol{v}}_t,\ \bar{\boldsymbol{v}}_t \sim \mathcal{N}\left(\mathbf{0}_{K+M}, \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & \Sigma_{\bar{v}}^{yy} \end{bmatrix}\right) \textit{iid}, \tag{5.7}$$

with transformed loadings matrix $\bar{\Lambda} = [\Lambda^f\ \Lambda^y]R^{-1}H^{-1}$, joint vector $\left[\bar{\boldsymbol{F}}_t', \boldsymbol{Y}_t'\right]' = HR\left[\boldsymbol{F}_t', \boldsymbol{Y}_t'\right]'$, coefficient matrices $\bar{\Phi}_i = HR\Phi_i R^{-1}H^{-1}$ for all $1 \le i \le p$ and idiosyncratic shocks $\bar{\boldsymbol{v}}_t = HR\boldsymbol{v}_t$.

Proof:
For the invertible matrix R in Lemma 5.1.2, the transformed FAVAR is given by:

$$\boldsymbol{X}_t = \begin{bmatrix} \Lambda^f R_1^{-1} & -\Lambda^f R_1^{-1} R_2 + \Lambda^y \end{bmatrix} \begin{bmatrix} \breve{\boldsymbol{F}}_t \\ \boldsymbol{Y}_t \end{bmatrix} + \boldsymbol{e}_t,\ \boldsymbol{e}_t \sim \mathcal{N}\left(\mathbf{0}_N, \Sigma_{\boldsymbol{e}}\right), \tag{5.8}$$

$$\begin{bmatrix} \breve{\boldsymbol{F}}_t \\ \boldsymbol{Y}_t \end{bmatrix} = \sum_{i=1}^{p} \underbrace{R\Phi_i R^{-1}}_{\breve{\Phi}_i} \begin{bmatrix} \breve{\boldsymbol{F}}_{t-i} \\ \boldsymbol{Y}_{t-i} \end{bmatrix} + \underbrace{R\boldsymbol{v}_t}_{\breve{\boldsymbol{v}}_t} = \sum_{i=1}^{p} \breve{\Phi}_i \begin{bmatrix} \breve{\boldsymbol{F}}_{t-i} \\ \boldsymbol{Y}_{t-i} \end{bmatrix} + \breve{\boldsymbol{v}}_t,\ \breve{\boldsymbol{v}}_t \sim \mathcal{N}\left(\mathbf{0}_{K+M}, \underbrace{\begin{bmatrix} \Sigma_{\breve{\boldsymbol{v}}}^{ff} & \Sigma_{\breve{\boldsymbol{v}}}^{fy} \\ \Sigma_{\breve{\boldsymbol{v}}}^{yf} & \Sigma_{\breve{\boldsymbol{v}}}^{yy} \end{bmatrix}}_{\Sigma_{\breve{\boldsymbol{v}}}}\right). \tag{5.9}$$

The matrices $\Sigma_{\breve{\boldsymbol{v}}|y}^{ff}$ and $\Sigma_{\breve{\boldsymbol{v}}}^{yy}$ are supposed to have full rank such that their inverse matrices exist and thus, the matrix H is well-defined. The inverse matrix of H is given by:

$$H^{-1} = \begin{bmatrix} \left(\Sigma_{\breve{\boldsymbol{v}}|y}^{ff}\right)^{\frac{1}{2}} & \Sigma_{\breve{\boldsymbol{v}}}^{fy} \left(\Sigma_{\breve{\boldsymbol{v}}}^{yy}\right)^{-1} \\ O_{M\times K} & I_M \end{bmatrix},$$

which is well-defined, too. Note, the matrix H is non-singular and has the shape Lemma 5.1.2 calls for. Hence, it also belongs to the mentioned class of transformation matrices. In particular, we can prove that the product of the matrices H and R remains in this transformation class:

$$HR = \begin{bmatrix} \left(\Sigma_{\breve{\boldsymbol{v}}|y}^{ff}\right)^{-\frac{1}{2}} R_1 & \left(\Sigma_{\breve{\boldsymbol{v}}|y}^{ff}\right)^{-\frac{1}{2}} \left(R_2 - \Sigma_{\breve{\boldsymbol{v}}}^{fy} \left(\Sigma_{\breve{\boldsymbol{v}}}^{yy}\right)^{-1}\right) \\ O_{M\times K} & I_M \end{bmatrix}. \tag{5.10}$$

With this in mind, we insert the matrix H in (5.8)-(5.9) and obtain:

$$\boldsymbol{X}_t = \begin{bmatrix} \Lambda^f R_1^{-1} & -\Lambda^f R_1^{-1} R_2 + \Lambda^y \end{bmatrix} H^{-1} H \begin{bmatrix} \breve{\boldsymbol{F}}_t \\ \boldsymbol{Y}_t \end{bmatrix} + \boldsymbol{e}_t,\ \boldsymbol{e}_t \sim \mathcal{N}\left(\mathbf{0}_N, \Sigma_{\boldsymbol{e}}\right) \text{iid}, \tag{5.11}$$

$$H \begin{bmatrix} \breve{\boldsymbol{F}}_t \\ \boldsymbol{Y}_t \end{bmatrix} = \sum_{i=1}^{p} \underbrace{H\breve{\Phi}_i H^{-1}}_{\bar{\Phi}_i} H \begin{bmatrix} \breve{\boldsymbol{F}}_{t-i} \\ \boldsymbol{Y}_{t-i} \end{bmatrix} + \underbrace{H\breve{\boldsymbol{v}}_t}_{\bar{\boldsymbol{v}}_t},\ \bar{\boldsymbol{v}}_t \sim \mathcal{N}\left(\mathbf{0}_{K+M}, \underbrace{H\Sigma_{\breve{\boldsymbol{v}}}H'}_{\Sigma_{\bar{\boldsymbol{v}}}}\right) \text{iid}. \tag{5.12}$$

Next, the covariance matrix $\Sigma_{\bar{\boldsymbol{v}}}$ can be simplified in the following manner:

$$\begin{aligned}\mathbb{V}\mathrm{ar}_{\Theta}\left[H\breve{\boldsymbol{v}}_t\right] = H\Sigma_{\breve{\boldsymbol{v}}}H' &= H\begin{bmatrix}\Sigma_{\breve{\boldsymbol{v}}}^{ff}\left(\Sigma_{\breve{\boldsymbol{v}}|y}^{ff}\right)^{-\frac{1}{2}} - \Sigma_{\breve{\boldsymbol{v}}}^{fy}\left(\Sigma_{\breve{\boldsymbol{v}}}^{yy}\right)^{-1}\Sigma_{\breve{\boldsymbol{v}}}^{yf}\left(\Sigma_{\breve{\boldsymbol{v}}|y}^{ff}\right)^{-\frac{1}{2}} & \Sigma_{\breve{\boldsymbol{v}}}^{fy} \\ \Sigma_{\breve{\boldsymbol{v}}}^{yf}\left(\Sigma_{\breve{\boldsymbol{v}}|y}^{ff}\right)^{-\frac{1}{2}} - \Sigma_{\breve{\boldsymbol{v}}}^{yf}\left(\Sigma_{\breve{\boldsymbol{v}}|y}^{ff}\right)^{-\frac{1}{2}} & \Sigma_{\breve{\boldsymbol{v}}}^{yy}\end{bmatrix} \\ &= H\begin{bmatrix}\left(\Sigma_{\breve{\boldsymbol{v}}|y}^{ff}\right)^{\frac{1}{2}} & \Sigma_{\breve{\boldsymbol{v}}}^{fy} \\ O_{M\times K} & \Sigma_{\breve{\boldsymbol{v}}}^{yy}\end{bmatrix} = \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & \Sigma_{\breve{\boldsymbol{v}}}^{yy}\end{bmatrix}.\end{aligned}$$

Finally, a comparison of the submatrices of $R\Sigma_{\boldsymbol{v}}R'$ and $\Sigma_{\breve{\boldsymbol{v}}}$ yields $\Sigma_{\breve{\boldsymbol{v}}}^{yy} = \Sigma_{\boldsymbol{v}}^{yy}$. □

As shown in Lemma 5.1.2, the FAVAR parameters in Definition 5.1.1 are unique except for a non-singular transformation with $K(K+M)$ degrees of freedom. To eliminate this parameter ambiguity the one-step estimation method of Bernanke et al. (2005) restricted the first K rows of the loadings matrix as follows:

$$\bar{\Lambda}^f = \begin{bmatrix} I_K \\ \bar{\Lambda}^f_{(N-K)\times K}\end{bmatrix} \quad \text{and} \quad \bar{\Lambda}^y = \begin{bmatrix} O_{K\times M} \\ \bar{\Lambda}^y_{(N-K)\times M}\end{bmatrix}, \tag{5.13}$$

with submatrices $\bar{\Lambda}^f_{(N-K)\times K}$ and $\bar{\Lambda}^y_{(N-K)\times M}$ as the unconstrained last $N-K$ rows of $\bar{\Lambda}^f$ and $\bar{\Lambda}^y$. The same method is used in Marcellino and Sivec (2016). Although Bork (2009) took over this idea, he gained flexibility such that he admitted the structure in (5.13) to be scattered across any K rows of the loadings matrix. That is, not necessarily the first K rows matter.

In this chapter, linear parameter constraints also guarantee parameter identifiability. For this purpose, we linearly constrain the loadings matrix in (5.6) or the VAR(p) coefficients in (5.7). The transformed model in (5.6)-(5.7) shows that the total mapping HR does not affect the observed variables $\boldsymbol{Y}_t$, but it simplifies the covariance matrix of the errors $\bar{\boldsymbol{v}}_t$ in the transition equation. Furthermore, the special shape of matrix H decreases the number of degrees of freedom to $K(K-1)/2$. To verify this let $\tilde{R} \in \mathbb{R}^{(K+M)\times(K+M)}$ be a non-singular matrix defined as follows:

$$\tilde{R} = \begin{pmatrix}\tilde{R}_1 & \tilde{R}_2 \\ O_{M\times K} & I_M\end{pmatrix}. \tag{5.14}$$

To preserve the structure of the covariance matrix $\Sigma_{\bar{\boldsymbol{v}}}$ any additional transformation has to satisfy:

$$\begin{aligned} \tilde{R}\Sigma_{\bar{\boldsymbol{v}}}\left(\tilde{R}\right)' &= \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & \Sigma_{\boldsymbol{v}}^{yy}\end{bmatrix} \\ \Leftrightarrow \begin{bmatrix}\tilde{R}_1\left(\tilde{R}_1\right)' + \tilde{R}_2\Sigma_{\boldsymbol{v}}^{yy}\left(\tilde{R}_2\right)' & \tilde{R}_2\Sigma_{\boldsymbol{v}}^{yy} \\ \Sigma_{\boldsymbol{v}}^{yy}\left(\tilde{R}_2\right)' & \Sigma_{\boldsymbol{v}}^{yy}\end{bmatrix} &= \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & \Sigma_{\boldsymbol{v}}^{yy}\end{bmatrix},\end{aligned}$$

which implies $\tilde{R}_2 = O_{K\times M}$ and $\tilde{R}_1(\tilde{R}_1)' = I_K$. Thus, the $K(K+1)/2$ restrictions for the orthonormality of the columns of $\tilde{R}_1$ reduce the K^2 degrees of freedom of an arbitrary matrix $\tilde{R}_1 \in \mathbb{R}^{K\times K}$ and we finally have $K(K-1)/2$ degrees of freedom left. Thus, the matrix $\tilde{R}_1$ must be a rotation matrix.

Therefore, we do not impose restrictions on matrix $\bar{\Lambda}^y$ in (5.6), when we talk about loadings constraints. However, for matrix $\bar{\Lambda}^f$ in (5.6), we propose the following formulation:

$$\bar{\Lambda}^f = \begin{bmatrix} * & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ * & \cdots & * & 0 \\ * & \cdots & * & * \\ & \bar{\Lambda}^f_{(N-K)\times K} & & \end{bmatrix}, \tag{5.15}$$

where the matrix $\bar{\Lambda}^f_{(N-K)\times K} \in \mathbb{R}^{(N-K)\times K}$ comprises the unconstrained last $N-K$ rows of $\bar{\Lambda}^f$ and the upper $K \times K$-dimensional submatrix of $\bar{\Lambda}^f$ represents a lower triangular matrix. Before we proceed, we explain in more detail how to obtain the shape of $\bar{\Lambda}^f$ in (5.15) using the rotation matrix $\tilde{R}$ in (5.14). Referring to Golub and Van Loan (1996, p. 215, Section 5.1.8), Givens Rotations enable us to zero entries of a vector and rank among the rotation matrices. In the sequel, let $\bar{\Lambda}^f_{K\times K} \in \mathbb{R}^{K\times K}$ be the unrestricted upper block matrix of $\bar{\Lambda}^f$ in (5.6) and let $G_{i,j} \in \mathbb{R}^{K\times K}$ be the Givens Rotation to zero the element in the i-th row and j-th column of $\bar{\Lambda}^f_{K\times K}$ for $1 \le i, j \le K$. Then, we have:

$$\begin{bmatrix} * & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ * & \cdots & * & 0 \\ * & \cdots & * & * \end{bmatrix} = \bar{\Lambda}^f_{K\times K} \underbrace{\left(\prod_{i=1}^{K-1} \left(\prod_{j=i+1}^{K} G_{i,j} \right) \right)}_{\tilde{G}},$$

with $\tilde{G}(\tilde{G})' = (\tilde{G})'\tilde{G} = I_K$. In this manner, we eliminate all remaining degrees of freedom. Furthermore, for matrix $\tilde{R}$ defined by:

$$\tilde{R} = \begin{bmatrix} \tilde{G} & O_{K\times M} \\ O_{M\times K} & I_M \end{bmatrix}$$

the FAVAR in (5.6)-(5.7) keeps the special structure of the covariance matrix $\Sigma_{\bar{\boldsymbol{v}}}$.

At the end, we put the FAVAR in (5.6)-(5.7) with loadings restrictions in (5.15) in relation to the results from Bai et al. (2015). If $\Sigma_{\boldsymbol{e}}$ is a diagonal matrix, “Assumptions A-D” and the identification restrictions “IRb” in Bai et al. (2015) are satisfied such that their asymptotic distributions of the factor loadings, the coefficient matrices and the IRFs remain valid. Unfortunately, their lengthy expressions of the distribution parameters appear cumbersome and unattraktive, when it comes to their implementation.

5.1.2 Estimation of FAVARs with Complete Panel Data

As in Bork (2009), we derive an EM for the estimation of FAVARs. For clarity reasons, we introduce the joint vector $\boldsymbol{C}_t = \left[\bar{\boldsymbol{F}}'_t, \boldsymbol{Y}'_t\right]' \in \mathbb{R}^{K+M}$ and rewrite the model in (5.6)-(5.7) as follows:

$$\boldsymbol{X}_t = \bar{\Lambda}\boldsymbol{C}_t + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \mathcal{N}\left(\boldsymbol{0}_N, \Sigma_{\boldsymbol{e}}\right) \text{iid}, \tag{5.16}$$

$$\boldsymbol{C}_t = \sum_{i=1}^{p} \bar{\Phi}_i \boldsymbol{C}_{t-i} + \bar{\boldsymbol{v}}_t, \quad \bar{\boldsymbol{v}}_t \sim \mathcal{N}\left(\boldsymbol{0}_{K+M}, \Sigma_{\bar{\boldsymbol{v}}}\right) \text{iid}. \tag{5.17}$$

With this in mind, we receive the subsequent conditional distributions of $\boldsymbol{X}_t$ and $\boldsymbol{C}_t$.

Lemma 5.1.4 (Conditional Vector Distributions of FAVARs)
For the FAVAR in (5.16)-(5.17)*, the vectors* $\boldsymbol{X}_t|\boldsymbol{C}_t$ *and* $\boldsymbol{C}_t|\boldsymbol{C}_{t-1}, \ldots, \boldsymbol{C}_{t-p}$ *are Gaussian as follows:*

$$\boldsymbol{X}_t|\boldsymbol{C}_t \sim \mathcal{N}\left(\bar{\Lambda}\boldsymbol{C}_t, \Sigma_{\boldsymbol{e}}\right), \tag{5.18}$$

$$\boldsymbol{C}_t|\boldsymbol{C}_{t-1}, ..., \boldsymbol{C}_{t-p} \sim \mathcal{N}\left(\sum_{i=1}^{p} \bar{\Phi}_i \boldsymbol{C}_{t-i}, \Sigma_{\bar{\boldsymbol{v}}}\right). \tag{5.19}$$

Proof:
Follows directly from (5.16)-(5.17). □

Next, we focus on the log-likelihood function of the FAVAR in (5.16)-(5.17).

Lemma 5.1.5 (Conditional Log-Likelihood Function of FAVARs)
Assume the FAVAR in (5.16)-(5.17) *with parameters* $\Theta = \{\bar{\Lambda}, \Sigma_{\boldsymbol{\epsilon}}, \bar{\Phi}_1, \ldots, \bar{\Phi}_p, \Sigma_{\bar{\boldsymbol{v}}}^{yy}\}$ *for lag order* $p > 0$. *Moreover, the matrices* $X = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T]' \in \mathbb{R}^{T \times N}$ *and* $Y = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T]' \in \mathbb{R}^{T \times M}$ *are the completely observed panel data and variables. For a sufficiently large sample size* $T > p$*, let* $\mathcal{L}(\Theta|X, C)$ *be the log-likelihood function for the observations* X *and partially hidden factors* $\boldsymbol{C}_{p+1}, ..., \boldsymbol{C}_T$ *conditioned on the factors* $\boldsymbol{C}_1, ..., \boldsymbol{C}_p$*. Then, it holds:*

$$\begin{aligned}
\mathcal{L}(\Theta|X, C) = &- \frac{TN}{2} \, ln(2\pi) - \frac{T}{2} \, ln(|\Sigma_{\boldsymbol{e}}|) - \frac{1}{2} \sum_{t=1}^{T} \left(\boldsymbol{X}_t - \bar{\Lambda}\boldsymbol{C}_t\right)' \Sigma_{\boldsymbol{e}}^{-1} \left(\boldsymbol{X}_t - \bar{\Lambda}\boldsymbol{C}_t\right) \\
&- \frac{(K+M)(T-p)}{2} \, ln(2\pi) - \frac{T-p}{2} \, ln(|\Sigma_{\bar{\boldsymbol{v}}}|) \\
&- \frac{1}{2} \sum_{t=p+1}^{T} \left(\boldsymbol{C}_t - \sum_{i=1}^{p} \bar{\Phi}_i \boldsymbol{C}_{t-i}\right)' \Sigma_{\bar{\boldsymbol{v}}}^{-1} \left(\boldsymbol{C}_t - \sum_{i=1}^{p} \bar{\Phi}_i \boldsymbol{C}_{t-i}\right),
\end{aligned} \tag{5.20}$$

with joint vector $\boldsymbol{C}_t = \left[\bar{\boldsymbol{F}}_t', \boldsymbol{Y}_t'\right]' \in \mathbb{R}^{K+M}$.

Proof:
Similar to Lemma 4.1.5, the Bayes' theorem yields for $f_\Theta(\boldsymbol{X}_T, ..., \boldsymbol{X}_1, \boldsymbol{C}_T, ..., \boldsymbol{C}_{p+1}|\boldsymbol{C}_p, ..., \boldsymbol{C}_1)$, i.e., the joint density of the observations X and factors $\boldsymbol{C}_{p+1}, ..., \boldsymbol{C}_T$ conditioned on the factors $\boldsymbol{C}_1, ..., \boldsymbol{C}_p$:

$$f_\Theta(\boldsymbol{X}_T, ..., \boldsymbol{X}_1, \boldsymbol{C}_T, ..., \boldsymbol{C}_{p+1}|\boldsymbol{C}_p, ..., \boldsymbol{C}_1) = \left(\prod_{t=1}^{T} f_\Theta(\boldsymbol{X}_t|\boldsymbol{C}_t)\right) \left(\prod_{t=p+1}^{T} f_\Theta(\boldsymbol{C}_t|\boldsymbol{C}_{t-1}, ..., \boldsymbol{C}_{t-p})\right).$$

Next, we substitute $f_\Theta(\boldsymbol{X}_t|\boldsymbol{C}_t)$ and $f_\Theta(\boldsymbol{C}_t|\boldsymbol{C}_{t-1}, ..., \boldsymbol{C}_{t-p})$ by the conditional distributions in (5.18) and (5.19), take the logarithm of the joint distribution and rearrange the terms:

$$\begin{aligned}
\mathcal{L}(\Theta|X, C) &= \sum_{t=1}^{T} ln\left((2\pi)^{-\frac{N}{2}} |\Sigma_{\boldsymbol{e}}|^{-\frac{1}{2}} exp\left(-\frac{1}{2}\left(\boldsymbol{X}_t - \bar{\Lambda}\boldsymbol{C}_t\right)' \Sigma_{\boldsymbol{e}}^{-1} \left(\boldsymbol{X}_t - \bar{\Lambda}\boldsymbol{C}_t\right)\right)\right) \\
&+ \sum_{t=p+1}^{T} ln\left((2\pi)^{-\frac{K+M}{2}} |\Sigma_{\bar{\boldsymbol{v}}}|^{-\frac{1}{2}} exp\left(-\frac{1}{2}\left(\boldsymbol{C}_t - \sum_{i=1}^{p} \bar{\Phi}_i \boldsymbol{C}_{t-i}\right)' \Sigma_{\bar{\boldsymbol{v}}}^{-1} \left(\boldsymbol{C}_t - \sum_{i=1}^{p} \bar{\Phi}_i \boldsymbol{C}_{t-i}\right)\right)\right) \\
&= -\frac{TN}{2} \, ln(2\pi) - \frac{T}{2} \, ln(|\Sigma_{\boldsymbol{e}}|) - \frac{1}{2} \sum_{t=1}^{T} \left(\boldsymbol{X}_t - \bar{\Lambda}\boldsymbol{C}_t\right)' \Sigma_{\boldsymbol{e}}^{-1} \left(\boldsymbol{X}_t - \bar{\Lambda}\boldsymbol{C}_t\right) \\
&- \frac{(K+M)(T-p)}{2} \, ln(2\pi) - \frac{T-p}{2} \, ln(|\Sigma_{\bar{\boldsymbol{v}}}|) \\
&- \frac{1}{2} \sum_{t=p+1}^{T} \left(\boldsymbol{C}_t - \sum_{i=1}^{p} \bar{\Phi}_i \boldsymbol{C}_{t-i}\right)' \Sigma_{\bar{\boldsymbol{v}}}^{-1} \left(\boldsymbol{C}_t - \sum_{i=1}^{p} \bar{\Phi}_i \boldsymbol{C}_{t-i}\right),
\end{aligned}$$

which proves the claim. □

Because of $p \geq 1$, Equation (5.20) represents a generalization of the log-likelihood functions in Shumway and Stoffer (1982), Bork (2009), Bańbura and Modugno (2014) and Jungbacker and Koopman (2015). The log-likelihood function $\mathcal{L}(\Theta|X, C)$ in (5.20) depends on the latent factors $\bar{\boldsymbol{F}}_t$ part of the vector $\boldsymbol{C}_t$. Therefore, we cannot compute it directly. However, to estimate the model in (5.16)-(5.17) in the maximum likelihood framework we use the EM of Dempster et al. (1977). Thereby, the expectation step integrates the hidden factors $\bar{\boldsymbol{F}}_t$ out by computing the conditional expectation of $\mathcal{L}(\Theta|X, C)$ with respect to the observed data X and Y.

Lemma 5.1.6 (Conditional Expectation of FAVAR Log-Likelihood)
With the notation in Lemma 5.1.5, the expectation of the log-likelihood function $\mathcal{L}(\Theta|X,C)$ conditioned on the completely observed panel data X and variables Y is given by:

$$\begin{aligned}
\mathbb{E}_\Theta\left[\mathcal{L}(\Theta|X,C)|X,Y\right] = &-\frac{TN}{2}\,ln\,(2\pi) - \frac{(K+M)(T-p)}{2}\,ln\,(2\pi) - \frac{T}{2}\,ln\,(|\Sigma_{\boldsymbol{e}}|) - \frac{T-p}{2}\,ln\,(|\Sigma_{\boldsymbol{v}}^{yy}|) \\
&-\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{X}_t'\Sigma_{\boldsymbol{e}}^{-1}\boldsymbol{X}_t + \frac{1}{2}\sum_{t=1}^{T}\boldsymbol{X}_t'\Sigma_{\boldsymbol{e}}^{-1}\bar{\Lambda}\mathbb{E}_\Theta\left[\boldsymbol{C}_t|X,Y\right] \\
&+\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_\Theta\left[\boldsymbol{C}_t|X,Y\right]'\bar{\Lambda}'\Sigma_{\boldsymbol{e}}^{-1}\boldsymbol{X}_t - \frac{1}{2}\sum_{t=1}^{T}tr\left(\bar{\Lambda}'\Sigma_{\boldsymbol{e}}^{-1}\bar{\Lambda}\mathbb{E}_\Theta\left[\boldsymbol{C}_t\boldsymbol{C}_t'|X,Y\right]\right) \\
&-\frac{1}{2}\sum_{t=p+1}^{T}tr\left(\begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & (\Sigma_{\boldsymbol{v}}^{yy})^{-1}\end{bmatrix}\mathbb{E}_\Theta\left[\boldsymbol{C}_t\boldsymbol{C}_t'|X,Y\right]\right) \\
&-\frac{1}{2}\sum_{t=p+1}^{T}\sum_{i,j=1}^{p}tr\left(\bar{\Phi}_i'\begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & (\Sigma_{\boldsymbol{v}}^{yy})^{-1}\end{bmatrix}\bar{\Phi}_j\mathbb{E}_\Theta\left[\boldsymbol{C}_{t-j}\boldsymbol{C}_{t-i}'|X,Y\right]\right) \\
&+\frac{1}{2}\sum_{t=p+1}^{T}\sum_{i=1}^{p}tr\left(\begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & (\Sigma_{\boldsymbol{v}}^{yy})^{-1}\end{bmatrix}\bar{\Phi}_i\mathbb{E}_\Theta\left[\boldsymbol{C}_{t-i}\boldsymbol{C}_t'|X,Y\right]\right) \\
&+\frac{1}{2}\sum_{t=p+1}^{T}\sum_{i=1}^{p}tr\left(\bar{\Phi}_i'\begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & (\Sigma_{\boldsymbol{v}}^{yy})^{-1}\end{bmatrix}\mathbb{E}_\Theta\left[\boldsymbol{C}_t\boldsymbol{C}_{t-i}'|X,Y\right]\right),
\end{aligned} \tag{5.21}$$

with

$$\begin{aligned}
\mathbb{E}_\Theta\left[\boldsymbol{C}_t|X,Y\right] &= \begin{bmatrix}\mathbb{E}_\Theta\left[\bar{\boldsymbol{F}}_t|X,Y\right] \\ \boldsymbol{Y}_t\end{bmatrix}, \quad 1 \le t \le T, \\
\mathbb{E}_\Theta\left[\boldsymbol{C}_{t-j}\boldsymbol{C}_{t-i}'|X,Y\right] &= \begin{bmatrix}\mathbb{E}_\Theta\left[\bar{\boldsymbol{F}}_{t-j}\bar{\boldsymbol{F}}_{t-i}'|X,Y\right] & \mathbb{E}_\Theta\left[\bar{\boldsymbol{F}}_{t-j}|X,Y\right]\boldsymbol{Y}_{t-i}' \\ \boldsymbol{Y}_{t-j}\mathbb{E}_\Theta\left[\bar{\boldsymbol{F}}_{t-i}|X,Y\right]' & \boldsymbol{Y}_{t-j}\boldsymbol{Y}_{t-i}'\end{bmatrix}, \quad \begin{matrix}1 \le i,j \le p, \\ p+1 \le t \le T.\end{matrix}
\end{aligned} \tag{5.22}$$

Proof:
Follows from the log-likelihood function $\mathcal{L}(\Theta|X,C)$ in (5.20), the definition of the covariance matrix $\Sigma_{\bar{\boldsymbol{v}}}$ in (5.7), the matrix trace properties in Lemma A.1.2 and the linearity of the conditional expectation. □

In Lemma 5.1.6, the unobserved factors $\bar{\boldsymbol{F}}_t$ are replaced by their conditional moments, which the Kalman Smoother in Section 5.1.4 yields for the latest parameter estimates $\hat{\Theta}_{(l)}$. Then, the second step of the EM maximizes $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}(\Theta|X,C)|X,Y\right]$ with respect to Θ subject to parameter constraints. In this context, we follow Bernanke et al. (2005), Bork (2009) as well as Bańbura and Modugno (2014) and consider loadings constraints. Additionally, we show later on how to incorporate linear restrictions for the VAR coefficients Φ_i. Although the moments of $\boldsymbol{C}_t$ in (5.22) are functions of $\hat{\Theta}_{(l)}$ and so, of Θ, our optimization routine ignores this fact and treats them as constants during each EM loop. That means, they are frozen at the latest level. As Bańbura and Modugno (2014), we pursue the approach of Bork (2009) and Bork et al. (2010), when we reformulate the loadings constraints in (5.15) as $H_\Lambda vec\left(\bar{\Lambda}\right) = \boldsymbol{\kappa}_\Lambda$ with vector $\boldsymbol{\kappa}_\Lambda \in \mathbb{R}^P$ and matrix $H_\Lambda \in \mathbb{R}^{P\times N(K+M)}$ such that $P \le N(K+M)$ and $\text{rank}(H_\Lambda) = P$.

Lemma 5.1.7 (EM for Complete FAVARs with Loadings Restrictions)
Let the panel data $X \in \mathbb{R}^{T\times N}$ and variables $Y \in \mathbb{R}^{T\times M}$ be complete. Assume the FAVAR in (5.16)-(5.17) with linear loadings constraints: $H_\Lambda vec\left(\bar{\Lambda}\right) = \boldsymbol{\kappa}_\Lambda$ for vector $\boldsymbol{\kappa}_\Lambda \in \mathbb{R}^P$ and matrix $H_\Lambda \in \mathbb{R}^{P\times N(K+M)}$ of full row rank $P \le N(K+M)$. Furthermore, let matrix $\bar{\Phi} = \left[\bar{\Phi}_1, ..., \bar{\Phi}_p\right] \in \mathbb{R}^{(K+M)\times p(K+M)}$ collect all

coefficients of the factor dynamics and let (l) be the current EM loop. For clarity reasons, we write $\Theta_{(l)}$ instead of $\hat{\Theta}_{(l)}$. Then, the parameters of the next loop $(l+1)$ are updated in the following way:

$$\begin{aligned} vec\left(\bar{\Lambda}\right)_{(l+1)} &= \left(\mathcal{B}^{-1}\otimes\Sigma_{\boldsymbol{e}(l)}\right)H_\Lambda'\left(H_\Lambda\left(\mathcal{B}^{-1}\otimes\Sigma_{\boldsymbol{e}(l)}\right)H_\Lambda'\right)^{-1}\left(\boldsymbol{\kappa}_\Lambda - H_\Lambda\ vec\left(\mathcal{A}\mathcal{B}^{-1}\right)\right) \\ &\quad + vec\left(\mathcal{A}\mathcal{B}^{-1}\right), && (5.23) \\ \Sigma_{\boldsymbol{e}(l+1)} &= \mathcal{C} - \mathcal{A}\bar{\Lambda}_{(l+1)}' - \bar{\Lambda}_{(l+1)}\mathcal{A}' + \bar{\Lambda}_{(l+1)}\mathcal{B}\bar{\Lambda}_{(l+1)}', && (5.24) \\ \bar{\Phi}_{(l+1)} &= \mathcal{D}\mathcal{E}^{-1}, && (5.25) \\ \Sigma_{\boldsymbol{v}(l+1)}^{yy} &= \begin{bmatrix} O_{M\times K} & I_M \end{bmatrix}\left(\mathcal{F} - \mathcal{D}\bar{\Phi}_{(l+1)}' - \bar{\Phi}_{(l+1)}\mathcal{D}' + \bar{\Phi}_{(l+1)}\mathcal{E}\bar{\Phi}_{(l+1)}'\right)\begin{bmatrix} O_{K\times M} \\ I_M \end{bmatrix}, && (5.26) \end{aligned}$$

where the matrices $\mathcal{A}\in\mathbb{R}^{N\times(K+M)}$, $\mathcal{B}\in\mathbb{R}^{(K+M)\times(K+M)}$, $\mathcal{C}\in\mathbb{R}^{N\times N}$, $\mathcal{D}\in\mathbb{R}^{(K+M)\times p(K+M)}$, $\mathcal{E}\in\mathbb{R}^{p(K+M)\times p(K+M)}$ and $\mathcal{F}\in\mathbb{R}^{(K+M)\times(K+M)}$ require the conditional moments in (5.22) *and are given by:*

$$\begin{aligned} \mathcal{A} &= \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]', & \mathcal{D} &= \frac{1}{T-p}\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t\bar{\boldsymbol{C}}_{t-1}'|X,Y\right], \\ \mathcal{B} &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t\boldsymbol{C}_t'|X,Y\right], & \mathcal{E} &= \frac{1}{T-p}\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\bar{\boldsymbol{C}}_{t-1}\bar{\boldsymbol{C}}_{t-1}'|X,Y\right], && (5.27) \\ \mathcal{C} &= \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t\boldsymbol{X}_t', & \mathcal{F} &= \frac{1}{T-p}\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t\boldsymbol{C}_t'|X,Y\right]. \end{aligned}$$

For the state vector $\bar{\boldsymbol{C}}_t\in\mathbb{R}^{p(K+M)}$ we have $\bar{\boldsymbol{C}}_t = \left[\boldsymbol{C}_t', ..., \boldsymbol{C}_{t-p+1}'\right]'$.

Proof:
We aim at maximizing $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,C\right)|X,Y\right]$ with respect to Θ subject to $H_\Lambda vec\left(\bar{\Lambda}\right) = \boldsymbol{\kappa}_\Lambda$. Therefore, we apply the method of Lagrange multipliers to the optimization problem:

$$\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,C\right)|X,Y\right] + \boldsymbol{\lambda}'\left(H_\Lambda vec\left(\bar{\Lambda}\right) - \boldsymbol{\kappa}_\Lambda\right), \tag{5.28}$$

with $\boldsymbol{\lambda}$ as Lagrange multiplier. Similar to Bork (2009) and Bańbura and Modugno (2014), the conditional moments in (5.22) are constants, when we derive the partial derivatives of (5.28) and solve the resulting system of matrix equations.

At first, we show that the optimization of the parameters $\bar{\Lambda}$ and $\Sigma_{\boldsymbol{e}}$ can be separated from the one of the matrices $\bar{\Phi}_i$, $1\le i\le p$, and $\Sigma_{\boldsymbol{v}}^{yy}$, since both problems do not affect each other. In the last case, the optimization of $\bar{\Phi}_i$, $1\le i\le p$, does not depend on $\Sigma_{\boldsymbol{v}}^{yy}$. Therefore, we start with the solutions for $\bar{\Phi}_i$, $1\le i\le p$. Thereafter, we insert these into the solution of $\Sigma_{\boldsymbol{v}}^{yy}$. The loadings restrictions cause that the partial derivatives of $\bar{\Lambda}$ depend on $\Sigma_{\boldsymbol{e}}$ in a non-linear manner, however, for simplicity reasons, we proceed as with $\bar{\Phi}_i$, $1\le i\le p$, and $\Sigma_{\boldsymbol{v}}^{yy}$.

For the partial derivatives of (5.28) with respect to $\bar{\Lambda}$, the derivatives of the matrix trace functions and the logarithm of the matrix determinant in Lemmata A.1.5 and A.1.8 provide:

$$\begin{aligned} &\frac{\partial\left(\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,C\right)|X,Y\right] + \boldsymbol{\lambda}'\left(H_\Lambda vec\left(\bar{\Lambda}\right) - \boldsymbol{\kappa}_\Lambda\right)\right)}{\partial\bar{\Lambda}} \\ &= \frac{\partial}{\partial\bar{\Lambda}}\frac{1}{2}\sum_{t=1}^{T}tr\left(\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]\boldsymbol{X}_t'\Sigma_{\boldsymbol{e}}^{-1}\bar{\Lambda}\right) + \frac{\partial}{\partial\bar{\Lambda}}\boldsymbol{\lambda}'H_\Lambda vec\left(\bar{\Lambda}\right) \\ &\quad + \frac{\partial}{\partial\bar{\Lambda}}\frac{1}{2}\sum_{t=1}^{T}tr\left(\Sigma_{\boldsymbol{e}}^{-1}\boldsymbol{X}_t\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]'\bar{\Lambda}'\right) - \frac{\partial}{\partial\bar{\Lambda}}\frac{1}{2}\sum_{t=1}^{T}tr\left(\bar{\Lambda}'\Sigma_{\boldsymbol{e}}^{-1}\bar{\Lambda}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t\boldsymbol{C}_t'|X,Y\right]\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{t=1}^{T}\Sigma_{\boldsymbol{e}}^{-1}\boldsymbol{X}_t\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]' + \frac{1}{2}\sum_{t=1}^{T}\Sigma_{\boldsymbol{e}}^{-1}\boldsymbol{X}_t\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]' \\
&\quad - \frac{1}{2}\sum_{t=1}^{T}\Sigma_{\boldsymbol{e}}^{-1}\bar{\Lambda}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t\boldsymbol{C}_t'|X,Y\right]\cdot 2 + \begin{bmatrix} (\boldsymbol{\lambda}'H_\Lambda)_1 & \cdots & (\boldsymbol{\lambda}'H_\Lambda)_{N(K+M-1)+1} \\ \vdots & & \vdots \\ (\boldsymbol{\lambda}'H_\Lambda)_N & \cdots & (\boldsymbol{\lambda}'H_\Lambda)_{N(K+M)} \end{bmatrix} \\
&= \Sigma_{\boldsymbol{e}}^{-1}\sum_{t=1}^{T}\boldsymbol{X}_t\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]' - \Sigma_{\boldsymbol{e}}^{-1}\bar{\Lambda}\sum_{t=1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t\boldsymbol{C}_t'|X,Y\right] \\
&\quad + \begin{bmatrix} (\boldsymbol{\lambda}'H_\Lambda)_1 & \cdots & (\boldsymbol{\lambda}'H_\Lambda)_{N(K+M-1)+1} \\ \vdots & & \vdots \\ (\boldsymbol{\lambda}'H_\Lambda)_N & \cdots & (\boldsymbol{\lambda}'H_\Lambda)_{N(K+M)} \end{bmatrix}.
\end{aligned}
$$

By setting these partial derivatives equal to the zero matrix $O_{N\times(K+M)}$ and solving it for $\bar{\Lambda}$, we get:

$$
\Sigma_{\boldsymbol{e}}^{-1}\sum_{t=1}^{T}\boldsymbol{X}_t\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]' + \begin{bmatrix} (\boldsymbol{\lambda}'H_\Lambda)_1 & \cdots & (\boldsymbol{\lambda}'H_\Lambda)_{N(K+M-1)+1} \\ \vdots & & \vdots \\ (\boldsymbol{\lambda}'H_\Lambda)_N & \cdots & (\boldsymbol{\lambda}'H_\Lambda)_{N(K+M)} \end{bmatrix} = \Sigma_{\boldsymbol{e}}^{-1}\bar{\Lambda}\sum_{t=1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t\boldsymbol{C}_t'|X,Y\right].
$$

For clarity reasons, we replace the sums by matrices $\mathcal{A}$ and $\mathcal{B}$ in (5.27), apply the $vec\,(\cdot)$ operator to both sides and profit from its properties in Lemma A.1.13 as follows:

$$
\begin{aligned}
vec\,(T\mathcal{A}) + vec\left(\Sigma_{\boldsymbol{e}}\begin{bmatrix} (\boldsymbol{\lambda}'H_\Lambda)_1 & \cdots & (\boldsymbol{\lambda}'H_\Lambda)_{N(K+M-1)+1} \\ \vdots & & \vdots \\ (\boldsymbol{\lambda}'H_\Lambda)_N & \cdots & (\boldsymbol{\lambda}'H_\Lambda)_{N(K+M)} \end{bmatrix}\right) &= vec\left(\bar{\Lambda}T\mathcal{B}\right) \\
\Leftrightarrow T\ vec\,(\mathcal{A}) + \left(I_{K+M}\otimes\Sigma_{\boldsymbol{e}}\right)H_\Lambda'\boldsymbol{\lambda} &= T\ vec\left(\bar{\Lambda}\mathcal{B}\right) \\
\Leftrightarrow T\ vec\,(\mathcal{A}) + \left(I_{K+M}\otimes\Sigma_{\boldsymbol{e}}\right)H_\Lambda'\boldsymbol{\lambda} &= T\left(\mathcal{B}\otimes I_N\right)\ vec\left(\bar{\Lambda}\right) \\
\Leftrightarrow \left(\mathcal{B}\otimes I_N\right)^{-1}vec\,(\mathcal{A}) + \frac{1}{T}\left(\mathcal{B}\otimes I_N\right)^{-1}\left(I_{K+M}\otimes\Sigma_{\boldsymbol{e}}\right)H_\Lambda'\boldsymbol{\lambda} &= vec\left(\bar{\Lambda}\right).
\end{aligned} \tag{5.29}
$$

Now, we substitute the constraint $H_\Lambda vec\left(\bar{\Lambda}\right) = \boldsymbol{\kappa}_\Lambda$ in (5.29) and benefit from the multiplication property of the Kronecker product in Lemma A.1.10:

$$
\begin{aligned}
(5.29) &\Leftrightarrow H_\Lambda\left(\mathcal{B}\otimes I_N\right)^{-1}vec\,(\mathcal{A}) + \frac{1}{T}H_\Lambda\left(\mathcal{B}\otimes I_N\right)^{-1}\left(I_{K+M}\otimes\Sigma_{\boldsymbol{e}}\right)H_\Lambda'\boldsymbol{\lambda} = \boldsymbol{\kappa}_\Lambda \\
&\Leftrightarrow H_\Lambda\left(\mathcal{B}\otimes I_N\right)^{-1}vec\,(\mathcal{A}) + \frac{1}{T}H_\Lambda\left(\mathcal{B}^{-1}\otimes\Sigma_{\boldsymbol{e}}\right)H_\Lambda'\boldsymbol{\lambda} = \boldsymbol{\kappa}_\Lambda \\
&\Leftrightarrow H_\Lambda\left(\mathcal{B}^{-1}\otimes\Sigma_{\boldsymbol{e}}\right)H_\Lambda'\boldsymbol{\lambda} = T\left(\boldsymbol{\kappa}_\Lambda - H_\Lambda\left(\mathcal{B}\otimes I_N\right)^{-1}vec\,(\mathcal{A})\right).
\end{aligned}
$$

Solving this expression for the Lagrangian multiplier $\boldsymbol{\lambda}$ leads to:

$$
\boldsymbol{\lambda} = T\left(H_\Lambda\left(\mathcal{B}^{-1}\otimes\Sigma_{\boldsymbol{e}}\right)H_\Lambda'\right)^{-1}\left(\boldsymbol{\kappa}_\Lambda - H_\Lambda\left(\mathcal{B}^{-1}\otimes I_N\right)vec\,(\mathcal{A})\right). \tag{5.30}
$$

Next, we insert the derived multiplier (5.30) in (5.29) and obtain:

$$
\begin{aligned}
vec\left(\bar{\Lambda}\right) &= \left(\mathcal{B}\otimes I_N\right)^{-1}vec\,(\mathcal{A}) + \frac{1}{T}\left(\mathcal{B}^{-1}\otimes\Sigma_{\boldsymbol{e}}\right)H_\Lambda'T\left(H_\Lambda\left(\mathcal{B}^{-1}\otimes\Sigma_{\boldsymbol{e}}\right)H_\Lambda'\right)^{-1} \\
&\qquad\cdot\left(\boldsymbol{\kappa}_\Lambda - H_\Lambda\left(\mathcal{B}^{-1}\otimes I_N\right)vec\,(\mathcal{A})\right) \\
&= vec\left(\mathcal{A}\mathcal{B}^{-1}\right) + \left(\mathcal{B}^{-1}\otimes\Sigma_{\boldsymbol{e}}\right)H_\Lambda'\left(H_\Lambda\left(\mathcal{B}^{-1}\otimes\Sigma_{\boldsymbol{e}}\right)H_\Lambda'\right)^{-1}\left(\boldsymbol{\kappa}_\Lambda - H_\Lambda\ vec\left(\mathcal{A}\mathcal{B}^{-1}\right)\right),
\end{aligned}
$$

where the last line applies Lemma A.1.13. Let (l) be the current loop of the EM. Then, the estimator of the restricted loadings matrix for the next loop $(l+1)$ is given by (5.23).

The estimator for $\Sigma_{\boldsymbol{e}}$ is obtained by differentiating $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,C\right)|X,Y\right]$ with respect to $\Sigma_{\boldsymbol{e}}$. Again see Lemmata A.1.5 and A.1.8 for the derivatives of the matrix traces and logarithm of the matrix determinant:

$$\begin{aligned}
&\frac{\partial \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,C\right)|X,Y\right]}{\partial \Sigma_{\boldsymbol{e}}} \\
&= -\frac{T}{2}\frac{\partial}{\partial \Sigma_{\boldsymbol{e}}} ln\left(|\Sigma_{\boldsymbol{e}}|\right) - \frac{1}{2}\frac{\partial}{\partial \Sigma_{\boldsymbol{e}}} tr\left(\sum_{t=1}^{T} \boldsymbol{X}_t \boldsymbol{X}_t' \Sigma_{\boldsymbol{e}}^{-1}\right) + \frac{1}{2}\frac{\partial}{\partial \Sigma_{\boldsymbol{e}}} tr\left(\bar{\Lambda}\sum_{t=1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]\boldsymbol{X}_t'\Sigma_{\boldsymbol{e}}^{-1}\right) \\
&+ \frac{1}{2}\frac{\partial}{\partial \Sigma_{\boldsymbol{e}}} tr\left(\sum_{t=1}^{T} \boldsymbol{X}_t \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]'\bar{\Lambda}'\Sigma_{\boldsymbol{e}}^{-1}\right) - \frac{1}{2}\frac{\partial}{\partial \Sigma_{\boldsymbol{e}}} tr\left(\bar{\Lambda}\sum_{t=1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t\boldsymbol{C}_t'|X,Y\right]\bar{\Lambda}'\Sigma_{\boldsymbol{e}}^{-1}\right) \\
&= -\frac{T}{2}\Sigma_{\boldsymbol{e}}^{-1} + \frac{1}{2}\left(\Sigma_{\boldsymbol{e}}^{-1}\sum_{t=1}^{T} \boldsymbol{X}_t \boldsymbol{X}_t' \Sigma_{\boldsymbol{e}}^{-1}\right)' - \frac{1}{2}\left(\Sigma_{\boldsymbol{e}}^{-1}\bar{\Lambda}\sum_{t=1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]\boldsymbol{X}_t'\Sigma_{\boldsymbol{e}}^{-1}\right)' \\
&- \frac{1}{2}\left(\Sigma_{\boldsymbol{e}}^{-1}\sum_{t=1}^{T} \boldsymbol{X}_t \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]'\bar{\Lambda}'\Sigma_{\boldsymbol{e}}^{-1}\right)' + \frac{1}{2}\left(\Sigma_{\boldsymbol{e}}^{-1}\bar{\Lambda}\sum_{t=1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t\boldsymbol{C}_t'|X,Y\right]\bar{\Lambda}'\Sigma_{\boldsymbol{e}}^{-1}\right)' \\
&= -\frac{T}{2}\Sigma_{\boldsymbol{e}}^{-1} + \frac{T}{2}\Sigma_{\boldsymbol{e}}^{-1}\mathcal{C}\Sigma_{\boldsymbol{e}}^{-1} - \frac{T}{2}\Sigma_{\boldsymbol{e}}^{-1}\mathcal{A}\bar{\Lambda}'\Sigma_{\boldsymbol{e}}^{-1} - \frac{T}{2}\Sigma_{\boldsymbol{e}}^{-1}\bar{\Lambda}\mathcal{A}'\Sigma_{\boldsymbol{e}}^{-1} + \frac{T}{2}\Sigma_{\boldsymbol{e}}^{-1}\bar{\Lambda}\mathcal{B}\bar{\Lambda}'\Sigma_{\boldsymbol{e}}^{-1},
\end{aligned}$$

where the last step substitutes the sums by matrices $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ in (5.27). For the zeros of the associated equation system we receive:

$$\Sigma_{\boldsymbol{e}} = \mathcal{C} - \mathcal{A}\bar{\Lambda}' - \bar{\Lambda}\mathcal{A}' + \bar{\Lambda}\mathcal{B}\bar{\Lambda}'.$$

This provides for loop $(l+1)$ the estimator of the covariance matrix $\Sigma_{\boldsymbol{e}}$ in (5.24), which coincides with Shumway and Stoffer (1982, p. 257, Eq. 14). Actually, an equation system covering $vec\left(\bar{\Lambda}\right)$ and $\Sigma_{\boldsymbol{e}}$ has to be solved. The loadings constraints cause a non-linear equation system with unknown solution, therefore, we use a simplification: We first update $\bar{\Lambda}$, before we proceed with $\Sigma_{\boldsymbol{e}}$. This is why $\bar{\Lambda}_{(l+1)}$ in (5.23) involves $\Sigma_{\boldsymbol{e}(l)}$ instead of $\Sigma_{\boldsymbol{e}(l+1)}$.

For the estimators of $\Sigma_{\boldsymbol{v}}^{yy}$ and $\bar{\Phi}_i$, $1 \leq i \leq p$, we use the same trick as in Bańbura and Modugno (2014). That is, we replace the sum in the transition equation by the coefficient matrix $\bar{\Phi} = \left[\bar{\Phi}_1, ..., \bar{\Phi}_p\right]$ and the vector $\bar{\boldsymbol{C}}_t = \left[\boldsymbol{C}_t', ..., \boldsymbol{C}_{t-p+1}'\right]'$, before we rewrite the log-likelihood function in (5.20). Then, we receive:

$$\begin{aligned}
\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,C\right)|X,Y\right] = &-\frac{1}{2}\left(TN + \left(T-p\right)\left(K+M\right)\right) ln\left(2\pi\right) - \frac{T}{2} ln\left(|\Sigma_{\boldsymbol{e}}|\right) - \frac{T-p}{2} ln\left(|\Sigma_{\boldsymbol{v}}^{yy}|\right) \\
&- \frac{1}{2}\sum_{t=1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\left(\boldsymbol{X}_t - \bar{\Lambda}\boldsymbol{C}_t\right)'\Sigma_{\boldsymbol{e}}^{-1}\left(\boldsymbol{X}_t - \bar{\Lambda}\boldsymbol{C}_t\right)|X,Y\right] \\
&- \frac{1}{2}\sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\left(\boldsymbol{C}_t - \bar{\Phi}\bar{\boldsymbol{C}}_{t-1}\right)'\Sigma_{\bar{\boldsymbol{v}}}^{-1}\left(\boldsymbol{C}_t - \bar{\Phi}\bar{\boldsymbol{C}}_{t-1}\right)|X,Y\right].
\end{aligned}$$

The estimators for $\bar{\Phi}_i$, $1 \leq i \leq p$, are obtained by differentiating this conditional expectation with respect to $\bar{\Phi}$. Using Lemmata A.1.5 and A.1.8, we obtain:

$$\begin{aligned}
&\frac{\partial \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,C\right)|X,Y\right]}{\partial \bar{\Phi}} \\
&= -\frac{1}{2}\frac{\partial}{\partial \bar{\Phi}}\sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\left(\boldsymbol{C}_t - \bar{\Phi}\bar{\boldsymbol{C}}_{t-1}\right)'\Sigma_{\bar{\boldsymbol{v}}}^{-1}\left(\boldsymbol{C}_t - \bar{\Phi}\bar{\boldsymbol{C}}_{t-1}\right)|X,Y\right] \\
&= -\frac{1}{2}\frac{\partial}{\partial \bar{\Phi}}\sum_{t=p+1}^{T}\left(\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\bar{\boldsymbol{C}}_{t-1}'\bar{\Phi}'\Sigma_{\bar{\boldsymbol{v}}}^{-1}\bar{\Phi}\bar{\boldsymbol{C}}_{t-1} - \boldsymbol{C}_t'\Sigma_{\bar{\boldsymbol{v}}}^{-1}\bar{\Phi}\bar{\boldsymbol{C}}_{t-1} - \bar{\boldsymbol{C}}_{t-1}'\bar{\Phi}'\Sigma_{\bar{\boldsymbol{v}}}^{-1}\boldsymbol{C}_t|X,Y\right]\right)
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{2}\frac{\partial}{\partial\bar{\Phi}}\ tr\left(\bar{\Phi}'\Sigma_{\bar{\boldsymbol{v}}}^{-1}\bar{\Phi}\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\bar{\boldsymbol{C}}_{t-1}\bar{\boldsymbol{C}}_{t-1}'|X,Y\right]\right)\\
&+\frac{1}{2}\frac{\partial}{\partial\bar{\Phi}}\ tr\left(\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\bar{\boldsymbol{C}}_{t-1}\boldsymbol{C}_t'|X,Y\right]\Sigma_{\bar{\boldsymbol{v}}}^{-1}\bar{\Phi}\right)\\
&+\frac{1}{2}\frac{\partial}{\partial\bar{\Phi}}\ tr\left(\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t\bar{\boldsymbol{C}}_{t-1}'|X,Y\right]\bar{\Phi}'\Sigma_{\bar{\boldsymbol{v}}}^{-1}\right)\\
&= -\frac{1}{2}\Sigma_{\bar{\boldsymbol{v}}}^{-1}\bar{\Phi}\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\bar{\boldsymbol{C}}_{t-1}\bar{\boldsymbol{C}}_{t-1}'|X,Y\right]\cdot 2+\frac{1}{2}\Sigma_{\bar{\boldsymbol{v}}}^{-1}\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t\bar{\boldsymbol{C}}_{t-1}'|X,Y\right]\\
&+\frac{1}{2}\Sigma_{\bar{\boldsymbol{v}}}^{-1}\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t\bar{\boldsymbol{C}}_{t-1}'|X,Y\right]\\
&= \Sigma_{\bar{\boldsymbol{v}}}^{-1}\left(-\bar{\Phi}\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\bar{\boldsymbol{C}}_{t-1}\bar{\boldsymbol{C}}_{t-1}'|X,Y\right]+\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t\bar{\boldsymbol{C}}_{t-1}'|X,Y\right]\right)\\
&= \Sigma_{\bar{\boldsymbol{v}}}^{-1}(T-p)\left(-\bar{\Phi}\mathcal{E}+\mathcal{D}\right),
\end{aligned}
$$

where the last step replaced the sums by matrices $\mathcal{D}$ and $\mathcal{E}$ in (5.27). By setting these partial derivatives equal to the zero matrix and solving it for the parameter $\bar{\Phi}$, it follows:

$$\bar{\Phi} = \mathcal{D}\mathcal{E}^{-1}. \tag{5.31}$$

which confirms (5.25) for loop $(l+1)$ and is in line with Shumway and Stoffer (1982, p. 257, Eq. 12).

Finally, the estimator of $\Sigma_{\boldsymbol{v}}^{yy}$ is left. Before we determine the partial derivatives of $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,C\right)|X,Y\right]$ with respect to $\Sigma_{\boldsymbol{v}}^{yy}$, we rewrite the inverse of the covariance matrix $\Sigma_{\bar{\boldsymbol{v}}}$ as follows:

$$
\begin{aligned}
\Sigma_{\bar{\boldsymbol{v}}}^{-1} &= \begin{bmatrix} I_K & O_{K\times M}\\ O_{M\times K} & (\Sigma_{\boldsymbol{v}}^{yy})^{-1}\end{bmatrix} = \begin{bmatrix} I_K & O_{K\times M}\\ O_{M\times K} & O_{M\times M}\end{bmatrix} + \begin{bmatrix} O_{K\times K} & O_{K\times M}\\ O_{M\times K} & (\Sigma_{\boldsymbol{v}}^{yy})^{-1}\end{bmatrix}\\
&= \begin{bmatrix} I_K & O_{K\times M}\\ O_{M\times K} & O_{M\times M}\end{bmatrix} + \begin{bmatrix} O_{K\times M}\\ I_M\end{bmatrix}(\Sigma_{\boldsymbol{v}}^{yy})^{-1}\begin{bmatrix} O_{M\times K} & I_M\end{bmatrix}.
\end{aligned}
\tag{5.32}
$$

Based on $\bar{\Phi}$ in (5.31) as well as Lemmata A.1.5 and A.1.8, we get:

$$
\begin{aligned}
&\frac{\partial\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,C\right)|X,Y\right]}{\partial\Sigma_{\boldsymbol{v}}^{yy}}\\
&= -\frac{T-p}{2}\frac{\partial}{\partial\Sigma_{\boldsymbol{v}}^{yy}}\ ln\left(|\Sigma_{\boldsymbol{v}}^{yy}|\right)-\frac{1}{2}\frac{\partial}{\partial\Sigma_{\boldsymbol{v}}^{yy}}\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\left(\boldsymbol{C}_t-\bar{\Phi}\bar{\boldsymbol{C}}_{t-1}\right)'\Sigma_{\bar{\boldsymbol{v}}}^{-1}\left(\boldsymbol{C}_t-\bar{\Phi}\bar{\boldsymbol{C}}_{t-1}\right)|X,Y\right]\\
&= -\frac{T-p}{2}(\Sigma_{\boldsymbol{v}}^{yy})^{-1}-\frac{1}{2}\frac{\partial}{\partial\Sigma_{\boldsymbol{v}}^{yy}}\ tr\left(\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\left(\boldsymbol{C}_t-\bar{\Phi}\bar{\boldsymbol{C}}_{t-1}\right)\left(\boldsymbol{C}_t-\bar{\Phi}\bar{\boldsymbol{C}}_{t-1}\right)'|X,Y\right]\Sigma_{\bar{\boldsymbol{v}}}^{-1}\right)\\
&= -\frac{T-p}{2}(\Sigma_{\boldsymbol{v}}^{yy})^{-1}-\frac{1}{2}\frac{\partial}{\partial\Sigma_{\boldsymbol{v}}^{yy}}\ tr\Bigg((T-p)\left(\mathcal{F}-\mathcal{D}\bar{\Phi}'-\bar{\Phi}\mathcal{D}'+\bar{\Phi}\mathcal{E}\bar{\Phi}'\right)\\
&\qquad\cdot\left(\begin{bmatrix} I_K & O_{K\times M}\\ O_{M\times K} & O_{M\times M}\end{bmatrix}+\begin{bmatrix} O_{K\times M}\\ I_M\end{bmatrix}(\Sigma_{\boldsymbol{v}}^{yy})^{-1}\begin{bmatrix} O_{M\times K} & I_M\end{bmatrix}\right)\Bigg)\\
&= -\frac{T-p}{2}(\Sigma_{\boldsymbol{v}}^{yy})^{-1}-\frac{T-p}{2}\frac{\partial}{\partial\Sigma_{\boldsymbol{v}}^{yy}}\ tr\left(\left(\mathcal{F}-\mathcal{D}\bar{\Phi}'-\bar{\Phi}\mathcal{D}'+\bar{\Phi}\mathcal{E}\bar{\Phi}'\right)\begin{bmatrix} O_{K\times M}\\ I_M\end{bmatrix}(\Sigma_{\boldsymbol{v}}^{yy})^{-1}\begin{bmatrix} O_{M\times K} & I_M\end{bmatrix}\right)\\
&= -\frac{T-p}{2}(\Sigma_{\boldsymbol{v}}^{yy})^{-1}-\frac{T-p}{2}\frac{\partial}{\partial\Sigma_{\boldsymbol{v}}^{yy}}\ tr\left(\begin{bmatrix} O_{M\times K} & I_M\end{bmatrix}\left(\mathcal{F}-\mathcal{D}\bar{\Phi}'-\bar{\Phi}\mathcal{D}'+\bar{\Phi}\mathcal{E}\bar{\Phi}'\right)\begin{bmatrix} O_{K\times M}\\ I_M\end{bmatrix}(\Sigma_{\boldsymbol{v}}^{yy})^{-1}\right)
\end{aligned}
$$

$$= -\frac{T-p}{2}\left(\Sigma_{\boldsymbol{v}}^{yy}\right)^{-1} + \frac{T-p}{2}\left(\left(\Sigma_{\boldsymbol{v}}^{yy}\right)^{-1}\begin{bmatrix}O_{M\times K} & I_M\end{bmatrix}\left(\mathcal{F}-\mathcal{D}\bar{\Phi}'-\bar{\Phi}\mathcal{D}'+\bar{\Phi}\mathcal{E}\bar{\Phi}'\right)\begin{bmatrix}O_{K\times M}\\ I_M\end{bmatrix}\left(\Sigma_{\boldsymbol{v}}^{yy}\right)^{-1}\right)',$$

where the third line substitutes the sum of the conditional expectations by matrices $\mathcal{D}$, $\mathcal{E}$ and $\mathcal{F}$ in (5.27). Searching for the zeros of the corresponding equation system provides for the parameter $\Sigma_{\boldsymbol{v}}^{yy}$:

$$\Sigma_{\boldsymbol{v}}^{yy} = \begin{bmatrix}O_{M\times K} & I_M\end{bmatrix}\left(\mathcal{F}-\mathcal{D}\bar{\Phi}'-\bar{\Phi}\mathcal{D}'+\bar{\Phi}\mathcal{E}\bar{\Phi}'\right)\begin{bmatrix}O_{K\times M}\\ I_M\end{bmatrix},$$

which justifies the solution in (5.26) for loop $(l+1)$ and $\bar{\Phi}_{(l+1)}$ as the estimator of the coefficient matrix in (5.25). Note, the term $\left(\mathcal{F}-\mathcal{D}\bar{\Phi}'-\bar{\Phi}\mathcal{D}'+\bar{\Phi}\mathcal{E}\bar{\Phi}'\right)$ in the above equation is in accordance with Shumway and Stoffer (1982, p. 257, Eq. 13). □

In Lemma 5.1.7, the matrix H_Λ is required to have full row rank P with $P \leq N(K+M)$. This condition guarantees that the inverse matrix $\left(H_\Lambda\left(\mathcal{B}^{-1}\otimes\Sigma_{\boldsymbol{e}(l)}\right)H_\Lambda'\right)^{-1} \in \mathbb{R}^{P\times P}$ is well-defined. Furthermore, this assumption especially prohibits to choose a larger matrix $H_\Lambda \in \mathbb{R}^{N(K+M)\times N(K+M)}$ with zero rows for all unrestricted parameters of the loadings matrix.

By similar reasoning, we receive an EM for complete data admitting restrictions of the coefficient matrices in (5.17). As before, let the matrix $H_\Phi \in \mathbb{R}^{Q\times p(K+M)^2}$ have full row rank $Q \leq p(K+M)^2$ implying that the inverse matrix $\left(H_\Phi\left(\mathcal{E}^{-1}\otimes\Sigma_{\bar{\boldsymbol{v}}(l)}\right)H_\Phi'\right)^{-1} \in \mathbb{R}^{Q\times Q}$ in Lemma 5.1.8 is well-defined. Then, it holds:

Lemma 5.1.8 (EM for Complete FAVARs with Restricted VAR Coefficients)
For complete panel data $X \in \mathbb{R}^{T\times N}$ and variables $Y \in \mathbb{R}^{T\times M}$, assume the FAVAR in (5.16)-(5.17) with linearly constrained coefficients $\bar{\Phi} = \left[\bar{\Phi}_1, ..., \bar{\Phi}_p\right] \in \mathbb{R}^{(K+M)\times p(K+M)}$, i.e.: $H_\Phi vec\left(\bar{\Phi}\right) = \boldsymbol{\kappa}_\Phi$ for vector $\boldsymbol{\kappa}_\Phi \in \mathbb{R}^Q$ and matrix $H_\Phi \in \mathbb{R}^{Q\times p(K+M)^2}$ of full row rank $Q \leq p(K+M)^2$. If (l) stands for the current EM loop, the parameters of the next loop $(l+1)$ are updated in the following way:

$$\bar{\Lambda}_{(l+1)} = \mathcal{A}\mathcal{B}^{-1}, \tag{5.33}$$

$$\Sigma_{\boldsymbol{e}(l+1)} = \mathcal{C} - \mathcal{A}\bar{\Lambda}_{(l+1)}' - \bar{\Lambda}_{(l+1)}\mathcal{A}' + \bar{\Lambda}_{(l+1)}\mathcal{B}\bar{\Lambda}_{(l+1)}', \tag{5.34}$$

$$\begin{aligned} vec\left(\bar{\Phi}_{(l+1)}\right) = &\left(\mathcal{E}^{-1}\otimes\Sigma_{\bar{\boldsymbol{v}}(l)}\right)H_\Phi'\left(H_\Phi\left(\mathcal{E}^{-1}\otimes\Sigma_{\bar{\boldsymbol{v}}(l)}\right)H_\Phi'\right)^{-1}\left(\boldsymbol{\kappa}_{\bar{\Phi}} - H_\Phi\ vec\left(\mathcal{D}\mathcal{E}^{-1}\right)\right)\\ &+ vec\left(\mathcal{D}\mathcal{E}^{-1}\right), \end{aligned} \tag{5.35}$$

$$\Sigma_{\boldsymbol{v}(l+1)}^{yy} = \begin{bmatrix}O_{M\times K} & I_M\end{bmatrix}\left(\mathcal{F}-\mathcal{D}\bar{\Phi}_{(l+1)}'-\bar{\Phi}_{(l+1)}\mathcal{D}'+\bar{\Phi}_{(l+1)}\mathcal{E}\bar{\Phi}_{(l+1)}'\right)\begin{bmatrix}O_{K\times M}\\ I_M\end{bmatrix}, \tag{5.36}$$

where the matrices $\mathcal{A} \in \mathbb{R}^{N\times(K+M)}$, $\mathcal{B} \in \mathbb{R}^{(K+M)\times(K+M)}$, $\mathcal{C} \in \mathbb{R}^{N\times N}$, $\mathcal{D} \in \mathbb{R}^{(K+M)\times p(K+M)}$, $\mathcal{E} \in \mathbb{R}^{p(K+M)\times p(K+M)}$ and $\mathcal{F} \in \mathbb{R}^{(K+M)\times(K+M)}$ are defined in (5.27). For the state vector $\bar{\boldsymbol{C}}_t \in \mathbb{R}^{p(K+M)}$, we have $\bar{\boldsymbol{C}}_t = \left[\boldsymbol{C}_t', ..., \boldsymbol{C}_{t-p+1}'\right]'$.

Proof:
We apply the method of Lagrange multipliers to the subsequent optimization problem:

$$\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,C\right)|X,Y\right] + \boldsymbol{\lambda}'\left(H_\Phi vec\left(\bar{\Phi}\right) - \boldsymbol{\kappa}_\Phi\right), \tag{5.37}$$

with $\boldsymbol{\lambda}$ as Lagrange multiplier. Then, Lemmata A.1.5 and A.1.8 provide for the partial derivative of (5.37) with respect to $\bar{\Lambda}$:

$$\frac{\partial\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,C\right)|X,Y\right]}{\partial\bar{\Lambda}}$$

$$
\begin{aligned}
&= \frac{\partial}{\partial \bar{\Lambda}} \frac{1}{2} \sum_{t=1}^{T} tr\left(\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right] \boldsymbol{X}_t' \Sigma_{\boldsymbol{e}}^{-1} \bar{\Lambda}\right) + \frac{\partial}{\partial \bar{\Lambda}} \frac{1}{2} \sum_{t=1}^{T} tr\left(\Sigma_{\boldsymbol{e}}^{-1} \boldsymbol{X}_t \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]' \bar{\Lambda}'\right) \\
&\quad - \frac{\partial}{\partial \bar{\Lambda}} \frac{1}{2} \sum_{t=1}^{T} tr\left(\bar{\Lambda}' \Sigma_{\boldsymbol{e}}^{-1} \bar{\Lambda} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t \boldsymbol{C}_t'|X,Y\right]\right) \\
&= \frac{1}{2}\left(\sum_{t=1}^{T} \Sigma_{\boldsymbol{e}}^{-1} \boldsymbol{X}_t \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]' + \sum_{t=1}^{T} \Sigma_{\boldsymbol{e}}^{-1} \boldsymbol{X}_t \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]' - 2\sum_{t=1}^{T} \Sigma_{\boldsymbol{e}}^{-1} \bar{\Lambda} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t \boldsymbol{C}_t'|X,Y\right]\right) \\
&= \Sigma_{\boldsymbol{e}}^{-1} \sum_{t=1}^{T} \boldsymbol{X}_t \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t|X,Y\right]' - \Sigma_{\boldsymbol{e}}^{-1} \bar{\Lambda} \sum_{t=1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t \boldsymbol{C}_t'|X,Y\right].
\end{aligned}
$$

Finally, the definitions of matrices $\mathcal{A}$ and $\mathcal{B}$ and searching for the zeros of the previous equation system yields the solution of $\bar{\Lambda}$ for loop $(l+1)$ in (5.33). The partial derivatives of the target functions in (5.28) and (5.37), respectively, with respect to the matrix $\Sigma_{\boldsymbol{e}}$ coincide, as the additional terms with the Lagrange multipliers do not depend on $\Sigma_{\boldsymbol{e}}$. Hence, the solutions of $\Sigma_{\boldsymbol{e}(l+1)}$ in (5.24) and (5.34) are the same. The same argument remains valid for the solutions of $\Sigma_{\boldsymbol{v}(l+1)}^{yy}$ in (5.26) and (5.36).

For the partial derivatives of (5.37) with respect to matrix $\bar{\Phi}$, Lemmata A.1.5 and A.1.8 result in:

$$
\begin{aligned}
&\frac{\partial\left(\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\Theta|X,C\right)|X,Y\right] + \boldsymbol{\lambda}'\left(H_{\Phi} vec\left(\bar{\Phi}\right) - \boldsymbol{\kappa}_{\Phi}\right)\right)}{\partial \bar{\Phi}} \\
&= -\frac{1}{2}\frac{\partial}{\partial \bar{\Phi}} \sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\left(\boldsymbol{C}_t - \bar{\Phi}\bar{\boldsymbol{C}}_{t-1}\right)' \Sigma_{\bar{\boldsymbol{v}}}^{-1}\left(\boldsymbol{C}_t - \bar{\Phi}\bar{\boldsymbol{C}}_{t-1}\right)|X,Y\right] + \frac{\partial}{\partial \bar{\Phi}} \boldsymbol{\lambda}'\left(H_{\Phi} vec\left(\bar{\Phi}\right) - \boldsymbol{\kappa}_{\Phi}\right) \\
&= -\frac{1}{2}\frac{\partial}{\partial \bar{\Phi}} \sum_{t=p+1}^{T} \left(\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\bar{\boldsymbol{C}}_{t-1}' \bar{\Phi}' \Sigma_{\bar{\boldsymbol{v}}}^{-1} \bar{\Phi} \bar{\boldsymbol{C}}_{t-1} - \boldsymbol{C}_t' \Sigma_{\bar{\boldsymbol{v}}}^{-1} \bar{\Phi} \bar{\boldsymbol{C}}_{t-1} - \bar{\boldsymbol{C}}_{t-1}' \bar{\Phi}' \Sigma_{\bar{\boldsymbol{v}}}^{-1} \boldsymbol{C}_t|X,Y\right]\right) + \frac{\partial}{\partial \bar{\Phi}} \boldsymbol{\lambda}' H_{\Phi} vec\left(\bar{\Phi}\right) \\
&= -\frac{1}{2}\frac{\partial}{\partial \bar{\Phi}}\, tr\left(\bar{\Phi}' \Sigma_{\bar{\boldsymbol{v}}}^{-1} \bar{\Phi} \sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\bar{\boldsymbol{C}}_{t-1} \bar{\boldsymbol{C}}_{t-1}'|X,Y\right]\right) + \frac{1}{2}\frac{\partial}{\partial \bar{\Phi}}\, tr\left(\sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\bar{\boldsymbol{C}}_{t-1} \boldsymbol{C}_t'|X,Y\right] \Sigma_{\bar{\boldsymbol{v}}}^{-1} \bar{\Phi}\right) \\
&\quad + \frac{1}{2}\frac{\partial}{\partial \bar{\Phi}}\, tr\left(\sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t \bar{\boldsymbol{C}}_{t-1}'|X,Y\right] \bar{\Phi}' \Sigma_{\bar{\boldsymbol{v}}}^{-1}\right) + \frac{\partial}{\partial \bar{\Phi}} \boldsymbol{\lambda}' H_{\Phi} vec\left(\bar{\Phi}\right) \\
&= -\frac{1}{2}\Sigma_{\bar{\boldsymbol{v}}}^{-1} \bar{\Phi} \sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\bar{\boldsymbol{C}}_{t-1} \bar{\boldsymbol{C}}_{t-1}'|X,Y\right] \cdot 2 + \frac{1}{2}\Sigma_{\bar{\boldsymbol{v}}}^{-1} \sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t \bar{\boldsymbol{C}}_{t-1}'|X,Y\right] \\
&\quad + \frac{1}{2}\Sigma_{\bar{\boldsymbol{v}}}^{-1} \sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t \bar{\boldsymbol{C}}_{t-1}'|X,Y\right] + \begin{bmatrix} \left(\boldsymbol{\lambda}' H_{\Phi}\right)_1 & \cdots & \left(\boldsymbol{\lambda}' H_{\Phi}\right)_{(p(K+M)-1)(K+M)+1} \\ \vdots & & \vdots \\ \left(\boldsymbol{\lambda}' H_{\Phi}\right)_{K+M} & \cdots & \left(\boldsymbol{\lambda}' H_{\Phi}\right)_{p(K+M)^2} \end{bmatrix} \\
&= \Sigma_{\bar{\boldsymbol{v}}}^{-1}\left(-\bar{\Phi} \sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\bar{\boldsymbol{C}}_{t-1} \bar{\boldsymbol{C}}_{t-1}'|X,Y\right] + \sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}}\left[\boldsymbol{C}_t \bar{\boldsymbol{C}}_{t-1}'|X,Y\right]\right) \\
&\quad + \begin{bmatrix} \left(\boldsymbol{\lambda}' H_{\Phi}\right)_1 & \cdots & \left(\boldsymbol{\lambda}' H_{\Phi}\right)_{(p(K+M)-1)(K+M)+1} \\ \vdots & & \vdots \\ \left(\boldsymbol{\lambda}' H_{\Phi}\right)_{K+M} & \cdots & \left(\boldsymbol{\lambda}' H_{\Phi}\right)_{p(K+M)^2} \end{bmatrix}.
\end{aligned}
$$

After the sums in the previous step have been replaced by matrices $\mathcal{D}$ and $\mathcal{E}$ in (5.27), we obtain for the zeros of the arising equation system:

$$
\bar{\Phi}\left(T-p\right)\mathcal{E} = \left(T-p\right)\mathcal{D} + \Sigma_{\bar{\boldsymbol{v}}} \begin{bmatrix} \left(\boldsymbol{\lambda}' H_{\Phi}\right)_1 & \cdots & \left(\boldsymbol{\lambda}' H_{\Phi}\right)_{(p(K+M)-1)(K+M)+1} \\ \vdots & & \vdots \\ \left(\boldsymbol{\lambda}' H_{\Phi}\right)_{K+M} & \cdots & \left(\boldsymbol{\lambda}' H_{\Phi}\right)_{p(K+M)^2} \end{bmatrix}
$$

Next, we apply the $vec(\cdot)$ operator to both sides and benefit from Lemma A.1.13 as follows:

$$
\begin{aligned}
(T-p)\ vec\left(\bar{\Phi}\mathcal{E}\right) &= (T-p)\ vec\left(\mathcal{D}\right) + vec\left(\Sigma_{\bar{\boldsymbol{v}}}\begin{bmatrix} \left(\boldsymbol{\lambda}' H_\Phi\right)_1 & \cdots & \left(\boldsymbol{\lambda}' H_\Phi\right)_{(p(K+M)-1)(K+M)+1} \\ \vdots & & \vdots \\ \left(\boldsymbol{\lambda}' H_\Phi\right)_{K+M} & \cdots & \left(\boldsymbol{\lambda}' H_\Phi\right)_{p(K+M)^2} \end{bmatrix}\right) \\
\Leftrightarrow (T-p)\left(\mathcal{E}\otimes I_{K+M}\right)\ vec\left(\bar{\Phi}\right) &= (T-p)\ vec\left(\mathcal{D}\right) + \left(I_{p(K+M)}\otimes\Sigma_{\bar{\boldsymbol{v}}}\right) H_\Phi'\boldsymbol{\lambda} \\
\Leftrightarrow\ vec\left(\bar{\Phi}\right) &= \left(\mathcal{E}\otimes I_{K+M}\right)^{-1}\ vec\left(\mathcal{D}\right) + \frac{1}{T-p}\left(\mathcal{E}\otimes I_{K+M}\right)^{-1}\left(I_{p(K+M)}\otimes\Sigma_{\bar{\boldsymbol{v}}}\right) H_\Phi'\boldsymbol{\lambda} \\
\Leftrightarrow\ vec\left(\bar{\Phi}\right) &= \left(\mathcal{E}^{-1}\otimes I_{K+M}\right)\ vec\left(\mathcal{D}\right) + \frac{1}{T-p}\left(\mathcal{E}^{-1}\otimes\Sigma_{\bar{\boldsymbol{v}}}\right) H_\Phi'\boldsymbol{\lambda}, \qquad (5.38)
\end{aligned}
$$

where the last transformation relies on the properties of the Kronecker product in Lemma A.1.10. With the restrictions of the coefficient matrices in mind, we have:

$$H_\Phi\ vec\left(\bar{\Phi}\right) = \boldsymbol{\kappa}_\Phi = H_\Phi\left(\mathcal{E}^{-1}\otimes I_{K+M}\right)\ vec\left(\mathcal{D}\right) + \frac{1}{T-p} H_\Phi\left(\mathcal{E}^{-1}\otimes\Sigma_{\bar{\boldsymbol{v}}}\right) H_\Phi'\boldsymbol{\lambda}.$$

Hence, we obtain for the Lagrange multiplier $\boldsymbol{\lambda}$:

$$\boldsymbol{\lambda} = (T-p)\left(H_\Phi\left(\mathcal{E}^{-1}\otimes\Sigma_{\bar{\boldsymbol{v}}}\right) H_\Phi'\right)^{-1}\left(\boldsymbol{\kappa}_\Phi - H_\Phi\left(\mathcal{E}^{-1}\otimes I_{K+M}\right)\ vec\left(\mathcal{D}\right)\right). \qquad (5.39)$$

Then, we insert (5.39) in (5.38) and get:

$$
\begin{aligned}
vec\left(\bar{\Phi}\right) = &\left(\mathcal{E}^{-1}\otimes I_{K+M}\right)\ vec\left(\mathcal{D}\right) + \left(\mathcal{E}^{-1}\otimes\Sigma_{\bar{\boldsymbol{v}}}\right) H_\Phi'\left(H_\Phi\left(\mathcal{E}^{-1}\otimes\Sigma_{\bar{\boldsymbol{v}}}\right) H_\Phi'\right)^{-1} \\
&\cdot\left(\boldsymbol{\kappa}_\Phi - H_\Phi\left(\mathcal{E}^{-1}\otimes I_{K+M}\right)\ vec\left(\mathcal{D}\right)\right).
\end{aligned}
$$

Eventually, we apply Lemma A.1.13 to the above solution and receive for loop $(l+1)$ the representation in (5.35). □

A comparison of Lemmata 5.1.7 and 5.1.8 shows that the inlcusion of linear parameter constraints follows the same procedure irrespective of whether the loadings matrix or the coefficient matrices are considered. Without a lot of effort we can also derive an EM for complete FAVARs with restricted loadings and VAR coefficient matrices. However, the restricted solutions for $\bar{\Lambda}$ in (5.23) and $\bar{\Phi}$ in (5.35) are more complicated than the unrestricted analogs in (5.33) and (5.25). This is why we propose to take either loadings or VAR coefficient restrictions into account, but not both at the same time. Moreover, if the loadings matrix and VAR coefficients are constrained together, one has to make sure that those do not exclude each other. In the sequel, we focus on the loadings constraints in (5.15) to compare our results and the ones in Bernanke et al. (2005) and Bork (2009), respectively, in Section 5.2.

The loadings constraints in (5.13) require a careful arrangement of $\boldsymbol{X}_t$, as the first entry of $\bar{\boldsymbol{F}}_t$ only affects the first element of $\boldsymbol{X}_t$. Similarly, the next $K-1$ components do. Therefore, Bernanke et al. (2005) and Bork (2009) performed a pre-analysis and chose the first K entries of $\boldsymbol{X}_t$ from the slow-moving variables, which were supposed not to simultaneously respond to a monetary policy shock. For instance, indicators for spending, wages and prices belong to them. In our empirical study, we follow Bork (2009) to sort the panel data. That is, we first run a usual PCA, before we determine for each times series in $\boldsymbol{X}_t$ and for each estimated factor the absolute value of their correlation coefficient. Finally, we select the K variables with the highest absolute value, sort them in descending order and take them as the first K entries of $\boldsymbol{X}_t$. If a fast-moving variable is the most correlated with a factor, we calculate for this and for each remaining slow-moving variable the absolute value of their correlation coefficient and choose the best slow-moving one instead.

For computational efficiency and to facilitate the implementation of our approach, we insert the parameter estimates from (5.23)-(5.26) in the conditional expectation of the log-likelihood function in Lemma 5.1.6 and summarize as many terms as possible. In total, we get the subsequent formulation.

Lemma 5.1.9 (Simple Formulation of Expected FAVAR Log-Likelihood)
For the MLE parameters in (5.23)-(5.26) *a simplified version of the conditional expectation of the FAVAR log-likelihood function in* (5.21) *is given by:*

$$\begin{aligned}
&\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l)}|X,C\right)|X,Y\right]\\
&\quad= -\frac{1}{2}\left(TN+(T-p)\left(K+M\right)\right)\ ln\left(2\pi\right) - \frac{1}{2}\left(T\ ln\left(\left|\Sigma_{\boldsymbol{e}(l)}\right|\right) + (T-p)\left(M+\ ln\left(\left|\Sigma^{yy}_{\boldsymbol{v}(l)}\right|\right)\right)\right)\\
&\quad\quad - \frac{TN}{2} - \frac{1}{2}\left(T-p\right)\ tr\left(\left(\mathcal{F}-\mathcal{D}\bar{\Phi}'-\bar{\Phi}\mathcal{D}'+\bar{\Phi}\mathcal{E}\bar{\Phi}'\right)\begin{bmatrix} I_K & O_{K\times M}\\ O_{M\times K} & O_{M\times M}\end{bmatrix}\right),
\end{aligned} \tag{5.40}$$

with matrices $\mathcal{D}\in\mathbb{R}^{(K+M)\times p(K+M)}, \mathcal{E}\in\mathbb{R}^{p(K+M)\times p(K+M)}$ *and* $\mathcal{F}\in\mathbb{R}^{(K+M)\times(K+M)}$ *as in* (5.27).

Proof:
The MLE parameters in (5.23)-(5.26) and matrices $\mathcal{A}-\mathcal{F}$ in (5.27) provide:

$$\begin{aligned}
\mathbb{E}_{\hat{\Theta}_{(l)}}\Big[\mathcal{L}\left(\hat{\Theta}_{(l)}|X,C\right)|X,Y\Big] &= -\frac{1}{2}\left(TN+(T-p)\left(K+M\right)\right) ln\left(2\pi\right) - \frac{T}{2} ln\left(\left|\Sigma_{\boldsymbol{e}(l)}\right|\right) - \frac{T-p}{2} ln\left(\left|\Sigma^{yy}_{\boldsymbol{v}(l)}\right|\right)\\
&\quad -\frac{1}{2}\ tr\left(\Sigma^{-1}_{\boldsymbol{e}(l)}\sum_{t=1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\left(\boldsymbol{X}_t-\bar{\Lambda}_{(l)}\boldsymbol{C}_t\right)\left(\boldsymbol{X}_t-\bar{\Lambda}_{(l)}\boldsymbol{C}_t\right)'|X,Y\right]\right)\\
&\quad -\frac{1}{2}\ tr\left(\Sigma^{-1}_{\bar{\boldsymbol{v}}(l)}\sum_{t=p+1}^{T}\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\left(\boldsymbol{C}_t-\bar{\Phi}_{(l)}\bar{\boldsymbol{C}}_{t-1}\right)\left(\boldsymbol{C}_t-\bar{\Phi}_{(l)}\bar{\boldsymbol{C}}_{t-1}\right)'|X,Y\right]\right)\\
&= -\frac{1}{2}\left(TN+(T-p)\left(K+M\right)\right) ln\left(2\pi\right) - \frac{T}{2} ln\left(\left|\Sigma_{\boldsymbol{e}(l)}\right|\right) - \frac{T-p}{2} ln\left(\left|\Sigma^{yy}_{\boldsymbol{v}(l)}\right|\right)\\
&\quad -\frac{1}{2}\ tr\left(\Sigma^{-1}_{\boldsymbol{e}(l)}\ T\left(\mathcal{C}-\mathcal{A}\bar{\Lambda}'-\bar{\Lambda}\mathcal{A}'+\bar{\Lambda}\mathcal{B}\bar{\Lambda}'\right)\right)\\
&\quad -\frac{1}{2}\ tr\left(\Sigma^{-1}_{\bar{\boldsymbol{v}}(l)}\left(T-p\right)\left(\mathcal{F}-\mathcal{D}\bar{\Phi}'-\bar{\Phi}\mathcal{D}'+\bar{\Phi}\mathcal{E}\bar{\Phi}'\right)\right).
\end{aligned}$$

Because of (5.24), the matrix $\left(\mathcal{C}-\mathcal{A}\bar{\Lambda}'-\bar{\Lambda}\mathcal{A}'+\bar{\Lambda}\mathcal{B}\bar{\Lambda}'\right)$ serves as estimator of $\Sigma_{\boldsymbol{e}(l)}$. Therefore, we have: $\Sigma^{-1}_{\boldsymbol{e}(l)}\left(\mathcal{C}-\mathcal{A}\bar{\Lambda}'-\bar{\Lambda}\mathcal{A}'+\bar{\Lambda}\mathcal{B}\bar{\Lambda}'\right)=I_N$. Using the matrix trace properties in Lemma A.1.2 and the decomposition of the inverse matrix $\Sigma^{-1}_{\bar{\boldsymbol{v}}}$ in (5.32) we proceed as follows:

$$\begin{aligned}
&\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l)}|X,C\right)|X,Y\right]\\
&\quad= -\frac{1}{2}\left(TN+(T-p)\left(K+M\right)\right) ln\left(2\pi\right) - \frac{T}{2} ln\left(\left|\Sigma_{\boldsymbol{e}(l)}\right|\right) - \frac{T-p}{2} ln\left(\left|\Sigma^{yy}_{\boldsymbol{v}(l)}\right|\right)\\
&\quad -\frac{TN}{2} - \frac{1}{2}\ tr\left(\begin{bmatrix} I_K & O_{K\times M}\\ O_{M\times K} & O_{M\times M}\end{bmatrix}\left(T-p\right)\left(\mathcal{F}-\mathcal{D}\bar{\Phi}'-\bar{\Phi}\mathcal{D}'+\bar{\Phi}\mathcal{E}\bar{\Phi}'\right)\right)\\
&\quad -\frac{1}{2}\ tr\left(\begin{bmatrix} O_{K\times M}\\ I_M\end{bmatrix}\left(\Sigma^{yy}_{\boldsymbol{v}(l)}\right)^{-1}\begin{bmatrix} O_{M\times K} & I_M\end{bmatrix}\left(T-p\right)\left(\mathcal{F}-\mathcal{D}\bar{\Phi}'-\bar{\Phi}\mathcal{D}'+\bar{\Phi}\mathcal{E}\bar{\Phi}'\right)\right).
\end{aligned}$$

Due to (5.36) and the matrix trace poperties in Lemma A.1.2, we obtain $-\frac{1}{2}\left(T-p\right)M$ for the last term of the above representation and so, prove the statement. □

For sufficiently long samples and certain regularity assumptions the sum of the conditional expectations converges to the covariance matrix $\Sigma_{\bar{\boldsymbol{v}}}$ with increasing T, that is:

$$\sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}} \left[\left(\boldsymbol{C}_t - \bar{\Phi}\bar{\boldsymbol{C}}_{t-1}\right) \left(\boldsymbol{C}_t - \bar{\Phi}\bar{\boldsymbol{C}}_{t-1}\right)' |X,Y\right] = \sum_{t=p+1}^{T} \mathbb{E}_{\hat{\Theta}_{(l)}} \left[\bar{\boldsymbol{v}}_t \bar{\boldsymbol{v}}_t' |X,Y\right] \xrightarrow{\text{large } T} (T-p)\,\Sigma_{\bar{\boldsymbol{v}}}.$$

In this case, the conditional expectation can be further simplified and is asymptotically equivalent to:

$$\begin{aligned}
&\mathbb{E}_{\hat{\Theta}_{(l)}} \left[\mathcal{L}\left(\hat{\Theta}_{(l)}|X,C\right)|X,Y\right] \\
&\approx -\frac{1}{2}\left(TN + (T-p)(K+M)\right) ln\left(2\pi\right) - \frac{1}{2}(T-p)\, tr\left(\begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & O_{M\times M}\end{bmatrix}\right) \\
&- \frac{1}{2}\left(T\left(N + ln\left(\left|\Sigma_{\boldsymbol{e}(l)}\right|\right)\right) + (T-p)\left(M + ln\left(\left|\Sigma^{yy}_{\boldsymbol{v}(l)}\right|\right)\right)\right) \\
&= -\frac{1}{2}\left(TN + (T-p)(K+M)\right) ln\left(2\pi\right) \\
&- \frac{1}{2}\left(T\left(N + ln\left(\left|\Sigma_{\boldsymbol{e}(l)}\right|\right)\right) + (T-p)\left(K + M + ln\left(\left|\Sigma^{yy}_{\boldsymbol{v}(l)}\right|\right)\right)\right).
\end{aligned}$$

As for the updates in Theorem 4.1.9, we must fix, when the EM in Lemma 5.1.7 stops. For this purpose, $\eta > 0$ is a prespecified error tolerance and $\mathbb{E}_{\hat{\Theta}_{(l+1)}} \left[\mathcal{L}\left(\hat{\Theta}_{(l+1)}|X,C\right)|X,Y\right]$ is the expected log-likelihood function from (5.40) based on the MLE parameters in (5.23)-(5.26) of loop (l+1). Then, the EM terminates as soon as it holds:

$$\frac{abs\left(\mathbb{E}_{\hat{\Theta}_{(l+1)}} \left[\mathcal{L}\left(\hat{\Theta}_{(l+1)}|X,C\right)|X,Y\right] - \mathbb{E}_{\hat{\Theta}_{(l)}} \left[\mathcal{L}\left(\hat{\Theta}_{(l)}|X,C\right)|X,Y\right]\right)}{\frac{1}{2}\left(abs\left(\mathbb{E}_{\hat{\Theta}_{(l+1)}} \left[\mathcal{L}\left(\hat{\Theta}_{(l+1)}|X,C\right)|X,Y\right]\right) + abs\left(\mathbb{E}_{\hat{\Theta}_{(l)}} \left[\mathcal{L}\left(\hat{\Theta}_{(l)}|X,C\right)|X,Y\right]\right)\right)} < \eta. \qquad (5.41)$$

That is to say, the EM stops as soon as the absolute value of the relative change in $\mathbb{E}_{\Theta}\left[\mathcal{L}\left(\Theta|X,C\right)|X,Y\right]$ between two successive iterations falls below the tolerance η. Hence, the termination criterion in (5.41) is similar to the one in (4.14) and the same embedding in the literature applies. Note that it would also be possible to work with a termination criterion checking, whether the absolute value of the relative change in the parameters Θ between two consecutive loops is smaller than η. For the initialization of the EM in Lemma 5.1.7, we reuse the two-step principal component approach of Bernanke et al. (2005), which is in accordance with Bork (2009) and Boivin et al. (2010).

The formulation and estimation of the FAVAR in (5.16)-(5.17) with loadings constraints in (5.15) require knowledge of the factor dimension K and lag order p. In empirical analyses, both are unknown and so, must be determined. Here, we adapt the usual AIC in the sense that we replace the log-likelihood function in (5.20) by the expected log-likelihood function in (5.40).

Lemma 5.1.10 (Selection of FAVARs with Loadings Restrictions)
For the FAVAR in (5.16)-(5.17) *with loadings constraints in* (5.15)*, let* $1 \leq \bar{p}$ *and* $1 \leq \bar{K}$ *be upper limits of the lag length and factor dimension, respectively, to be analyzed. Furthermore, let* $\hat{\Theta}$ *denote the estimated model parameters. Then, we choose the pair* (p^*, K^*)*, which obeys:*

$$\begin{aligned}
(p^*, K^*) = \underset{\substack{1\leq p\leq \bar{p} \\ 1\leq K\leq \bar{K}}}{\arg\min} &\left\{-2\,\mathbb{E}_{\hat{\Theta}} \left[\mathcal{L}\left(\hat{\Theta}|X,C\right)|X,Y\right] + 2N(K+M) + N(N+1)\right. \\
&\left. +2p(K+M)^2 + M(M+1) - K(K-1)\right\},
\end{aligned}$$

where the expected log-likelihood function $\mathbb{E}_{\hat{\Theta}} \left[\mathcal{L}\left(\hat{\Theta}|X,C\right)|X,Y\right]$ *is defined in* (5.40).

Proof:
The usual AIC satisfies (Akaike, 1987):

$$(q^*, K^*) = \underset{(q,K)}{\arg\min} \left\{ -2\tilde{\mathcal{L}}\,(q, K) + 2\,(\text{ number of estimated parameters })\right\},$$

with $\tilde{\mathcal{L}}\,(q, K)$ as the maximized log-likelihood function based on the estimated FAVAR parameters. Here, we replace $\tilde{\mathcal{L}}\,(q, K)$ by $\mathbb{E}_{\hat{\Theta}}\left[\mathcal{L}\left(\hat{\Theta}|X, C\right)|X, Y\right]$ to remove the factors $\boldsymbol{F}_t$ from the log-likelihood function. The constraints in (5.15) decrease the $N\,(K + M)$ degrees of freedom of the unrestricted loadings matrix by $K(K-1)/2$. In addition, the covariance matrix $\Sigma_{\boldsymbol{e}}$ has $N\,(N+1)\,/2$ parameters and the coefficent matrix $\bar{\Phi}$ comprises $p\,(K+M)^2$ parameters. Finally, the covariance matrix $\Sigma_{\bar{\boldsymbol{v}}}$ provides $M\,(M+1)\,/2$ degrees of freedom in the form of the matrix $\Sigma_{\boldsymbol{v}}^{yy}$. □

As an alternative to the AIC in Lemma 5.1.10, the information criteria in Bai and Ng (2002, 2008b) or Hallin and Liška (2007) could be deployed for model selection. However, for clarity reasons, we proceed with the presented AIC approach. As in previous sections, we summarize all steps in Algorithm 5.1.1.

Algorithm 5.1.1: Estimate FAVARs with loadings constraints in (5.15) based on complete data

```
Set relative termination criterion η > 0;
Define upper limits of factor dimension K̄ ≥ 1 and lag order p̄ ≥ 1;
Initialize overall parameter set Θ̂_ov = ∅;
Initialize overall AIC by AIC_ov = ∞ (or any sufficiently large number);
for K = 1 to K̄ do
    for p = 1 to p̄ do
        Initialize model parameters using PCA and OLS regression;
        Run EM in Lemma 5.1.7 with η, store (K, p) and estimated parameters Θ̂;
        Determine temporary AIC, i.e., AIC_tmp, using Lemma 5.1.10;
        if AIC_tmp < AIC_ov then
            Renew overall AIC value by AIC_ov = AIC_tmp;
            Update overall parameter set by Θ̂_ov = Θ̂;
        end
    end
end
```

5.1.3 Kalman Filter for FAVARs

The need for loadings constraints in Bork (2009) and Marcellino and Sivec (2016) differs from ours. Here, they are more or less arbitrary, since we are looking for $K(K-1)/2$ identification restrictions. Instead of fixing the loadings structure, we can use these degrees of freedom in Lemma 5.1.8 to change the shape of the coefficient matrix $\bar{\Phi}$. By contrast, Bork (2009) and Marcellino and Sivec (2016) must keep some of their loadings restrictions to estimate FAVARs. They actually considered ADFMs, however, the special loadings matrix, properly sorted panel data and the fact that the observable variables $\boldsymbol{Y}_t$ were a subset of the panel data $\boldsymbol{X}_t$ caused their ADFMs to align itself to the FAVAR in Bernanke et al. (2005). Thereby, both papers applied the standard Kalman Filter and Smoother for estimating the moments of the latent factors. From our point of view, the standard KF does not take into account that the variables $\boldsymbol{Y}_t$ are observed. Therefore, we derive KF and KS equations for the FAVAR in (5.16)-(5.17) that explicitly treat the observability of $\boldsymbol{Y}_t$. For this purpose, we convert the original VAR(p) into a VAR(1).

Lemma 5.1.11 (VAR(1) Formulation of FAVARs)
The FAVAR in (5.16)-(5.17) with autoregressive order $p > 0$ can be rewritten as:

$$\boldsymbol{X}_t = \begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{F}}_t \\ \boldsymbol{Y}_t \end{bmatrix} + \boldsymbol{e}_t = \begin{bmatrix} \bar{\Lambda}^f & O_{N\times(p-1)K} & \bar{\Lambda}^y & O_{N\times(p-1)M} \end{bmatrix} \begin{bmatrix} \bar{\mathbb{F}}_t \\ \mathbb{Y}_t \end{bmatrix} + \boldsymbol{e}_t, \tag{5.42}$$

$$\begin{bmatrix} \bar{\mathbb{F}}_t \\ \mathbb{Y}_t \end{bmatrix} = \mathbb{A} \begin{bmatrix} \bar{\mathbb{F}}_{t-1} \\ \mathbb{Y}_{t-1} \end{bmatrix} + \mathbb{v}_t = \begin{bmatrix} \mathbb{A}^f \\ \mathbb{A}^y \end{bmatrix} \begin{bmatrix} \bar{\mathbb{F}}_{t-1} \\ \mathbb{Y}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbb{v}_t^f \\ \mathbb{v}_t^y \end{bmatrix}, \tag{5.43}$$

with the vectors $\bar{\mathbb{F}}_t \in \mathbb{R}^{pK}$ and $\mathbb{Y}_t \in \mathbb{R}^{pM}$ defined as $\bar{\mathbb{F}}_t = \left[\bar{\boldsymbol{F}}_t', ..., \bar{\boldsymbol{F}}_{t-p+1}'\right]'$ and $\mathbb{Y}_t = \left[\boldsymbol{Y}_t', ..., \boldsymbol{Y}_{t-p+1}'\right]'$, respectively. Hence, both represent the p stacked latent factors and observed variables, respectively, up to time t. Furthermore, we have for matrix $\mathbb{A} \in \mathbb{R}^{p(K+M)\times p(K+M)}$ and vector $\mathbb{v}_t \in \mathbb{R}^{p(K+M)}$:

$$\mathbb{A} = \begin{bmatrix} \mathbb{A}^f \\ \mathbb{A}^y \end{bmatrix} = \left[\begin{array}{ccccccccc} \bar{\Phi}_1^{ff} & \bar{\Phi}_2^{ff} & \cdots & \bar{\Phi}_p^{ff} & \bar{\Phi}_1^{fy} & \bar{\Phi}_2^{fy} & \cdots & \bar{\Phi}_p^{fy} \\ I_K & O_{K\times K} & \cdots & O_{K\times K} & O_{K\times M} & O_{K\times M} & \cdots & O_{K\times M} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ O_{K\times K} & \cdots & I_K & O_{K\times K} & O_{K\times M} & O_{K\times M} & \cdots & O_{K\times M} \\ \hline \bar{\Phi}_1^{yf} & \bar{\Phi}_2^{yf} & \cdots & \bar{\Phi}_p^{yf} & \bar{\Phi}_1^{yy} & \bar{\Phi}_2^{yy} & \cdots & \bar{\Phi}_p^{yy} \\ O_{M\times K} & O_{M\times K} & \cdots & O_{M\times K} & I_M & O_{M\times M} & \cdots & O_{M\times M} \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \ddots & \vdots \\ O_{M\times K} & O_{M\times K} & \cdots & O_{M\times K} & O_{M\times M} & \cdots & I_M & O_{M\times M} \end{array}\right], \mathbb{v}_t = \begin{bmatrix} \mathbb{v}_t^f \\ \mathbb{v}_t^y \end{bmatrix} = \left[\begin{array}{c} \bar{\boldsymbol{v}}_t^f \\ \mathbf{0}_K \\ \vdots \\ \mathbf{0}_K \\ \hline \bar{\boldsymbol{v}}_t^y \\ \mathbf{0}_M \\ \vdots \\ \mathbf{0}_M \end{array}\right].$$

The covariance matrix $\Sigma_{\mathbb{v}} \in \mathbb{R}^{p(K+M)\times p(K+M)}$ of the error term $\mathbb{v}_t$ has the following shape:

$$\Sigma_{\mathbb{v}} = \begin{bmatrix} \Sigma_{\mathbb{v}}^{ff} & \Sigma_{\mathbb{v}}^{fy} \\ \Sigma_{\mathbb{v}}^{yf} & \Sigma_{\mathbb{v}}^{yy} \end{bmatrix} = \left[\begin{array}{cc|cc} I_K & O_{K\times(p-1)K} & \multicolumn{2}{c}{O_{pK\times pM}} \\ O_{(p-1)K\times K} & O_{(p-1)K\times(p-1)K} & & \\ \hline \multicolumn{2}{c|}{O_{pM\times pK}} & \Sigma_{\boldsymbol{v}}^{yy} & O_{M\times(p-1)M} \\ & & O_{(p-1)M\times M} & O_{(p-1)M\times(p-1)M} \end{array}\right]. \tag{5.44}$$

Proof:
Rearrange the FAVAR in (5.16)-(5.17). □

As usual for the Kalman Filter, we assume the model parameters $\Theta = \{\bar{\Lambda}, \Sigma_{\boldsymbol{e}}, \bar{\Phi}_1, \ldots, \bar{\Phi}_p, \Sigma_{\boldsymbol{v}}^{yy}\}$ as known and collect all observed data up to time $t \geq 0$ as follows:

$$\begin{aligned} &\Omega_0 = \emptyset, \\ &\Omega_t = \{\boldsymbol{X}_1, ..., \boldsymbol{X}_t, \boldsymbol{Y}_1, ..., \boldsymbol{Y}_t\}, \ \forall\ t > 0. \end{aligned}$$

For clarity reasons, we omit the index Θ in case of expectations and covariance matrices in this section and deploy the following abbreviations:

$$\begin{aligned} \hat{\bar{\mathbb{F}}}_{t|t-1} &= \mathbb{E}\left[\bar{\mathbb{F}}_t|\Omega_{t-1}\right] \in \mathbb{R}^{pK}, \\ \hat{P}_{t|t-1}^{\bar{\mathbb{F}}} &= \mathbb{V}\mathrm{ar}\left[\bar{\mathbb{F}}_t|\Omega_{t-1}\right] \in \mathbb{R}^{pK\times pK}, \\ \hat{P}_{(t,t-1)|t}^{\bar{\mathbb{F}},\bar{\mathbb{F}}} &= \mathbb{C}\mathrm{ov}\left[\bar{\mathbb{F}}_t, \bar{\mathbb{F}}_{t-1}|\Omega_t\right] \in \mathbb{R}^{pK\times pK}. \end{aligned}$$

Analogously, we abbreviate the expectations and covariance matrices of $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$ conditioned on Ω_{t-1}. Consequently, Ω_T equals the overall sample $\{X, Y\}$.

The Kalman Filter sequentially updates linear projections of a system. The factor $\bar{\mathbb{F}}_t$ is unobservable and so, the filter estimates it using the measurement variable $\boldsymbol{X}_t$ and the observed transition variable $\boldsymbol{Y}_t$. Its moment estimators up to time $t-1$ are the conditional expectation $\hat{\bar{\mathbb{F}}}_{t|t-1}$ and the conditional covariance matrix $\hat{P}^{\bar{\mathbb{F}}}_{t|t-1}$. As soon as new information in the form of $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$ arrives, the moments $\hat{\bar{\mathbb{F}}}_{t|t}$ and $\hat{P}^{\bar{\mathbb{F}}}_{t|t}$ serve as updated estimators. By this means, the Kalman Filter estimates the distribution of $\bar{\mathbb{F}}_t$ given the data Ω_t. The model in (5.42)-(5.43) is still Gaussian such that it holds:

$$\bar{\mathbb{F}}_t \big| \, \Omega_t \sim \mathcal{N}\left(\hat{\bar{\mathbb{F}}}_{t|t}, \hat{P}^{\bar{\mathbb{F}}}_{t|t}\right).$$

Before we obtain the moments $\hat{\bar{\mathbb{F}}}_{t|t}$ and $\hat{P}^{\bar{\mathbb{F}}}_{t|t}$, we determine the distributions of $\bar{\mathbb{F}}_t$, $\boldsymbol{Y}_t$ and $\boldsymbol{X}_t$ given Ω_{t-1}. In doing so, we receive for the distribution of the hidden factors $\bar{\mathbb{F}}_t$ given Ω_{t-1}:

Lemma 5.1.12 (Distribution of Hidden Factors)
For the FAVAR in Lemma 5.1.11, the unobserved factors $\bar{\mathbb{F}}_t$ given the information Ω_{t-1} are Gaussian:

$$\bar{\mathbb{F}}_t \big| \, \Omega_{t-1} \sim \mathcal{N}\left(\hat{\bar{\mathbb{F}}}_{t|t-1}, \hat{P}^{\bar{\mathbb{F}}}_{t|t-1}\right), \tag{5.45}$$

with mean and covariance matrix given by:

$$\begin{aligned}
\hat{\bar{\mathbb{F}}}_{t|t-1} &= \mathbb{A}^f \begin{bmatrix} \hat{\bar{\mathbb{F}}}_{t-1|t-1} \\ \mathbb{Y}_{t-1} \end{bmatrix}, \\
\hat{P}^{\bar{\mathbb{F}}}_{t|t-1} &= \mathbb{A}^f \begin{bmatrix} \hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & O_{pK\times pM} \\ O_{pM\times pK} & O_{pM\times pM} \end{bmatrix} \left(\mathbb{A}^f\right)' + \Sigma^{ff}_{\mathbb{v}}.
\end{aligned}$$

Proof:
At first, we plug (5.43) into the conditional expectation. Thereby, we gain from the observability of $\mathbb{Y}_{t-1}$, since it holds: $\hat{\mathbb{Y}}_{t-1|t-1} = \mathbb{Y}_{t-1}$. Furthermore, the independence of the error term $\mathbb{v}_t$ provides:

$$\hat{\bar{\mathbb{F}}}_{t|t-1} = \mathbb{E}\left[\bar{\mathbb{F}}_t|\Omega_{t-1}\right] = \mathbb{E}\left[\mathbb{A}^f \begin{bmatrix} \bar{\mathbb{F}}_{t-1} \\ \mathbb{Y}_{t-1} \end{bmatrix} + \mathbb{v}^f_t|\Omega_{t-1}\right] = \mathbb{A}^f \begin{bmatrix} \hat{\bar{\mathbb{F}}}_{t-1|t-1} \\ \mathbb{Y}_{t-1} \end{bmatrix}.$$

The covariance matrix of $\bar{\mathbb{F}}_t$ given Ω_{t-1} relies on (5.43), the properties of the covariance matrix, the fact that $\mathbb{Y}_{t-1}$ is known given Ω_{t-1} and that the factor $\bar{\mathbb{F}}_t$ is uncorrelated with the error term $\mathbb{v}_t$. Moreover, the error term $\mathbb{v}_t$ is independent of Ω_{t-1} and thus, we have:

$$\begin{aligned}
\hat{P}^{\bar{\mathbb{F}}}_{t|t-1} &= \mathbb{V}\mathrm{ar}\left[\mathbb{A}^f \begin{bmatrix} \bar{\mathbb{F}}_{t-1} \\ \mathbb{Y}_{t-1} \end{bmatrix} + \mathbb{v}_t|\Omega_{t-1}\right] = \mathbb{A}^f \mathbb{V}\mathrm{ar}\left[\begin{bmatrix} \bar{\mathbb{F}}_{t-1} \\ \mathbb{Y}_{t-1} \end{bmatrix}|\Omega_{t-1}\right]\left(\mathbb{A}^f\right)' + \begin{bmatrix} I_K & O_{K\times(p-1)K} \\ O_{(p-1)K\times M} & O_{(p-1)K\times(p-1)K} \end{bmatrix} \\
&= \mathbb{A}^f \begin{bmatrix} \hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & \hat{P}^{\bar{\mathbb{F}},\mathbb{Y}}_{t-1|t-1} \\ \hat{P}^{\mathbb{Y},\bar{\mathbb{F}}}_{t-1|t-1} & \hat{P}^{\mathbb{Y}}_{t-1|t-1} \end{bmatrix}\left(\mathbb{A}^f\right)' + \Sigma^{ff}_{\mathbb{v}} = \mathbb{A}^f \begin{bmatrix} \hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & O_{pK\times pM} \\ O_{pM\times pK} & O_{pM\times pM} \end{bmatrix}\left(\mathbb{A}^f\right)' + \Sigma^{ff}_{\mathbb{v}},
\end{aligned}$$

where $\hat{P}^{\bar{\mathbb{F}},\mathbb{Y}}_{t-1|t-1} = O_{pK\times pM}$ and $\hat{P}^{\mathbb{Y}}_{t-1|t-1} = O_{pM\times pM}$ arise from the observability of $\mathbb{Y}_{t-1}$. □

In addition, we obtain for the distributions of $\boldsymbol{Y}_t$ and $\boldsymbol{X}_t$ given Ω_{t-1} the following expressions.

Lemma 5.1.13 (Distribution of Future Observations)
Assume the FAVAR in Lemma 5.1.11. Then, the panel data $\boldsymbol{X}_t$ and variables $\boldsymbol{Y}_t$ given Ω_{t-1} are Gaussian:

$$\boldsymbol{Y}_t|\Omega_{t-1} \sim \mathcal{N}\left(\hat{\boldsymbol{Y}}_{t|t-1}, \hat{P}^{\boldsymbol{Y}}_{t|t-1}\right), \tag{5.46}$$

$$\boldsymbol{X}_t|\Omega_{t-1} \sim \mathcal{N}\left(\hat{\boldsymbol{X}}_{t|t-1}, \hat{P}^{\boldsymbol{X}}_{t|t-1}\right). \tag{5.47}$$

For the distribution parameters, we have:

$$\begin{aligned}
\hat{\boldsymbol{Y}}_{t|t-1} &= \begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}\begin{bmatrix}\hat{\bar{\mathbb{F}}}_{t-1|t-1}\\ \mathbb{Y}_{t-1}\end{bmatrix},\\
\hat{\boldsymbol{X}}_{t|t-1} &= \begin{bmatrix}\bar{\Lambda}^f & O_{N\times(p-1)K} & \bar{\Lambda}^y\end{bmatrix}\begin{bmatrix}\hat{\bar{\mathbb{F}}}_{t|t-1}\\ \hat{\boldsymbol{Y}}_{t|t-1}\end{bmatrix},\\
\hat{P}^{\boldsymbol{Y}}_{t|t-1} &= \begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}\begin{bmatrix}\hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & O_{pK\times pM}\\ O_{pM\times pK} & O_{pM\times pM}\end{bmatrix}\\
&\quad\cdot\begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}' + \Sigma^{yy}_{\boldsymbol{v}},\\
\hat{P}^{\boldsymbol{X}}_{t|t-1} &= \begin{bmatrix}\bar{\Lambda}^f & O_{N\times(p-1)K} & \bar{\Lambda}^y\end{bmatrix}\begin{bmatrix}\hat{P}^{\bar{\mathbb{F}}}_{t|t-1} & \hat{P}^{\bar{\mathbb{F}},\boldsymbol{Y}}_{t|t-1}\\ \hat{P}^{\boldsymbol{Y},\bar{\mathbb{F}}}_{t|t-1} & \hat{P}^{\boldsymbol{Y}}_{t|t-1}\end{bmatrix}\begin{bmatrix}\left(\bar{\Lambda}^f\right)'\\ O_{(p-1)K\times N}\\ \left(\bar{\Lambda}^y\right)'\end{bmatrix} + \Sigma_{\boldsymbol{e}}.
\end{aligned}$$

Proof:
Now, by similar reasoning as in Lemma 5.1.12, we prove the stated expectations and covariance matrices. Besides the former arguments, the independence of the error term $\boldsymbol{e}_t$ and the relation in (5.42) are required to obtain the following:

$$\begin{aligned}
\hat{\boldsymbol{Y}}_{t|t-1} &= \mathbb{E}\left[\boldsymbol{Y}_t|\Omega_{t-1}\right] = \mathbb{E}\left[\begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}\begin{bmatrix}\bar{\mathbb{F}}_{t-1}\\ \mathbb{Y}_{t-1}\end{bmatrix} + \bar{\boldsymbol{v}}^y_t|\Omega_{t-1}\right]\\
&= \begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}\begin{bmatrix}\hat{\bar{\mathbb{F}}}_{t-1|t-1}\\ \mathbb{Y}_{t-1}\end{bmatrix},\\
\hat{\boldsymbol{X}}_{t|t-1} &= \mathbb{E}\left[\boldsymbol{X}_t|\Omega_{t-1}\right] = \mathbb{E}\left[\begin{bmatrix}\bar{\Lambda}^f & \bar{\Lambda}^y\end{bmatrix}\begin{bmatrix}\bar{\boldsymbol{F}}_t\\ \boldsymbol{Y}_t\end{bmatrix} + \boldsymbol{e}_t|\Omega_{t-1}\right] = \begin{bmatrix}\bar{\Lambda}^f & \bar{\Lambda}^y\end{bmatrix}\begin{bmatrix}\hat{\bar{\boldsymbol{F}}}_{t|t-1}\\ \hat{\boldsymbol{Y}}_{t|t-1}\end{bmatrix}\\
&= \begin{bmatrix}\bar{\Lambda}^f & \bar{\Lambda}^y\end{bmatrix}\begin{bmatrix}\begin{bmatrix}I_K & O_{K\times(p-1)K}\end{bmatrix}\hat{\bar{\mathbb{F}}}_{t|t-1}\\ \hat{\boldsymbol{Y}}_{t|t-1}\end{bmatrix} = \begin{bmatrix}\bar{\Lambda}^f & O_{N\times(p-1)K} & \bar{\Lambda}^y\end{bmatrix}\begin{bmatrix}\hat{\bar{\mathbb{F}}}_{t|t-1}\\ \hat{\boldsymbol{Y}}_{t|t-1}\end{bmatrix}.
\end{aligned}$$

The conditional covariance matrices of $\boldsymbol{Y}_t$ and $\boldsymbol{X}_t$ are derived from (5.42), (5.43) and the independence of the error terms. Thereby, we have:

$$\begin{aligned}
\hat{P}^{\boldsymbol{Y}}_{t|t-1} &= \mathbb{V}\text{ar}\left[\boldsymbol{Y}_t|\Omega_{t-1}\right] = \mathbb{V}\text{ar}\left[\begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}\begin{bmatrix}\bar{\mathbb{F}}_{t-1}\\ \mathbb{Y}_{t-1}\end{bmatrix} + \bar{\boldsymbol{v}}^y_t|\Omega_{t-1}\right]\\
&= \begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}\begin{bmatrix}\hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & \hat{P}^{\bar{\mathbb{F}},\mathbb{Y}}_{t-1|t-1}\\ \hat{P}^{\mathbb{Y},\bar{\mathbb{F}}}_{t-1|t-1} & \hat{P}^{\mathbb{Y}}_{t-1|t-1}\end{bmatrix}\\
&\quad\cdot\begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}' + \Sigma^{yy}_{\boldsymbol{v}}\\
&= \begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}\begin{bmatrix}\hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & O_{pK\times pM}\\ O_{pM\times pK} & O_{pM\times pM}\end{bmatrix}\\
&\quad\cdot\begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}' + \Sigma^{yy}_{\boldsymbol{v}},
\end{aligned}$$

$$\hat{P}^{\boldsymbol{X}}_{t|t-1} = \mathbb{V}\text{ar}\left[\boldsymbol{X}_t|\Omega_{t-1}\right] = \mathbb{V}\text{ar}\left[\begin{bmatrix}\bar{\Lambda}^f & O_{N\times(p-1)K} & \bar{\Lambda}^y\end{bmatrix}\begin{bmatrix}\bar{\mathbb{F}}_t \\ \boldsymbol{Y}_t\end{bmatrix} + \boldsymbol{e}_t|\Omega_{t-1}\right]$$
$$= \begin{bmatrix}\bar{\Lambda}^f & O_{N\times(p-1)K} & \bar{\Lambda}^y\end{bmatrix}\begin{bmatrix}\hat{P}^{\bar{\mathbb{F}}}_{t|t-1} & \hat{P}^{\bar{\mathbb{F}},\boldsymbol{Y}}_{t|t-1} \\ \hat{P}^{\boldsymbol{Y},\bar{\mathbb{F}}}_{t|t-1} & \hat{P}^{\boldsymbol{Y}}_{t|t-1}\end{bmatrix}\begin{bmatrix}\left(\bar{\Lambda}^f\right)' \\ O_{(p-1)K\times N} \\ \left(\bar{\Lambda}^y\right)'\end{bmatrix} + \Sigma_{\boldsymbol{e}}.$$

which proves the claim. □

So far, Lemmata 5.1.12 and 5.1.13 provide the distributions of $\bar{\mathbb{F}}_t$, $\boldsymbol{Y}_t$ and $\boldsymbol{X}_t$ given Ω_{t-1}. However, for the distribution of the vector $\left[\bar{\mathbb{F}}'_t, \boldsymbol{Y}'_t, \boldsymbol{X}'_t\right]' \in \mathbb{R}^{K+M+N}$ given Ω_{t-1}, their covariance matrices are missing. For this purpose, we determine them next.

Lemma 5.1.14 (Covariance Matrices of FAVAR components)
For the FAVAR in Lemma 5.1.11, we get the following covariance matrices:

$$\hat{P}^{\bar{\mathbb{F}},\boldsymbol{Y}}_{t|t-1} = \mathbb{A}^f \begin{bmatrix}\hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & O_{pK\times pM} \\ O_{pM\times pK} & O_{pM\times pM}\end{bmatrix}\begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}',$$

$$\hat{P}^{\bar{\mathbb{F}},\boldsymbol{X}}_{t|t-1} = \begin{bmatrix}\hat{P}^{\bar{\mathbb{F}}}_{t|t-1} & \hat{P}^{\bar{\mathbb{F}},\boldsymbol{Y}}_{t|t-1}\end{bmatrix}\begin{bmatrix}\left(\bar{\Lambda}^f\right)' \\ O_{(p-1)K\times N} \\ \left(\bar{\Lambda}^y\right)'\end{bmatrix},$$

$$\hat{P}^{\boldsymbol{Y},\boldsymbol{X}}_{t|t-1} = \begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}\begin{bmatrix}\hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & O_{pK\times pM} \\ O_{pM\times pK} & O_{pM\times pM}\end{bmatrix}\mathbb{A}'\begin{bmatrix}\left(\bar{\Lambda}^f\right)' \\ O_{(p-1)K\times N} \\ \left(\bar{\Lambda}^y\right)' \\ O_{(p-1)M\times N}\end{bmatrix}.$$

Proof:
Based on (5.42)-(5.43), the observability of $\boldsymbol{X}_{t-1}$ and $\boldsymbol{Y}_{t-1}$ at time $t-1$ as well as the independence of the error terms provide:

$$\hat{P}^{\bar{\mathbb{F}},\boldsymbol{Y}}_{t|t-1} = \mathbb{C}\text{ov}\left[\mathbb{A}^f\begin{bmatrix}\bar{\mathbb{F}}_{t-1} \\ \mathbb{Y}_{t-1}\end{bmatrix} + \mathbb{v}^f_t, \begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}\begin{bmatrix}\bar{\mathbb{F}}_{t-1} \\ \mathbb{Y}_{t-1}\end{bmatrix} + \bar{\boldsymbol{v}}^y_t|\Omega_{t-1}\right]$$
$$= \mathbb{A}^f\begin{bmatrix}\hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & \hat{P}^{\bar{\mathbb{F}},\mathbb{Y}}_{t-1|t-1} \\ \hat{P}^{\mathbb{Y},\bar{\mathbb{F}}}_{t-1|t-1} & \hat{P}^{\mathbb{Y}}_{t-1|t-1}\end{bmatrix}\begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}'$$
$$= \mathbb{A}^f\begin{bmatrix}\hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & O_{pK\times pM} \\ O_{pM\times pK} & O_{pM\times pM}\end{bmatrix}\begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}',$$

$$\hat{P}^{\bar{\mathbb{F}},\boldsymbol{X}}_{t|t-1} = \mathbb{C}\text{ov}\left[\bar{\mathbb{F}}_t, \begin{bmatrix}\bar{\Lambda}^f & O_{N\times(p-1)K} & \bar{\Lambda}^y\end{bmatrix}\begin{bmatrix}\bar{\mathbb{F}}_t \\ \boldsymbol{Y}_t\end{bmatrix} + \boldsymbol{e}_t|\Omega_{t-1}\right] = \begin{bmatrix}\hat{P}^{\bar{\mathbb{F}}}_{t|t-1} & \hat{P}^{\bar{\mathbb{F}},\boldsymbol{Y}}_{t|t-1}\end{bmatrix}\begin{bmatrix}\left(\bar{\Lambda}^f\right)' \\ O_{(p-1)K\times N} \\ \left(\bar{\Lambda}^y\right)'\end{bmatrix},$$

$$\hat{P}^{\boldsymbol{Y},\boldsymbol{X}}_{t|t-1} = \mathbb{C}\text{ov}\left[\begin{bmatrix}\bar{\Phi}^{yf}_1 & \dots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \dots & \bar{\Phi}^{yy}_p\end{bmatrix}\begin{bmatrix}\bar{\mathbb{F}}_{t-1} \\ \mathbb{Y}_{t-1}\end{bmatrix} + \bar{\boldsymbol{v}}^y_t,\right.$$
$$\left.\begin{bmatrix}\bar{\Lambda}^f & O_{N\times(p-1)K} & \bar{\Lambda}^y & O_{N\times(p-1)M}\end{bmatrix}\begin{bmatrix}\bar{\mathbb{F}}_t \\ \mathbb{Y}_t\end{bmatrix} + \boldsymbol{e}_t|\Omega_{t-1}\right]$$

$$= \begin{bmatrix} \bar{\Phi}_1^{yf} & \cdots & \bar{\Phi}_p^{yf} & \bar{\Phi}_1^{yy} & \cdots & \bar{\Phi}_p^{yy} \end{bmatrix} \mathbb{C}\text{ov}\left[\begin{bmatrix} \bar{\mathbb{F}}_{t-1} \\ \mathbb{Y}_{t-1} \end{bmatrix}, \begin{bmatrix} \bar{\mathbb{F}}_t \\ \mathbb{Y}_t \end{bmatrix} |\Omega_{t-1}\right] \begin{bmatrix} \left(\bar{\Lambda}^f\right)' \\ O_{(p-1)K\times N} \\ \left(\bar{\Lambda}^y\right)' \\ O_{(p-1)M\times N} \end{bmatrix}$$

$$= \begin{bmatrix} \bar{\Phi}_1^{yf} & \cdots & \bar{\Phi}_p^{yf} & \bar{\Phi}_1^{yy} & \cdots & \bar{\Phi}_p^{yy} \end{bmatrix} \mathbb{C}\text{ov}\left[\begin{bmatrix} \bar{\mathbb{F}}_{t-1} \\ \mathbb{Y}_{t-1} \end{bmatrix}, \mathbb{A} \begin{bmatrix} \bar{\mathbb{F}}_{t-1} \\ \mathbb{Y}_{t-1} \end{bmatrix} + \mathbb{v}_t |\Omega_{t-1}\right] \begin{bmatrix} \left(\bar{\Lambda}^f\right)' \\ O_{(p-1)K\times N} \\ \left(\bar{\Lambda}^y\right)' \\ O_{(p-1)M\times N} \end{bmatrix}$$

$$= \begin{bmatrix} \bar{\Phi}_1^{yf} & \cdots & \bar{\Phi}_p^{yf} & \bar{\Phi}_1^{yy} & \cdots & \bar{\Phi}_p^{yy} \end{bmatrix} \begin{bmatrix} \hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & \hat{P}^{\bar{\mathbb{F}},\mathbb{Y}}_{t-1|t-1} \\ \hat{P}^{\mathbb{Y},\bar{\mathbb{F}}}_{t-1|t-1} & \hat{P}^{\mathbb{Y}}_{t-1|t-1} \end{bmatrix} \mathbb{A}' \begin{bmatrix} \left(\bar{\Lambda}^f\right)' \\ O_{(p-1)K\times N} \\ \left(\bar{\Lambda}^y\right)' \\ O_{(p-1)M\times N} \end{bmatrix}$$

$$= \begin{bmatrix} \bar{\Phi}_1^{yf} & \cdots & \bar{\Phi}_p^{yf} & \bar{\Phi}_1^{yy} & \cdots & \bar{\Phi}_p^{yy} \end{bmatrix} \begin{bmatrix} \hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & O_{pK\times pM} \\ O_{pM\times pK} & O_{pM\times pM} \end{bmatrix} \mathbb{A}' \begin{bmatrix} \left(\bar{\Lambda}^f\right)' \\ O_{(p-1)K\times N} \\ \left(\bar{\Lambda}^y\right)' \\ O_{(p-1)M\times N} \end{bmatrix}.$$

which completes the proof. □

Eventually, Lemmata 5.1.12-5.1.14 yield for the joint distribution of $\bar{\mathbb{F}}_t, \boldsymbol{Y}_t$ and $\boldsymbol{X}_t$ given Ω_{t-1} :

$$\begin{pmatrix} \bar{\mathbb{F}}_t \\ \boldsymbol{Y}_t \\ \boldsymbol{X}_t \end{pmatrix} \Bigg| \Omega_{t-1} \sim \mathcal{N}\left(\begin{pmatrix} \hat{\bar{\mathbb{F}}}_{t|t-1} \\ \hat{\boldsymbol{Y}}_{t|t-1} \\ \hat{\boldsymbol{X}}_{t|t-1} \end{pmatrix}, \begin{pmatrix} \hat{P}^{\bar{\mathbb{F}}}_{t|t-1} & \hat{P}^{\bar{\mathbb{F}},\boldsymbol{Y}}_{t|t-1} & \hat{P}^{\bar{\mathbb{F}},\boldsymbol{X}}_{t|t-1} \\ \hat{P}^{\boldsymbol{Y},\bar{\mathbb{F}}}_{t|t-1} & \hat{P}^{\boldsymbol{Y}}_{t|t-1} & \hat{P}^{\boldsymbol{Y},\boldsymbol{X}}_{t|t-1} \\ \hat{P}^{\boldsymbol{X},\bar{\mathbb{F}}}_{t|t-1} & \hat{P}^{\boldsymbol{X},\boldsymbol{Y}}_{t|t-1} & \hat{P}^{\boldsymbol{X}}_{t|t-1} \end{pmatrix} \right). \tag{5.48}$$

Furthermore, we can accomplish our original objective, i.e., the distribution of $\bar{\mathbb{F}}_t$ given Ω_t. In this context, we benefit in Theorem 5.1.15 from the fact: $\Omega_t = \Omega_{t-1} \cup \{\boldsymbol{X}_t, \boldsymbol{Y}_t\}$ and derive the following:

Theorem 5.1.15 (Updated Distribution of Hidden Factors)
For the FAVAR in Lemma 5.1.11, the hidden factors $\bar{\mathbb{F}}_t$ given the information Ω_t are Gaussian:

$$\bar{\mathbb{F}}_t \big| \boldsymbol{Y}_t, \boldsymbol{X}_t, \Omega_{t-1} = \bar{\mathbb{F}}_t \big| \Omega_t \sim \mathcal{N}\left(\hat{\bar{\mathbb{F}}}_{t|t}, \hat{P}^{\bar{\mathbb{F}}}_{t|t}\right), \tag{5.49}$$

with mean and covariance matrix defined as follows:

$$\hat{\bar{\mathbb{F}}}_{t|t} = \hat{\bar{\mathbb{F}}}_{t|t-1} + \Gamma_t^{KF} \begin{pmatrix} \boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|t-1} \\ \boldsymbol{X}_t - \hat{\boldsymbol{X}}_{t|t-1} \end{pmatrix},$$

$$\hat{P}^{\bar{\mathbb{F}}}_{t|t} = \hat{P}^{\bar{\mathbb{F}}}_{t|t-1} - \Gamma_t^{KF} \begin{bmatrix} \hat{P}^{\boldsymbol{Y}}_{t|t-1} & \hat{P}^{\boldsymbol{Y},\boldsymbol{X}}_{t|t-1} \\ \hat{P}^{\boldsymbol{X},\boldsymbol{Y}}_{t|t-1} & \hat{P}^{\boldsymbol{X}}_{t|t-1} \end{bmatrix} \left(\Gamma_t^{KF}\right)',$$

where we have for the Kalman Filter gain Γ_t^{KF}:

$$\Gamma_t^{KF} = \begin{bmatrix} \hat{P}^{\bar{\mathbb{F}},\boldsymbol{Y}}_{t|t-1} & \hat{P}^{\bar{\mathbb{F}},\boldsymbol{X}}_{t|t-1} \end{bmatrix} \begin{bmatrix} \hat{P}^{\boldsymbol{Y}}_{t|t-1} & \hat{P}^{\boldsymbol{Y},\boldsymbol{X}}_{t|t-1} \\ \hat{P}^{\boldsymbol{X},\boldsymbol{Y}}_{t|t-1} & \hat{P}^{\boldsymbol{X}}_{t|t-1} \end{bmatrix}^{-1}.$$

Proof:
Due to its definition, it holds $\Omega_t = \Omega_{t-1} \cup \{\boldsymbol{X}_t, \boldsymbol{Y}_t\}$. Finally, the properties of the conditional multivariate

normal distribution (Greene, 2003, pp. 871-872, Theorem B.7) yield the stated parameters. □

As mentioned in Section 5.1.2, we initialize the model parameters in the EM using the two-step principal component approach of Bernanke et al. (2005). In addition, the Kalman Filter requires an initial guess of the covariance matrix $\hat{P}^{\bar{\mathbb{F}}}_{p|p}$. The VAR(1) in (5.43) is supposed to be covariance-stationary. Therefore, we take the upper left submatrix of its overall covariance matrix $\Sigma_{\bar{\mathbb{F}},\mathbb{Y}}$ for this purpose. For $\Sigma_{\bar{\mathbb{F}},\mathbb{Y}}$, we have:

$$\begin{aligned}\Sigma_{\bar{\mathbb{F}},\mathbb{Y}} = \mathbb{A}\Sigma_{\bar{\mathbb{F}},\mathbb{Y}}\mathbb{A}' + \Sigma_{\mathbb{v}} &\Leftrightarrow vec\left(\Sigma_{\bar{\mathbb{F}},\mathbb{Y}}\right) = (\mathbb{A}\otimes\mathbb{A})\, vec\left(\Sigma_{\bar{\mathbb{F}},\mathbb{Y}}\right) + vec\left(\Sigma_{\mathbb{v}}\right)\\ &\Leftrightarrow \left[I_{(p(K+M))^2} - (\mathbb{A}\otimes\mathbb{A})\right] vec\left(\Sigma_{\bar{\mathbb{F}},\mathbb{Y}}\right) = vec\left(\Sigma_{\mathbb{v}}\right)\\ &\Leftrightarrow vec\left(\Sigma_{\bar{\mathbb{F}},\mathbb{Y}}\right) = \left[I_{(p(K+M))^2} - (\mathbb{A}\otimes\mathbb{A})\right]^{-1} vec\left(\Sigma_{\mathbb{v}}\right). \end{aligned} \tag{5.50}$$

The Jordan decomposition of the matrix $\mathbb{A}\otimes\mathbb{A}$ (Hamilton, 1994, p. 731, Eq. A.4.25) provides that the matrix $\left[I_{(p(K+M))^2} - (\mathbb{A}\otimes\mathbb{A})\right]$ has full rank such that its inverse is well-defined. As in previous sections, Algorithm 5.1.2 summarizes all findings in a compact form.

Algorithm 5.1.2: Kalman Filter for FAVARs with complete panel data

Initialization

$\bar{\Lambda}, \Sigma_{\boldsymbol{e}}, \bar{\Phi}_i, \Sigma_{\bar{\boldsymbol{v}}}, \boldsymbol{X}_t, \boldsymbol{Y}_t$ for $t = 1, ..., T$ and $i = 1, ..., p$ are known;

$\hat{\bar{\mathbb{F}}}_{p|p} = \left[\left(\bar{\boldsymbol{F}}_p^{\mathrm{PCA}}\right)', \ldots, \left(\bar{\boldsymbol{F}}_1^{\mathrm{PCA}}\right)'\right]'$;

$\hat{P}^{\bar{\mathbb{F}}}_{p|p} = \begin{bmatrix} I_{pK} & O_{pK\times pM}\end{bmatrix} \Sigma_{\bar{\mathbb{F}},\mathbb{Y}} \begin{bmatrix} I_{pK} \\ O_{pM\times pK}\end{bmatrix}$;

$\hat{\mathbb{Y}}_{p|p} = \mathbb{Y}_p$;

$\hat{\boldsymbol{X}}_{p|p} = \boldsymbol{X}_p$;

Forward recursion

for $t = p+1$ **to** T **do**

`# Prediction step`

$\hat{\bar{\mathbb{F}}}_{t|t-1} = \mathbb{A}^f \begin{bmatrix} \hat{\bar{\mathbb{F}}}_{t-1|t-1} \\ \mathbb{Y}_{t-1}\end{bmatrix}$;

$\hat{P}^{\bar{\mathbb{F}}}_{t|t-1} = \mathbb{A}^f \begin{bmatrix} \hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & O_{pK\times pM} \\ O_{pM\times pK} & O_{pM\times pM}\end{bmatrix} \left(\mathbb{A}^f\right)' + \Sigma^{ff}_{\mathbf{v}}$;

$\hat{\boldsymbol{Y}}_{t|t-1} = \begin{bmatrix} \bar{\Phi}^{yf}_1 & \ldots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \ldots & \bar{\Phi}^{yy}_p\end{bmatrix} \begin{bmatrix} \hat{\bar{\mathbb{F}}}_{t-1|t-1} \\ \mathbb{Y}_{t-1}\end{bmatrix}$;

$\hat{P}^{\boldsymbol{Y}}_{t|t-1} = \begin{bmatrix} \bar{\Phi}^{yf}_1 & \ldots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \ldots & \bar{\Phi}^{yy}_p\end{bmatrix} \begin{bmatrix} \hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & O_{pK\times pM} \\ O_{pM\times pK} & O_{pM\times pM}\end{bmatrix} \begin{bmatrix} \bar{\Phi}^{yf}_1 & \ldots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \ldots & \bar{\Phi}^{yy}_p\end{bmatrix}' + \Sigma^{yy}_{\boldsymbol{v}}$;

$\hat{P}^{\bar{\mathbb{F}},\boldsymbol{Y}}_{t|t-1} = \mathbb{A}^f \begin{bmatrix} \hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & O_{pK\times pM} \\ O_{pM\times pK} & O_{pM\times pM}\end{bmatrix} \begin{bmatrix} \bar{\Phi}^{yf}_1 & \ldots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \ldots & \bar{\Phi}^{yy}_p\end{bmatrix}'$;

$\hat{\boldsymbol{X}}_{t|t-1} = \begin{bmatrix} \bar{\Lambda}^f & O_{N\times(p-1)K} & \bar{\Lambda}^y\end{bmatrix} \begin{bmatrix} \hat{\bar{\mathbb{F}}}_{t|t-1} \\ \hat{\boldsymbol{Y}}_{t|t-1}\end{bmatrix}$;

$\hat{P}^{\boldsymbol{X}}_{t|t-1} = \begin{bmatrix} \bar{\Lambda}^f & O_{N\times(p-1)K} & \bar{\Lambda}^y\end{bmatrix} \begin{bmatrix} \hat{P}^{\bar{\mathbb{F}}}_{t|t-1} & \hat{P}^{\bar{\mathbb{F}},\boldsymbol{Y}}_{t|t-1} \\ \hat{P}^{\boldsymbol{Y},\bar{\mathbb{F}}}_{t|t-1} & \hat{P}^{\boldsymbol{Y}}_{t|t-1}\end{bmatrix} \begin{bmatrix} (\bar{\Lambda}^f)' \\ O_{(p-1)K\times N} \\ (\bar{\Lambda}^y)'\end{bmatrix} + \Sigma_{\boldsymbol{e}}$;

$\hat{P}^{\bar{\mathbb{F}},\boldsymbol{X}}_{t|t-1} = \begin{bmatrix} \hat{P}^{\bar{\mathbb{F}}}_{t|t-1} & \hat{P}^{\bar{\mathbb{F}},\boldsymbol{Y}}_{t|t-1}\end{bmatrix} \begin{bmatrix} (\bar{\Lambda}^f)' \\ O_{(p-1)K\times N} \\ (\bar{\Lambda}^y)'\end{bmatrix}$;

$\hat{P}^{\boldsymbol{Y},\boldsymbol{X}}_{t|t-1} = \begin{bmatrix} \bar{\Phi}^{yf}_1 & \ldots & \bar{\Phi}^{yf}_p & \bar{\Phi}^{yy}_1 & \ldots & \bar{\Phi}^{yy}_p\end{bmatrix} \begin{bmatrix} \hat{P}^{\bar{\mathbb{F}}}_{t-1|t-1} & O_{pK\times pM} \\ O_{pM\times pK} & O_{pM\times pM}\end{bmatrix} \mathbb{A}' \begin{bmatrix} (\bar{\Lambda}^f)' \\ O_{(p-1)K\times N} \\ (\bar{\Lambda}^y)' \\ O_{(p-1)M\times N}\end{bmatrix}$;

`# Correction step`

$\Gamma^{KF}_t = \begin{bmatrix} \hat{P}^{\bar{\mathbb{F}},\boldsymbol{Y}}_{t|t-1} & \hat{P}^{\bar{\mathbb{F}},\boldsymbol{X}}_{t|t-1}\end{bmatrix} \begin{bmatrix} \hat{P}^{\boldsymbol{Y}}_{t|t-1} & \hat{P}^{\boldsymbol{Y},\boldsymbol{X}}_{t|t-1} \\ \hat{P}^{\boldsymbol{X},\boldsymbol{Y}}_{t|t-1} & \hat{P}^{\boldsymbol{X}}_{t|t-1}\end{bmatrix}^{-1}$;

$\hat{\bar{\mathbb{F}}}_{t|t} = \hat{\bar{\mathbb{F}}}_{t|t-1} + \Gamma^{KF}_t \begin{pmatrix} \boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|t-1} \\ \boldsymbol{X}_t - \hat{\boldsymbol{X}}_{t|t-1}\end{pmatrix}$;

$\hat{P}^{\bar{\mathbb{F}}}_{t|t} = \hat{P}^{\bar{\mathbb{F}}}_{t|t-1} - \Gamma^{KF}_t \begin{bmatrix} \hat{P}^{\boldsymbol{Y}}_{t|t-1} & \hat{P}^{\boldsymbol{Y},\boldsymbol{X}}_{t|t-1} \\ \hat{P}^{\boldsymbol{X},\boldsymbol{Y}}_{t|t-1} & \hat{P}^{\boldsymbol{X}}_{t|t-1}\end{bmatrix} (\Gamma^{KF}_t)'$;

end

5.1.4 Kalman Smoother for FAVARs

In (5.42)-(5.43) we separate the hidden factors from the observable variables and hence, do not work with the most common VAR(1) state-space representation, which simply stacks the original VAR(p) vectors irrespective of their observability. Therefore, we check whether the standard Kalman Smoother equations remain valid for our FAVAR formulation in (5.42)-(5.43). For clarity reasons, we drop again the index Θ for expectations as well as variance and covariance matrices in this section.

Theorem 5.1.16 (Smoothed Distribution of Hidden Factors)
Assume the FAVAR in (5.42)-(5.43) *with complete sample data $\Omega_T = \{X, Y\}$ for $T \geq p$. Then, for all points in time $p \leq t \leq T$ the hidden factors $\overline{\mathbb{F}}_t$ given the overall information Ω_T are Gaussian:*

$$\overline{\mathbb{F}}_t \big| \, \Omega_T \sim \mathcal{N}\left(\hat{\overline{\mathbb{F}}}_{t|T}, \hat{P}^{\overline{\mathbb{F}}}_{t|T}\right),$$

with distribution parameters defined as follows:

$$\hat{\overline{\mathbb{F}}}_{t|T} = \hat{\overline{\mathbb{F}}}_{t|t} + \Gamma_t^{KS}\left(\hat{\overline{\mathbb{F}}}_{t+1|T} - \hat{\overline{\mathbb{F}}}_{t+1|t}\right), \tag{5.51}$$

$$\hat{P}^{\overline{\mathbb{F}}}_{t|T} = \hat{P}^{\overline{\mathbb{F}}}_{t|t} - \Gamma_t^{KS}\left(\hat{P}^{\overline{\mathbb{F}}}_{t+1|t} - \hat{P}^{\overline{\mathbb{F}}}_{t+1|T}\right)\left(\Gamma_t^{KS}\right)'. \tag{5.52}$$

In this context, the Kalman Smoother gain Γ_t^{KS} is given by:

$$\Gamma_t^{KS} = \hat{P}^{\overline{\mathbb{F}},\overline{\mathbb{F}}}_{(t,t+1)|t}\left(\hat{P}^{\overline{\mathbb{F}}}_{t+1|t}\right)^{-1}. \tag{5.53}$$

Proof:
By virtue of the tower rule for conditional expectations (Durrett, 2010, p. 228, Theorem 5.1.6), we have:

$$\hat{\overline{\mathbb{F}}}_{t|T} = \mathbb{E}\left[\overline{\mathbb{F}}_t|\Omega_T\right] = \mathbb{E}\left[\mathbb{E}\left[\overline{\mathbb{F}}_t|\overline{\mathbb{F}}_{t+1}, \Omega_T\right]|\Omega_T\right]. \tag{5.54}$$

If we define $\Omega_{t+1}^T = \mathrm{g}\left(\{\boldsymbol{e}_s, \boldsymbol{v}_s\}_{s=t+1}^T\right)$ for some function g, the information set Ω_T coincides with the sets $\overline{\mathbb{F}}_{t+1}$, Ω_t and Ω_{t+1}^T. Consequently, we obtain for the inner expectation of (5.54):

$$\begin{aligned}\mathbb{E}\left[\overline{\mathbb{F}}_t|\overline{\mathbb{F}}_{t+1}, \Omega_T\right] &= \mathbb{E}\left[\overline{\mathbb{F}}_t|\overline{\mathbb{F}}_{t+1}, \Omega_t, \Omega_{t+1}^T\right] = \mathbb{E}\left[\overline{\mathbb{F}}_t|\overline{\mathbb{F}}_{t+1}, \Omega_t, \mathrm{g}\left(\{\boldsymbol{e}_s, \boldsymbol{v}_s\}_{s=t+1}^T\right)\right] \\ &= \mathbb{E}\left[\overline{\mathbb{F}}_t|\overline{\mathbb{F}}_{t+1}, \Omega_t\right],\end{aligned} \tag{5.55}$$

where the last step requires the independence of the factors $\overline{\mathbb{F}}_t$ and $\{\boldsymbol{e}_s, \boldsymbol{v}_s\}_{s=t+1}^T$.

Similarly, the law of the total conditional variance provides for the the covariance matrix $\hat{P}^{\overline{\mathbb{F}}}_{t|T}$:

$$\hat{P}^{\overline{\mathbb{F}}}_{t|T} = \mathbb{V}\mathrm{ar}\left[\overline{\mathbb{F}}_t|\Omega_T\right] = \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\overline{\mathbb{F}}_t|\overline{\mathbb{F}}_{t+1}, \Omega_T\right]|\Omega_T\right] + \mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\overline{\mathbb{F}}_t|\overline{\mathbb{F}}_{t+1}, \Omega_T\right]|\Omega_T\right]. \tag{5.56}$$

The conditional covariance matrix in the inside of the first term and the inner conditional expectation of the second term can be simplified by the same method as in (5.55). Therefore, we can write (5.56) as:

$$\hat{P}^{\overline{\mathbb{F}}}_{t|T} = \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\overline{\mathbb{F}}_t|\overline{\mathbb{F}}_{t+1}, \Omega_t\right]|\Omega_T\right] + \mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\overline{\mathbb{F}}_t|\overline{\mathbb{F}}_{t+1}, \Omega_t\right]|\Omega_T\right]. \tag{5.57}$$

The calculation of (5.54) and (5.56) calls for the conditional expectation in (5.55) and the conditional covariance matrix part of (5.57). We derive both from the distribution of the vector $(\overline{\mathbb{F}}_t', \overline{\mathbb{F}}_{t+1}')'$ conditioned on Ω_t, which is given by:

$$\begin{pmatrix}\overline{\mathbb{F}}_t \\ \overline{\mathbb{F}}_{t+1}\end{pmatrix}\Bigg|\,\Omega_t \sim \mathcal{N}\left(\begin{pmatrix}\hat{\overline{\mathbb{F}}}_{t|t} \\ \hat{\overline{\mathbb{F}}}_{t+1|t}\end{pmatrix}, \begin{pmatrix}\hat{P}^{\overline{\mathbb{F}}}_{t|t} & \hat{P}^{\overline{\mathbb{F}},\overline{\mathbb{F}}}_{(t,t+1)|t} \\ \hat{P}^{\overline{\mathbb{F}},\overline{\mathbb{F}}}_{(t+1,t)|t} & \hat{P}^{\overline{\mathbb{F}}}_{t+1|t}\end{pmatrix}\right), \tag{5.58}$$

where the vectors $\hat{\bar{\mathbb{F}}}_{t|t}$ and $\hat{\bar{\mathbb{F}}}_{t+1|t}$ and the matrices $\hat{P}^{\bar{\mathbb{F}}}_{t|t}$ and $\hat{P}^{\bar{\mathbb{F}}}_{t+1|t}$ are known from the KF in Algorithm 5.1.2. Regarding the unknown covariance matrix $\hat{P}^{\bar{\mathbb{F}},\bar{\mathbb{F}}}_{(t,t+1)|t}$ the relation in (5.43), the independence of the factors $\bar{\mathbb{F}}_t$ and errors $\mathbb{v}_t$ as well as the observability of $\mathbb{Y}_t$ yield:

$$\begin{aligned}\hat{P}^{\bar{\mathbb{F}},\bar{\mathbb{F}}}_{(t,t+1)|t} &= \mathbb{C}\mathrm{ov}\left[\bar{\mathbb{F}}_t, \bar{\mathbb{F}}_{t+1}|\Omega_t\right] = \mathbb{C}\mathrm{ov}\left[\bar{\mathbb{F}}_t, \mathbb{A}^f \begin{bmatrix}\bar{\mathbb{F}}_t \\ \mathbb{Y}_t\end{bmatrix} + \mathbb{v}_t|\Omega_t\right] \\ &= \begin{bmatrix}\hat{P}^{\bar{\mathbb{F}}}_{t|t} & \hat{P}^{\bar{\mathbb{F}},\mathbb{Y}}_{t|t}\end{bmatrix}\left(\mathbb{A}^f\right)' = \begin{bmatrix}\hat{P}^{\bar{\mathbb{F}}}_{t|t} & O_{pK\times pM}\end{bmatrix}\left(\mathbb{A}^f\right)'.\end{aligned}$$

Now, the properties of the conditional Gaussian distribution (Greene, 2003, pp. 871-872, Theorem B.7) result in the below expectation and variance of $\bar{\mathbb{F}}_t$ conditioned on $\bar{\mathbb{F}}_{t+1}$ and Ω_t:

$$\mathbb{E}\left[\bar{\mathbb{F}}_t|\bar{\mathbb{F}}_{t+1},\Omega_t\right] = \hat{\bar{\mathbb{F}}}_{t|t} + \Gamma^{KS}_t\left(\bar{\mathbb{F}}_{t+1} - \hat{\bar{\mathbb{F}}}_{t+1|t}\right), \tag{5.59}$$

$$\mathbb{V}\mathrm{ar}\left[\bar{\mathbb{F}}_t|\bar{\mathbb{F}}_{t+1},\Omega_t\right] = \hat{P}^{\bar{\mathbb{F}}}_{t|t} - \Gamma^{KS}_t\hat{P}^{\bar{\mathbb{F}}}_{t+1|t}\left(\Gamma^{KS}_t\right)', \tag{5.60}$$

where the Kalman Smoother gain Γ^{KS}_t is defined in (5.53).

Inserting (5.55) and (5.59) into (5.54) as well as (5.59) and (5.60) into (5.57) leads to:

$$\begin{aligned}\hat{\bar{\mathbb{F}}}_{t|T} &= \mathbb{E}\left[\hat{\bar{\mathbb{F}}}_{t|t} + \Gamma^{KS}_t\left(\bar{\mathbb{F}}_{t+1} - \hat{\bar{\mathbb{F}}}_{t+1|t}\right)|\Omega_T\right], \\ \hat{P}^{\bar{\mathbb{F}}}_{t|T} &= \mathbb{E}\left[\hat{P}^{\bar{\mathbb{F}}}_{t|t} - \Gamma^{KS}_t\hat{P}^{\bar{\mathbb{F}}}_{t+1|t}\left(\Gamma^{KS}_t\right)'|\Omega_T\right] + \mathbb{V}\mathrm{ar}\left[\bar{\mathbb{F}}_{t|t} + \Gamma^{KS}_t\left(\bar{\mathbb{F}}_{t+1} - \hat{\bar{\mathbb{F}}}_{t+1|t}\right)|\Omega_T\right].\end{aligned}$$

Note, the vectors $\hat{\bar{\mathbb{F}}}_{t|t}$ and $\hat{\bar{\mathbb{F}}}_{t+1|t}$ and the matrix Γ^{KS}_t are deterministic such that the above solutions can be further simplified and we end up with the expressions in (5.51)-(5.52). □

Until now, we smooth for each point in time the expectation and covariance matrix of the hidden factors. However, the transition from one point in time to another is still missing. For this purpose, we use the subsequent lag-one autocovariance smoother.

Lemma 5.1.17 (Lag-One Autocovariance Smoother)
Assume the FAVAR in (5.42)-(5.43) *based on complete sample data* $\Omega_T = \{X, Y\}$ *with* $T \geq p+1$. *Then, for all points in time* $p+1 \leq t \leq T$, *a lag-one autocovariance smoother is given by:*

$$\hat{P}^{\bar{\mathbb{F}},\bar{\mathbb{F}}}_{(t,t-1)|T} = \mathbb{C}ov\left[\bar{\mathbb{F}}_t, \bar{\mathbb{F}}_{t-1}|\Omega_T\right] = \hat{P}^{\bar{\mathbb{F}}}_{t|T}\left(\Gamma^{KS}_{t-1}\right)'. \tag{5.61}$$

Proof:
At first, we apply the algebraic formula for the conditional covariance matrix:

$$\hat{P}^{\bar{\mathbb{F}},\bar{\mathbb{F}}}_{(t,t-1)|T} = \mathbb{C}\mathrm{ov}\left[\bar{\mathbb{F}}_t, \bar{\mathbb{F}}_{t-1}|\Omega_T\right] = \mathbb{E}\left[\bar{\mathbb{F}}_t\bar{\mathbb{F}}'_{t-1}|\Omega_T\right] - \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t-1|T}. \tag{5.62}$$

The values $\hat{\bar{\mathbb{F}}}_{t|T}$ are provided in Theorem 5.1.16. Therefore, we focus on the first term. As in the proof of $\hat{\bar{\mathbb{F}}}_{t|T}$ and $\hat{P}^{\bar{\mathbb{F}}}_{t|T}$, we apply the tower rule and (5.55) to justify the following steps:

$$\begin{aligned}\mathbb{E}\left[\bar{\mathbb{F}}_t\bar{\mathbb{F}}'_{t-1}|\Omega_T\right] &= \mathbb{E}\left[\mathbb{E}\left[\bar{\mathbb{F}}_t\bar{\mathbb{F}}'_{t-1}|\bar{\mathbb{F}}_t,\Omega_T\right]|\Omega_T\right] = \mathbb{E}\left[\bar{\mathbb{F}}_t\mathbb{E}\left[\bar{\mathbb{F}}'_{t-1}|\bar{\mathbb{F}}_t,\Omega_T\right]|\Omega_T\right] \\ &= \mathbb{E}\left[\bar{\mathbb{F}}_t\mathbb{E}\left[\bar{\mathbb{F}}'_{t-1}|\bar{\mathbb{F}}_t,\Omega_{t-1}\right]|\Omega_T\right].\end{aligned}$$

Next, (5.59) and the fact that $\hat{\bar{\mathbb{F}}}_{t-1|t-1}$, Γ^{KS}_{t-1} and $\hat{\bar{\mathbb{F}}}_{t|t-1}$ are constants yield:

$$\mathbb{E}\left[\bar{\mathbb{F}}_t\bar{\mathbb{F}}'_{t-1}|\Omega_T\right] = \mathbb{E}\left[\bar{\mathbb{F}}_t\left(\hat{\bar{\mathbb{F}}}'_{t-1|t-1} + \left(\bar{\mathbb{F}}_t - \hat{\bar{\mathbb{F}}}_{t|t-1}\right)'\left(\Gamma^{KS}_{t-1}\right)'\right)|\Omega_T\right]$$

$$
\begin{aligned}
&= \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t-1|t-1} + \mathbb{E}\left[\bar{\mathbb{F}}_t\left(\bar{\mathbb{F}}_t - \hat{\bar{\mathbb{F}}}_{t|t-1}\right)'|\Omega_T\right]\left(\Gamma^{KS}_{t-1}\right)' \\
&= \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t-1|t-1} + \left(\mathbb{E}\left[\bar{\mathbb{F}}_t\bar{\mathbb{F}}'_t|\Omega_T\right] - \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t|t-1}\right)\left(\Gamma^{KS}_{t-1}\right)' \\
&= \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t-1|t-1} + \left(\hat{P}^{\bar{\mathbb{F}}}_{t|T} + \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t|T} - \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t|t-1}\right)\left(\Gamma^{KS}_{t-1}\right)'. \qquad (5.63)
\end{aligned}
$$

If we combine (5.62) and (5.63), we have for the lag-one autocovariance smoother:

$$
\hat{P}^{\bar{\mathbb{F}},\bar{\mathbb{F}}}_{(t,t-1)|T} = \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t-1|t-1} + \left(\hat{P}^{\bar{\mathbb{F}}}_{t|T} + \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t|T} - \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t|t-1}\right)\left(\Gamma^{KS}_{t-1}\right)' - \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t-1|T}.
$$

Eventually, the smoothed expectation in (5.51) for time $t-1$ provides:

$$
\begin{aligned}
\hat{P}^{\bar{\mathbb{F}},\bar{\mathbb{F}}}_{(t,t-1)|T} &= \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t-1|t-1} + \left(\hat{P}^{\bar{\mathbb{F}}}_{t|T} + \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t|T} - \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t|t-1}\right)\left(\Gamma^{KS}_{t-1}\right)' \\
&\quad - \hat{\bar{\mathbb{F}}}_{t|T}\left(\hat{\bar{\mathbb{F}}}_{t-1|t-1} + \Gamma^{KS}_{t-1}\left(\hat{\bar{\mathbb{F}}}_{t|T} - \hat{\bar{\mathbb{F}}}_{t|t-1}\right)\right)' \\
&= \left(\hat{P}^{\bar{\mathbb{F}}}_{t|T} + \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t|T} - \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t|t-1} - \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t|T} + \hat{\bar{\mathbb{F}}}_{t|T}\hat{\bar{\mathbb{F}}}'_{t|t-1}\right)'\left(\Gamma^{KS}_{t-1}\right)' \\
&= \hat{P}^{\bar{\mathbb{F}}}_{t|T}\left(\Gamma^{KS}_{t-1}\right)', \qquad (5.64)
\end{aligned}
$$

which finishes the proof. □

For reasons of implementation efficiency, Algorithm 5.1.3 summarizes the Kalman Smoother and Lag-One Autocovariance Smoother as single routine.

Algorithm 5.1.3: Kalman and lag-one autocovariance smoother for FAVARs with complete data

Initialization
$\hat{\bar{\mathbb{F}}}_{t+1|t}$ and $\hat{P}^{\bar{\mathbb{F}}}_{t+1|t}$ are provided by the Kalman Filter for $p \le t \le T-1$;
$\hat{\bar{\mathbb{F}}}_{t|t}$ and $\hat{P}^{\bar{\mathbb{F}}}_{t|t}$ are provided by the Kalman Filter for $p \le t \le T$;

Backward recursion
for $t = T-1$ **to** p **do**

$\quad \hat{P}^{\bar{\mathbb{F}},\bar{\mathbb{F}}}_{(t,t+1)|t} = \left[\hat{P}^{\bar{\mathbb{F}}}_{t|t} \quad O_{pK\times pM}\right]\left(\mathbb{A}^f\right)'$;

$\quad \Gamma^{KS}_t = \hat{P}^{\bar{\mathbb{F}},\bar{\mathbb{F}}}_{(t,t+1)|t}\left(\hat{P}^{\bar{\mathbb{F}}}_{t+1|t}\right)^{-1}$;

$\quad \hat{\bar{\mathbb{F}}}_{t|T} = \hat{\bar{\mathbb{F}}}_{t|t} + \Gamma^{KS}_t\left(\hat{\bar{\mathbb{F}}}_{t+1|T} - \hat{\bar{\mathbb{F}}}_{t+1|t}\right)$;

$\quad \hat{P}^{\bar{\mathbb{F}}}_{t|T} = \hat{P}^{\bar{\mathbb{F}}}_{t|t} - \Gamma^{KS}_t\left(\hat{P}^{\bar{\mathbb{F}}}_{t+1|t} - \hat{P}^{\bar{\mathbb{F}}}_{t+1|T}\right)\left(\Gamma^{KS}_t\right)'$;

$\quad \hat{P}^{\bar{\mathbb{F}},\bar{\mathbb{F}}}_{(t+1,t)|T} = \hat{P}^{\bar{\mathbb{F}}}_{t+1|T}\left(\Gamma^{KS}_t\right)'$;

end

5.1.5 Estimation of FAVARs with Incomplete Panel Data

The FAVARs in Sections 5.1.1-5.1.4 rely on complete data, that is, panel data and observed variables $\boldsymbol{Y}_t$ are of the same frequency and do not comprise any gaps. To allow for incomplete panel data we pursue the ansatz in Definition 2.2.1. Each univariate time series in the panel data is supposed to be standardized, therefore, some minor adjustments are necessary. Unfortunately, the proposed method does not permit data incompleteness for the observed variables $\boldsymbol{Y}_t$ and hence, leaves this for the future research.

Definition 5.1.18 (FAVARs with Incomplete Panel Data)
Similar to Definition 2.2.1, for $1 \le i \le N$ *the vectors* $\boldsymbol{X}^i_{obs} \in \mathbb{R}^{T(i)}$ *and* $\bar{\boldsymbol{X}}^i \in \mathbb{R}^T$ *with* $T(i) \le T$ *collect the*

observations and complete analogs of signal i with matrix $Q_i \in \mathbb{R}^{T(i)\times T}$ modeling their linear relation. The panel data in (5.16) is supposed to consist of standardized time series, thus, we set: $\boldsymbol{X}^i = (\bar{\boldsymbol{X}}^i - \mu_{\bar{X}_i}\mathbb{1}_T)\sigma_{\bar{X}_i}^{-1}$ for each time series i with mean $\mu_{\bar{X}_i}$ and variance $\sigma_{\bar{X}_i}^2$. If the index $1 \le t \le T$ covers the points in time, when new data arrives, the FAVAR with incomplete data is defined as follows:

$$\begin{aligned}
\boldsymbol{X}_{obs}^i &= Q_i \bar{\boldsymbol{X}}^i, && \forall\ 1 \le i \le N, && (5.65)\\
\boldsymbol{X}^i &= (\bar{\boldsymbol{X}}^i - \mu_{\bar{X}_i}\mathbb{1}_T)\sigma_{\bar{X}_i}^{-1}, && \forall\ 1 \le i \le N, && (5.66)\\
\boldsymbol{X}_t &= \bar{\Lambda}\boldsymbol{C}_t + \boldsymbol{e}_t, && \boldsymbol{e}_t \sim \mathcal{N}(\boldsymbol{0}_N, \Sigma_{\boldsymbol{e}})\ iid, && (5.67)\\
\boldsymbol{C}_t &= \sum_{i=j}^{p} \bar{\Phi}_j \boldsymbol{C}_{t-j} + \bar{\boldsymbol{v}}_t, && \bar{\boldsymbol{v}}_t \sim \mathcal{N}(\boldsymbol{0}_{K+M}, \Sigma_{\bar{\boldsymbol{v}}})\ iid, && (5.68)
\end{aligned}$$

with joint vector $\boldsymbol{C}_t = \left[\bar{\boldsymbol{F}}_t', \boldsymbol{Y}_t'\right]' \in \mathbb{R}^{K+M}$ and constant matrices $\bar{\Lambda} \in \mathbb{R}^{N\times(K+M)}, \Sigma_{\boldsymbol{e}} \in \mathbb{R}^{N\times N}, \bar{\Phi}_j \in \mathbb{R}^{(K+M)\times(K+M)}, 1 \le j \le p$, and $\Sigma_{\bar{\boldsymbol{v}}} \in \mathbb{R}^{(K+M)\times(K+M)}$.

Note, the FAVARs in Definition 5.1.18 support the treatment of stock, flow and change in flow variables in Section 2.2. Neither the choice of the matrices Q_i nor the existence of constraints for the loadings and VAR coefficients matter at this stage. Therefore, we continue with the general form in (5.65)-(5.68). As in Lemma 3.1.7, we modify the reconstruction formula of Stock and Watson (1999a, 2002b) for FAVARs.

Lemma 5.1.19 (Conditional Distribution of Complete Panel Data)
For the FAVAR in Definition 5.1.18 with incomplete panel data, the matrices $\bar{F} = \left[\bar{\boldsymbol{F}}_1, ..., \bar{\boldsymbol{F}}_T\right]' \in \mathbb{R}^{T\times K}$, $Y = [\boldsymbol{Y}_1, ..., \boldsymbol{Y}_T]' \in \mathbb{R}^{T\times M}$ and $E = [\boldsymbol{e}_1, ..., \boldsymbol{e}_T]' \in \mathbb{R}^{T\times N}$ collect all factors, standardized observations and errors, respectively. Then, for $1 \le i \le N$ the vector $\bar{\boldsymbol{X}}^i$ given the matrices $\bar{F}$ and Y as well as the observations $\boldsymbol{X}_{obs}^i$ is Gaussian with the following parameters:

$$\begin{aligned}
\mathbb{E}_\Theta\left[\bar{\boldsymbol{X}}^i | \bar{F}, Y, \boldsymbol{X}_{obs}^i\right] &= \mu_{\bar{X}_i}\mathbb{1}_T + \sigma_{\bar{X}_i}\bar{F}\left(\bar{\boldsymbol{\Lambda}}_i^f\right)' + \sigma_{\bar{X}_i} Y\left(\bar{\boldsymbol{\Lambda}}_i^y\right)' \\
&\quad + Q_i'\left(Q_iQ_i'\right)^{-1}\left[\boldsymbol{X}_{obs}^i - Q_i\left(\mu_{\bar{X}_i}\mathbb{1}_T + \sigma_{\bar{X}_i}\bar{F}\left(\bar{\boldsymbol{\Lambda}}_i^f\right)' + \sigma_{\bar{X}_i} Y\left(\bar{\boldsymbol{\Lambda}}_i^y\right)'\right)\right], \quad (5.69)\\
\mathbb{V}ar_\Theta\left[\bar{\boldsymbol{X}}^i | \bar{F}, Y, \boldsymbol{X}_{obs}^i\right] &= \sigma_{\bar{X}_i}^2\sigma_{\boldsymbol{e}_i}^2\left(I_T - Q_i'\left(Q_iQ_i'\right)^{-1}Q_i\right). \quad (5.70)
\end{aligned}$$

Proof:
First we rearrange (5.67), that is, we consider this equation for a univariate time series $1 \le i \le N$:

$$\begin{aligned}
\boldsymbol{X}^i &= \bar{F}\left(\bar{\boldsymbol{\Lambda}}_i^f\right)' + Y\left(\bar{\boldsymbol{\Lambda}}_i^y\right)' + \boldsymbol{E}^i \\
\Leftrightarrow \quad \frac{\bar{\boldsymbol{X}}^i - \mu_{\bar{X}_i}\mathbb{1}_T}{\sigma_{\bar{X}_i}} &= \bar{F}\left(\bar{\boldsymbol{\Lambda}}_i^f\right)' + Y\left(\bar{\boldsymbol{\Lambda}}_i^y\right)' + \boldsymbol{E}^i \\
\Leftrightarrow \quad \bar{\boldsymbol{X}}^i &= \mu_{\bar{X}_i}\mathbb{1}_T + \sigma_{\bar{X}_i}\bar{F}\left(\bar{\boldsymbol{\Lambda}}_i^f\right)' + \sigma_{\bar{X}_i} Y\left(\bar{\boldsymbol{\Lambda}}_i^y\right)' + \sigma_{\bar{X}_i}\boldsymbol{E}^i, \\
\Leftrightarrow \quad \boldsymbol{X}_{obs}^i &= Q_i\mu_{\bar{X}_i}\mathbb{1}_T + Q_i\sigma_{\bar{X}_i}\bar{F}\left(\bar{\boldsymbol{\Lambda}}_i^f\right)' + Q_i\sigma_{\bar{X}_i} Y\left(\bar{\boldsymbol{\Lambda}}_i^y\right)' + Q_i\sigma_{\bar{X}_i}\boldsymbol{E}^i,
\end{aligned}$$

where $\bar{\boldsymbol{\Lambda}}_i^f, \bar{\boldsymbol{\Lambda}}_i^y$ and $\boldsymbol{E}^i$ denote the i-th row of $\bar{\Lambda}^f$ and $\bar{\Lambda}^y$ and the i-th column of E, respectively. Because of $\boldsymbol{e}_t \sim \mathcal{N}(\boldsymbol{0}_N, \Sigma_{\boldsymbol{e}})$ iid, for all $1 \le i \le N$ we get $\boldsymbol{E}^i \sim \left(\boldsymbol{0}_T, \sigma_{\boldsymbol{e}_i}^2 I_T\right)$ resulting in:

$$\left.\begin{pmatrix} \bar{\boldsymbol{X}}^i \\ \boldsymbol{X}_{obs}^i \end{pmatrix}\right| \bar{F}, Y \sim \mathcal{N}\left(\begin{pmatrix} \mu_{\bar{X}_i}\mathbb{1}_T + \sigma_{\bar{X}_i}\bar{F}\left(\bar{\boldsymbol{\Lambda}}_i^f\right)' + \sigma_{\bar{X}_i} Y\left(\bar{\boldsymbol{\Lambda}}_i^y\right)' \\ Q_i\mu_{\bar{X}_i}\mathbb{1}_T + Q_i\sigma_{\bar{X}_i}\bar{F}\left(\bar{\boldsymbol{\Lambda}}_i^f\right)' + Q_i\sigma_{\bar{X}_i} Y\left(\bar{\boldsymbol{\Lambda}}_i^y\right)' \end{pmatrix}, \sigma_{\bar{X}_i}^2\sigma_{\boldsymbol{e}_i}^2\begin{pmatrix} I_T & Q_i' \\ Q_i & Q_iQ_i' \end{pmatrix}\right).$$

Eventually, the properties of the conditional multivariate normal distribution (Greene, 2003, pp. 871-872, Theorem B.7) prove the claim. □

With Lemma 5.1.19 in mind, we extend Algorithm 5.1.1 to incomplete panel data. To be precise, in each loop $(l) \geq 0$, we update for all $1 \leq i \leq N$ the univariate complete time series $\bar{\boldsymbol{X}}^i_{(l+1)}$ as follows:

$$\begin{aligned}\bar{\boldsymbol{X}}^i_{(l+1)} &= \mathbb{E}_{\Theta_{(l)}}\left[\bar{\boldsymbol{X}}^i|\bar{F}_{(l)}, Y, \boldsymbol{X}^i_{obs}\right] = \mu_{\bar{X}_i(l)}\mathbb{1}_T + \sigma_{\bar{X}_i(l)}\bar{F}_{(l)}\left(\bar{\boldsymbol{\Lambda}}^f_{i(l)}\right)' + \sigma_{\bar{X}_i(l)}Y\left(\bar{\boldsymbol{\Lambda}}^y_{i(l)}\right)' \\ &+ Q_i'\left(Q_iQ_i'\right)^{-1}\left[\boldsymbol{X}^i_{obs} - Q_i\left(\mu_{\bar{X}_i(l)}\mathbb{1}_T + \sigma_{\bar{X}_i(l)}\bar{F}_{(l)}\left(\bar{\boldsymbol{\Lambda}}^f_{i(l)}\right)' + \sigma_{\bar{X}_i(l)}Y\left(\bar{\boldsymbol{\Lambda}}^y_{i(l)}\right)'\right)\right]. \qquad (5.71)\end{aligned}$$

In our empirical application, we replace the univariate means $\mu_{\bar{X}_i(l)}$ and variances $\sigma^2_{\bar{X}_i(l)}$ by their empirical estimates. Similar to Algorithms 3.1.2 and 4.1.2, the updates (5.71) stop as soon as the absolute value of the relative change in the expected log-likelihood function $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l)}|X, C\right)|X, Y\right]$ from Lemma 5.1.9 between two consecutive iterations falls below the prespecified error threshold ξ.

As usual, we collect all findings in the form of an algorithm. In total, Algorithm 5.1.4 admits the estimation of FAVARs with incomplete panel data. After the general initialization, it jointly estimates the unknown model parameters, latent factor moments and missing data for fixed factor dimension K and autoregressive order p. In opposition to Algorithms 3.1.2 and 4.1.2, which clearly separate the construction of complete data from model selection and parameter estimation, Algorithm 5.1.4 estimates for fixed factor dimension and lag length the overall model based on incomplete panel data. Hence, there are no changes in (K, p) that may affect the termination of the EMs. This procedure performed well in our analyses, in particular, it provided more stable results. However, as in previous chapters, the termination criteria η and ξ still control the absolute value of relative changes in the expected log-likelihood instead of the model parameters. This also weakens the impact of the sample size on the termination of the algorithm. In case of the initialization the same argumentation as for Algorithms 3.1.2 and 4.1.2 holds.

Algorithm 5.1.4: Estimation of FAVARs with constraints in (5.15) for incomplete panel data

Set relative termination criteria $\eta > 0$ and $\xi > 0$;
Define upper limits of factor dimension $\bar{K} \geq 1$ and lag order $\bar{p} \geq 1$;
Initialize overall parameter set $\hat{\Theta}_{\text{ov}} = \emptyset$;
Initialize overall AIC by $\text{AIC}_{\text{ov}} = \infty$ (or any sufficiently large number);
for $K = 1$ **to** $\bar{K}$ **do**
  for $p = 1$ **to** $\bar{p}$ **do**
    ### **Initialization**
    for $i = 1$ **to** N **do**
      Initialize $\bar{\boldsymbol{X}}^i_{(0)}$ (if necessary, fill gaps);
      Define Q_i;
    end
    Determine standardized panel data $X_{(0)}$;
    Set loop $(l) = 0$;
    Initialize model parameters using PCA and OLS;
    Run EM in Lemma 5.1.7 with $X_{(l)}$, η, store (K, p) and estimated parameters $\hat{\Theta}_{(l)}$;
    Determine $\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l)}|X_{(l)}, C\right)|X_{(l)}, Y\right]$ in Lemma 5.1.9;
    for $i = 1$ **to** N **do**
      Use (5.71) such that $\bar{\boldsymbol{X}}^i_{(l+1)}$ is obtained;
    end
    Determine standardized panel data $X_{(l+1)}$;
    Run EM in Lemma 5.1.7 with $X_{(l+1)}$, η, store (K, p) and estimated parameters $\hat{\Theta}_{(l+1)}$;
    Determine $\mathbb{E}_{\hat{\Theta}_{(l+1)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l+1)}|X_{(l+1)}, C\right)|X_{(l+1)}, Y\right]$ in Lemma 5.1.9;
    ### **Alternating EMs**
    while $\dfrac{abs\left(\mathbb{E}_{\hat{\Theta}_{(l+1)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l+1)}|X_{(l+1)},C\right)|X_{(l+1)},Y\right]-\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l)}|X_{(l)},C\right)|X_{(l)},Y\right]\right)}{\frac{1}{2}\left(abs\left(\mathbb{E}_{\hat{\Theta}_{(l+1)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l+1)}|X_{(l+1)},C\right)|X_{(l+1)},Y\right]\right)+abs\left(\mathbb{E}_{\hat{\Theta}_{(l)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l)}|X_{(l)},C\right)|X_{(l)},Y\right]\right)\right)} > \xi$ **do**
      Set loop index $(l) = (l + 1)$;
      for $i = 1$ **to** N **do**
        Use (5.71) such that $\bar{\boldsymbol{X}}^i_{(l+1)}$ is obtained;
      end
      Determine standardized panel data $X_{(l+1)}$;
      Run EM in Lemma 5.1.7 with $X_{(l+1)}$, η, store (K, p) and estimated parameters $\hat{\Theta}_{(l+1)}$;
      Determine $\mathbb{E}_{\hat{\Theta}_{(l+1)}}\left[\mathcal{L}\left(\hat{\Theta}_{(l+1)}|X_{(l+1)}, C\right)|X_{(l+1)}, Y\right]$ in Lemma 5.1.9;
    end
    Determine temporary AIC, i.e., AIC_{tmp}, using Lemma 5.1.10;
    if $AIC_{tmp} < AIC_{ov}$ **then**
      Renew overall AIC value by $\text{AIC}_{\text{ov}} = \text{AIC}_{\text{tmp}}$;
      Update overall parameter set by $\hat{\Theta}_{\text{ov}} = \hat{\Theta}_{(l+1)}$;
    end
  end
end

5.1.6 Shock Analysis

FAVARs rank among VARs, therefore, we may apply standard tools for analyzing VARs such as Impulse Response Functions (IRFs) and Forecast Error Variance Decomposition (FEVD). For instance, we measure the effects of the US monetary policy on financial markets and the real economy in Section 5.3 with them. An advantage of FAVARs is that we can derive both for all observed data and hidden factors. This allows us to examine the reactions of the economy to structural shocks on a much broader set of variables compared to usual small-scale VARs. Algorithm 5.1.4 enables the estimation of FAVARs with incomplete panel data, this is why we also have IRFs and FEVDs for low-frequency indicators as GDP.

In the sequel, we follow Hamilton (1994) and Lütkepohl (2005) to obtain IRFs and FEVD for the FAVAR in (5.16)-(5.17). Thereby, we do not discuss loadings or VAR constraints, since IRFs and FEVD call for known model parameters and so, are calculated, after a FAVAR has been estimated. By similar reasoning we can argue that Algorithm 5.1.4 offers for each time series with missing elements or of lower frequency a complete analog. So, it is sufficient to derive IRFs and FEVD for FAVARs based on complete data.

Every covariance-stationary VAR has a MA(∞) representation of its errors, which is also known as Wold representation. Thus, the covariance-stationarity of the transition equation in (5.17) and Hamilton (1994, p. 260, Eq. 10.1.15) provide:

$$\begin{bmatrix} \bar{\boldsymbol{F}}_t \\ \boldsymbol{Y}_t \end{bmatrix} = \sum_{i=1}^{p} \bar{\Phi}_i \begin{bmatrix} \bar{\boldsymbol{F}}_{t-i} \\ \boldsymbol{Y}_{t-i} \end{bmatrix} + \bar{\boldsymbol{v}}_t \Rightarrow \begin{bmatrix} \bar{\boldsymbol{F}}_t \\ \boldsymbol{Y}_t \end{bmatrix} = \sum_{k=0}^{\infty} \bar{\Phi}^k \bar{\boldsymbol{v}}_{t-k}, \tag{5.72}$$

where for all $k \geq 0$ we define:

$$\bar{\Phi}^k = \begin{bmatrix} \bar{\Phi}_1 & \cdots & \bar{\Phi}_p \end{bmatrix} \begin{bmatrix} \bar{\Phi}^{k-1} \\ \vdots \\ \bar{\Phi}^{k-p} \end{bmatrix} \quad \text{with} \quad \bar{\Phi}^0 = I_{K+M} \quad \text{and} \quad \bar{\Phi}^{k-p} = O_{(K+M)\times(K+M)}, \forall k-p<0.$$

For the IRFs, we would like to have the (i, j)-th element of $\bar{\Phi}^s$ denoted by $\left(\bar{\Phi}^s\right)_{ij}$ as the impulse response:

$$\frac{\partial \begin{bmatrix} \bar{\boldsymbol{F}}_{t+s} \\ \boldsymbol{Y}_{t+s} \end{bmatrix}_i}{\partial \left(\bar{\boldsymbol{v}}_t\right)_j} = \frac{\partial \begin{bmatrix} \bar{\boldsymbol{F}}_{t} \\ \boldsymbol{Y}_{t} \end{bmatrix}_i}{\partial \left(\bar{\boldsymbol{v}}_{t-s}\right)_j} = \left(\bar{\Phi}^s\right)_{ij},$$

with $1 \leq i, j \leq K + M$. But, this requires the covariance matrix $\Sigma_{\bar{\boldsymbol{v}}}$ of the error terms $\bar{\boldsymbol{v}}_t$ to be diagonal such that the elements of $\bar{\boldsymbol{v}}_t$ are uncorrelated. Otherwise, the shocks cannot be uniquely identified. Here, we use a recursive identification scheme to achieve uncorrelated errors. Thereby, we assume that the first factor is only affected by its own innovation. Besides its own innovation, the second factor is affected by the first innovation. If we proceed similarly, all previous innovations drive the monetary policy variables in Section 5.3, since the observable vector $\boldsymbol{Y}_t$ is ordered last. In the below lemma we state the IRF for our rotated FAVAR in (5.16)-(5.17).

Lemma 5.1.20 (Impulse Response Functions for FAVARs)
Let $P \in \mathbb{R}^{M\times M}$ be a lower triangular matrix such that the Cholesky decomposition of the submatrix $\Sigma_{\boldsymbol{v}}^{yy}$ part of the covariance matrix $\Sigma_{\bar{\boldsymbol{v}}}$ in (5.17) is given by $\Sigma_{\boldsymbol{v}}^{yy} = PP'$. Then, the new innovations $\boldsymbol{\xi}_t \in \mathbb{R}^{K+M}$ defined by:

$$\boldsymbol{\xi}_t = \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix}^{-1} \bar{\boldsymbol{v}}_t, \tag{5.73}$$

are orthogonal and for all $1 \le i, j \le K+M$ *and* $1 \le n \le N$ *the impulse responses to their shocks obey:*

$$\frac{\partial \begin{bmatrix} \bar{\boldsymbol{F}}_{t+s} \\ \boldsymbol{Y}_{t+s} \end{bmatrix}_i}{\partial \left(\boldsymbol{\xi}_t\right)_j} = \frac{\partial \begin{bmatrix} \bar{\boldsymbol{F}}_{t} \\ \boldsymbol{Y}_{t} \end{bmatrix}_i}{\partial \left(\boldsymbol{\xi}_{t-s}\right)_j} = \left(\bar{\Phi}^s \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \right)_{ij}, \tag{5.74}$$

$$\frac{\partial \left(\boldsymbol{X}_{t+s}\right)_n}{\partial \left(\boldsymbol{\xi}_t\right)_j} = \frac{\partial \left(\boldsymbol{X}_{t}\right)_n}{\partial \left(\boldsymbol{\xi}_{t-s}\right)_j} = \left(\begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix} \bar{\Phi}^s \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \right)_{nj}. \tag{5.75}$$

Proof:
The upper left submatrix of the covariance matrix $\Sigma_{\bar{\boldsymbol{v}}}$ in (5.17) is the identity matrix, whereas its upper right and lower left submatrices are zero matrices. Hence, this special structure simplifies our problem such that only the submatrix $\Sigma_{\boldsymbol{v}}^{yy}$ has to be adjusted to achieve the desired recursive scheme. Let P be the lower triangular matrix arising from the Cholesky decomposition of $\Sigma_{\boldsymbol{v}}^{yy}$. Then, the new shocks $\boldsymbol{\xi}_t$ are orthogonal, since we have for their covariance matrix:

$$\begin{aligned}\Sigma_{\boldsymbol{\xi}} &= \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix}^{-1} \Sigma_{\bar{\boldsymbol{v}}} \left(\begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix}^{-1} \right)' = \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P^{-1} \end{bmatrix} \Sigma_{\bar{\boldsymbol{v}}} \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & \left(P'\right)^{-1} \end{bmatrix} \\ &= \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P^{-1}\Sigma_{\boldsymbol{v}}^{yy} \end{bmatrix} \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & \left(P'\right)^{-1} \end{bmatrix} = \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P^{-1}\Sigma_{\boldsymbol{v}}^{yy}\left(P'\right)^{-1} \end{bmatrix} = I_{(K+M)}.\end{aligned}$$

Next, we rewrite the MA(∞) representation in (5.72) as follows:

$$\begin{bmatrix} \bar{\boldsymbol{F}}_t \\ \boldsymbol{Y}_t \end{bmatrix} = \sum_{k=0}^{\infty} \bar{\Phi}^k \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix}^{-1} \bar{\boldsymbol{v}}_{t-k} = \sum_{k=0}^{\infty} \bar{\Phi}^k \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \boldsymbol{\xi}_{t-k}, \tag{5.76}$$

which results in the solution in (5.74). For the IRFs of the panel data $\boldsymbol{X}_t$, the MA(∞) representation in (5.76) and the observation equation in (5.16) lead us to:

$$\boldsymbol{X}_t = \begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix} \begin{pmatrix} \bar{\boldsymbol{F}}_t \\ \boldsymbol{Y}_t \end{pmatrix} + \boldsymbol{e}_t = \begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix} \left(\sum_{k=0}^{\infty} \bar{\Phi}^k \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \boldsymbol{\xi}_{t-k} \right) + \boldsymbol{e}_t. \tag{5.77}$$

Eventually, the partial derivatives of $\boldsymbol{X}_t$ in (5.77) with respect to the shocks $\boldsymbol{\xi}_{t-s}$ yield (5.75). □

Due to Lemma 5.1.20, the IRFs of $\begin{bmatrix} \bar{\boldsymbol{F}}_t', \boldsymbol{Y}_t' \end{bmatrix}$ and $\boldsymbol{X}_t$ with respect to $\boldsymbol{\xi}_{t-s}$ are received by plotting (5.74) or (5.75). Before we continue with the FEVD, we are interested in the shock weight $\boldsymbol{z} \in \mathbb{R}^{K+M}$, ceteris paribus, causing a change in the vector $\boldsymbol{Y}_T$ of size $\Delta\boldsymbol{Y}$. For instance, which shock increases the Effective Federal Funds Rate in Section 5.3 by 25 basis points (bps) at point in time $s=0$, but leaves everything else as it is. For this purpose, the vector $\boldsymbol{z}$ has to satisfy:

$$\begin{bmatrix} \bar{\boldsymbol{F}}_T \\ D_Y^{-1}\left(\boldsymbol{Y}_T + \Delta\boldsymbol{Y} - \boldsymbol{\mu}_Y\right) \end{bmatrix} = \bar{\Phi}^0 \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \boldsymbol{z},$$

where the vector $\Delta\boldsymbol{Y} \in \mathbb{R}^M$ contains the planned changes in $\boldsymbol{Y}_T$. The vector $\boldsymbol{\mu}_Y \in \mathbb{R}^M$ collects the means of the observed variables, while the diagonal matrix $D_Y = \operatorname{diag}\left(\sigma_{Y_1}, \ldots, \sigma_{Y_M}\right) \in \mathbb{R}^{M\times M}$ summarizes their standard deviations. Hence, the standardized shock $\boldsymbol{z}$ is given by:

$$\boldsymbol{z} = \begin{bmatrix} \bar{\boldsymbol{F}}_T \\ P^{-1}D_Y^{-1}\left(\boldsymbol{Y}_T + \Delta\boldsymbol{Y} - \boldsymbol{\mu}_Y\right) \end{bmatrix}. \tag{5.78}$$

Finally, we derive the FEVD of the FAVAR in (5.16)-(5.17) assigning the contributions of the innovations to the forecasting error of a variable for a given time horizon. As for IRFs, FAVARs admit the FEVD for all variables. In the below lemma we state the FEVD for our rotated FAVAR in (5.16)-(5.17).

Lemma 5.1.21 (Forecast Error Variance Decomposition for FAVARs)
For any $1 \leq i \leq N, 1 \leq j \leq K+M$ and $s > 0$, the proportion of the j-th innovation regarding the s-step ahead forecast error variance of the i-th element of variable $\boldsymbol{X}_t$ is given by:

$$\omega_{ij}^s = \frac{\sum_{k=0}^{s-1} \left(\left(\begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix}_i \bar{\Phi}^k \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \right)_j \right)^2}{\sum_{j=1}^{K+M} \sum_{k=0}^{s-1} \left(\left(\begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix}_i \bar{\Phi}^k \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \right)_j \right)^2 + (\Sigma_{\boldsymbol{e}})_{ii}}. \tag{5.79}$$

Proof:
Following Lütkepohl (2005, pp. 63-66) and Bork (2009), we receive the FEVD of the measurement equation from the MA(∞) representation in (5.77). For the optimal s-step ahead forecast $\hat{\boldsymbol{X}}_{t+s}$ we have:

$$\hat{\boldsymbol{X}}_{t+s} = \begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix} \sum_{k=s}^{\infty} \bar{\Phi}^k \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \boldsymbol{\xi}_{t-k+s}.$$

Therefore, the error of the optimal s-step ahead forecast $\boldsymbol{err}_s$ is:

$$\boldsymbol{err}_s = \boldsymbol{X}_{t+s} - \hat{\boldsymbol{X}}_{t+s} = \begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix} \sum_{k=0}^{s-1} \bar{\Phi}^k \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \boldsymbol{\xi}_{t-k+s} + \boldsymbol{e}_{t+s}.$$

Hence, this means for its i-th component:

$$\begin{aligned}
(\boldsymbol{err}_s)_i &= \begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix}_i \sum_{k=0}^{s-1} \bar{\Phi}^k \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \boldsymbol{\xi}_{t-k+s} + (\boldsymbol{e}_{t+s})_i \\
&= \sum_{k=0}^{s-1} \left(\begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix}_i \bar{\Phi}^k \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \boldsymbol{\xi}_{t-k+s} \right) + (\boldsymbol{e}_{t+s})_i \\
&= \sum_{k=0}^{s-1} \sum_{j=1}^{K+M} \left(\left(\begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix}_i \bar{\Phi}^k \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \right)_j \left(\boldsymbol{\xi}_{t-k+s}\right)_j \right) + (\boldsymbol{e}_{t+s})_i \\
&= \sum_{j=1}^{K+M} \sum_{k=0}^{s-1} \left(\left(\begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix}_i \bar{\Phi}^k \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \right)_j \left(\boldsymbol{\xi}_{t-k+s}\right)_j \right) + (\boldsymbol{e}_{t+s})_i\,,
\end{aligned}$$

where $\begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix}_i$ denotes the i-th row of the loadings matrix and $(\cdot)_j$ is the j-th element of the respective vector. The last equation clearly confirms that the forecast error of the i-th entry of $\boldsymbol{X}_t$ is driven by all shocks $\left(\boldsymbol{\xi}_{t-k+s}\right)_j$ with $0 \leq k \leq s-1$ and $1 \leq j \leq K+M$. Perhaps, some innovations $\left(\boldsymbol{\xi}_{t-k+s}\right)_j$ have no impact, if the corresponding elements of the loadings matrix are zeros.

The errors $\left(\boldsymbol{\xi}_{t-k+s}\right)_j$ are uncorrelated and multivariate standard normally distributed. So, the algebraic formula for variances provides for the Mean-Squared Error of $(\boldsymbol{err}_s)_i$ denoted by $MSE((\hat{\boldsymbol{X}}_{t+s})_i)$:

$$MSE\left(\left(\hat{\boldsymbol{X}}_{t+s}\right)_i\right) = \mathbb{E}_\Theta\left[(\boldsymbol{err}_s)_i^2\right] = \mathbb{V}\mathrm{ar}_\Theta\left[(\boldsymbol{err}_s)_i\right] + \underbrace{\mathbb{E}_\Theta\left[(\boldsymbol{err}_s)_i\right]^2}_{=0} = \mathbb{V}\mathrm{ar}_\Theta\left[(\boldsymbol{err}_s)_i\right]$$

$$= \sum_{j=1}^{K+M} \sum_{k=0}^{s-1} \left(\left(\begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix}_i \bar{\Phi}^k \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \right)_j \right)^2 + (\Sigma_{\boldsymbol{e}})_{ii} \, .$$

Thus, the contribution of the j-th innovation to the forecast error variance of $(\boldsymbol{X}_{t+s})_i$ is given by:

$$\sum_{k=0}^{s-1} \left(\left(\begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix}_i \bar{\Phi}^k \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \right)_j \right)^2 .$$

In relative terms the above contribution is equal to the ratio in (5.79). □

Note, the FEVD is often tabulated or plotted as function of the forecast horizon s for fixed dimensions i, j. For an example, see Figures 5.5-5.8.

5.2 Simulation Study

In the scope of a MC simulation study, we now compare the estimation accuracy of our two-step estimation method using the modified KF and KS from Section 5.1 and three alternative approaches. Besides a non-parametric ansatz based on PCA and OLS, we test two parametric estimation methods treating FAVARs as special ADFMs more precisely described in the sequel. For all procedures, an outer EM reconstructs complete panel data from the observations and latest parameter estimates. In this manner, we concentrate on the estimation quality of the modified KF, but also address the issue of incomplete panel data.

As in Section 4.4, we first describe how we generate our sample data. For $a, b \in \mathbb{R}$ with $a < b$, let $\mathcal{U}(a, b)$ denote the uniform distribution on the interval $[a, b]$, while $\operatorname{diag}(\boldsymbol{z}) \in \mathbb{R}^{N\times N}$ is a diagonal matrix with elements $\boldsymbol{z} = [z_1, \ldots, z_N] \in \mathbb{R}^N$. Furthermore, let $V_i \in \mathbb{R}^{(K+M)\times(K+M)}, 1 \le i \le p, V_{\boldsymbol{v}} \in \mathbb{R}^{(K+M)\times(K+M)}$ and $V_{\boldsymbol{e}} \in \mathbb{R}^{N\times N}$ represent arbitrary orthonormal matrices for fixed dimensions (T, N, K, M, p). Then, we obtain the parameters of the FAVARs in Definition 5.1.1 as follows:

$$\begin{aligned}
\Phi_i &= V_i \operatorname{diag}\left(\frac{z_{i,1}}{i}, \ldots, \frac{z_{i,K+M}}{i}\right)(V_i'), & z_{i,j} &\sim \mathcal{U}(0.25, 0.75) \operatorname{iid}, 1 \le i \le p, 1 \le j \le K+M, \\
\Sigma_{\boldsymbol{v}} &= V_{\boldsymbol{v}} \operatorname{diag}(z_{\boldsymbol{v},1}, \ldots, z_{\boldsymbol{v},K+M})(V_{\boldsymbol{v}}'), & z_{\boldsymbol{v},j} &\sim \mathcal{U}(0.75, 1.25) \operatorname{iid}, 1 \le j \le K+M, \\
\Lambda &= (\lambda_{n,j})_{n,j}, & \lambda_{n,j} &\sim \mathcal{U}(0, 1) \operatorname{iid}, 1 \le n \le N, 1 \le j \le K+M, \\
\Sigma_{\boldsymbol{e}} &= V_{\boldsymbol{e}} \operatorname{diag}(z_{\boldsymbol{e},1}, \ldots, z_{\boldsymbol{e},N})(V_{\boldsymbol{e}}'), & z_{\boldsymbol{e},n} &\sim \mathcal{U}(0.5, 1.5) \operatorname{iid}, 1 \le n \le N.
\end{aligned} \tag{5.80}$$

Hence, the parameters in (5.80) specify the general FAVAR formulation in Definition 5.1.1 instead of its rotated simplification in (5.16)-(5.17). To ensure covariance-stationarity of the factor process $\{[\boldsymbol{F}_t', \boldsymbol{Y}_t']'\}$, we check, if all matrices $\Phi_i, 1 \le i \le p$, satisfy the conditions in Lemma A.2.3, before we proceed. If not, the matrices $\Phi_i, 1 \le i \le p$, are drawn again. To prevent us from the case that we implicitly construct matrices Φ_i, whose eigenvalues are close to zero, their eigenvalues are taken from the range of $[0.25/i, 0.75/i]$. Here, the division by i reduces the impact of lagged factors. As before, the restriction to matrices Φ_i with positive eigenvalues and the division by i are made for simplicity and so, can be removed. In contrast to Section 4.4, the eigenvalues of matrix $\Sigma_{\boldsymbol{v}}$ lie in the range of $[0.75, 1.25]$ instead of $[0.25, 0.50]$, which yields more noise. With the help of (5.2), we construct the factor sample $[F, Y] \in \mathbb{R}^{T\times(K+M)}$. The univariate time series in Y are assumed to have zero mean and variance of one. This is why we standardize all univariate time series in $[F, Y]$ and adjust the matrices $\Phi_i, 1 \le i \le p$, and $\Sigma_{\boldsymbol{v}}$ accordingly. Next, we simulate the panel data $X \in \mathbb{R}^{T\times N}$ based on (5.1). Thereby, we restrict ourselves to matrices W and $\Sigma_{\boldsymbol{e}}$ of full column rank. Further, we standardize all univariate time series in X, since those are again supposed to have zero mean and variance of one, and adapt the matrices W and $\Sigma_{\boldsymbol{e}}$ correspondingly.

At this stage, we have complete panel data X. For incomplete panel data, we pursue the same approach as in Section 4.4. That is, let $\rho_m \in [0,1]$ be the ratio of gaps. Then, we randomly delete $\lceil \rho_m T \rceil$ elements from each times series serving as stock variable to get a scattered pattern. By contrast, we aggregate the entries of each times series, which represents a flow or change in flow variable, as given in Definition 2.2.2 or Lemma 2.2.3. This results in observations at times $t = \lceil 1 + s/(1-\rho_m) \rceil$ with $0 \leq s \leq \lfloor (T-1)(1-\rho_m) \rfloor$ and $s \in \mathbb{N}_0$ such that we receive a regular pattern for flow and change in flow variables. None of the four methods estimates hidden factors for points in time without any observation. Therefore, we always check, whether the obtained matrix of incomplete panel data comprises an empty row, before we proceed. If so, we reapply our routine for preparing incomplete data to the full panel data X. Furthermore, we combine $\lceil N/2 \rceil$ stock and $\lfloor N/2 \rfloor$ flow (change in flow) variables in the second (third) column of Tables 5.1-5.7 to avoid such cases, where the incomplete panel data matrix has an empty row.

Tables 5.1-5.7 compare four methods for estimating the hidden factors F. In doing so, the trace R^2 from Definition A.3.1 evaluates the quality of the estimated factors. Here, we focus on the hidden factors, since the variables Y are observed in full and therefore, do not call for estimation. Note, all trace R^2 means in Tables 5.1-5.4 and thus, all ratios of trace R^2 means in Tables 5.5-5.7 arise from the same MC simulations. To be precise, the same combination of incomplete panel data and observed variables Y enters all four estimation methods to ensure that our results are comparable.

In addition, all four estimation methods have the same outer EM, which applies the updates in (5.71) to construct complete panel data from the observations and latest parameter estimates. In this context, for all estimation methods the updates in (5.71) stop, as soon as the absolute value of the relative change in the expected log-likelihood function in (5.41) is below 10^{-2}. To guarantee that all approaches have the same starting conditions, we initialize $\bar{X}_{(0)}$, i.e., the first guess of the complete panel data, for all in the same way. That is, for each univariate time series in the panel data, we fill its gaps by the empirical mean of its observations.

Finally, in Tables 5.1-5.7, the dimension of the hidden factors, i.e., K, and the lag order of the joint factor dynamics, i.e., p, are supposed to be known. So, we exclude any impact of model selection on the quality of the estimated factors. Next, we describe the alternative estimation methods in more detail, before we discuss the trace R^2 means in Tables 5.1-5.7.

In Table 5.1, we estimate the simulated FAVARs with the non-parametric method of Boivin and Giannoni (2008) and Boivin et al. (2010). That is, let $\bar{X}_{(l)}$ be the complete panel data of the outer EM constructed in loop $(l-1)$ from the observations and latest parameter estimates. Then, we run the following algorithm.

Algorithm 5.2.1: Estimate FAVARs with PCA and OLS based on complete panel data

Determine $\hat{F} = [\boldsymbol{F}_1, \ldots, \boldsymbol{F}_T]'$ as first K principal components of $\bar{X}_{(l)}$;
An OLS of $\bar{X}_{(l)}$ on $\hat{F}$ and Y provides $\hat{\Lambda}^f$ and $\hat{\Lambda}^y$;
Remove impact of Y on $\bar{X}_{(l)}$ by $\tilde{X}_{(l)} = \bar{X}_{(l)} - Y(\hat{\Lambda}^y)'$;
Reestimte $\hat{F} = [\boldsymbol{F}_1, \ldots, \boldsymbol{F}_T]'$ as first K principal components of $\tilde{X}_{(l)}$;
An OLS of $\bar{X}_{(l)}$ on $\hat{F}$ and Y provides new $\hat{\Lambda}^f$ and $\hat{\Lambda}^y$;
Compute $\hat{\Sigma}_{\boldsymbol{e}} = (E')E/T$ with $E = \bar{X}_{(l)} - F(\hat{\Lambda}^f)' - Y(\hat{\Lambda}^y)'$;
An OLS on $[\hat{F}, Y]$ results in $\hat{\Phi}_i, 1 \leq i \leq p$, and $\hat{\Sigma}_{\boldsymbol{v}}$;

Note, Algorithm 5.2.1 applies PCA and OLS several times to enforce Y as observed factors. Here, we do not have the non-parametric estimation method of Bernanke et al. (2005) as benchmark approach, since there is no distinction between slow- and fast-moving variables for simulated data. Hence, the missing economic meaning of simulated data matters in this context.

In Tables 5.2 and 5.3, the inner EM coincides with the EM in Bork (2009). In doing so, let $\vec{X}_{(l)}$ denote the complete panel data of the outer EM constructed in loop $(l-1)$. In addition, let $\boldsymbol{X}_t^{1:K}$ and $\boldsymbol{X}_t^{K+1:N}$ be the first K and last $N-K$ entries of vector $\boldsymbol{X}_t$. Next, we add the observed variables $\boldsymbol{Y}_t$ to the panel data and modify the FAVAR from Definition 5.1.1 as follows:

$$\underbrace{\begin{bmatrix} \boldsymbol{X}_t^{1:K} \\ \boldsymbol{Y}_t \\ \boldsymbol{X}_t^{K+1:N} \end{bmatrix}}_{\vec{\boldsymbol{X}}_t} = \underbrace{\begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & I_M \\ \bar{\Lambda}^f_{(N-K)\times K} & \bar{\Lambda}^y_{(N-K)\times M} \end{bmatrix}}_{\vec{\Lambda}} \underbrace{\begin{bmatrix} \boldsymbol{F}_t \\ \boldsymbol{Y}_t \end{bmatrix}}_{\boldsymbol{C}_t} + \vec{\boldsymbol{\epsilon}}_t, \vec{\boldsymbol{\epsilon}}_t \sim \mathcal{N}\left(\boldsymbol{0}_{N+M}, \Sigma_{\vec{\boldsymbol{\epsilon}}}\right) \text{iid}, \tag{5.81}$$

$$\boldsymbol{C}_t = \sum_{i=1}^{p} \Phi_i \boldsymbol{C}_{t-i} + \boldsymbol{v}_t,\ \boldsymbol{v}_t \sim \mathcal{N}\left(\boldsymbol{0}_{K+M}, \Sigma_{\boldsymbol{v}}\right) \text{iid}, \tag{5.82}$$

with submatrices $\bar{\Lambda}^f_{(N-K)\times K}$ and $\bar{\Lambda}^y_{(N-K)\times M}$ as the unconstrained last $N-K$ rows of $\vec{\Lambda}$. The restrictions in the first K rows of the loadings matrices in (5.81) and (5.13) coincide to ensure unique parameters for all non-parametric estimation methods in Tables 5.2-5.3. The enlarged panel data $\vec{\boldsymbol{X}}_t$ calls for additional $M\,(K+M)$ loadings restrictions forcing the variables $\boldsymbol{Y}_t$ part of $\vec{\boldsymbol{X}}_t$ to become a factor in $\boldsymbol{C}_t$. For this purpose, $[O_{M\times K}, I_M]$ in rows $K+1$ to $K+M$ of $\vec{\Lambda}$ is a natural choice. In this way, Bork (2009) ignored that $\boldsymbol{Y}_t$ part of $\boldsymbol{C}_t$ is observed, treated the overall vector $\boldsymbol{C}_t$ as hidden and estimated the FAVAR in (5.81)-(5.82) as ADFM.

Here, we pursue the same approach. That is, the inner EM of the estimation methods in Tables 5.2-5.3 considers the FAVAR in (5.81)-(5.82) as ADFM with the corresponding loadings constraints. Whenever we estimate the moments of the hidden factors $\boldsymbol{C}_t$ using the standard KF, KS and Lag-One Autocovariance Smoother in Lemmata 2.1.8-2.1.10, we rewrite the ADFM in (5.81)-(5.82) with VAR(p) factor dynamics as ADFM with VAR(1) factor dynamics. Otherwise, we keep the general VAR(p) factor dynamics. As for all parametric estimation methods, the inner EMs terminate as soon as the absolute value of the relative change in the expected log-likelihood function is below 10^{-2}. Moreover, all inner EMs deploy PCA and OLS for parameter initialization. For the initialization of the KF, the covariance-stationary covariance matrix of the factors serves as starting point $\hat{P}^C_{p|p}$ or $\hat{P}^F_{p|p}$ in Lemma 2.1.8. Thereby, we truncate their MA(∞) representations as soon as the absolute value of the relative contribution of a new term is below 10^{-6}. Note that the estimation methods in Tables 5.2-5.3 do not assume the specially designed covariance matrix $\Sigma_{\bar{\boldsymbol{v}}}$ of the rotated FAVAR from (5.16)-(5.17).

The inner EMs of the estimation methods in Tables 5.2-5.3 are the same. So, we obtain the same smoothed factor means $\hat{\boldsymbol{C}}_{t|T}, 1 \le t \le T$, for both cases. As mentioned before, the reconstruction formula in (5.71) defines the outer EMs. However, in this regard, there is a difference between Tables 5.2 and 5.3. In Table 5.2, the updates completely rely on the estimated means $\hat{\boldsymbol{C}}_{t|T}, 1 \le t \le T$, and thus, neglect the observed variables Y. That means, for variable $1 \le i \le N+M$ and loop $l \ge 0$ we have:

$$\vec{\boldsymbol{X}}^i_{(l+1)} = \mu_{\vec{X}_i(l)} \mathbb{1}_T + \sigma_{\vec{X}_i(l)} \hat{C}_{(l)} \left(\vec{\boldsymbol{\Lambda}}_{i(l)}\right)' + Q_i' \left(Q_i Q_i'\right)^{-1} \left[\boldsymbol{X}^i_{obs} - Q_i \left(\mu_{\vec{X}_i(l)} \mathbb{1}_T + \sigma_{\vec{X}_i(l)} \hat{C}_{(l)} \left(\vec{\boldsymbol{\Lambda}}_{i(l)}\right)'\right)\right],$$

where $\hat{C}_{(l)} = \left[\hat{\boldsymbol{C}}_{1|T}, \ldots, \hat{\boldsymbol{C}}_{T|T}\right]' \in \mathbb{R}^{T\times(K+M)}$ are the estimated factor means, which a run of the standard KF and KS in loop (l) provided. By contrast, the panel data updates in Table 5.3 use the reconstruction formula in (5.71), with $\bar{F}_{(l)}$ as the first K columns of $\hat{C}_{(l)}$ and Y as the actual observations. This implies that the last M columns of $\hat{C}_{(l)}$ are discarded.

Eventually, Table 5.4 displays the trace R^2 means for the two-step estimation method from Section 5.1. To be precise, we estimate the rotated FAVAR in (5.16)-(5.17) with loadings constraints in (5.15). Although our MC simulations provide general FAVARs, we consider these special FAVARs to demonstrate that our model transformations do not exceed the class of possible linear factor transformations.

As already mentioned, the trace R^2 means in Tables 5.1-5.4 and so, the ratios of trace R^2 means in Tables 5.5-5.7 come from the same MC simulations. Moreover, all tables cover the same tuples (T, N, K, M, p). In this regard, we examine $T \in \{600, 700, 800\}$, $N \in \{80, 100, 120\}$ and $\rho_m \in \{0, 0.05, 0.10, 0.15\}$, which are close to the data dimensions of the empirical study in Section 5.3. That is, $N = 108, T = 682, \rho_m = 0.07$ and $M = 3$. For simplicity, we treat factor dimensions $K \in \{1, 3\}$ and $M \in \{1, 3\}$ for lag orders $p \in \{1, 2\}$.

A comparison of Tables 5.1-5.4 shows: First, irrespective of the estimation method, there are no obvious differences between the trace R^2 means of the three data types. Second, a higher percentage of data gaps, ceteris paribus, deteriorates the trace R^2 means. As the current settings for T, N and ρ_m are more close together than those in Section 4.4, the described patterns might be less clear than in Section 4.4. Third, longer samples, i.e., larger T, improve the trace R^2 means. The same holds for panel data covering more variables, i.e., larger N. Fourth, higher lag orders improve the trace R^2 means, which is rather surprising. So far, all findings are in place for all four estimation methods.

However, there are some differences between them. First, the estimation methods in Tables 5.1-5.3 do not explicitly take into account that the variables $\boldsymbol{Y}_t$ are observed. Thus, they offer some kind of work-around solution. For instance, the non-parametric approach repeatedly applies PCA and OLS for separating the impacts of $\boldsymbol{Y}_t$ and $\boldsymbol{F}_t$ on $\boldsymbol{X}_t$ from each other. In this context, the dimensions of the vectors $\boldsymbol{Y}_t$ and $\boldsymbol{F}_t$ matter. With a view to Tables 5.1-5.3, the pairs $(K = 1, M = 1)$ and $(K = 3, M = 3)$ have smaller trace R^2 means than the combination $(K = 3, M = 1)$. By contrast, the estimation method with our modified KF in Table 5.4 offers for $(K = 1, M = 1, p = 1)$ larger trace R^2 means than for $(K = 3, M = 1, p = 1)$. Second, the trace R^2 means in Table 5.4 are generally better than their counterparts in Tables 5.1-5.3. To verify this Tables 5.5-5.7 display the corresponding ratios of trace R^2 means. Thereby, ratios larger than one confirm that the estimation method based on our modified KF outperforms the respective alternative. Note that all ratios in Tables 5.5-5.7 are larger than one, but for the previously mentioned combinations $(K = 1, M = 1)$ and $(K = 3, M = 3)$ they exceed one by far. This clearly highlights, why it makes sense to take into account that the variables $\boldsymbol{Y}_t$ represent observed factors.

Model selection with AIC is quite common in the literature. Therefore, we do not address the selection of the factor dimension K and autoregressive order p in the scope of this MC simulation study. Since the variables $\boldsymbol{Y}_t$ are observable, their dimension M is known a priori and does not call for being estimated.

Table 5.1: *Means of trace R^2 based on hidden factors for random FAVARs using PCA and OLS*

		stock[a]				**stock/flow (average)**[b]				**stock/change in flow (average)**[c]			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%
$K=1, M=1, p=1$													
80	600	0.49	0.49	0.48	0.49	0.49	0.49	0.48	0.50	0.49	0.49	0.49	0.50
80	700	0.50	0.48	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
80	800	0.49	0.49	0.50	0.49	0.50	0.49	0.48	0.50	0.49	0.49	0.49	0.48
100	600	0.49	0.50	0.50	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
100	700	0.49	0.50	0.50	0.49	0.50	0.50	0.50	0.49	0.50	0.49	0.49	0.50
100	800	0.50	0.50	0.49	0.49	0.49	0.50	0.49	0.49	0.49	0.49	0.50	0.50
120	600	0.50	0.49	0.50	0.50	0.50	0.49	0.48	0.50	0.50	0.49	0.50	0.48
120	700	0.49	0.50	0.50	0.49	0.49	0.49	0.49	0.50	0.49	0.49	0.50	0.48
120	800	0.49	0.50	0.49	0.50	0.49	0.49	0.50	0.49	0.49	0.49	0.49	0.49
$K=3, M=1, p=1$													
80	600	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.73	0.74	0.74	0.73	0.73
80	700	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.73	0.74	0.74	0.73	0.72
80	800	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.73	0.74	0.74	0.73	0.73
100	600	0.75	0.76	0.76	0.75	0.76	0.75	0.75	0.75	0.76	0.75	0.75	0.74
100	700	0.76	0.75	0.75	0.75	0.75	0.75	0.75	0.74	0.75	0.75	0.75	0.74
100	800	0.75	0.76	0.75	0.75	0.75	0.75	0.75	0.74	0.75	0.75	0.75	0.74
120	600	0.76	0.77	0.76	0.76	0.77	0.77	0.76	0.76	0.77	0.77	0.76	0.75
120	700	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.75	0.76	0.76	0.76	0.75
120	800	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.75	0.76	0.76	0.75	0.75
$K=3, M=3, p=1$													
80	600	0.55	0.56	0.57	0.56	0.56	0.56	0.56	0.56	0.55	0.56	0.56	0.55
80	700	0.55	0.55	0.56	0.57	0.55	0.55	0.56	0.56	0.55	0.56	0.56	0.55
80	800	0.56	0.56	0.56	0.56	0.55	0.56	0.55	0.56	0.55	0.55	0.55	0.55
100	600	0.57	0.56	0.57	0.56	0.56	0.57	0.56	0.56	0.56	0.57	0.57	0.56
100	700	0.56	0.57	0.56	0.57	0.56	0.56	0.57	0.56	0.57	0.57	0.56	0.56
100	800	0.56	0.56	0.57	0.57	0.56	0.57	0.57	0.56	0.56	0.56	0.56	0.55
120	600	0.57	0.57	0.57	0.58	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.56
120	700	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.56
120	800	0.56	0.57	0.57	0.57	0.57	0.56	0.57	0.57	0.57	0.57	0.57	0.56
$K=3, M=1, p=2$													
80	600	0.79	0.79	0.79	0.78	0.79	0.79	0.78	0.78	0.79	0.79	0.78	0.77
80	700	0.79	0.79	0.79	0.78	0.80	0.79	0.78	0.78	0.79	0.79	0.78	0.77
80	800	0.79	0.79	0.78	0.78	0.79	0.79	0.78	0.78	0.79	0.79	0.78	0.78
100	600	0.80	0.80	0.80	0.79	0.80	0.80	0.80	0.80	0.81	0.81	0.79	0.79
100	700	0.81	0.81	0.80	0.80	0.80	0.81	0.80	0.79	0.80	0.80	0.80	0.79
100	800	0.81	0.80	0.80	0.80	0.81	0.80	0.80	0.80	0.80	0.80	0.79	0.79
120	600	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.80	0.81	0.81	0.80	0.80
120	700	0.81	0.82	0.81	0.81	0.81	0.81	0.81	0.80	0.81	0.81	0.81	0.80
120	800	0.82	0.82	0.81	0.81	0.82	0.81	0.81	0.80	0.81	0.81	0.81	0.80
$K=3, M=3, p=2$													
80	600	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65
80	700	0.65	0.65	0.65	0.64	0.65	0.65	0.66	0.65	0.65	0.65	0.65	0.64
80	800	0.65	0.65	0.65	0.65	0.66	0.66	0.65	0.65	0.65	0.65	0.65	0.65
100	600	0.66	0.66	0.66	0.65	0.66	0.66	0.66	0.66	0.66	0.66	0.65	0.66
100	700	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.65	0.66	0.66	0.66	0.65
100	800	0.67	0.66	0.66	0.67	0.66	0.66	0.66	0.65	0.66	0.66	0.66	0.66
120	600	0.67	0.67	0.67	0.67	0.67	0.67	0.66	0.66	0.67	0.67	0.66	0.66
120	700	0.66	0.66	0.67	0.67	0.66	0.66	0.66	0.66	0.67	0.67	0.66	0.66
120	800	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.66	0.67	0.67	0.67	0.66

The displayed means are derived from 500 MC simulations for known dimensions K and p.

[a] For incomplete time series a stock variable is assumed.
[b] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series are stock and flow (average formulation) variables, respectively.
[c] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series serve as stock or change in flow (average formulaton) variables.

Table 5.2: *Means of trace R^2 based on hidden factors for random FAVARs using standard KF and KS, when complete panel data relies on estimated factors instead of observed variables*

		stock[a]				**stock/flow (average)**[b]				**stock/change in flow (average)**[c]			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%
$K=1, M=1, p=1$													
80	600	0.46	0.46	0.45	0.46	0.46	0.46	0.45	0.44	0.46	0.47	0.45	0.43
80	700	0.47	0.45	0.46	0.47	0.46	0.46	0.45	0.45	0.47	0.46	0.44	0.43
80	800	0.47	0.46	0.47	0.46	0.47	0.47	0.45	0.45	0.47	0.46	0.43	0.43
100	600	0.46	0.47	0.48	0.46	0.47	0.47	0.44	0.40	0.47	0.47	0.41	0.40
100	700	0.47	0.47	0.47	0.47	0.48	0.47	0.44	0.41	0.48	0.46	0.43	0.42
100	800	0.48	0.48	0.46	0.47	0.46	0.47	0.44	0.41	0.46	0.47	0.43	0.40
120	600	0.48	0.47	0.48	0.47	0.48	0.47	0.42	0.42	0.48	0.46	0.42	0.40
120	700	0.47	0.48	0.48	0.47	0.47	0.47	0.41	0.41	0.47	0.46	0.42	0.42
120	800	0.47	0.48	0.47	0.48	0.47	0.47	0.45	0.40	0.47	0.46	0.42	0.42
$K=3, M=1, p=1$													
80	600	0.68	0.67	0.67	0.66	0.68	0.67	0.65	0.63	0.68	0.67	0.64	0.56
80	700	0.68	0.67	0.66	0.66	0.68	0.67	0.66	0.63	0.68	0.67	0.64	0.56
80	800	0.68	0.67	0.67	0.66	0.68	0.67	0.66	0.63	0.68	0.67	0.64	0.57
100	600	0.70	0.69	0.69	0.67	0.70	0.69	0.67	0.63	0.70	0.69	0.66	0.56
100	700	0.70	0.69	0.69	0.68	0.69	0.69	0.68	0.64	0.69	0.69	0.65	0.57
100	800	0.69	0.69	0.68	0.68	0.70	0.69	0.68	0.63	0.70	0.69	0.66	0.57
120	600	0.71	0.70	0.70	0.69	0.71	0.71	0.69	0.63	0.71	0.70	0.66	0.56
120	700	0.71	0.70	0.70	0.69	0.71	0.70	0.69	0.64	0.71	0.70	0.67	0.57
120	800	0.71	0.70	0.70	0.70	0.71	0.70	0.69	0.65	0.71	0.70	0.66	0.57
$K=3, M=3, p=1$													
80	600	0.38	0.37	0.37	0.35	0.38	0.37	0.35	0.31	0.38	0.37	0.33	0.29
80	700	0.38	0.36	0.35	0.35	0.38	0.37	0.35	0.31	0.38	0.36	0.33	0.30
80	800	0.38	0.38	0.36	0.35	0.38	0.36	0.34	0.31	0.37	0.36	0.32	0.29
100	600	0.41	0.39	0.38	0.37	0.41	0.39	0.36	0.31	0.40	0.38	0.33	0.30
100	700	0.41	0.39	0.38	0.37	0.40	0.39	0.36	0.31	0.41	0.38	0.33	0.29
100	800	0.40	0.39	0.38	0.37	0.40	0.39	0.36	0.31	0.40	0.38	0.33	0.30
120	600	0.42	0.41	0.40	0.39	0.42	0.41	0.36	0.31	0.42	0.40	0.33	0.28
120	700	0.42	0.41	0.40	0.39	0.42	0.41	0.36	0.31	0.42	0.40	0.33	0.28
120	800	0.41	0.41	0.40	0.39	0.42	0.40	0.37	0.31	0.42	0.40	0.34	0.29
$K=3, M=1, p=2$													
80	600	0.75	0.74	0.74	0.73	0.75	0.74	0.73	0.69	0.75	0.74	0.73	0.65
80	700	0.75	0.74	0.74	0.73	0.76	0.74	0.72	0.70	0.75	0.74	0.72	0.66
80	800	0.75	0.74	0.73	0.73	0.75	0.75	0.73	0.69	0.75	0.74	0.72	0.66
100	600	0.77	0.76	0.76	0.74	0.76	0.76	0.74	0.70	0.77	0.76	0.72	0.65
100	700	0.77	0.77	0.76	0.75	0.77	0.76	0.74	0.70	0.77	0.76	0.73	0.66
100	800	0.77	0.76	0.75	0.75	0.77	0.76	0.75	0.71	0.76	0.76	0.73	0.67
120	600	0.78	0.78	0.76	0.76	0.78	0.77	0.74	0.70	0.78	0.77	0.73	0.65
120	700	0.78	0.78	0.77	0.76	0.78	0.77	0.75	0.71	0.78	0.77	0.73	0.65
120	800	0.78	0.78	0.77	0.76	0.78	0.77	0.75	0.71	0.78	0.77	0.74	0.66
$K=3, M=3, p=2$													
80	600	0.54	0.52	0.52	0.51	0.55	0.54	0.50	0.47	0.54	0.52	0.50	0.45
80	700	0.54	0.53	0.52	0.50	0.54	0.53	0.52	0.48	0.54	0.53	0.50	0.45
80	800	0.54	0.52	0.53	0.52	0.55	0.54	0.49	0.48	0.55	0.52	0.50	0.47
100	600	0.56	0.55	0.53	0.52	0.56	0.54	0.51	0.47	0.56	0.54	0.48	0.45
100	700	0.55	0.55	0.54	0.54	0.56	0.54	0.51	0.46	0.56	0.54	0.51	0.44
100	800	0.57	0.55	0.55	0.54	0.55	0.55	0.52	0.46	0.55	0.54	0.50	0.45
120	600	0.57	0.57	0.56	0.55	0.57	0.57	0.50	0.45	0.58	0.55	0.49	0.42
120	700	0.57	0.56	0.56	0.55	0.56	0.55	0.51	0.45	0.57	0.56	0.49	0.43
120	800	0.57	0.57	0.55	0.55	0.57	0.57	0.52	0.46	0.57	0.56	0.49	0.43

The displayed means are derived from 500 MC simulations for known dimensions K and p.

[a] For incomplete time series a stock variable is assumed.
[b] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series are stock and flow (average formulation) variables, respectively.
[c] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series serve as stock or change in flow (average formulaton) variables.

Table 5.3: *Means of trace R^2 based on hidden factors for random FAVARs using standard KF and KS, when complete panel data takes observed variables into account*

		stock[a]				**stock/flow (average)**[b]				**stock/change in flow (average)**[c]			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%
$K=1, M=1, p=1$													
80	600	0.46	0.46	0.45	0.46	0.46	0.46	0.46	0.47	0.46	0.47	0.46	0.46
80	700	0.47	0.45	0.47	0.47	0.46	0.46	0.46	0.46	0.47	0.46	0.46	0.45
80	800	0.47	0.46	0.47	0.47	0.47	0.47	0.46	0.47	0.47	0.46	0.46	0.45
100	600	0.46	0.48	0.48	0.46	0.47	0.47	0.47	0.46	0.47	0.47	0.47	0.46
100	700	0.47	0.47	0.48	0.47	0.48	0.47	0.47	0.47	0.48	0.47	0.46	0.47
100	800	0.48	0.48	0.47	0.47	0.46	0.47	0.46	0.47	0.46	0.47	0.47	0.47
120	600	0.48	0.47	0.48	0.48	0.48	0.47	0.46	0.47	0.48	0.47	0.48	0.46
120	700	0.47	0.48	0.48	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.46
120	800	0.47	0.48	0.47	0.48	0.47	0.47	0.48	0.47	0.47	0.47	0.47	0.46
$K=3, M=1, p=1$													
80	600	0.68	0.67	0.67	0.66	0.68	0.68	0.66	0.66	0.68	0.67	0.66	0.64
80	700	0.68	0.68	0.67	0.66	0.68	0.68	0.67	0.65	0.68	0.67	0.66	0.64
80	800	0.68	0.67	0.67	0.66	0.68	0.67	0.67	0.66	0.68	0.67	0.66	0.64
100	600	0.70	0.69	0.69	0.68	0.70	0.69	0.68	0.67	0.70	0.69	0.68	0.66
100	700	0.70	0.69	0.69	0.68	0.69	0.69	0.68	0.67	0.69	0.69	0.68	0.66
100	800	0.69	0.69	0.68	0.68	0.70	0.69	0.69	0.67	0.70	0.69	0.68	0.66
120	600	0.71	0.71	0.70	0.69	0.71	0.71	0.70	0.69	0.71	0.71	0.69	0.67
120	700	0.71	0.71	0.70	0.70	0.71	0.70	0.70	0.69	0.71	0.70	0.69	0.68
120	800	0.71	0.71	0.70	0.70	0.71	0.70	0.70	0.68	0.71	0.70	0.69	0.67
$K=3, M=3, p=1$													
80	600	0.38	0.38	0.38	0.37	0.38	0.38	0.38	0.37	0.38	0.39	0.38	0.36
80	700	0.38	0.37	0.37	0.37	0.38	0.38	0.38	0.36	0.38	0.38	0.38	0.36
80	800	0.38	0.39	0.38	0.37	0.38	0.38	0.37	0.37	0.37	0.38	0.37	0.36
100	600	0.41	0.40	0.40	0.39	0.41	0.40	0.39	0.39	0.40	0.40	0.40	0.38
100	700	0.41	0.40	0.39	0.39	0.40	0.40	0.40	0.39	0.41	0.40	0.39	0.38
100	800	0.40	0.40	0.39	0.39	0.40	0.40	0.40	0.39	0.40	0.39	0.39	0.38
120	600	0.42	0.42	0.42	0.41	0.42	0.42	0.41	0.41	0.42	0.42	0.42	0.39
120	700	0.42	0.42	0.42	0.41	0.42	0.42	0.41	0.41	0.42	0.41	0.41	0.39
120	800	0.41	0.41	0.41	0.41	0.42	0.41	0.41	0.41	0.42	0.42	0.42	0.39
$K=3, M=1, p=2$													
80	600	0.75	0.74	0.74	0.73	0.75	0.74	0.74	0.73	0.75	0.74	0.74	0.72
80	700	0.75	0.75	0.74	0.73	0.76	0.75	0.74	0.73	0.75	0.75	0.73	0.72
80	800	0.75	0.74	0.74	0.73	0.75	0.75	0.74	0.72	0.75	0.75	0.73	0.72
100	600	0.77	0.76	0.76	0.74	0.76	0.76	0.76	0.75	0.77	0.77	0.75	0.74
100	700	0.77	0.77	0.76	0.75	0.77	0.77	0.76	0.74	0.77	0.76	0.75	0.74
100	800	0.77	0.76	0.76	0.75	0.77	0.76	0.76	0.75	0.76	0.76	0.75	0.74
120	600	0.78	0.78	0.76	0.77	0.78	0.78	0.77	0.76	0.78	0.78	0.76	0.75
120	700	0.78	0.78	0.77	0.76	0.78	0.78	0.77	0.76	0.78	0.77	0.77	0.75
120	800	0.78	0.78	0.77	0.76	0.78	0.77	0.77	0.76	0.78	0.78	0.77	0.76
$K=3, M=3, p=2$													
80	600	0.54	0.53	0.53	0.53	0.55	0.55	0.53	0.53	0.54	0.53	0.54	0.53
80	700	0.54	0.54	0.53	0.52	0.54	0.53	0.55	0.53	0.54	0.54	0.53	0.52
80	800	0.54	0.53	0.54	0.53	0.55	0.54	0.52	0.54	0.55	0.53	0.54	0.54
100	600	0.56	0.55	0.54	0.53	0.56	0.55	0.55	0.55	0.56	0.55	0.54	0.54
100	700	0.55	0.55	0.55	0.55	0.56	0.55	0.55	0.54	0.56	0.55	0.56	0.54
100	800	0.57	0.56	0.56	0.56	0.55	0.56	0.55	0.53	0.55	0.55	0.55	0.54
120	600	0.57	0.57	0.57	0.57	0.57	0.58	0.56	0.56	0.58	0.57	0.56	0.55
120	700	0.57	0.57	0.57	0.56	0.56	0.56	0.56	0.56	0.57	0.57	0.56	0.55
120	800	0.57	0.57	0.56	0.56	0.57	0.58	0.57	0.56	0.57	0.57	0.56	0.56

The displayed means are derived from 500 MC simulations for known dimensions K and p.

[a] For incomplete time series a stock variable is assumed.
[b] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series are stock and flow (average formulation) variables, respectively.
[c] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series serve as stock or change in flow (average formulaton) variables.

Table 5.4: *Means of trace R^2 based on hidden factors for random FAVARs using new KF and KS*

		stock[a]				**stock/flow (average)**[b]				**stock/change in flow (average)**[c]			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%
$K = 1, M = 1, p = 1$													
80	600	0.92	0.91	0.91	0.91	0.92	0.91	0.90	0.90	0.92	0.91	0.90	0.89
80	700	0.92	0.91	0.91	0.91	0.92	0.91	0.91	0.90	0.92	0.91	0.90	0.89
80	800	0.92	0.91	0.91	0.91	0.92	0.91	0.91	0.90	0.92	0.91	0.90	0.89
100	600	0.93	0.93	0.92	0.92	0.93	0.92	0.92	0.91	0.93	0.92	0.91	0.91
100	700	0.93	0.93	0.92	0.92	0.93	0.93	0.92	0.91	0.93	0.93	0.91	0.91
100	800	0.93	0.93	0.92	0.92	0.93	0.92	0.92	0.91	0.93	0.92	0.91	0.91
120	600	0.94	0.94	0.93	0.93	0.94	0.94	0.93	0.92	0.94	0.94	0.92	0.92
120	700	0.94	0.94	0.93	0.93	0.94	0.93	0.93	0.92	0.94	0.93	0.92	0.92
120	800	0.94	0.94	0.93	0.93	0.94	0.94	0.93	0.92	0.94	0.94	0.92	0.92
$K = 3, M = 1, p = 1$													
80	600	0.83	0.81	0.81	0.79	0.83	0.82	0.81	0.79	0.83	0.82	0.80	0.77
80	700	0.83	0.82	0.80	0.79	0.83	0.82	0.81	0.79	0.83	0.82	0.80	0.77
80	800	0.82	0.82	0.80	0.80	0.83	0.82	0.81	0.80	0.83	0.82	0.80	0.77
100	600	0.84	0.84	0.83	0.82	0.85	0.84	0.82	0.82	0.85	0.84	0.82	0.80
100	700	0.85	0.84	0.83	0.82	0.84	0.84	0.83	0.82	0.84	0.84	0.83	0.79
100	800	0.85	0.84	0.83	0.82	0.85	0.84	0.84	0.82	0.85	0.84	0.83	0.80
120	600	0.86	0.85	0.85	0.84	0.87	0.86	0.84	0.83	0.87	0.86	0.84	0.81
120	700	0.86	0.86	0.85	0.83	0.86	0.85	0.84	0.83	0.86	0.85	0.84	0.81
120	800	0.87	0.86	0.85	0.84	0.87	0.85	0.85	0.83	0.87	0.86	0.84	0.81
$K = 3, M = 3, p = 1$													
80	600	0.76	0.76	0.75	0.73	0.77	0.75	0.73	0.72	0.77	0.76	0.73	0.68
80	700	0.77	0.76	0.74	0.73	0.77	0.76	0.75	0.73	0.76	0.76	0.74	0.68
80	800	0.78	0.75	0.75	0.73	0.77	0.76	0.74	0.73	0.77	0.75	0.73	0.69
100	600	0.79	0.77	0.77	0.75	0.79	0.78	0.76	0.75	0.78	0.78	0.76	0.71
100	700	0.79	0.78	0.76	0.76	0.79	0.78	0.77	0.75	0.80	0.78	0.75	0.72
100	800	0.79	0.78	0.78	0.75	0.79	0.78	0.78	0.74	0.80	0.78	0.76	0.71
120	600	0.81	0.80	0.77	0.78	0.80	0.79	0.78	0.77	0.80	0.79	0.77	0.72
120	700	0.81	0.80	0.78	0.77	0.81	0.80	0.78	0.76	0.81	0.80	0.78	0.73
120	800	0.81	0.80	0.79	0.77	0.81	0.80	0.79	0.77	0.81	0.80	0.78	0.73
$K = 3, M = 1, p = 2$													
80	600	0.85	0.84	0.83	0.82	0.85	0.84	0.83	0.82	0.85	0.84	0.83	0.80
80	700	0.85	0.84	0.83	0.82	0.85	0.84	0.83	0.82	0.85	0.83	0.83	0.81
80	800	0.85	0.84	0.83	0.82	0.85	0.84	0.83	0.82	0.85	0.84	0.83	0.81
100	600	0.86	0.86	0.85	0.84	0.86	0.85	0.85	0.84	0.87	0.86	0.85	0.83
100	700	0.87	0.86	0.85	0.85	0.87	0.86	0.85	0.84	0.86	0.85	0.85	0.83
100	800	0.87	0.86	0.86	0.85	0.87	0.86	0.85	0.85	0.87	0.86	0.85	0.83
120	600	0.88	0.87	0.87	0.86	0.87	0.87	0.87	0.85	0.88	0.87	0.86	0.85
120	700	0.88	0.87	0.87	0.86	0.88	0.88	0.87	0.86	0.88	0.87	0.86	0.84
120	800	0.88	0.88	0.87	0.86	0.88	0.87	0.87	0.86	0.88	0.88	0.86	0.85
$K = 3, M = 3, p = 2$													
80	600	0.79	0.79	0.77	0.76	0.79	0.78	0.77	0.76	0.79	0.78	0.77	0.75
80	700	0.80	0.79	0.78	0.77	0.80	0.79	0.77	0.77	0.80	0.78	0.78	0.76
80	800	0.80	0.79	0.78	0.77	0.80	0.79	0.79	0.77	0.80	0.79	0.77	0.76
100	600	0.81	0.81	0.80	0.78	0.81	0.80	0.79	0.78	0.81	0.80	0.79	0.77
100	700	0.82	0.81	0.80	0.79	0.81	0.81	0.79	0.79	0.82	0.81	0.79	0.78
100	800	0.82	0.81	0.81	0.79	0.82	0.81	0.80	0.79	0.82	0.82	0.80	0.78
120	600	0.83	0.82	0.81	0.81	0.83	0.82	0.81	0.80	0.83	0.82	0.81	0.79
120	700	0.83	0.82	0.82	0.81	0.83	0.82	0.81	0.80	0.83	0.83	0.81	0.79
120	800	0.84	0.83	0.82	0.81	0.84	0.83	0.82	0.81	0.84	0.83	0.82	0.79

The displayed means are derived from 500 MC simulations for known dimensions K and p.

[a] For incomplete time series a stock variable is assumed.
[b] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series are stock and flow (average formulation) variables, respectively.
[c] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series serve as stock or change in flow (average formulaton) variables.

Table 5.5: *Ratios of trace R^2 means based on hidden factors for random FAVARs using new KF and KS versus PCA and OLS*

		stock[a]				stock/flow (average)[b]				stock/change in flow (average)[c]			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%
$K=1, M=1, p=1$													
80	600	1.88	1.86	1.88	1.85	1.88	1.86	1.87	1.81	1.88	1.84	1.82	1.80
80	700	1.84	1.91	1.85	1.83	1.87	1.86	1.86	1.83	1.86	1.86	1.83	1.84
80	800	1.86	1.86	1.84	1.84	1.83	1.84	1.87	1.81	1.85	1.85	1.85	1.85
100	600	1.91	1.87	1.85	1.90	1.88	1.89	1.86	1.87	1.88	1.88	1.85	1.86
100	700	1.90	1.87	1.86	1.86	1.87	1.87	1.86	1.85	1.87	1.89	1.87	1.82
100	800	1.86	1.85	1.90	1.86	1.91	1.86	1.89	1.85	1.91	1.87	1.84	1.83
120	600	1.89	1.92	1.87	1.87	1.90	1.91	1.92	1.85	1.90	1.91	1.85	1.90
120	700	1.92	1.88	1.86	1.89	1.92	1.89	1.88	1.85	1.92	1.89	1.86	1.89
120	800	1.91	1.89	1.89	1.87	1.91	1.89	1.85	1.87	1.91	1.89	1.87	1.87
$K=3, M=1, p=1$													
80	600	1.12	1.10	1.10	1.07	1.12	1.10	1.10	1.08	1.12	1.10	1.09	1.06
80	700	1.12	1.11	1.09	1.08	1.12	1.11	1.09	1.08	1.12	1.11	1.09	1.07
80	800	1.12	1.11	1.09	1.08	1.12	1.11	1.10	1.09	1.12	1.11	1.09	1.07
100	600	1.12	1.11	1.10	1.09	1.12	1.11	1.10	1.09	1.12	1.11	1.10	1.07
100	700	1.13	1.11	1.10	1.09	1.12	1.11	1.11	1.10	1.12	1.11	1.11	1.08
100	800	1.13	1.11	1.11	1.10	1.13	1.12	1.11	1.10	1.13	1.12	1.11	1.08
120	600	1.13	1.11	1.11	1.10	1.13	1.12	1.11	1.10	1.13	1.12	1.11	1.08
120	700	1.13	1.12	1.12	1.10	1.13	1.12	1.12	1.10	1.13	1.12	1.11	1.08
120	800	1.14	1.12	1.12	1.11	1.14	1.13	1.11	1.11	1.14	1.13	1.11	1.08
$K=3, M=3, p=1$													
80	600	1.38	1.35	1.31	1.31	1.37	1.34	1.31	1.29	1.39	1.34	1.31	1.23
80	700	1.39	1.37	1.33	1.29	1.39	1.36	1.33	1.31	1.38	1.35	1.31	1.24
80	800	1.39	1.35	1.34	1.31	1.39	1.36	1.34	1.31	1.39	1.36	1.32	1.25
100	600	1.39	1.37	1.35	1.32	1.40	1.37	1.35	1.33	1.39	1.37	1.33	1.26
100	700	1.41	1.37	1.36	1.33	1.41	1.39	1.37	1.33	1.40	1.38	1.34	1.28
100	800	1.41	1.38	1.36	1.33	1.41	1.38	1.37	1.32	1.41	1.38	1.35	1.29
120	600	1.43	1.39	1.36	1.35	1.41	1.39	1.37	1.34	1.41	1.39	1.34	1.29
120	700	1.43	1.40	1.37	1.35	1.42	1.39	1.38	1.35	1.43	1.40	1.37	1.30
120	800	1.44	1.40	1.38	1.35	1.42	1.41	1.39	1.35	1.43	1.41	1.36	1.29
$K=3, M=1, p=2$													
80	600	1.07	1.07	1.05	1.04	1.07	1.07	1.06	1.05	1.07	1.07	1.05	1.04
80	700	1.07	1.07	1.06	1.05	1.07	1.07	1.06	1.05	1.07	1.06	1.06	1.05
80	800	1.07	1.07	1.06	1.06	1.07	1.07	1.06	1.06	1.07	1.06	1.06	1.04
100	600	1.08	1.07	1.06	1.06	1.08	1.07	1.06	1.05	1.07	1.07	1.07	1.05
100	700	1.07	1.07	1.07	1.06	1.08	1.07	1.07	1.06	1.07	1.07	1.06	1.05
100	800	1.07	1.07	1.07	1.06	1.08	1.07	1.07	1.06	1.08	1.07	1.07	1.05
120	600	1.08	1.07	1.08	1.07	1.08	1.07	1.07	1.06	1.08	1.07	1.07	1.06
120	700	1.08	1.07	1.07	1.07	1.08	1.08	1.07	1.07	1.08	1.08	1.07	1.06
120	800	1.08	1.08	1.07	1.07	1.08	1.08	1.08	1.06	1.09	1.08	1.07	1.06
$K=3, M=3, p=2$													
80	600	1.22	1.22	1.19	1.17	1.21	1.20	1.19	1.17	1.22	1.21	1.18	1.15
80	700	1.22	1.21	1.19	1.19	1.22	1.20	1.18	1.19	1.23	1.21	1.19	1.17
80	800	1.22	1.22	1.20	1.18	1.22	1.21	1.22	1.18	1.22	1.22	1.18	1.16
100	600	1.23	1.22	1.22	1.20	1.23	1.22	1.21	1.19	1.23	1.22	1.22	1.17
100	700	1.25	1.23	1.22	1.19	1.23	1.23	1.21	1.20	1.24	1.23	1.20	1.20
100	800	1.23	1.23	1.21	1.18	1.24	1.22	1.22	1.21	1.25	1.24	1.21	1.18
120	600	1.24	1.22	1.22	1.21	1.24	1.22	1.23	1.21	1.24	1.23	1.22	1.20
120	700	1.25	1.24	1.22	1.21	1.26	1.24	1.22	1.22	1.25	1.24	1.22	1.20
120	800	1.25	1.23	1.23	1.22	1.26	1.23	1.22	1.23	1.25	1.24	1.23	1.20

The displayed means are derived from 500 MC simulations for known dimensions K and p. Thereby, each figure represents the mean of the trace R^2 in Table 5.4 divided by its counterpart in Table 5.1.

[a] For incomplete time series a stock variable is assumed.
[b] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series are stock and flow (average formulation) variables, respectively.
[c] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series serve as stock or change in flow (average formulaton) variables.

Table 5.6: *Ratios of trace R^2 means based on hidden factors for random FAVARs using new KF and KS versus standard KF and KS, when complete panel data relies on estimated factors instead of observed variables*

		stock[a]				**stock/flow (average)**[b]				**stock/change in flow (average)**[c]			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%
						$K=1, M=1, p=1$							
80	600	1.99	1.97	2.01	1.98	1.99	1.97	2.01	2.03	1.99	1.95	2.01	2.06
80	700	1.95	2.03	1.97	1.95	1.98	1.98	2.00	2.00	1.97	1.97	2.04	2.07
80	800	1.97	1.98	1.95	1.96	1.93	1.95	2.00	2.01	1.96	1.97	2.07	2.09
100	600	2.00	1.96	1.94	2.00	1.97	1.99	2.08	2.25	1.97	1.99	2.21	2.27
100	700	1.99	1.96	1.95	1.96	1.96	1.97	2.07	2.22	1.96	2.01	2.15	2.15
100	800	1.95	1.93	2.00	1.96	2.01	1.95	2.09	2.20	2.00	1.98	2.11	2.25
120	600	1.96	2.00	1.95	1.96	1.97	2.00	2.24	2.21	1.97	2.04	2.18	2.28
120	700	1.99	1.95	1.94	1.98	1.99	1.98	2.25	2.25	1.99	2.02	2.21	2.18
120	800	1.99	1.97	1.98	1.95	1.98	1.97	2.08	2.31	1.98	2.02	2.18	2.19
						$K=3, M=1, p=1$							
80	600	1.22	1.22	1.22	1.20	1.22	1.21	1.23	1.26	1.22	1.22	1.24	1.39
80	700	1.22	1.22	1.21	1.20	1.22	1.22	1.22	1.26	1.22	1.23	1.24	1.37
80	800	1.22	1.22	1.20	1.20	1.22	1.22	1.23	1.27	1.22	1.23	1.25	1.37
100	600	1.20	1.21	1.20	1.21	1.22	1.21	1.23	1.30	1.22	1.22	1.25	1.43
100	700	1.22	1.22	1.21	1.20	1.22	1.21	1.22	1.28	1.22	1.22	1.26	1.38
100	800	1.22	1.22	1.22	1.21	1.23	1.22	1.23	1.29	1.23	1.22	1.27	1.40
120	600	1.22	1.21	1.21	1.20	1.22	1.22	1.23	1.31	1.22	1.22	1.28	1.44
120	700	1.22	1.22	1.22	1.20	1.22	1.21	1.23	1.29	1.22	1.22	1.27	1.43
120	800	1.23	1.22	1.21	1.21	1.23	1.22	1.23	1.29	1.23	1.23	1.26	1.43
						$K=3, M=3, p=1$							
80	600	2.02	2.04	2.04	2.10	1.99	2.03	2.12	2.33	2.02	2.05	2.22	2.33
80	700	2.02	2.10	2.09	2.07	2.02	2.06	2.15	2.38	2.00	2.09	2.21	2.31
80	800	2.03	2.00	2.07	2.10	2.03	2.09	2.17	2.39	2.06	2.07	2.28	2.36
100	600	1.95	1.97	2.01	2.01	1.95	1.98	2.15	2.39	1.93	2.03	2.28	2.37
100	700	1.95	1.98	2.03	2.03	1.97	2.03	2.16	2.44	1.96	2.06	2.26	2.46
100	800	1.96	1.99	2.04	2.05	1.98	1.99	2.17	2.36	1.96	2.06	2.28	2.39
120	600	1.94	1.94	1.92	1.97	1.91	1.94	2.16	2.48	1.90	1.98	2.30	2.59
120	700	1.93	1.96	1.94	1.97	1.90	1.96	2.18	2.48	1.94	2.01	2.35	2.58
120	800	1.95	1.98	1.99	1.99	1.93	1.99	2.16	2.50	1.92	2.00	2.27	2.53
						$K=3, M=1, p=2$							
80	600	1.12	1.13	1.12	1.12	1.13	1.14	1.14	1.19	1.12	1.13	1.14	1.23
80	700	1.12	1.13	1.13	1.13	1.12	1.13	1.14	1.18	1.12	1.12	1.16	1.24
80	800	1.13	1.13	1.13	1.13	1.12	1.13	1.14	1.19	1.13	1.13	1.15	1.21
100	600	1.13	1.13	1.12	1.13	1.13	1.13	1.15	1.20	1.12	1.13	1.18	1.27
100	700	1.12	1.12	1.13	1.12	1.13	1.13	1.15	1.19	1.12	1.13	1.16	1.26
100	800	1.12	1.13	1.14	1.14	1.13	1.13	1.14	1.20	1.14	1.14	1.16	1.25
120	600	1.13	1.13	1.14	1.13	1.12	1.13	1.16	1.21	1.13	1.13	1.18	1.30
120	700	1.13	1.12	1.13	1.13	1.13	1.14	1.15	1.21	1.13	1.13	1.19	1.29
120	800	1.12	1.13	1.13	1.13	1.12	1.13	1.16	1.20	1.14	1.14	1.17	1.29
						$K=3, M=3, p=2$							
80	600	1.47	1.50	1.48	1.48	1.45	1.45	1.54	1.62	1.46	1.51	1.54	1.66
80	700	1.47	1.50	1.49	1.53	1.47	1.49	1.48	1.62	1.47	1.48	1.56	1.68
80	800	1.47	1.52	1.48	1.49	1.47	1.48	1.59	1.61	1.46	1.52	1.54	1.62
100	600	1.45	1.47	1.50	1.52	1.45	1.48	1.56	1.68	1.46	1.50	1.65	1.72
100	700	1.49	1.49	1.48	1.47	1.44	1.49	1.55	1.71	1.46	1.50	1.57	1.77
100	800	1.44	1.48	1.47	1.45	1.48	1.48	1.55	1.70	1.49	1.51	1.61	1.73
120	600	1.44	1.44	1.46	1.47	1.45	1.44	1.61	1.80	1.44	1.49	1.65	1.86
120	700	1.47	1.47	1.46	1.46	1.48	1.49	1.58	1.77	1.46	1.49	1.66	1.84
120	800	1.46	1.46	1.49	1.47	1.47	1.45	1.57	1.78	1.46	1.49	1.67	1.84

The displayed means are derived from 500 MC simulations for known dimensions K and p. Thereby, each figure represents the mean of the trace R^2 in Table 5.4 divided by its counterpart in Table 5.2.

[a] For incomplete time series a stock variable is assumed.
[b] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series are stock and flow (average formulation) variables, respectively.
[c] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series serve as stock or change in flow (average formulaton) variables.

Table 5.7: *Ratios of trace R^2 means based on hidden factors for random FAVARs using new KF and KS versus standard KF and KS, when complete panel data takes observed variables into account*

		stock[a]				**stock/flow (average)**[b]				**stock/change in flow (average)**[c]			
		ratio of missing data				*ratio of missing data*				*ratio of missing data*			
N	T	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%
						$K=1, M=1, p=1$							
80	600	1.99	1.96	2.00	1.96	1.99	1.97	1.98	1.92	1.99	1.96	1.94	1.93
80	700	1.95	2.03	1.96	1.93	1.98	1.98	1.98	1.95	1.97	1.97	1.95	1.98
80	800	1.97	1.97	1.94	1.94	1.93	1.95	1.99	1.92	1.96	1.96	1.98	1.98
100	600	2.00	1.95	1.93	1.98	1.97	1.98	1.96	1.98	1.97	1.97	1.95	1.98
100	700	1.99	1.95	1.94	1.95	1.96	1.96	1.95	1.96	1.96	1.99	1.97	1.94
100	800	1.95	1.93	1.99	1.94	2.01	1.95	1.98	1.95	2.00	1.97	1.94	1.94
120	600	1.96	1.99	1.95	1.95	1.97	1.99	2.00	1.95	1.97	1.99	1.93	2.01
120	700	1.99	1.95	1.93	1.97	1.99	1.97	1.96	1.94	1.99	1.97	1.95	2.01
120	800	1.99	1.97	1.97	1.94	1.98	1.97	1.93	1.97	1.98	1.97	1.96	1.98
						$K=3, M=1, p=1$							
80	600	1.22	1.21	1.21	1.20	1.22	1.21	1.22	1.21	1.22	1.21	1.20	1.21
80	700	1.22	1.22	1.21	1.20	1.22	1.21	1.21	1.21	1.22	1.22	1.21	1.21
80	800	1.22	1.22	1.20	1.20	1.22	1.22	1.22	1.21	1.22	1.22	1.21	1.22
100	600	1.20	1.21	1.20	1.21	1.22	1.21	1.21	1.21	1.22	1.21	1.20	1.21
100	700	1.22	1.22	1.21	1.20	1.22	1.20	1.21	1.22	1.22	1.21	1.21	1.21
100	800	1.22	1.22	1.22	1.21	1.23	1.22	1.22	1.21	1.23	1.22	1.22	1.21
120	600	1.22	1.21	1.21	1.20	1.22	1.21	1.21	1.21	1.22	1.21	1.22	1.21
120	700	1.22	1.21	1.22	1.19	1.22	1.21	1.21	1.21	1.22	1.21	1.21	1.20
120	800	1.23	1.21	1.22	1.21	1.23	1.22	1.21	1.22	1.23	1.22	1.21	1.20
						$K=3, M=3, p=1$							
80	600	2.02	1.99	1.94	1.99	1.99	1.97	1.95	1.95	2.02	1.96	1.94	1.89
80	700	2.02	2.04	2.01	1.96	2.02	1.99	1.97	2.01	2.00	2.00	1.95	1.87
80	800	2.03	1.95	2.00	1.99	2.03	2.02	2.00	1.98	2.06	1.99	2.01	1.94
100	600	1.95	1.92	1.93	1.89	1.95	1.93	1.94	1.92	1.93	1.94	1.90	1.84
100	700	1.95	1.93	1.95	1.94	1.97	1.97	1.94	1.92	1.96	1.96	1.91	1.88
100	800	1.96	1.95	1.96	1.95	1.98	1.94	1.96	1.90	1.96	1.98	1.95	1.89
120	600	1.94	1.90	1.85	1.88	1.91	1.88	1.88	1.85	1.90	1.89	1.85	1.86
120	700	1.93	1.92	1.88	1.89	1.90	1.92	1.91	1.87	1.94	1.93	1.89	1.88
120	800	1.95	1.94	1.92	1.89	1.93	1.93	1.92	1.87	1.92	1.92	1.87	1.89
						$K=3, M=1, p=2$							
80	600	1.12	1.13	1.12	1.12	1.13	1.13	1.12	1.13	1.12	1.13	1.12	1.12
80	700	1.12	1.13	1.13	1.13	1.12	1.13	1.13	1.13	1.12	1.11	1.14	1.13
80	800	1.13	1.13	1.13	1.13	1.12	1.12	1.12	1.13	1.13	1.12	1.12	1.12
100	600	1.13	1.13	1.12	1.13	1.13	1.12	1.12	1.12	1.12	1.12	1.14	1.12
100	700	1.12	1.12	1.13	1.12	1.13	1.12	1.13	1.13	1.12	1.12	1.13	1.12
100	800	1.12	1.13	1.13	1.13	1.13	1.13	1.12	1.13	1.14	1.13	1.13	1.12
120	600	1.13	1.12	1.14	1.13	1.12	1.12	1.13	1.12	1.13	1.12	1.13	1.13
120	700	1.13	1.12	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.12	1.12
120	800	1.12	1.13	1.12	1.13	1.12	1.13	1.13	1.13	1.14	1.13	1.12	1.13
						$K=3, M=3, p=2$							
80	600	1.47	1.48	1.45	1.44	1.45	1.43	1.45	1.43	1.46	1.47	1.43	1.42
80	700	1.47	1.48	1.46	1.48	1.47	1.47	1.41	1.45	1.47	1.46	1.45	1.45
80	800	1.47	1.51	1.45	1.45	1.47	1.46	1.50	1.43	1.46	1.50	1.42	1.42
100	600	1.45	1.46	1.47	1.47	1.45	1.46	1.45	1.43	1.46	1.46	1.47	1.42
100	700	1.49	1.47	1.46	1.44	1.44	1.47	1.44	1.45	1.46	1.48	1.42	1.44
100	800	1.44	1.46	1.45	1.42	1.48	1.46	1.45	1.47	1.49	1.48	1.45	1.43
120	600	1.44	1.43	1.43	1.43	1.45	1.41	1.45	1.44	1.44	1.44	1.43	1.44
120	700	1.47	1.46	1.44	1.43	1.48	1.47	1.44	1.43	1.46	1.45	1.44	1.43
120	800	1.46	1.45	1.47	1.44	1.47	1.44	1.43	1.44	1.46	1.46	1.45	1.43

The displayed means are derived from 500 MC simulations for known dimensions K and p. Thereby, each figure represents the mean of the trace R^2 in Table 5.4 divided by its counterpart in Table 5.3.

[a] For incomplete time series a stock variable is assumed.

[b] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series are stock and flow (average formulation) variables, respectively.

[c] For incomplete data, $\lceil N/2 \rceil$ and $\lfloor N/2 \rfloor$ time series serve as stock or change in flow (average formulaton) variables.

5.3 Empirical Application

The US economy ranks among the biggest and most important in the world. Moreover, after many years of declining interest rates, in December 2015 the US Federal Reserve decided to raise its key interest rate, i.e., the Effective Federal Funds Rate (FEDFUNDS), by 25 bps. In this way, it was the first large central bank to leave the path of an extremely relaxed monetary policy. Due to this actuality and, of course, for comparisons with the results in Bernanke et al. (2005), Bork (2009, 2015) and Bork et al. (2010), we deal with the impact of the US monetary policy on its real economy in the sequel.

At the beginning, we describe the underlying data. In doing so, we state which data serves as panel data and which specifies the observable factors. Then, we provide some technical details about the termination criteria, the upper limits of the factor dimension and autoregressive order, etc. Eventually, we discuss the derived Impulse Response Functions and Forecast Error Variance Decomposition.

Our data essentially is an updated version of the data in Bernanke et al. (2005), except for 24 variables, which we could not find anymore. This is why we have 96 of the original 120 time series over the period from January 1959 until October 2015. Besides the 96 monthly time series, we add 15 partially incomplete time series to our data. Among other things, we are interested in how monetary policy decisions may affect quarterly indices. For this purpose, the quarterly growth rates of GDP, Governmental Total Expenditures, Real Exports of Goods and Services as well as Real Imports of Goods and Services are part of these 15 new time series. With Section 2.2 in mind, we regard the four quarterly growth rates as sum versions of flow variables, while all other time series are treated as stock variables.[1]

Monetary policy actions, especially unexpected ones, can significantly move Foreign Exchange (FX) rates. As the European Union trades a lot with the US, our data comprises the USD-EUR exchange rate, which started in January 1999. Furthermore, USD FX rates against the German Mark, French Franc and Italian Lire serve as an approximation for the USD-EUR FX rate before January 1999. By this means, our data set consists of time series, which are discontinued and start late, respectively. Finally, 4 of the 15 new time series offer information about the balance sheets of the Federal Reserve Banks, which have dramatically increased since the financial crisis in the years 2007-2008. In total, we have 111 macroeconomic indicators for diverse areas of the US economy from January 1959 until October 2015. Besides a detailed overview of variables, Appendix B.3 states the data sources, the performed preprocessing as well as the classification in slow- and fast-moving ones based on Bernanke et al. (2005).

The “Quantitative Easing” programs QE1-QE3 were the response of the Federal Reserve to the problems arising from the financial crisis, after stimulating the economy by lowering the Effective Federal Funds Rate reached its limits in December 2008. For instance, the Federal Reserve massively bought Treasuries and mortgage-backed securities. To obtain a picture of the monetary policy actions as comprehensive as possible, the observable factor $\boldsymbol{Y}_t$ is given by the Currency in Circulation (CURRCIR), the St. Louis Adjusted Monetary Base (AMBSL) and the Effective Federal Funds Rate (FEDFUNDS). Our estimation method for FAVARs requires the time series $\{\boldsymbol{Y}_t\}$ to be complete. Therefore, holdings of Treasuries and mortgage-backed securities, which were only available for the years 2002-2015, belong to the panel data instead of the observed variables $\boldsymbol{Y}_t$.

Due to the loadings constraints in (5.15), the ranking of the first K variables matters. So far, the optimal

[1] In terms of the 107 monthly time series the distinction between stock, flow and change in flow variables does not matter, since there are no missing values. Although some time series start at a later point in time, e.g., the USD-EUR exchange rate, or are discontinued, e.g., the German Mark-USD exchange rate, there are no intermediately missing observations. Hence, for each monthly time series and each subsample, the matrix Q_i in (2.9) consists of ones and zeros only.

factor dimension K^* is unknown, therefore, we conduct a pre-analysis as in Bork (2009), which is described in Section 5.1.2, to properly arrange the panel data. After the complete, slow-moving variables have been sorted, we do the same with the fast-moving ones, before we add all ragged time series in arbitrary order.

Our technical settings are: $T = 682, M = 3, \bar{K} = 10, \bar{p} = 5, \eta = 0.01$ and $\xi = 0.01$. Thus, the termination criteria are not too strict and the run time of Algorithm 5.1.4 remains reasonable. The AIC-based model selection in Lemma 5.1.10 yields: $(K^*, p^*) = (9, 1)$. In this way, we have larger factor dimensions K and M, but a smaller lag order than Bork (2009). Because of this, Table 5.8 compares the first nine variables of our sorted panel data with their counterparts in Bork (2009). Thereby, we keep the long expressions of Bork (2009) in the second column, while we apply our abbreviations from Appendix B.3 in the third column. At first glance, both subsets cover the same areas. That means, Bork (2009) has four time series of the group "Real Output and Income", three time series belonging to "(Un)employment and Hours", one time series from "Consumption" and one from "Price Indices". Similarly, our subset consists of one, four, one and three, respectively, time series of the same groups. The main deviation arises from the fact that we are more strongly driven by price indices instead of production data. However, we should keep in mind that some differences possibly arise from that fact that some time has passed between the work of Bork (2009) and ours. Furthermore, the underlying data does not completely match. Note, the different loadings constraints in (5.13) and (5.15) are irrelevant, since the pre-analysis is independent of both.

Table 5.8: *Comparison of panel data sorted by performed pre-analysis*

No.	Bork (2009)	Our Data (Ticker)
1	Industrial production: manufacturing (1992 = 100, SA)	PAYEMS
2	Unemploy. by duration: average (mean) duration in weeks (SA)	CPILFESL
3	Purchasing managers' index (SA)	PPIFCG
4	Avg. weekly hrs. of prod. wkrs.: mfg., overtime hrs. (SA)	UNRATE
5	CPI-u: commodities (82-84 = 100, SA)	USFIRE
6	Employment: ratio; help-wanted ads: no. unemployed clf	IPCONGD
7	Capacity util rate: manufac., total (% of capacity, SA) (frb)	AWOTMAN
8	Pers cons exp (chained) - tot. dur. (bil 96$, SAAR)	PCE
9	Industrial production: total index (1992 = 100, SA)	PPICRM

Next, we focus on the impact of shocks on the included variables. The recursive structure in (5.76) implies that each factor is only driven by its own innovations and the ones of preceding factors. Based on (5.78), we obtain the subsequent innovation weight:

$$\boldsymbol{z} = [0.38, 0.30, -5.49, 4.53, 5.50, -6.70, 1.66, -1.04, -2.36, -0.06, 0.39, -9.94],$$

which causes an increase in FEDFUNDS of 25 bps at time $t = 0$, but leaves everything else unchanged. As in Bernanke et al. (2005), Bork et al. (2010) and Bork (2015), we derive confidence intervals for the IRFs. In doing so, there are diverse methods to construct those. E.g., Bernanke et al. (2005) and Boivin et al. (2010) used the bias-adjusted bootstrap approach of Kilian (1998). In this sense, Yamamoto (2012) also showed bootstrap routines with bias correction. Due to its unknown asymptotic properties, Benkwitz et al. (1999) rised doubts concerning the approach of Kilian (1998) and recommended the use of standard bootstrap techniques instead. For instance, Bork et al. (2010) applied the standard bootstrap method. Alternatively, Bai et al. (2015) derived closed-from expressions for the asymptotic distributions of IRFs. Since the idiosyncratic errors of their measurement equation are uncorrelated, we cannot use the findings of Bai et al. (2015) here. For simplicity reasons, we revert to a non-parametric bootstrap method without any bias correction.

Reestimation of latent factors and data incompleteness offer some flexibilty, this is why we briefly sketch our bootstrap method: We first estimate the parameters in (5.16)-(5.17) with loadings constraints in (5.15) and so, receive the residuals. To gain reliable confidence intervals we run 10,000 bootstrap simulations. For each path, we randomly draw with replacement from the recentered errors of (5.17) and keep the first p estimates and observations, respectively, of the vector $[\bar{\boldsymbol{F}}'_t, \boldsymbol{Y}'_t]'$ to generate a new sample $[(\bar{\boldsymbol{F}}^*_t)', (\boldsymbol{Y}^*_t)']'$ using standard non-parametric bootstrap. Next, we reestimate the coefficient matrices of the transition equation based on $[(\bar{\boldsymbol{F}}^*_t)', (\boldsymbol{Y}^*_t)']'$. Thereby, no model selection takes place, that is, a VAR(1) is estimated. With the help of (5.74), we then derive the IRFs of $[\bar{\boldsymbol{F}}'_t, \boldsymbol{Y}'_t]'_i$ for $1 \leq i \leq K+M$. For the IRFs of $\boldsymbol{X}_t$ in (5.75), we fix the initially estimated loadings matrix. In this manner, we ignore the uncertainty inherent in the bootstrapped panel data.

Similar to Bernanke et al. (2005), Bork et al. (2010) and Bork (2015), Figures 5.1-5.4 illustrate the impact of the shock $\boldsymbol{z}$ on the standardized variables. Our confidence intervals cover confidence levels of 68% (light gray) and 90% (dark gray). The underlying time horizon is 48 months. For the shock $\boldsymbol{z}$ and (5.74)-(5.75), Figure 5.1-5.4 display for the time series $1 \leq i \leq N$ or the factors $1 \leq j \leq K+M$:

$$\left(\bar{\Phi}^s \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \boldsymbol{z}\right)_j \quad \text{or} \quad \left(\begin{bmatrix} \bar{\Lambda}^f & \bar{\Lambda}^y \end{bmatrix} \bar{\Phi}^s \begin{bmatrix} I_K & O_{K\times M} \\ O_{M\times K} & P \end{bmatrix} \boldsymbol{z}\right)_i .$$

In Figure 5.1, an increase in FEDFUNDS weakens the industrial production (IPFINAL, IPCONGD, IPDCONGD, IPNCONGD, IPBUSEQ, IPMAT, IPB53100N, IPB53200N, IPMANSICS, INDPRO, NAPM, NAPMPI) in the short term, but has no long-term effects. At the same time, capacities (CUMFNS) are less utilized, personal income (RPI, W875RX1) decreases and unemployment (CE16OV, UNRATE, UEMPMEAN, UEMPLT5, UEMP5TO14, UEMP15OV, UEMP15T26) rises. Similarly, the number of employees across diverse business areas (PAYEMS, USPRIV, USGOOD, CES1021000001, USCONS, MANEMP, DMANEMP, NDMANEMP, CES0800000001, USTPU, USWTRADE, USFIRE, USPBS, USGOVT) and the average production time (AWHMAN, AWOTMAN) decline in the short run. As shown in Figures 5.1-5.2, these declines do not necessarily recovery. Higher unemployment rates together with lower incomes let the reduced personal expenditures (PCE, PCEDG, PCEND, PCES) appear reasonable.

In Figure 5.2, housing starts (HOUST, HOUSTNE, HOUSTMW, HOUSTS, HOUSTW, PERMITNSA) are supposed to increase over the next 48 months. Perhaps, this reflects that people are afraid of additional interest rate hikes and therefore, bring such projects forward. Since the Effective Federal Funds Rate applies for the whole US, regional aspects in case of housing starts do not matter. In the short term, less new orders (NAPMNOI) increase manufacturing inventories (NAPMII), which also confirms a reduction in consumption. In the long run, higher interest rates require companies to offer higher dividends (FSDXP), but boost their costs, too. E.g., the same amout of debt calls for higher interest rate payments. In total, the price-earnings ratio (FSPXE) naturally decreases.

Except for EXCAUS and EXITUS, the USD becomes stronger compared to foreign currencies (EXSZUS, EXJPUS, EXUSUK, EXGEUS, EXFRUS, EXUSEU). Note, EXITUS is the FX rate between Italian Lira, which the Euro succeeded, and USD. Thus, it is not relevant anymore. Here, it is part of our panel data, as we have EXGEUS, EXFRUS and EXITUS as proxies for EXUSEU, before the Euro was introduced on January 1, 1999. A stronger USD possibly comes from an increased demand for USD, when investors increase their exposure to US fixed income products. As shown in Figure 5.3, the yields of US Treasuries (TB3MS, TB6MS, GS1, GS5, GS10, TB3SMFFM, TB6SMFFM, T1YFFM, T5YFFM, T10YFFM) and

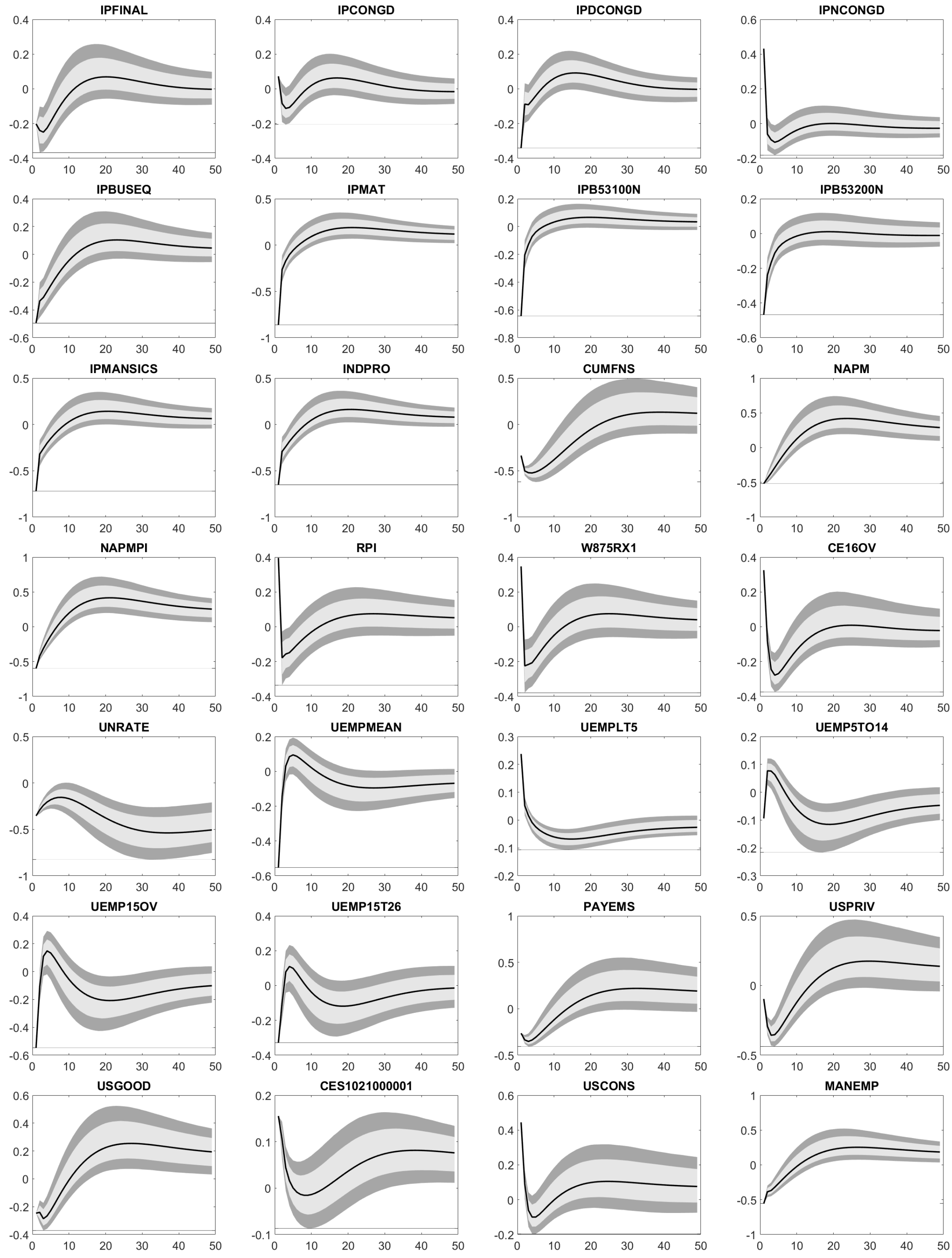

Figure 5.1: *IRFs (black lines) of standardized time series 1-28 in Appendix B.3 arising from an increase in FEDFUNDS by 0.25%. Light gray areas show the 68%-confidence intervals, i.e., the 1-sigma interval, while dark gray areas display the 90%-confidence intervals. In both cases, the intervals are based on 10,000 non-parametric bootstrap simulations of the transition equation, but the estimated loadings matrix is kept fixed.*

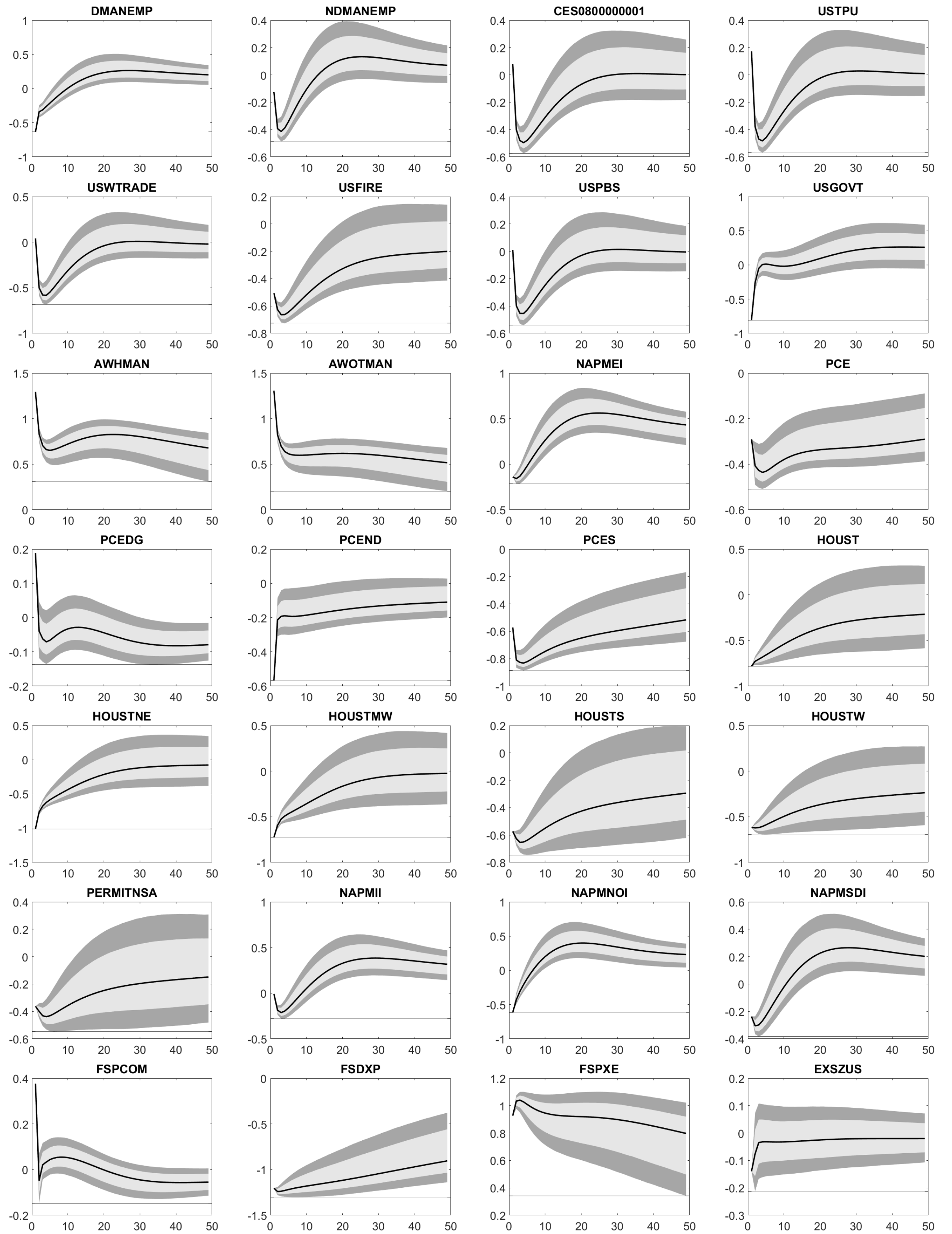

Figure 5.2: *IRFs (black lines) of standardized time series 29-56 in Appendix B.3 arising from an increase in FEDFUNDS by 0.25%. Light gray areas show the 68%-confidence intervals, i.e., the 1-sigma interval, while dark gray areas display the 90%-confidence intervals. In both cases, the intervals are based on 10,000 non-parametric bootstrap simulations of the transition equation, but the estimated loadings matrix is kept fixed.*

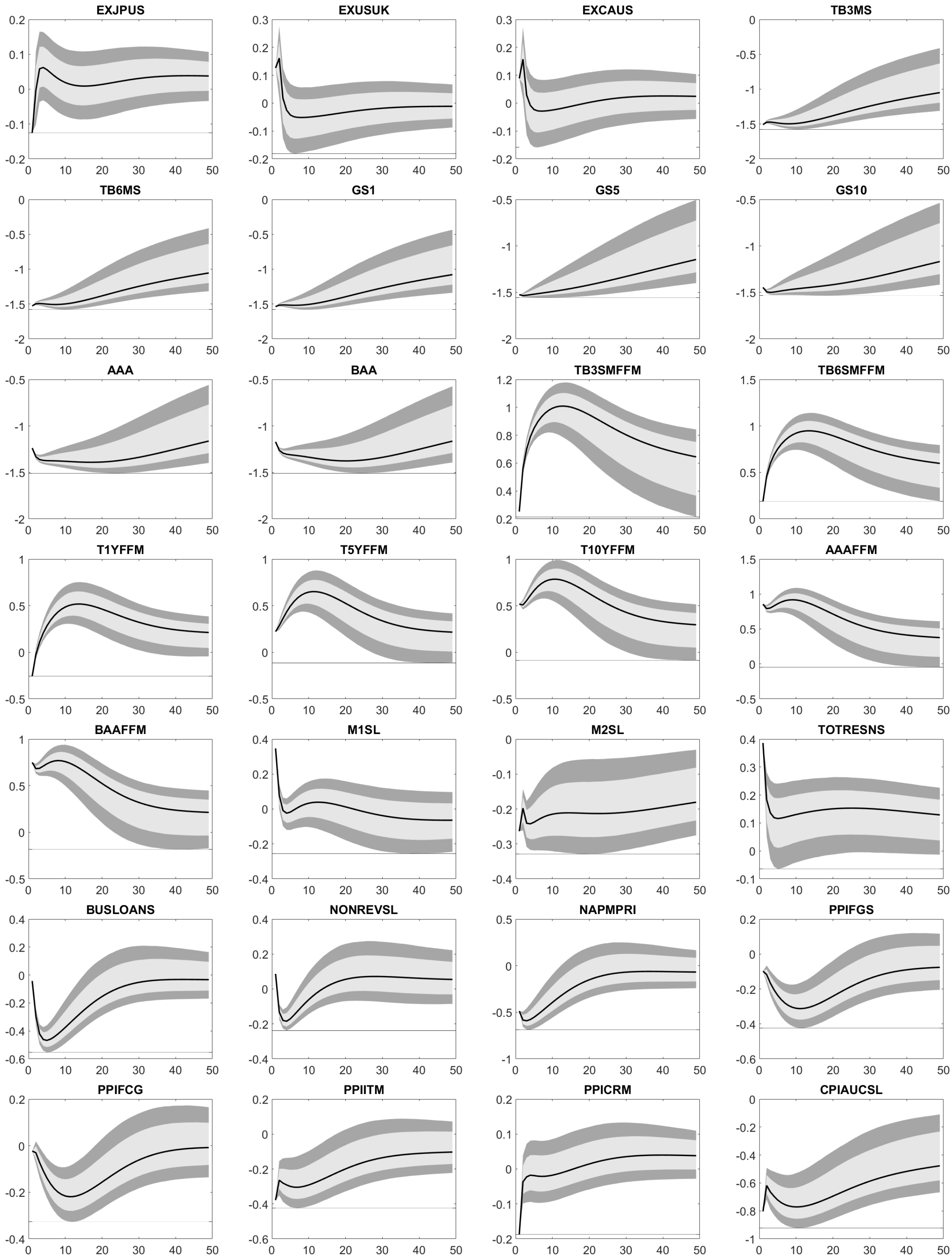

Figure 5.3: *IRFs (black lines) of standardized time series 57-84 in Appendix B.3 arising from an increase in FEDFUNDS by 0.25%. Light gray areas show the 68%-confidence intervals, i.e., the 1-sigma interval, while dark gray areas display the 90%-confidence intervals. In both cases, the intervals are based on 10,000 non-parametric bootstrap simulations of the transition equation, but the estimated loadings matrix is kept fixed.*

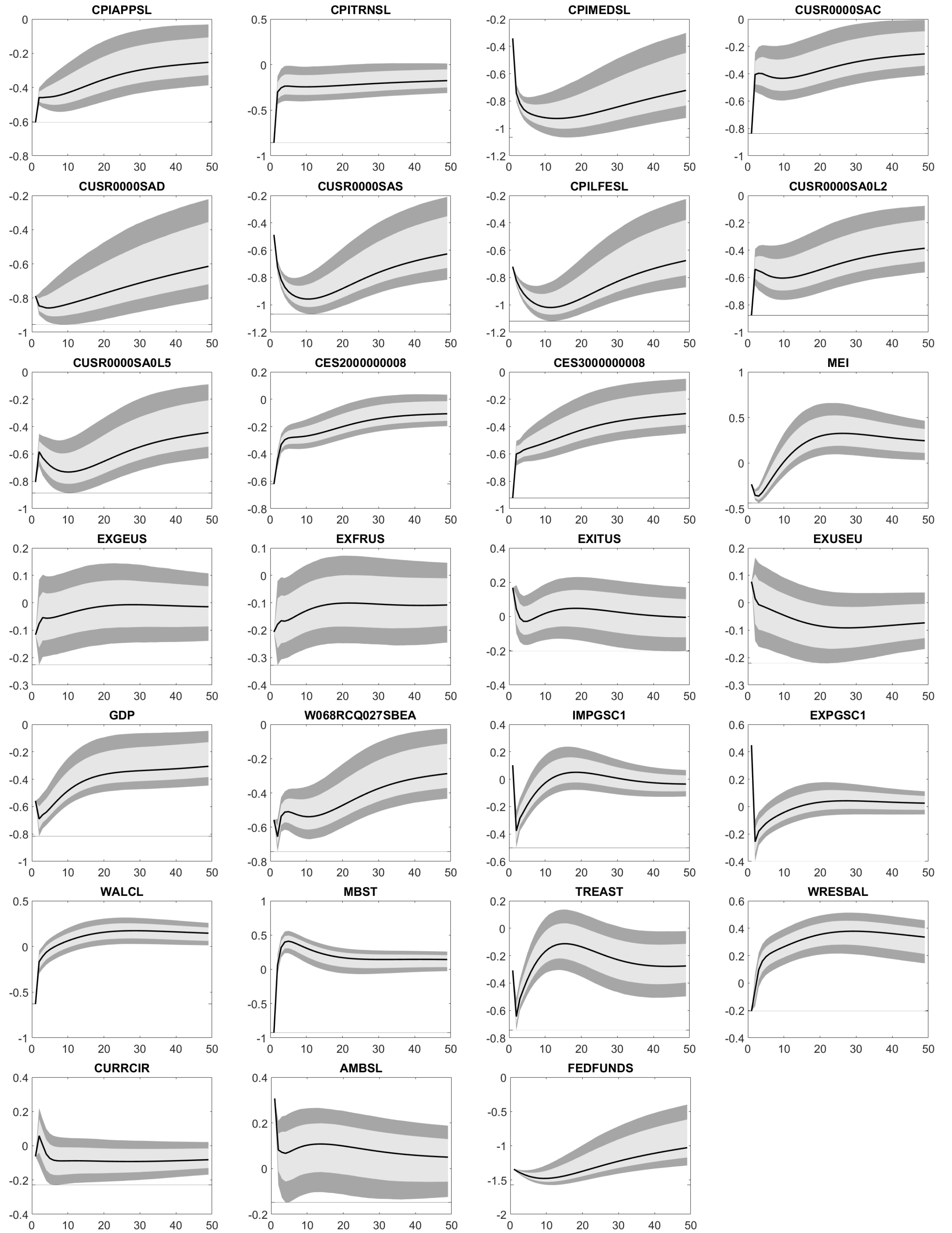

Figure 5.4: *IRFs (black lines) of standardized time series 85-111 in Appendix B.3 arising from an increase in FEDFUNDS by 0.25%. Light gray areas show the 68%-confidence intervals, i.e., the 1-sigma interval, while dark gray areas display the 90%-confidence intervals. In both cases, the intervals are based on 10,000 non-parametric bootstrap simulations of the transition equation, but the estimated loadings matrix is kept fixed.*

corporate bond spreads (AAA, BAA, AAAFFM, BAAFFM) follow an increase in FEDFUNDS.

In Figure 5.3, the drops in M1SL, TOTRESNS, BUSLOANS and NONREVSL let the available liquidity shrink, what the US Federal Reserve is exactly aiming at. In addition, prices and inflation (NAPMPRI, PPIFGS, PPIITM, PPICRM, CPIAUCSL, CPIAPPSL, CPITRNSL, CUSR0000SAC, CUSR0000SAD, CUSR0000SA0L2, CUSR0000SA0L5) climb in the long term such that the US economy eventually leaves its crisis mode and comes back to normal. This assumption is supported by the raising composite leading indicator MEI and GDP in Figure 5.4. Although there are no long-term effects on the export and import of goods and serices (EXPGSC1, IMPGSC1), both decrease after an increase of FEDFUNDS. The reduced export might arise from the strong USD, which makes US products more expensive abroad. By contrast, the strong USD reduces the USD prices of foreign products. Hence, the drop in USD prices is not balanced by a bigger amount of imported products. Finally, Figure 5.4 suggests changes in the assets and reserves of the Federal Reserve (WALCL, MBST, TREAST, WRESBAL, AMBSL).

Besides the IRFs in Figures 5.1-5.4, Figures 5.5-5.8 show the FEVD of all panel data variables. That is, each plot displays the contributions of the shocks, which belong to the variables $\boldsymbol{Y}_t$, as defined in (5.79). To be more precise, we stack the single ratios of the innovations in CURRCIR, AMBSL and FEDFUNDS and conclude: First, the total contribution as well as the single contributions of the shocks in CURRCIR, AMBSL and FEDFUNDS considerably change over time and depend on the chosen panel data variable. Second, CURRCIR innovations heavily affect the forecast error variance of IPB53100N, IPB53200N, RPI, W875RX1, HOUST, HOUSTS, HOUSTW, PERMITNSA, EXPGSC1 and MBST, which rank among the macroeconomic data. In case of AMBSL, we have a scattered picture. On the one hand, its shocks drive the forecast error variance of production data (IPFINAL, IPBUSEQ, IPMAT, INDPRO, CUMFNS, GDP, IMPGSC1), income (RPI, W875RX1), employment (PAYEMS, USGOOD, MANEMP, NDMANEMP, CES0800000001, USTPU, USWTRADE, USFIRE, USPBS), consumption (PCE, PCEND, PCES) and inflation (NAPMPRI, PPIFGS, PPIFCG, PPIITM, PPICRM, CPIAPPSL, CPITRNSL, CUSR0000SAC, CES3000000008). On the other hand, they also influence the forecast error variance of financial data (FSP-COM, EXJPUS, EXUSUK, EXCAUS, EXGEUS, EXFRUS, EXITUS, EXUSEU) and liquidity measures (M1SL, M2SL, TOTRESNS, BUSLOANS). Similarly, the impact of FEDFUNDS shocks covers all areas, but in case of US Treasuries (TB3MS, TB6MS, GS1, GS5, GS10, TB3SMFFM, TB6SMFFM, T1YFFM, T5YFFM, T10YFFM) and corporate bond spreads (AAA, BAA, AAAFFM, BAAFFM) FEDFUNDS innovations are most prominent. Note, for these variables the total contribution of the CURRCIR, AMBSL and FEDFUNDS innovations is largest. Besides the observable factors, the variance of the idiosyncratic error $(\Sigma_{\boldsymbol{e}})_{ii}$ usually represents another important driver of the forecast error variance.

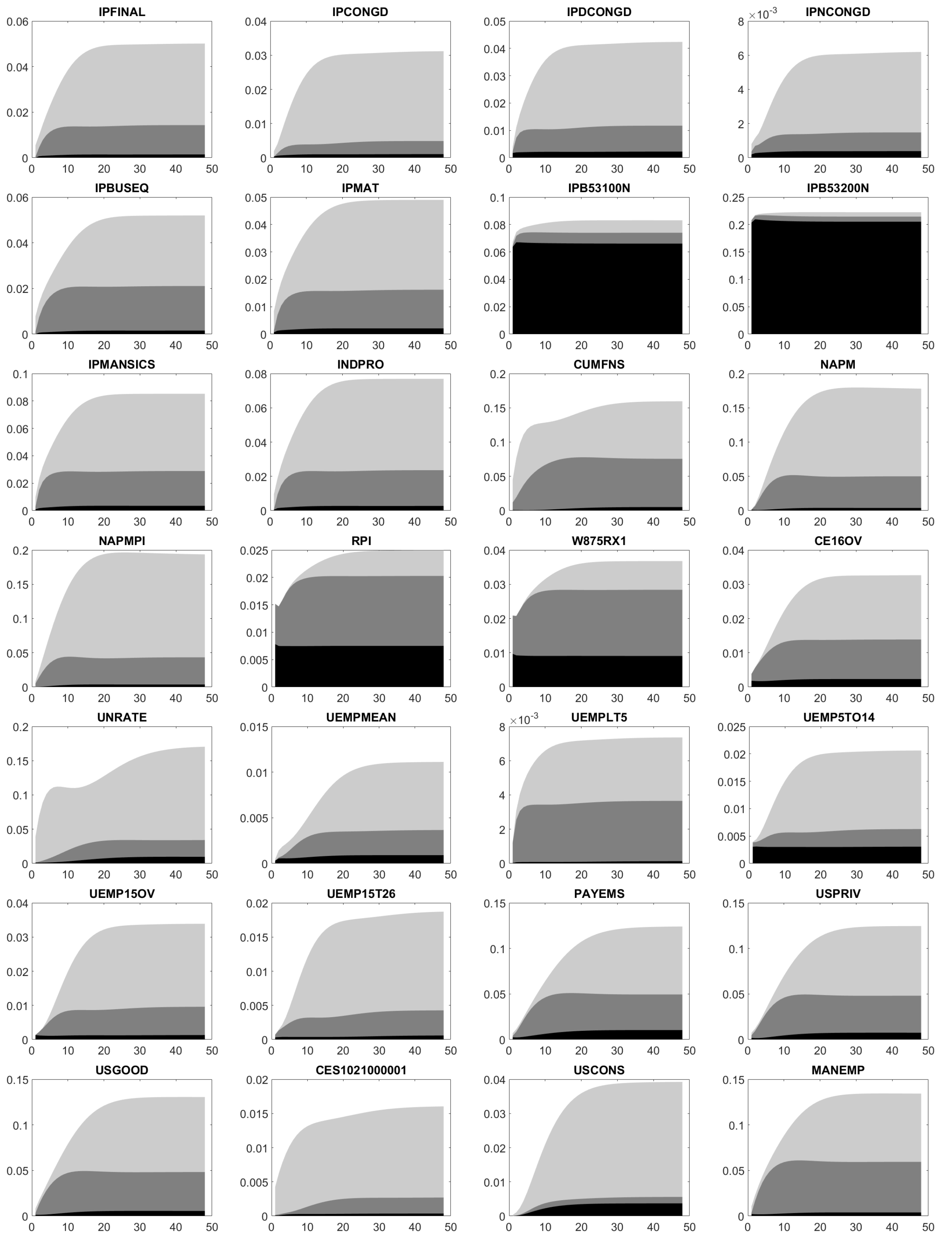

Figure 5.5: *Contributions of CURRCIR (black area), AMBSL (dark gray area) and FEDFUNDS (light gray area) to the forecast error variance of the standardized variables 1-28 in Appendix B.3 over the next 48 months.*

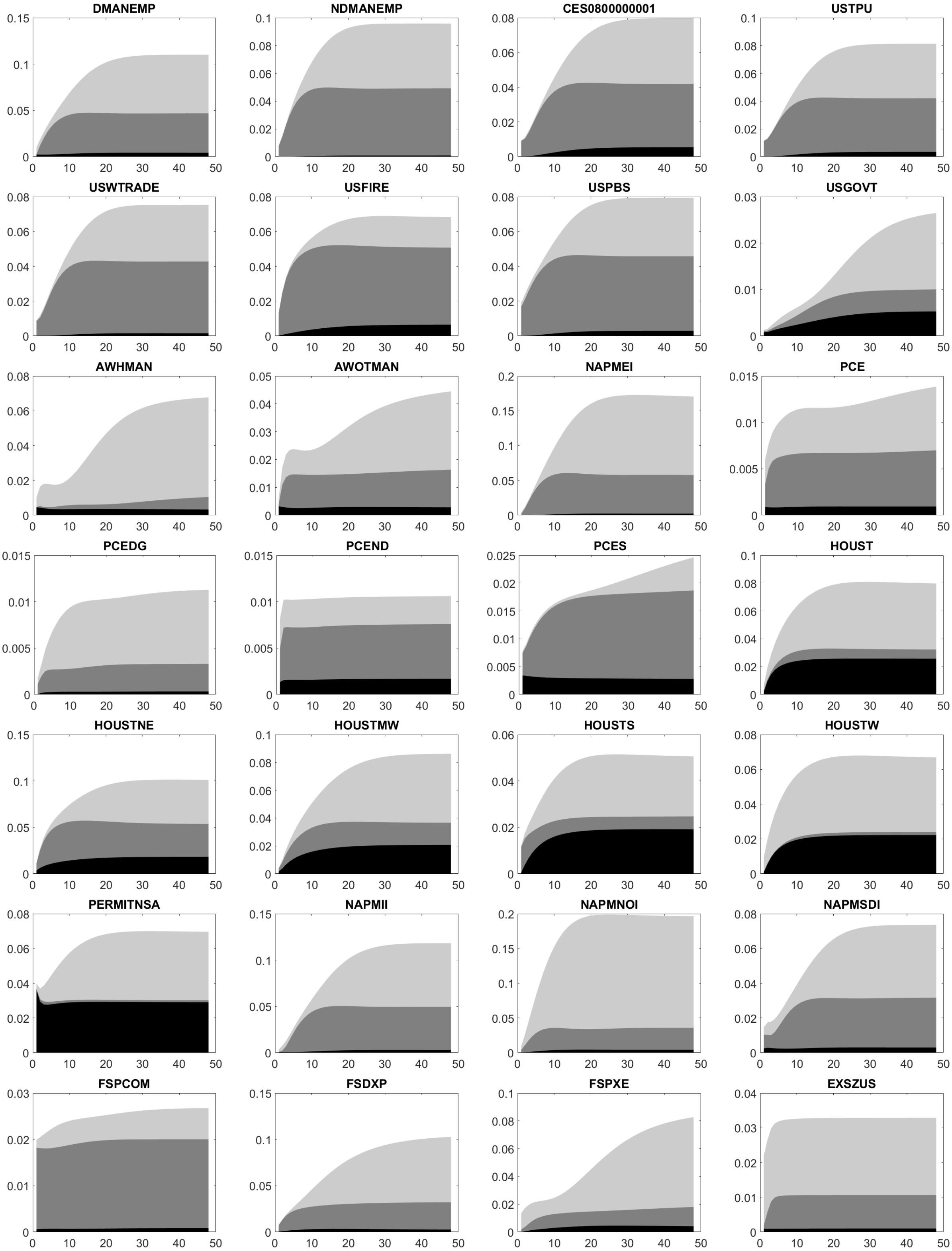

Figure 5.6: *Contributions of CURRCIR (black area), AMBSL (dark gray area) and FEDFUNDS (light gray area) to the forecast error variance of the standardized variables 29-56 in Appendix B.3 over the next 48 months.*

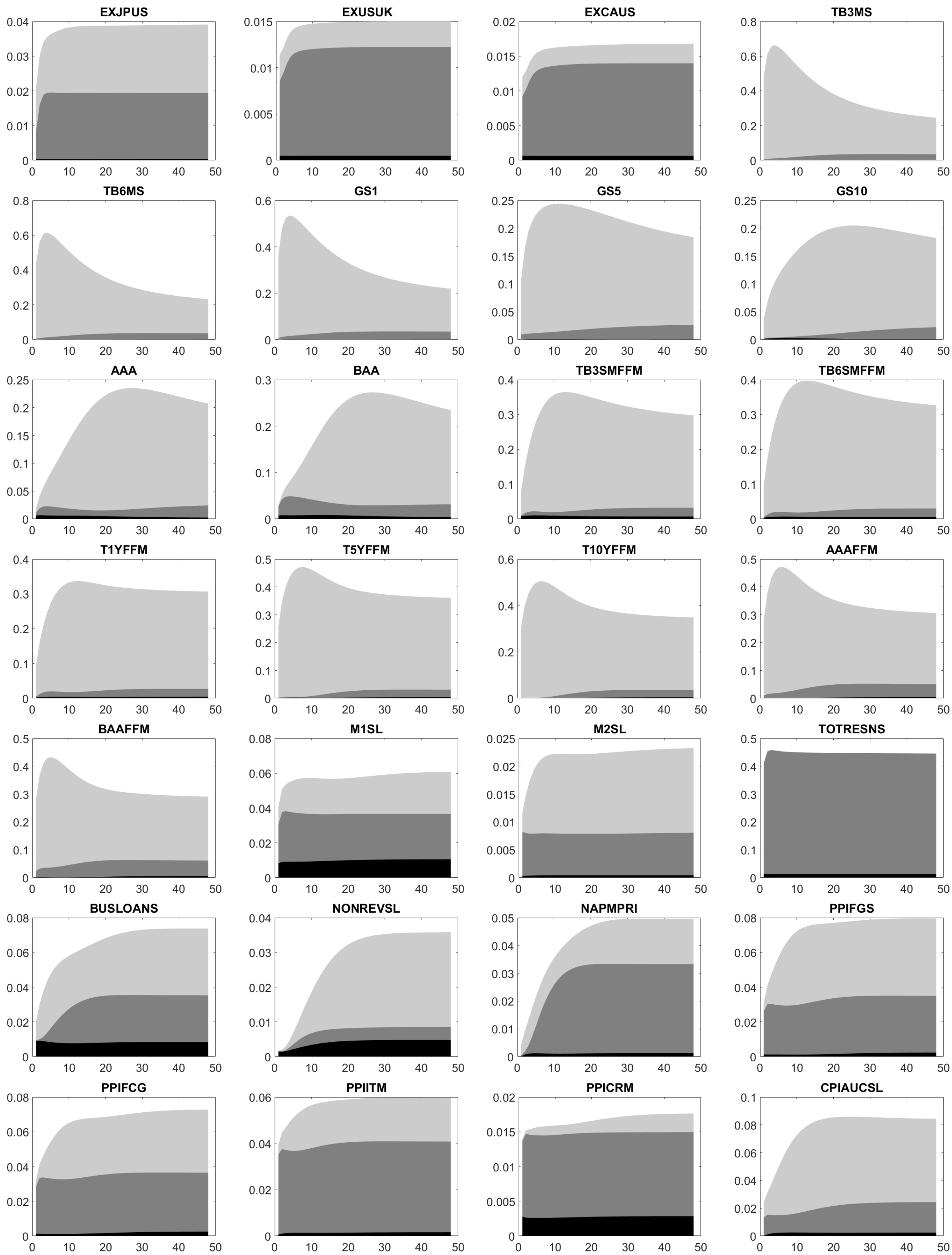

Figure 5.7: *Contributions of CURRCIR (black area), AMBSL (dark gray area) and FEDFUNDS (light gray area) to the forecast error variance of the standardized variables 57-84 in Appendix B.3 over the next 48 months.*

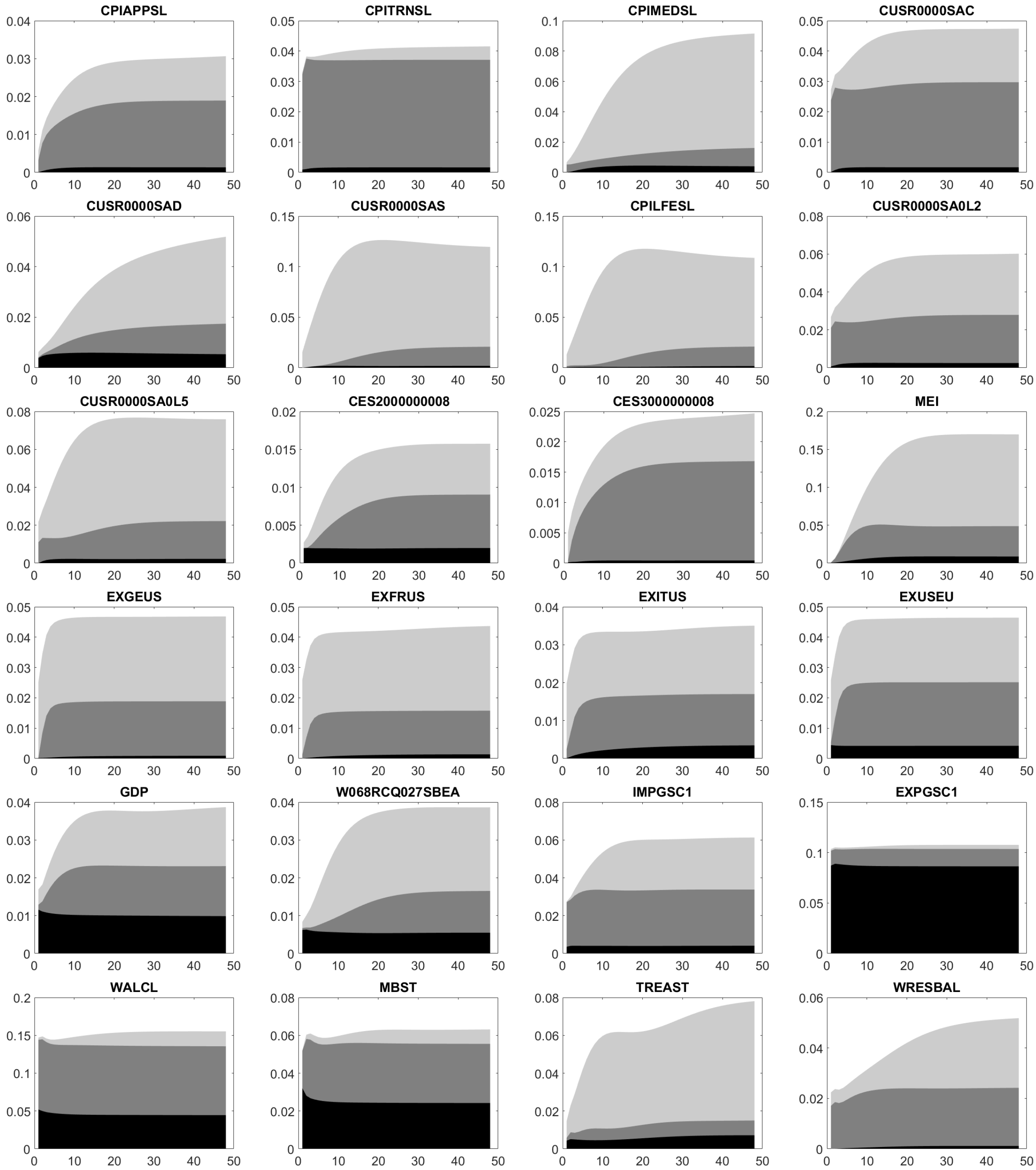

Figure 5.8: *Contributions of CURRCIR (black area), AMBSL (dark gray area) and FEDFUNDS (light gray area) to the forecast error variance of the standardized variables 85-108 in Appendix B.3 over the next 48 months.*

5.4 Conclusion and Future Research

This chapter estimated the FAVARs of Bernanke et al. (2005). Thereby, the panel data can be incomplete, but the time series $\{\boldsymbol{Y}_t\}$, i.e., the observable factor components, must be observed in full. Note, in case of FAVARs, a joint vector of hidden factors $\boldsymbol{F}_t$ and observed variables $\boldsymbol{Y}_t$, whose dynamics obeys a VAR(p), $p \geq 1$, describes the panel data $\boldsymbol{X}_t$. Hence, a direction of the future research is to extend our estimation approach such that incomplete time series $\{\boldsymbol{Y}_t\}$ are possible, too.

Regarding our estimation method, a fully parametric two-step routine simultaneously estimates the unknown model parameters and missing data in a maximum likelihood framework. In a nutshell, two EMs are alternately applied until the absolute value of relative change in the expected log-likelihood function is negligible. In this context, the first EM derives complete data from the observations and latest parameter estimates, while the second EM reestimates the parameters as soon as the balanced data set is updated. Here, we also discuss the selection of the factor dimension and autoregressive order, which is important, in particular, for empirical studies.

The main contributions of this chapter to the existing literature are stated in the sequel: First, we extend the FAVARs of Bernanke et al. (2005) to incomplete panel data. Although Marcellino and Sivec (2016) did the same, their approach requires the observed variables $\boldsymbol{Y}_t$ to be part of the panel data. By contrast, we are free of this condition. Moreover, our estimation method treats any autoregressive order $p \geq 1$ of the factor dynamics explicitly, as we do not stick to the argument that any VAR$(p), p \geq 1$, coincides with a VAR(1) of higher dimension. On the one hand, this is a nice feature. On the other hand, it admits that our method can be applied without any adjustments.

Second, we modify the standard Kalman Filter to take the partially observed factors, i.e., the variables $\boldsymbol{Y}_t$, into account. For this purpose, we choose a state-space representation of FAVARs, which is not most common, and repeat the steps from the proof of the KF. For the Kalman Smoother, we do the same, but finally obtain the standard KS equations. Therefore, we do not have to treat FAVARs as special ADFMs, which goes to Bork (2009). This was why he had to add the time series $\{\boldsymbol{Y}_t\}$ to the panel data $\{\boldsymbol{X}_t\}$ and to include certain loadings constraints. By contrast, we can do the same, but are not obliged to do so.

Third, the new estimation method offers more flexibility than classical ones. Besides the loadings matrix, it permits linear restrictions of the VAR(p) coefficients of the factor dynamics. Note, with the help of the model transformations in Bai et al. (2015), we reduce the degrees of freedom without a need for linear parameter constraints. As an alternative solution, we adjust our EM accordingly. That is, it automatically incorporates the linear constraints and only updates the remaining model parameters.

Fourth, the inclusion of mixed-frequency data enables us to investigate the impact of the monetary policy on quarterly indicators, e.g., the GDP. This is why our empirical study analyzes how an increase in the Effective Federal Funds Rates by 0.25% affects the US economy. Based on a data sample, which covers 108 macroeconomic variables as panel data and a three-dimensional $\boldsymbol{Y}_t$ over a period from January 1959 until October 2015, we conclude that the GDP gains from an increase in the Effectice Federal Funds Rate by 0.25% in the long term.

Possible directions of the future research are as follows: Regime-switching concepts became more popular in the recent literature. Hence, the formulation and estimation of FAVARs with incomplete panel data in a regime-switching framework is natural from our point of view. Furthermore, our two-step estimation method requires completely observable factor components. Therefore, one could consider FAVARs with observable factors of mixed frequencies. Eventually, Copula Autoregressive Models have been successfully used for multivariate time series. The combination of those and FAVARs could capture non-linear factor dependencies and might lead to better shock analyses.

Chapter 6

Conclusion and Future Research

The central theme of this thesis is twofold. On the one hand, we improve the estimation of Factor Models with incomplete panel data. In this regard, we link well-known concepts in a two-step estimation method, which alternately applies two Expectation-Maximization Algorithms until a certain termination criterion is reached. In contrast to alternative approaches, our parametric estimation method explicitly takes cross-sectional correlation of idiosyncratic shocks into account. As shown in the scope of a MC simulation study, this is an import feature for small, incomplete samples with cross-sectionally correlated errors. Moreover, depending on the characteristics of the underlying FM formulation, e.g., for the partially observed factors in FAVARs, we accordingly adjust our estimation routine. This triggered additional contributions to the existing literature as discussed below.

On the other hand, we show practical applications of FMs in the field of asset allocation decisions and risk management. That is, we set up a framework for the monitoring of financial markets, which delivers point and interval forecasts of returns for future periods of time, breaks the expected returns down into the single contributions of the input data, provides nowcasts of low-frequency signals and finally, determines means and covariance matrices of the predicted returns. Then, these means and covariance matrices enter classical mean-variance or marginal-risk-parity portfolio optimizations to enhance the performance of the associated portfolios. In addition to portfolio strategies, we develop single-market trading strategies, which are deterministic functions of the obtained prediction intervals.

In total, this work contributes to the existing literature as follows: First, we estimate Exact Static Factor Models with incomplete panel data to improve the performance of mean-variance or marginal-risk-parity optimal portfolios. That is, we apply models and techniques from statistics to problems in the area of asset allocation decisions and risk management. Thereby, we work with US data.

Second, we utilize closed-form expressions for the factor means and covariance matrices in Approximate Dynamic Factor Models. Since these enter our estimation method for ADFMs based on mixed-frequency panel data with missing observations, we estimate ADFMs of diverse factor dimensions and lag orders for various sample sizes and ratios of data gaps in an intense MC simulation study in two ways: On the one hand, we apply our closed-form solutions for the factor means and covariance matrices as part of our two-step estimation method. On the other hand, a run of the standard Kalman Filter and Smoother, ceteris paribus, provides the factor means and covariance matrices. Thereby, a comparison of both approaches confirms the superiority of the closed-form solutions and so, the usage of the standard KF and KS becomes optional instead of mandatory.

Third, Bańbura and Modugno (2014) presented an estimation method for ADFMs, which admitted arbitrary patterns of ragged panel data and extended the work in Dempster et al. (1977), Rubin and Thayer (1982) and Shumway and Stoffer (1982). However, they actually derived their estimation procedure for EDFMs and referred to Doz et al. (2012) to justify its validity for ADFMs. Note, Doz et al. (2012) consistently estimated factors in ADFMs with cross-sectionally and serially correlated shocks as missspecified EDFMs in a maximum likelihood framework. Thereby, they deemphasized cross-sectional error correlation. In this thesis, we develop a two-step estimation procedure for ADFMs with incomplete panel data that explicitly takes cross-sectional error correlation into account and show through MC simulations that for small, ragged panel data with many gaps cross-sectional error correlation may matter. In addition, we demonstrate that our two-step approach outperforms the one-step estimation method of Bańbura and Modugno (2014) in such cases. Moreover, our two-step estimation method facilitates the combination of irregular patterns, e.g., weeks and months, which requires further state-space modifications regarding the one-step approach of Bańbura and Modugno (2014).

Fourth, we address the problem of selecting the factor dimension and lag order in ADFMs with incomplete panel data. In this sense, we recommend a two-step model selection routine that automatically determines both parameters. Hence, our estimation method for ADFMs simultaneously performs model selection and parameter estimation. This also explains why our derivations rely on general factor dynamics of order $p \geq 1$ instead of the simple case with $p = 1$. Finally, we examine the proposed model selection criteria in an intensive MC simulation study.

Fifth, we provide point and interval forecasts for returns of future periods of time. Thereby, we decompose our point forecasts. That is, we figure out for all panel data variables to what extent and in which direction they contribute to the predicted returns. Moreover, we generate interval forecasts in the form of prediction intervals, which incorporate uncertainties arising from factor and parameter estimation. In this context, we also introduce single-market trading strategies, which are deterministic functions of our prediction intervals. The idea behind such trading strategies is to convert prediction intervals into concrete market trades.

Sixth, Bork (2009) and Marcellino and Sivec (2016) treat the FAVARs of Bernanke et al. (2005) as special ADFMs to apply standard ADFM techniques for their estimation. Therefore, they are obliged to treat the observable factors as subset of the panel data and to incorporate appropriate loadings constraints. In doing so, they use the standard Kalman Filter and Smoother, which ignore that factors in FAVARs are partially observed. By contrast, we adapt the original KF such that it includes the observed factors and confirm the validity of the usual KS.

Seventh, due to the modified KF and KS, we extend the FAVARs of Bernanke et al. (2005) to incomplete panel data. Parameters and factors in FMs and FAVARs are unique except for rotation. For identification purposes, we therefore transform the general FAVAR formulation as in Bai et al. (2015). Then, we deploy parameter constraints to remove left degrees of freedom. Besides loadings constraints, which are common for FAVARs, our estimation method admits the inclusion of linear restrictions for the VAR(p) coefficients of the factor dynamics. Hence, we become more flexible in case of parameter constraints, which prevents us from parameter ambiguity.

Among other things, possible directions for the future research are as follows: First, our model for weekly S&P500 log-returns stumbles a little during the financial crisis of the years 2008/2009. That is, there are more interval outliers than in the run-up and sequel of the crisis. Thus, the inclusion of regime-switching concepts could improve the forecasting accuracy. Second, we considered ADFMs with cross-sectionally, but not serially correlated errors. Therefore, estimation methods for ADFMs with homoscedastic, cross-sectionally and serially correlated shocks and incomplete panel data represent another research area. If this works, research could proceed with heteroscedastic errors. Third, our two-step estimation method for FAVARs admits incomplete panel data, but requires the observable factors to be observed in full without any gaps. Consequently, an estimation method for FAVARs that also permits data incompleteness for the panel data as well as the observable factor components might outline another research area.

Appendix A

Additional Definitions and Proofs

This appendix pursues two targets. On the one hand, it lists well-known mathematical principles, which are important for this thesis, and states corresponding references for their proofs. This is for the readers' convenience to reduce the need for external references. Moreover, if there are differences in the literature, we specify the notation we work with. On the other hand, this appendix shortens our main chapters, since some lengthy calculations were moved here. If applicable, alternative proofs and additional comments are also stated here to avoid unnecessary confusion.

A.1 Mathematical Principles

We briefly repeat the main mathematical fundamentals entering our calculations. As these are well-known in the literature, we state appropriate references instead of proving their validity once again.

Definition A.1.1 (Matrix Trace)
The trace of a square matrix $A = (a_{ij})_{ij} \in \mathbb{R}^{K \times K}$ coincides with the sum of its diagonal elements, i.e.:

$$tr(A) = \sum_{k=1}^{K} a_{kk}.$$

The above definition is in line with Rao and Toutenburg (1999, p. 355, Definition A.12).

Lemma A.1.2 (Properties of Matrix Trace)
Assume matrices $A \in \mathbb{R}^{K \times K}, B \in \mathbb{R}^{K \times K}, C \in \mathbb{R}^{N \times K}, D \in \mathbb{R}^{K \times N}$ and scalar $s \in \mathbb{R}$. Then, it holds:

(i) $tr(s) = s,$

(ii) $tr(A \pm B) = tr(A) \pm tr(B),$

(iii) $tr(sA) = s\ tr(A),$

(iv) $tr(A') = tr(A)$*, with* A' *as the transpose of matrix* $A,$

(v) $tr(CD) = tr(DC),$

(vi) $tr(ADC) = tr(CAD) = tr(DCA)$.

Proof:
The relation in *(i)* results from Definition A.1.1, while *(ii)*-*(v)* are given in Rao and Toutenburg (1999, p. 355, Theorem A.13). For more details see, e.g., the references at the start of Appendix A in Rao and Toutenburg (1999, p. 353). The claim in *(vi)* follows from *(v)*, if the matrix arising from the multiplication of two adjacent matrices is regarded, that is, the initial product of three matrices simplifies to a product of two matrices (Harville, 1997, p. 51, Eq. 2.9). □

Remark A.1.3 (Invariance under Cyclical Permutation of Matrix Trace)
The proof of (vi) in Lemma A.1.2 works for matrix products of more than three matrices, too. In total, this yields that the matrix trace is invariant unter cyclical permutation for any finite number of matrices.

With regard to Definition A.1.1, the matrix trace is a real-valued, differentiable function $tr : \mathbb{R}^{K\times K} \to \mathbb{R}$. Unfortunately, there are distinct conventions of how to differentiate a scalar function with respect to a matrix. Here, we take the version of Rao and Toutenburg (1999, p. 384, Definition A.90).

Definition A.1.4 (Differentiation of Scalar Functions)
Let $\partial f/\partial x_{ij}$ with $1 \leq i \leq K, 1 \leq j \leq N$ be the partial derivatives of a real-valued function $f : \mathbb{R}^{K\times N} \to \mathbb{R}$. Then, the partial differential of function f with respect to X is defined as:

$$\frac{\partial f(X)}{\partial X} = \begin{pmatrix} \frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1N}} \\ \vdots & & \vdots \\ \frac{\partial f}{\partial x_{K1}} & \cdots & \frac{\partial f}{\partial x_{KN}} \end{pmatrix}.$$

With Definition A.1.4 in mind, it follows for the matrix trace:

Lemma A.1.5 (Derivatives of Matrix Trace Functions)
For any non-singular matrix Y, let Y^{-1} be its inverse matrix. Furthermore, let matrix Z' be the transpose of matrix Z. Then, the subsequent equalities remain valid:

(i) $\frac{\partial}{\partial X}(tr(AX)) = A'$, *for* $X \in \mathbb{R}^{K\times N}, A \in \mathbb{R}^{N\times K}$,

(ii) $\frac{\partial}{\partial X}(tr(X'AX)) = (A + A')X$, *for* $X \in \mathbb{R}^{K\times N}, A \in \mathbb{R}^{K\times K}$,

(iii) $\frac{\partial}{\partial X}(tr(XAX)) = X'A + A'X'$, *for* $X \in \mathbb{R}^{K\times K}, A \in \mathbb{R}^{K\times K}$,

(iv) $\frac{\partial}{\partial X}(tr(XAX')) = X(A + A')$, *for* $X \in \mathbb{R}^{K\times N}, A \in \mathbb{R}^{N\times N}$,

(v) $\frac{\partial}{\partial X}(tr(X'AX')) = AX' + X'A$, *for* $X \in \mathbb{R}^{K\times K}, A \in \mathbb{R}^{K\times K}$,

(vi) $\frac{\partial}{\partial X}(tr(X'AXB)) = AXB + A'XB'$, *for* $X \in \mathbb{R}^{K\times N}, A \in \mathbb{R}^{K\times K}, B \in \mathbb{R}^{N\times N}$,

(vii) $\frac{\partial}{\partial X}\left(tr\left(AX^{-1}\right)\right) = -\left(X^{-1}AX^{-1}\right)'$, *for non-singular* $X \in \mathbb{R}^{K\times K}$ *and* $A \in \mathbb{R}^{K\times K}$.

(viii) $\frac{\partial}{\partial X}\left(tr\left(X^{-1}AX^{-1}B\right)\right) = -\left(X^{-1}AX^{-1}BX^{-1} + X^{-1}BX^{-1}AX^{-1}\right)'$,
for non-singular $X \in \mathbb{R}^{K\times K}$ *and* $A, B \in \mathbb{R}^{K\times K}$.

Proof:
For *(i)-(vi)*, see Rao and Toutenburg (1999, p. 386, Theorem A.97). The latter points are stated in Rao and Toutenburg (1999, p. 386, Theorem A.97). □

Besides the matrix trace, our calculations include the determinant of a matrix. For its definition see, e.g., Harville (1997, p. 192, Theorem 13.5.1) and Rao and Toutenburg (1999, p. 356, Definition A.14).

Definition A.1.6 (Matrix Determinant)
The determinant of a square matrix $A = (a_{ij})_{ij} \in \mathbb{R}^{K\times K}$ is defined as:

$$|A| = \sum_{i=1}^{K} (-1)^{i+j} \, a_{ij}|M_{ij}|, \quad \textit{for any fixed } 1 \le j \le K,$$

with $|M_{ij}|$ denoting the first minor of the element a_{ij}. Here, the first minor $|M_{ij}|$ of a_{ij} is the determinant of the reduced matrix arising after the i-th row and the j-th column of A have been removed. Let $A_{ij} = (-1)^{i+j}\,|M_{ij}|$ be the sign adjusted minor, then, A_{ij} is called the cofactor of a_{ij}. For any scalar $s \in \mathbb{R}$, it holds: $|s| = s$, which completes this recursive definition of the matrix determinant.

Lemma A.1.7 (Properties of Matrix Determinant)
Let $A, B \in \mathbb{R}^{K\times K}$ and $C \in \mathbb{R}^{N\times N}$ be square matrices. Assume matrix $D \in \mathbb{R}^{K\times N}$, scalar $s \in \mathbb{R}$ and let $O_{N\times K} \in \mathbb{R}^{N\times K}$ be a zero matrix. Then, the matrix determinant satisfies the following:

(i) $|A'| = |A|$,

(ii) $|s\,A| = s^K|A|$,

(iii) $|AB| = |A||B|$,

(iv) $|A| = \prod_{k=1}^{K} a_{kk}$, *if A is a diagonal or triangular matrix,*

(v) $\left|\begin{pmatrix} A & D \\ O_{N\times K} & C \end{pmatrix}\right| = |A||C| = \left|\begin{pmatrix} A' & O'_{N\times K} \\ D' & C' \end{pmatrix}\right|$,

(vi) $|A^{-1}| = \frac{1}{|A|}$, *if A is a non-singular matrix.*

Proof:
Points *(i)-(v)* are taken from Rao and Toutenburg (1999, p. 356, Theorem A.16), while the last is given in Rao and Toutenburg (1999, p. 358, Theorem A.18). □

Lemma A.1.8 (Derivatives of Matrix Determinant Functions)
For any positive scalar $s \in \mathbb{R}_+$, let $ln\,(s)$ denote its natural logarithm and let $X \in \mathbb{R}^{K\times K}$ be a non-singular square matrix. Then, it follows:

(i) $\frac{\partial}{\partial X}|X| = |X|\,(X')^{-1}$,

(ii) $\frac{\partial}{\partial X}\, ln\,(|X|) = (X')^{-1}$.

Proof:
See Rao and Toutenburg (1999, p. 387, Theorem A.98). $\square$

To simplify some expressions we use the Kronecker product as in Harville (1997, p. 337, Section 16.1), which is defined as follows:

Definition A.1.9 (Kronecker Product)
Let $A = (a_{ij})_{ij} \in \mathbb{R}^{K\times N}$ and $B = (b_{ij})_{ij} \in \mathbb{R}^{J\times L}$ be two matrices. Then, the Kronecker product of A and B is given by:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1N}B \\ a_{21}B & a_{22}B & \cdots & a_{2N}B \\ \vdots & \vdots & & \vdots \\ a_{K1}B & a_{K2}B & \cdots & a_{KN}B \end{pmatrix} \in \mathbb{R}^{KJ\times NL}.$$

Lemma A.1.10 (Properties of Kronecker Product)
Assume matrices $A \in \mathbb{R}^{K\times N}, B \in \mathbb{R}^{J\times L}, C \in \mathbb{R}^{K\times N}, D \in \mathbb{R}^{U\times V}, E \in \mathbb{R}^{N\times U}, F \in \mathbb{R}^{L\times V}$ and scalar $s \in \mathbb{R}$. Furthermore, let $G \in \mathbb{R}^{K\times K}$ and $H \in \mathbb{R}^{N\times N}$ be square matrices. Then, it holds:

(i) $s \otimes A = A \otimes s = s\ A$,

(ii) $(s\ A) \otimes B = A \otimes (s\ B) = s\ (A \otimes B)$,

(iii) $(A + C) \otimes B = (A \otimes B) + (C \otimes B)$,

(iv) $B \otimes (A + C) = (B \otimes A) + (B \otimes C)$,

(v) $(A \otimes B)' = A' \otimes B'$,

(vi) $(A \otimes B) \otimes D = A \otimes (B \otimes D)$,

(vii) $(A \otimes B)(E \otimes F) = (AE) \otimes (BF)$,

(viii) $(G \otimes H)^{-1} = G^{-1} \otimes H^{-1}$, *if G and H are non-singular matrices,*

(ix) $tr(G \otimes H) = tr(G)\ tr(H)$,

(x) $|G \otimes H| = |G|^N\ |H|^K$.

Proof:
In Harville (1997), see Equation (1.1) on p. 338 for *(i)*. The results in *(ii)*-*(iv)* are stated in Equations (1.10)-(1.12) on p. 339. The transpose in *(v)* is given by Equation (1.15) on p. 340. For the associativity law in *(vi)*, see Lemma 16.1.1 on p. 340, whereas *(vii)* is shown in Lemma 16.1.2 on p. 341. For the inverse in *(viii)*, see Equation (1.23) on p. 342. The trace in *(ix)* is provided by Equation (1.25) on p. 342. Finally, for the matrix determinant in *(x)*, see Equation (3.18) on p. 354. $\square$

Lemma A.1.11 (Eigenvalues of Kronecker product)
For matrices $A \in \mathbb{R}^{K\times K}$ and $B \in \mathbb{R}^{N\times N}$ with possibly nondistinct eigenvalues $\lambda_1, \ldots, \lambda_K$ and $\mu_1, \ldots, \mu_N$, respectively, the eigenvalues of $A \otimes B$ are given by the products $\lambda_i\, \mu_j$ with $1 \leq i \leq K$ and $1 \leq j \leq N$.

Proof:
See Section "Eigenvalues of a Kronecker Product" on p. 733 in Hamilton (1994). □

Sometimes, reshaping matrices supports solving linear equation systems. For this purpose, we introduce the vectorization operator (Harville, 1997, p. 343, Eq. 2.1) as given below:

Definition A.1.12 (Vectorization of Matrix)
The vectorization of matrix $A = (a_{ij})_{ij} \in \mathbb{R}^{K \times N}$ stacks one column of A after the other such that the first is on top and the last is at the bottom, that is:

$$vec\,(A) = (a_{11}, \ldots, a_{K1}, a_{12}, \ldots, a_{K2}, \ldots, a_{1N}, \ldots, a_{KN})' \in \mathbb{R}^{KN \times 1}.$$

Lemma A.1.13 (Properties of Vectorization Operator)
For matrices $A \in \mathbb{R}^{K \times N}, B \in \mathbb{R}^{N \times U}$ and $C \in \mathbb{R}^{U \times V}$, which admit the matrix product ABC, we obtain:

$$vec\,(ABC) = (C' \otimes A)\, vec\,(B)\,.$$

Proof:
See Harville (1997, p. 345, Theorem 16.2.1). □

A.2 Factor Models

This section contains basic results in the area of FMs that are well-known in the literature. If applicable, we refer to relevant books and papers for further details. In addition, lenghty calculations and alternative proofs are part of this section to improve the clarity of the main chapters. In case of covariance-stationarity, we take the definition in Hamilton (1994, p. 258).

Definition A.2.1 (Covariance-Stationarity of Vector Process)
A vector process $\{\boldsymbol{X}_t\}$ is said to be covariance-stationary, if its first and second order moments, i.e., $\mathbb{E}[\boldsymbol{X}_t]$ and $\mathbb{E}\left[\boldsymbol{X}_t \boldsymbol{X}'_{t-j}\right]$, for any integer j, are independent of time t.

Alternative proof of Theorem 4.1.4:
From Definition 2.1.4 we derive the covariance matrix of the vectors $\boldsymbol{X}_t$ and $\boldsymbol{F}_t$ as follows:

$$\mathbb{C}\mathrm{ov}_\Theta\left[\boldsymbol{X}_t, \boldsymbol{F}_t\right] = \mathbb{C}\mathrm{ov}_\Theta\left[W\boldsymbol{F}_t + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \boldsymbol{F}_t\right] = W\ \mathbb{C}\mathrm{ov}_\Theta\left[\boldsymbol{F}_t, \boldsymbol{F}_t\right] = W\Sigma_{\boldsymbol{F}}.$$

This and the marginal distributions in Lemma 4.1.3 provide for the joint vector $(\boldsymbol{F}'_t, \boldsymbol{X}'_t)' \in \mathbb{R}^{K+N}$:

$$\begin{pmatrix} \boldsymbol{F}_t \\ \boldsymbol{X}_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{0}_K \\ \boldsymbol{\mu} \end{pmatrix}, \begin{pmatrix} \Sigma_{\boldsymbol{F}} & \Sigma_{\boldsymbol{F}} W' \\ W\Sigma_{\boldsymbol{F}} & W\Sigma_{\boldsymbol{F}} W' + \Sigma_{\boldsymbol{\epsilon}} \end{pmatrix}. \right)$$

Then, we get for the conditional normal distribution (Greene, 2003, pp. 871-872, Theorem B.7):

$$\boldsymbol{F}_t | \boldsymbol{X}_t \sim \mathcal{N}\left(\Sigma_{\boldsymbol{F}} W' \left(W\Sigma_{\boldsymbol{F}} W' + \Sigma_{\boldsymbol{\epsilon}} \right)^{-1} (\boldsymbol{X}_t - \boldsymbol{\mu})\,, \Sigma_{\boldsymbol{F}} - \Sigma_{\boldsymbol{F}} W' \left(W\Sigma_{\boldsymbol{F}} W' + \Sigma_{\boldsymbol{\epsilon}} \right)^{-1} W\Sigma_{\boldsymbol{F}} \right).$$

If we set $M = W'\Sigma_{\boldsymbol{\epsilon}}^{-1}W + \Sigma_{\boldsymbol{F}}^{-1} \in \mathbb{R}^{K\times K}$, we obtain for the covariance matrix:

$$\begin{aligned}
&M\left(\Sigma_{\boldsymbol{F}} - \Sigma_{\boldsymbol{F}}W'\left(W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}\right)^{-1}W\Sigma_{\boldsymbol{F}}\right)\\
&= \left(W'\Sigma_{\boldsymbol{\epsilon}}^{-1}W + \Sigma_{\boldsymbol{F}}^{-1}\right)\left(\Sigma_{\boldsymbol{F}} - \Sigma_{\boldsymbol{F}}W'\left(W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}\right)^{-1}W\Sigma_{\boldsymbol{F}}\right)\\
&= W'\Sigma_{\boldsymbol{\epsilon}}^{-1}W\Sigma_{\boldsymbol{F}} + I_K - W'\left[\Sigma_{\boldsymbol{\epsilon}}^{-1}W\Sigma_{\boldsymbol{F}}W' + I_K\right]\left(W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}\right)^{-1}W\Sigma_{\boldsymbol{F}}\\
&= W'\Sigma_{\boldsymbol{\epsilon}}^{-1}W\Sigma_{\boldsymbol{F}} + I_K - W'\left[\Sigma_{\boldsymbol{\epsilon}}^{-1}\left(W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}\right)\right]\left(W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}\right)^{-1}W\Sigma_{\boldsymbol{F}}\\
&= W'\Sigma_{\boldsymbol{\epsilon}}^{-1}W\Sigma_{\boldsymbol{F}} + I_K - W'\Sigma_{\boldsymbol{\epsilon}}^{-1}W\Sigma_{\boldsymbol{F}} = I_K,
\end{aligned}$$

which proves the covariance matrix representation in Theorem 4.1.4. By similar reasoning, we receive for the conditional factor mean:

$$\begin{aligned}
M^{-1}W'\Sigma_{\boldsymbol{\epsilon}}^{-1} &= \left(\Sigma_{\boldsymbol{F}} - \Sigma_{\boldsymbol{F}}W'\left(W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}\right)^{-1}W\Sigma_{\boldsymbol{F}}\right)W'\Sigma_{\boldsymbol{\epsilon}}^{-1}\\
&= \Sigma_{\boldsymbol{F}}W'\left[\Sigma_{\boldsymbol{\epsilon}}^{-1} - \left(W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}\right)^{-1}W\Sigma_{\boldsymbol{F}}W'\Sigma_{\boldsymbol{\epsilon}}^{-1}\right]\\
&= \Sigma_{\boldsymbol{F}}W'\left[I_N - \left(W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}\right)^{-1}W\Sigma_{\boldsymbol{F}}W'\right]\Sigma_{\boldsymbol{\epsilon}}^{-1}\\
&= \Sigma_{\boldsymbol{F}}W'\left[I_N - \left(W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}\right)^{-1}\left(W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}\right) + \left(W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}\right)^{-1}\Sigma_{\boldsymbol{\epsilon}}\right]\Sigma_{\boldsymbol{\epsilon}}^{-1}\\
&= \Sigma_{\boldsymbol{F}}W'\left(W\Sigma_{\boldsymbol{F}}W' + \Sigma_{\boldsymbol{\epsilon}}\right)^{-1},
\end{aligned}$$

which eventually proves the assertion for the conditional factor mean. □

Lemma A.2.2 (Conversion of VAR(p) with $p \geq 1$ into VAR(1))
Let the process $\{\boldsymbol{F}_t\}$ with $\boldsymbol{F}_t \in \mathbb{R}^K$ be the VAR(p) in (2.5) *with $p \geq 1$. Then, it can be converted into the following VAR(1):*

$$\tilde{\boldsymbol{F}}_t = \tilde{\mathbb{A}}\tilde{\boldsymbol{F}}_{t-1} + \tilde{\boldsymbol{\delta}}_t, \tag{A.1}$$

where $\tilde{\boldsymbol{F}}_t \in \mathbb{R}^{pK}, \tilde{\boldsymbol{\delta}}_t \in \mathbb{R}^{pK}$ and $\tilde{\mathbb{A}} \in \mathbb{R}^{pK\times pK}$ are defined by:

$$\tilde{\boldsymbol{F}}_t = \begin{bmatrix} \boldsymbol{F}_t \\ \vdots \\ \boldsymbol{F}_{t-p+1} \end{bmatrix}, \quad \tilde{\boldsymbol{\delta}}_t = \begin{bmatrix} \boldsymbol{\delta}_t \\ \boldsymbol{0}_K \\ \vdots \\ \boldsymbol{0}_K \end{bmatrix} \quad \textit{and} \quad \tilde{\mathbb{A}} = \begin{bmatrix} A_1 & \cdots & A_{p-1} & A_p \\ I_K & & O_K & O_K \\ & \ddots & & \vdots \\ O_K & & I_K & O_K \end{bmatrix}.$$

Furthermore, we have:

$$\tilde{\boldsymbol{\delta}}_t\tilde{\boldsymbol{\delta}}_s' = \begin{cases} \Sigma_{\tilde{\boldsymbol{\delta}}} & \textit{if } t = s \\ O_{pK} & \textit{otherwise} \end{cases} \quad \textit{and} \quad \Sigma_{\tilde{\boldsymbol{\delta}}} = \begin{bmatrix} \Sigma_{\boldsymbol{\delta}} & O_K & \cdots & O_K \\ O_K & O_K & & \\ \vdots & & \ddots & \\ O_K & & & O_K \end{bmatrix} \in \mathbb{R}^{pK\times pK}.$$

Proof:
See Section "Rewriting a VAR(p) as a VAR(1)" in Hamilton (1994, p. 259). □

In Lemma A.2.2, the transformation of the VAR(p), $p \geq 1$, into a VAR(1) does not call for the covariance-stationarity of $\{\boldsymbol{F}_t\}$. However, for the covariance-stationarity of $\{\boldsymbol{F}_t\}$ and thus, $\{\tilde{\boldsymbol{F}}_t\}$ a certain structure has to exist. Here, $abs\,(\cdot)$ denotes the absolute value of a real number.

Lemma A.2.3 (Conditions for Covariance-Stationarity)
Let $\{\boldsymbol{F}_t\}$ and $\{\tilde{\boldsymbol{F}}_t\}$ be the processes in Lemma A.2.2. If one of two following conditions is satisfied, both processes are said to be covariance-stationary:

(i) If all $\lambda \in \mathbb{C}$ with $|I_K\lambda^p - \sum_{i=1}^{p} A_i\lambda^{p-i}| = 0$ lie inside the unit circle, i.e., $abs\,(\lambda) < 1$.

(ii) If all $z \in \mathbb{C}$ with $|I_K - \sum_{i=1}^{K} A_i z^i| = 0$ lie outside the unit circle, i.e., $abs\,(z) > 1$.

Proof:
See Proposition 10.1 in Hamilton (1994, p. 259). □

Remark A.2.4 (Covariance-Stationarity and Eigenvalues of $\tilde{\mathbb{A}}$)
Because of the assumed covariance-stationarity, the eigenvalues of the matrix $\tilde{\mathbb{A}}$ in Lemma A.2.2 satisfy $|I_K\lambda^p - \sum_{i=1}^{p} A_i\lambda^{p-i}| = 0$ (Hamilton, 1994, p. 259, Proposition 10.1). Hence, condition (i) in Lemma A.2.3 implies that $\tilde{\mathbb{A}}^k \to O_{pK\times pK}$ as $k \to \infty$ (Hamilton, 1994, p. 260, statement before Eq. 10.1.15).

Lemma A.2.5 (Vector MA(∞) Representation of $\tilde{\boldsymbol{F}}_t$)
Let $\{\tilde{\boldsymbol{F}}_t\}$ be the covariance-stationary VAR(1) process in Lemma A.2.2. Then, the infinite moving average, abbreviated by MA(∞), representation of $\tilde{\boldsymbol{F}}_t$, that is,

$$\tilde{\boldsymbol{F}}_t = \sum_{k=0}^{\infty}\left(\tilde{\mathbb{A}}^k\tilde{\boldsymbol{\delta}}_{t-k}\right) \tag{A.2}$$

converges. Moreover, it meets the absolute summability condition, i.e., it holds:

$$\sum_{k=0}^{\infty} abs\left(\tilde{\mathbb{A}}^k_{ij}\right) < \infty, \quad \text{for all} \quad 1 \leq i, j \leq pK,$$

where $\tilde{\mathbb{A}}^k_{ij}$ is the element in the i-th row and j-th column of the k-th power of $\tilde{\mathbb{A}}$.

Proof:
The covariance-stationarity of $\{\tilde{\boldsymbol{F}}_t\}$ provides $\tilde{\mathbb{A}}^k \to O_{pK\times pK}$ as $k \to \infty$ and so, justifies the representation of $\tilde{\boldsymbol{F}}_t$ in (A.2) as convergent series of the errors (Hamilton, 1994, p. 260, Eq. 10.1.15). The process $\{\tilde{\boldsymbol{F}}_t\}$ obeys Equation [10.1.4] in Hamilton (1994, p. 257) and is covariance-stationary. So, the MA(∞) representation of $\tilde{\boldsymbol{F}}_t$ meets the absolute summability condition (Hamilton, 1994, p. 263, 2nd paragraph after Proposition 10.2). For the definition of absolute summability, see Hamilton (1994, p. 263, Eq. 10.2.7). □

Lemma A.2.6 (Mean and Covariance Matrix of $\tilde{\boldsymbol{F}}_t$)
Let $\{\tilde{\boldsymbol{F}}_t\}$ be the covariance-stationary VAR(1) process in Lemma A.2.2. Then, we obtain for its mean $\boldsymbol{\mu}_{\tilde{\boldsymbol{F}}}$ and covariance matrix $\Sigma_{\tilde{\boldsymbol{F}}}$:

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{F}}} = \boldsymbol{0}_{pK}, \tag{A.3}$$

$$\Sigma_{\tilde{\boldsymbol{F}}} = \sum_{k=0}^{\infty}\left(\tilde{\mathbb{A}}^k\Sigma_{\tilde{\boldsymbol{\delta}}}\left(\tilde{\mathbb{A}}^k\right)'\right). \tag{A.4}$$

Proof:
Using the MA(∞) representation of $\tilde{\boldsymbol{F}}_t$ in (A.2), we have for the mean:

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{F}}} = \mathbb{E}_\Theta\left[\tilde{\boldsymbol{F}}_t\right] = \mathbb{E}_\Theta\left[\sum_{k=0}^{\infty}\left(\tilde{\mathbb{A}}^k\tilde{\boldsymbol{\delta}}_{t-k}\right)\right] = \sum_{k=0}^{\infty}\left(\tilde{\mathbb{A}}^k\mathbb{E}_\Theta\left[\tilde{\boldsymbol{\delta}}_{t-k}\right]\right) = \mathbf{0}_{pK}.$$

Here, the absolute summability of the matrix sequence $(\tilde{\mathbb{A}}^k)_{k\geq 0}$ admits the interchange of the expectation operator and the series. For the proof of (A.4), see Hamilton (1994, p. 263, Proposition 10.2). □

The representation in (A.4) guarantees that $\Sigma_{\tilde{\boldsymbol{F}}}$ is symmetric and positive semi-definite, which are the characteristics of a covariance matrix. As mentioned at the beginning, we provide alternative proofs, if applicable. In particular, whenever there are some nice conclusions. For this purpose, we derive $\boldsymbol{\mu}_{\tilde{\boldsymbol{F}}}$ and $\Sigma_{\tilde{\boldsymbol{F}}}$ once again, however, without using the MA(∞) formulation and the absolute summability property.

Lemma A.2.7 (Mean and Vectorized Covariance Matrix of $\tilde{\boldsymbol{F}}_t$)
Let $\{\tilde{\boldsymbol{F}}_t\}$ be the covariance-stationary VAR(1) process in Lemma A.2.2. Then, it follows for its mean $\boldsymbol{\mu}_{\tilde{\boldsymbol{F}}}$ and vectorized covariance matrix $vec(\Sigma_{\tilde{\boldsymbol{F}}})$:

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{F}}} = \mathbf{0}_{pK}, \tag{A.5}$$

$$vec\left(\Sigma_{\tilde{\boldsymbol{F}}}\right) = \left[I_{(pK)^2} - \left(\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}}\right)\right]^{-1} vec\left(\Sigma_{\tilde{\boldsymbol{\delta}}}\right). \tag{A.6}$$

Proof:
The expectation of both sides in (A.1) yields:

$$\begin{aligned}\mathbb{E}_\Theta\left[\tilde{\boldsymbol{F}}_t\right] &= \tilde{\mathbb{A}}\,\mathbb{E}_\Theta\left[\tilde{\boldsymbol{F}}_{t-1}\right] + \mathbb{E}_\Theta\left[\tilde{\boldsymbol{\delta}}_t\right]\\ \boldsymbol{\mu}_{\tilde{\boldsymbol{F}}} &= \tilde{\mathbb{A}}\,\boldsymbol{\mu}_{\tilde{\boldsymbol{F}}} + \mathbf{0}_{pK}\\ \left(I_{pK} - \tilde{\mathbb{A}}\right)\boldsymbol{\mu}_{\tilde{\boldsymbol{F}}} &= \mathbf{0}_{pK}.\end{aligned}$$

The Jordan decomposition of $\tilde{\mathbb{A}}$ (Hamilton, 1994, p. 731, Eq. A.4.25) results in:

$$\begin{aligned}\left(I_{pK} - M_{\tilde{\mathbb{A}}}J_{\tilde{\mathbb{A}}}M_{\tilde{\mathbb{A}}}^{-1}\right)\boldsymbol{\mu}_{\tilde{\boldsymbol{F}}} &= \mathbf{0}_{pK}\\ M_{\tilde{\mathbb{A}}}\left(I_{pK} - J_{\tilde{\mathbb{A}}}\right)M_{\tilde{\mathbb{A}}}^{-1}\boldsymbol{\mu}_{\tilde{\boldsymbol{F}}} &= \mathbf{0}_{pK}.\end{aligned}$$

The matrix $M_{\tilde{\mathbb{A}}}$ is non-singular and hence, has full rank. $J_{\tilde{\mathbb{A}}}$ is an upper triangular matrix, whose diagonal contains the eigenvalues of $\tilde{\mathbb{A}}$. The covariance-stationarity of $\{\tilde{\boldsymbol{F}}_t\}$ implies that all eigenvalues of $\tilde{\mathbb{A}}$ lie inside the unit circle (see Lemma A.2.3 and Remark A.2.4). Thus, the matrix $I_{pK} - J_{\tilde{\mathbb{A}}}$ also has full rank and the unique solution $\boldsymbol{\mu}_{\tilde{\boldsymbol{F}}} = \mathbf{0}_{pK}$ follows.

For the covariance matrix $\Sigma_{\tilde{\boldsymbol{F}}}$, we obtain with the help of (A.1):

$$\begin{aligned}\mathbb{V}\mathrm{ar}_\Theta\left[\tilde{\boldsymbol{F}}_t\right] &= \mathbb{E}_\Theta\left[\tilde{\boldsymbol{F}}_t\tilde{\boldsymbol{F}}_t'\right] = \mathbb{E}_\Theta\left[\left(\tilde{\mathbb{A}}\tilde{\boldsymbol{F}}_{t-1} + \tilde{\boldsymbol{\delta}}_t\right)\left(\tilde{\mathbb{A}}\tilde{\boldsymbol{F}}_{t-1} + \tilde{\boldsymbol{\delta}}_t\right)'\right]\\ &= \tilde{\mathbb{A}}\,\mathbb{E}_\Theta\left[\tilde{\boldsymbol{F}}_{t-1}\tilde{\boldsymbol{F}}_{t-1}'\right]\tilde{\mathbb{A}}' + \mathbb{E}_\Theta\left[\tilde{\boldsymbol{\delta}}_t\tilde{\boldsymbol{\delta}}_t'\right],\end{aligned}$$

where we benefit from the independence of $\tilde{\boldsymbol{F}}_{t-1}$ and $\tilde{\boldsymbol{\delta}}_t$. The expectation of $\tilde{\boldsymbol{\delta}}_t$ is zero (see Equation (2.5) and Lemma A.2.2) such that the last term is replaced by its covariance matrix. If we vectorize both sides, we get:

$$vec\left(\Sigma_{\tilde{\boldsymbol{F}}}\right) = vec\left(\tilde{\mathbb{A}}\Sigma_{\tilde{\boldsymbol{F}}}\tilde{\mathbb{A}}'\right) + vec\left(\Sigma_{\tilde{\boldsymbol{\delta}}}\right).$$

Then, Lemma A.1.13 provides:

$$vec\left(\Sigma_{\tilde{\boldsymbol{F}}}\right) = \left(\tilde{\mathbb{A}} \otimes \tilde{\mathbb{A}}\right) vec\left(\Sigma_{\tilde{\boldsymbol{F}}}\right) + vec\left(\Sigma_{\tilde{\boldsymbol{\delta}}}\right)$$
$$\left[I_{(pK)^2} - \left(\tilde{\mathbb{A}} \otimes \tilde{\mathbb{A}}\right)\right] vec\left(\Sigma_{\tilde{\boldsymbol{F}}}\right) = vec\left(\Sigma_{\tilde{\boldsymbol{\delta}}}\right).$$

The Jordan decomposition of $\tilde{\mathbb{A}} \otimes \tilde{\mathbb{A}}$ (Hamilton, 1994, p. 731, Eq. A.4.25) yields:

$$\left[I_{(pK)^2} - M_{(\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}})} J_{(\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}})} M^{-1}_{(\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}})}\right] vec\left(\Sigma_{\tilde{\boldsymbol{F}}}\right) = vec\left(\Sigma_{\tilde{\boldsymbol{\delta}}}\right)$$
$$M_{(\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}})}\left[I_{(pK)^2} - J_{(\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}})}\right] M^{-1}_{(\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}})}\, vec\left(\Sigma_{\tilde{\boldsymbol{F}}}\right) = vec\left(\Sigma_{\tilde{\boldsymbol{\delta}}}\right).$$

The matrix $M_{(\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}})}$ is non-singular. $J_{(\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}})}$ is an upper triangular matrix with all eigenvalues of $\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}}$ on its diagonal. Lemma A.1.11 provides that the $(pK)^2$ eigenvalues of $\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}}$ are given by the $(pK)^2$ products of the pK eigenvalues of $\tilde{\mathbb{A}}$. Due to the covariance-stationarity of $\{\tilde{\boldsymbol{F}}_t\}$, we have: $abs\,(\lambda_i) < 1, 1 \leq i \leq pK$, for all eigenvalues λ_i of $\tilde{\mathbb{A}}$. Thus, we get: $abs\,(\lambda_i\lambda_j) = abs\,(\lambda_i)\, abs\,(\lambda_j) < 1, 1 \leq i, j \leq pK$. All in all, the matrix $I_{(pK)^2} - J_{(\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}})}$ has full rank such that the inverse of $[I_{(pK)^2} - (\tilde{\mathbb{A}} \otimes \tilde{\mathbb{A}})]$ exists and the solution in (A.6) is well-defined. □

Remark A.2.8 (Comparison of $\tilde{\boldsymbol{F}}_t$ Moments)
Lemmata A.2.6 and A.2.7 derive the same mean, i.e., $\boldsymbol{\mu}_{\tilde{\boldsymbol{F}}} = \mathbf{0}_{pK}$. *Perhaps not at first glance, they also provide the same covariance matrix. If we apply the vectorization operator to* (A.4)*, we obtain:*

$$vec\left(\Sigma_{\tilde{\boldsymbol{F}}}\right) = vec\left(\sum_{k=0}^{\infty}\left(\tilde{\mathbb{A}}^k \Sigma_{\tilde{\boldsymbol{\delta}}}\left(\tilde{\mathbb{A}}^k\right)'\right)\right)$$
$$= \sum_{k=0}^{\infty} vec\left(\tilde{\mathbb{A}}^k \Sigma_{\tilde{\boldsymbol{\delta}}}\left(\tilde{\mathbb{A}}^k\right)'\right).$$

We may move the vectorization operator in the series, as a matrix sequence "is absolutely summable if each of its elements forms an absolutely summable scalar sequence" (Hamilton, 1994, p. 263, 1st sentence). Next, we use Lemma A.1.13:

$$vec\left(\Sigma_{\tilde{\boldsymbol{F}}}\right) = \sum_{k=0}^{\infty}\left(\left(\tilde{\mathbb{A}}^k \otimes \tilde{\mathbb{A}}^k\right) vec\left(\Sigma_{\tilde{\boldsymbol{\delta}}}\right)\right)$$
$$= \left(\sum_{k=0}^{\infty}\left(\tilde{\mathbb{A}}^k \otimes \tilde{\mathbb{A}}^k\right)\right) vec\left(\Sigma_{\tilde{\boldsymbol{\delta}}}\right).$$

The last step requires the absolute summability of each scalar series once again, when we separate $vec(\Sigma_{\tilde{\boldsymbol{\delta}}})$ *from the series. The iterative use of (vii) in Lemma A.1.10 yields for* $k \geq 2$*:*

$$\tilde{\mathbb{A}}^k \otimes \tilde{\mathbb{A}}^k = \left(\tilde{\mathbb{A}}^{k-1}\tilde{\mathbb{A}}\right) \otimes \left(\tilde{\mathbb{A}}^{k-1}\tilde{\mathbb{A}}\right) = \left(\tilde{\mathbb{A}}^{k-1} \otimes \tilde{\mathbb{A}}^{k-1}\right)\left(\tilde{\mathbb{A}} \otimes \tilde{\mathbb{A}}\right) = \ldots = \left(\tilde{\mathbb{A}} \otimes \tilde{\mathbb{A}}\right)^k.$$

The cases $k = 0$ *and* $k = 1$ *do not require any adjustment. In total, we have:*

$$vec\left(\Sigma_{\tilde{\boldsymbol{F}}}\right) = \left(\sum_{k=0}^{\infty}\left(\tilde{\mathbb{A}} \otimes \tilde{\mathbb{A}}\right)^k\right) vec\left(\Sigma_{\tilde{\boldsymbol{\delta}}}\right).$$

Since $\mathbb{R}^{(pK)^2}$ *is a Banach space, e.g., with respect to the matrix norm, the matrix* $\tilde{\mathbb{A}} \otimes \tilde{\mathbb{A}}$ *ranks among its linear operators and the covariance-stationarity of* $\{\tilde{\boldsymbol{F}}_t\}$ *guarantees the necessary boundedness constraint,* *the series may be regarded as Neumann series (Alt, 2012, p. 153, Section 3.7), which satisfies:*

$$\sum_{k=0}^{\infty}\left(\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}}\right)^{k}=\left[I_{(pK)^2}-\left(\tilde{\mathbb{A}}\otimes\tilde{\mathbb{A}}\right)\right]^{-1}.$$

Eventually, we end up with the vectorized version of $\Sigma_{\tilde{\boldsymbol{F}}}$ *in* (A.6).

Lemma A.2.9 (Vector MA(∞) Representation of $\tilde{\boldsymbol{B}}_t$)
Let the process $\{\tilde{\boldsymbol{B}}_t\}$ obey the VAR(1) representation in Lemma 4.2.5 based on the covariance-stationary processes $\{r_t\}$ in Definition 4.2.1 and $\{\boldsymbol{F}_t\}$ in Definitions 2.1.3 and 2.1.4, respectively. Then, the process $\{\tilde{\boldsymbol{B}}_t\}$ is covariance-stationary such that its subsequent MA(∞) representation is well-defined:

$$\tilde{\boldsymbol{B}}_t=\sum_{k=0}^{\infty}\left(\mathbb{H}^k\left(\boldsymbol{a}+\boldsymbol{e}_{t-k}\right)\right).$$

Moreover, it satisfies the absolute summability condition, i.e., it holds:

$$\sum_{k=0}^{\infty} abs\left(\mathbb{H}^k_{ij}\right)<\infty,\quad \textit{for all}\ \ 1\le i,j\le\max\left(1,\tilde{q}\right)+\max\left(1,p\right)K,$$

where $\mathbb{H}^k_{ij}$ is the element in the i-th row and j-th column of the k-th power of $\mathbb{H}$.

Proof:
At first, we subdivide the matrix $\mathbb{H}$ in Lemma 4.2.5 into the following four block matrices:

$$\mathbb{H}=\begin{bmatrix}\mathbb{B} & * \\ O_{\max(1,p)K\times\max(1,\tilde{q})} & \tilde{\mathbb{A}}\end{bmatrix}.$$

For $p\ge 1$, the lower right block matrix of $\mathbb{H}$ is $\tilde{\mathbb{A}}$ in Lemma A.2.2 with eigenvalues inside the unit circle, since the process $\{\boldsymbol{F}_t\}$ shall be covariance-stationary. Thus, for all $\lambda_{\tilde{\mathbb{A}}}$ satisfying $|\tilde{\mathbb{A}}-\lambda_{\tilde{\mathbb{A}}}I_{pK}|=0$ we have $abs\left(\lambda_{\tilde{\mathbb{A}}}\right)<1$. If $p=0$ holds, $\tilde{\mathbb{A}}=O_K$ yields $\lambda_{\tilde{\mathbb{A}}}=0$ for all of its eigenvalues. The covariance-stationarity of the ARX implies that all $\lambda_{\mathbb{B}}$ with $|\mathbb{B}-\lambda_{\mathbb{B}}I_{\tilde{q}}|=0$ lie inside the unit circle, i.e., $abs\left(\lambda_{\mathbb{B}}\right)<1$. The lower left block matrix of $\mathbb{H}$ is a zero matrix, thus, Lemma A.1.7 provides: $|\mathbb{H}-\lambda_{\mathbb{H}}I_{pK}|=|\mathbb{A}-\lambda_{\mathbb{A}}I_{pK}||\mathbb{B}-\lambda_{\mathbb{B}}I_{\tilde{q}}|$ such that all eigenvalues of $\mathbb{H}$ lie inside the unit circle and the covariance-stationarity of $\{\tilde{\boldsymbol{B}}_t\}$ follows. Further, it holds: $\mathbb{H}^k\to O_{(\max(1,\tilde{q})+\max(1,p)K)\times(\max(1,\tilde{q})+\max(1,p)K)}$ as $k\to\infty$, which justifies the MA(∞) representation as convergent series (Hamilton, 1994, p. 260, Eq. 10.1.15). The dynamics of process $\{\tilde{\boldsymbol{B}}_t\}$ coincides with Equation [10.1.4] in Hamilton (1994), therefore, the MA(∞) formulation of $\tilde{\boldsymbol{F}}_t$ meets the absolute summability condition (Hamilton, 1994, p. 263, 2nd paragraph after Proposition 10.2). □

Lemma A.2.10 (Mean and Covariance Matrix of $\tilde{\boldsymbol{B}}_t$)
Assume $\{\tilde{\boldsymbol{B}}_t\}$ as the covariance-stationary process in Lemma A.2.9. Then, the vector $\tilde{\boldsymbol{B}}_t$ is Gaussian, i.e., $\tilde{\boldsymbol{B}}_t\sim\mathcal{N}\left(\boldsymbol{\mu}_{\tilde{\boldsymbol{B}}},\Sigma_{\tilde{\boldsymbol{B}}}\right)$, with mean and covariance matrix given by:

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{B}}}=\sum_{k=0}^{\infty}\left(\mathbb{H}^k\boldsymbol{a}\right),\tag{A.7}$$

$$\Sigma_{\tilde{\boldsymbol{B}}}=\sum_{k=0}^{\infty}\left(\mathbb{H}^k\Sigma_{\boldsymbol{e}}\left(\mathbb{H}^k\right)'\right),\tag{A.8}$$

with covariance matrix $\Sigma_{\boldsymbol{e}}$ as in Lemma 4.2.5.

Proof:
The absolute summability of the matrix sequence $(\mathbb{H}^k)_{k\geq 0}$ admits the interchange of the expectation and the series. Thus, the distribution of $\boldsymbol{e}_t$ in Lemma 4.2.5 yields for the mean of $\tilde{\boldsymbol{B}}_t$:

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{B}}} = \mathbb{E}_\Theta\left[\tilde{\boldsymbol{B}}_t\right] = \mathbb{E}_\Theta\left[\sum_{k=0}^{\infty}\left(\mathbb{H}^k\left(\boldsymbol{a}+\boldsymbol{e}_{t-k}\right)\right)\right] = \sum_{k=0}^{\infty}\left(\mathbb{H}^k\left(\boldsymbol{a}+\mathbb{E}_\Theta\left[\boldsymbol{e}_{t-k}\right]\right)\right) = \sum_{k=0}^{\infty}\left(\mathbb{H}^k\boldsymbol{a}\right).$$

The proof of $\Sigma_{\tilde{\boldsymbol{B}}}$ is provided in Hamilton (1994, p. 263, Proposition 10.2). □

Lemma A.2.11 (Mean and Vectorized Covariance Matrix of $\tilde{\boldsymbol{B}}_t$)
Assume $\{\tilde{\boldsymbol{B}}_t\}$ as the covariance-stationary VAR(1) process in Lemma 4.2.5 of dimension $d = \max(1,\tilde{q}) + \max(1,p)\,K$. Then, we obtain for its mean $\boldsymbol{\mu}_{\tilde{\boldsymbol{B}}}$ and vectorized covariance matrix $vec\left(\Sigma_{\tilde{\boldsymbol{B}}}\right)$:

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{B}}} = \left(I_d - \mathbb{H}\right)^{-1}\boldsymbol{a} \tag{A.9}$$

$$vec\left(\Sigma_{\tilde{\boldsymbol{B}}}\right) = \left(I_d - \mathbb{H}\right)^{-1} vec\left(\Sigma_{\boldsymbol{e}}\right). \tag{A.10}$$

Proof:
For $\boldsymbol{\mu}_{\tilde{\boldsymbol{B}}}$, see Hamilton (1994, p. 258, last unnamed Equation before Eq. 10.1.8). For $vec(\Sigma_{\tilde{\boldsymbol{B}}})$, the same steps as in Lemma A.2.7 are applied. Similarly, we may transfer the findings in Remark A.2.8 to $\{\tilde{\boldsymbol{B}}_t\}$. □

Lemma A.2.12 (Mean and Variance of r_t Without Any Information)
Let $\{r_t\}$ be the covariance-stationary ARX$(\tilde{q},\tilde{p})$ in Definition 4.2.1. Then, it holds: $r_t \sim \mathcal{N}\left(\mu_r, \sigma_r^2\right)$ with:

$$\mu_r = [1,0,\ldots,0]\,\boldsymbol{\mu}_{\tilde{\boldsymbol{B}}} \quad \textit{and} \quad \sigma_r^2 = [1,0,\ldots,0]\,\Sigma_{\tilde{\boldsymbol{B}}}\,[1,0,\ldots,0]',$$

with $\boldsymbol{\mu}_{\tilde{\boldsymbol{B}}}$ and $\Sigma_{\tilde{\boldsymbol{B}}}$ as in Lemmata A.2.10 and A.2.11, respectively.

Proof:
The definition of $\tilde{\boldsymbol{B}}_t$ in Lemma 4.2.5 results in:

$$\mu_r = \mathbb{E}_\Theta\left[r_t\right] = \mathbb{E}_\Theta\left[[1,0,\ldots,0]\,\tilde{\boldsymbol{B}}_t\right] = [1,0,\ldots,0]\,\boldsymbol{\mu}_{\tilde{\boldsymbol{B}}}$$
$$\sigma_r^2 = \mathbb{V}\mathrm{ar}_\Theta\left[r_t\right] = \mathbb{V}\mathrm{ar}_\Theta\left[[1,0,\ldots,0]\,\tilde{\boldsymbol{B}}_t\right] = [1,0,\ldots,0]\,\Sigma_{\tilde{\boldsymbol{B}}}\,[1,0,\ldots,0]',$$

which proves the assertion. □

Lemma A.2.13 (Mean and Variance of r_t Based on Full History)
Let $\boldsymbol{r} = [r_1,\ldots,r_T]$ and $F^c = [\boldsymbol{F}_1^c,\ldots,\boldsymbol{F}_T^c]$ be return and factor samples, respectively. For $\tilde{m} = \max(\tilde{q},\tilde{p})$, the returns are supposed to obey the ARX in Definition 4.2.1 based on F^c, i.e.:

$$r_t = \alpha + \sum_{i=1}^{\tilde{q}}\left(\beta_i r_{t-i}\right) + \sum_{i=1}^{\tilde{p}}\left(\boldsymbol{\gamma}_i'\boldsymbol{F}_{t-i}^c\right) + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}\left(0,\sigma_\varepsilon^2\right)\ iid.$$

Then, for all points in time $\tilde{m}+1 \leq t \leq T$, we have: $r_t | r_{t-1},\ldots,r_{t-\tilde{q}}, F^c \sim \mathcal{N}\left(\mu_{r_t|Full}, \sigma^2_{r_t|Full}\right)$ with mean and variance given by:

$$\mu_{r_t|Full} = \alpha + \sum_{i=1}^{\tilde{q}}\left(\beta_i r_{t-i}\right) + \sum_{i=1}^{\tilde{p}}\left(\boldsymbol{\gamma}_i'\boldsymbol{F}_{t-i}^c\right) \quad \textit{and} \quad \sigma^2_{r_t|Full} = \sigma_\varepsilon^2.$$

Proof:
The properties of the conditional expectation result in:

$$\begin{aligned}
\mu_{r_t|Full} &= \mathbb{E}_\Theta\left[r_t|r_{t-1},\ldots,r_{t-\tilde{q}},F^c\right] = \mathbb{E}_\Theta\left[\alpha+\sum_{i=1}^{\tilde{q}}(\beta_i r_{t-i})+\sum_{i=1}^{\tilde{p}}\left(\boldsymbol{\gamma}_i'\boldsymbol{F}_{t-i}^c\right)+\varepsilon_t\middle| r_{t-1},\ldots,r_{t-\tilde{q}},F^c\right]\\
&= \alpha+\sum_{i=1}^{\tilde{q}}(\beta_i r_{t-i})+\sum_{i=1}^{\tilde{p}}\left(\boldsymbol{\gamma}_i'\boldsymbol{F}_{t-i}^c\right),\\
\sigma^2_{r_t|Full} &= \mathbb{V}\mathrm{ar}_\Theta\left[r_t|r_{t-1},\ldots,r_{t-\tilde{q}},F^c\right]\\
&= \mathbb{V}\mathrm{ar}_\Theta\left[\alpha+\sum_{i=1}^{\tilde{q}}(\beta_i r_{t-i})+\sum_{i=1}^{\tilde{p}}\left(\boldsymbol{\gamma}_i'\boldsymbol{F}_{t-i}^c\right)+\varepsilon_t\middle| r_{t-1},\ldots,r_{t-\tilde{q}},F^c\right]=\sigma_\varepsilon^2,
\end{aligned}$$

which proves the claim. □

Lemma A.2.14 (Mean and Variance of r_t Based on Partial History)
Let $\boldsymbol{r}=[r_1,\ldots,r_T]$ and $F^c=[\boldsymbol{F}_1^c,\ldots,\boldsymbol{F}_T^c]$ be return and factor samples, respectively. For $\tilde{m}=\max(\tilde{q},\tilde{p})$, the returns are supposed to obey the ARX in Definition 4.2.1 based on F^c, i.e.:

$$r_t=\alpha+\sum_{i=1}^{\tilde{q}}(\beta_i r_{t-i})+\sum_{i=1}^{\tilde{p}}\left(\boldsymbol{\gamma}_i'\boldsymbol{F}_{t-i}^c\right)+\varepsilon_t,\ \varepsilon_t\sim\mathcal{N}\left(0,\sigma_\varepsilon^2\right)\ iid.$$

Then, for all points in time $2\le t\le\tilde{m}$, we have: $r_t|r_{t-1},\ldots,r_{\max(1,t-\tilde{q})},F^c\sim\mathcal{N}\left(\mu_{r_t|Part},\sigma^2_{r_t|Part}\right)$ with mean and variance as follows:

$$\begin{aligned}
\mu_{r_t|Part} &= \alpha+\sum_{i=1}^{\min(\tilde{q},t-1)}(\beta_i r_{t-i})+\sum_{i=1}^{\min(\tilde{p},t-1)}\left(\boldsymbol{\gamma}_i'\boldsymbol{F}_{t-i}^c\right)+\boldsymbol{\kappa}'_{(t,\tilde{q},\tilde{p},\boldsymbol{\beta},\boldsymbol{\gamma})}\boldsymbol{\mu}_{\tilde{\boldsymbol{B}}},\\
\sigma^2_{r_t|Part} &= \boldsymbol{\kappa}'_{(t,\tilde{q},\tilde{p},\boldsymbol{\beta},\boldsymbol{\gamma})}\Sigma_{\tilde{\boldsymbol{B}}}\boldsymbol{\kappa}_{(t,\tilde{q},\tilde{p},\boldsymbol{\beta},\boldsymbol{\gamma})}+\sigma_\varepsilon^2,
\end{aligned}$$

with $\boldsymbol{\mu}_{\tilde{\boldsymbol{B}}}$ and $\Sigma_{\tilde{\boldsymbol{B}}}$ as in Lemmata A.2.10 or A.2.11 and $\boldsymbol{\kappa}_{(t,\tilde{q},\tilde{p},\boldsymbol{\beta},\boldsymbol{\gamma})}\in\mathbb{R}^{\max(1,\tilde{q})+\max(1,p)K}$ given by:

$$\boldsymbol{\kappa}_{(t,\tilde{q},\tilde{p},\boldsymbol{\beta},\boldsymbol{\gamma})}=\left[\mathbb{1}_{\{t\le\tilde{q}\}}\beta_t,\ldots,\mathbb{1}_{\{t\le\tilde{q}\}}\beta_{\tilde{q}},0\ldots,0,\mathbb{1}_{\{t\le\tilde{p}\}}\boldsymbol{\gamma}_t',\ldots,\mathbb{1}_{\{t\le\tilde{p}\}}\boldsymbol{\gamma}_{\tilde{p}}',0\ldots,0\right]'.$$

Proof:
The idea behind is to change to the stationary distribution as soon as there is no information in the form of return and factor samples available. Thereby, the properties of the conditional expectation provide:

$$\begin{aligned}
\mu_{r_t|Part} &= \mathbb{E}_\Theta\left[r_t|r_{t-1},\ldots,r_{\max(1,t-\tilde{q})},F^c\right]\\
&= \alpha+\sum_{i=1}^{\min(\tilde{q},t-1)}(\beta_i r_{t-i})+\mathbb{1}_{\{t\le\tilde{q}\}}\sum_{i=t}^{\tilde{q}}\left(\beta_i\mathbb{E}_\Theta\left[r_{t-i}\right]\right)+\sum_{i=1}^{\min(\tilde{p},t-1)}\left(\boldsymbol{\gamma}_i'\boldsymbol{F}_{t-i}^c\right)+\mathbb{1}_{\{t\le\tilde{p}\}}\sum_{i=t}^{\tilde{p}}\left(\boldsymbol{\gamma}_i'\mathbb{E}_\Theta\left[\boldsymbol{F}_{t-i}\right]\right)\\
&= \alpha+\sum_{i=1}^{\min(\tilde{q},t-1)}(\beta_i r_{t-i})+\sum_{i=1}^{\min(\tilde{p},t-1)}\left(\boldsymbol{\gamma}_i'\boldsymbol{F}_{t-i}^c\right)+\boldsymbol{\kappa}'_{(t,\tilde{q},\tilde{p},\boldsymbol{\beta},\boldsymbol{\gamma})}\mathbb{E}_\Theta\left[\tilde{\boldsymbol{B}}_0\right],\\
\sigma^2_{r_t|Part} &= \mathbb{V}\mathrm{ar}_\Theta\left[r_t|r_{t-1},\ldots,r_{\max(1,t-\tilde{q})},F^c\right]\\
&= \mathbb{E}_\Theta\left[\left(\mathbb{1}_{\{t\le\tilde{q}\}}\sum_{i=t}^{\tilde{q}}\beta_i\left(r_{t-i}-\mathbb{E}_\Theta\left[r_{t-i}\right]\right)+\mathbb{1}_{\{t\le\tilde{p}\}}\sum_{i=t}^{\tilde{p}}\boldsymbol{\gamma}_i'\left(\boldsymbol{F}_{t-i}-\mathbb{E}_\Theta\left[\boldsymbol{F}_{t-i}\right]\right)+\varepsilon_t\right)^2\middle| F^c\right]\\
&= \boldsymbol{\kappa}'_{(t,\tilde{q},\tilde{p},\boldsymbol{\beta},\boldsymbol{\gamma})}\mathbb{V}\mathrm{ar}_\Theta\left[\tilde{\boldsymbol{B}}_0\right]\boldsymbol{\kappa}_{(t,\tilde{q},\tilde{p},\boldsymbol{\beta},\boldsymbol{\gamma})}+\sigma_\varepsilon^2.
\end{aligned}$$

Finally, we have $\mathbb{E}_\Theta\left[\tilde{\boldsymbol{B}}_0\right] = \boldsymbol{\mu}_{\tilde{\boldsymbol{B}}}$ and $\mathbb{V}\text{ar}_\Theta\left[\tilde{\boldsymbol{B}}_0\right] = \Sigma_{\tilde{\boldsymbol{B}}}$, which completes the proof. □

Lemma A.2.15 (Vector MA(∞) Representation of $\tilde{\boldsymbol{S}}_t$)
Let the process $\{\tilde{\boldsymbol{S}}_t\}$ obey the VAR(1) in Lemma 4.2.11 with shift $s \geq 1$ based on the covariance-stationary processes $\{r_t\}$ in Definition 4.2.1 and $\{\boldsymbol{F}_t\}$ in Definitions 2.1.3 and 2.1.4, respectively. Then, the process $\{\tilde{\boldsymbol{S}}_t\}$ is covariance-stationary such that its subsequent MA(∞) representation is well-defined:

$$\tilde{\boldsymbol{S}}_t = \sum_{k=0}^{\infty} \left(\mathbb{H}_s^k \left(\boldsymbol{a} + \boldsymbol{e}_{t-k}\right)\right).$$

Moreover, it satisfies the absolute summability condition, i.e., it holds:

$$\sum_{k=0}^{\infty} abs\left(\mathbb{H}_{s,ij}^k\right) < \infty, \quad \textit{for all} \;\; 1 \leq i,j \leq \max\left(1,\tilde{q}\right) + \max\left(1,p\right)K + (s-1)(K+1),$$

where $\mathbb{H}_{s,ij}^k$ is the element in the i-th row and j-th column of the k-th power of $\mathbb{H}_s$.

Proof:
The statement is shown in exactly the same way as in the proof of Lemma A.2.9. □

Lemma A.2.16 (Mean and Covariance Matrix of $\tilde{\boldsymbol{S}}_t$)
Let $\{\tilde{\boldsymbol{S}}_t\}$ be the covariance-stationary process in Lemma A.2.15 with shift in time $s \geq 1$. Then, the vector $\tilde{\boldsymbol{S}}_t$ is Gaussian, i.e., $\tilde{\boldsymbol{S}}_t \sim \mathcal{N}\left(\boldsymbol{\mu}_{\tilde{\boldsymbol{S}}}, \Sigma_{\tilde{\boldsymbol{S}}}\right)$, with mean and covariance matrix defined as follows:

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{S}}} = \sum_{k=0}^{\infty} \left(\mathbb{H}_s^k \boldsymbol{a}\right),$$
$$\Sigma_{\tilde{\boldsymbol{S}}} = \sum_{k=0}^{\infty} \left(\mathbb{H}_s^k \Sigma_{\boldsymbol{e}} \left(\mathbb{H}_s^k\right)'\right),$$

with covariance matrix $\Sigma_{\boldsymbol{e}}$ as in Lemma 4.2.11.

Proof:
The absolute summability of the matrix sequence $(\mathbb{H}_s^k)_{k\geq 0}$ admits the interchange of the expectation and the series. Thus, the distribution of $\boldsymbol{e}_t$ in Lemma 4.2.11 yields for the mean of $\tilde{\boldsymbol{S}}_t$:

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{S}}} = \mathbb{E}_\Theta\left[\tilde{\boldsymbol{S}}_t\right] = \mathbb{E}_\Theta\left[\sum_{k=0}^{\infty} \left(\mathbb{H}_s^k \left(\boldsymbol{a} + \boldsymbol{e}_{t-k}\right)\right)\right] = \sum_{k=0}^{\infty} \left(\mathbb{H}_s^k \left(\boldsymbol{a} + \mathbb{E}_\Theta\left[\boldsymbol{e}_{t-k}\right]\right)\right) = \sum_{k=0}^{\infty} \left(\mathbb{H}_s^k \boldsymbol{a}\right).$$

The proof of $\Sigma_{\tilde{\boldsymbol{S}}}$ is provided in Hamilton (1994, p. 263, Proposition 10.2). □

Lemma A.2.17 (Mean and Vectorized Covariance Matrix of $\tilde{\boldsymbol{S}}_t$)
Assume $\{\tilde{\boldsymbol{S}}_t\}$ as the covariance-stationary VAR(1) process in Lemma 4.2.11 with shift $s \geq 1$ of dimension $d = \max\left(1,\tilde{q}\right) + \max\left(1,p\right)K + (s-1)(K+1)$. Then, we obtain for its mean $\boldsymbol{\mu}_{\tilde{\boldsymbol{S}}}$ and vectorized covariance matrix $vec(\Sigma_{\tilde{\boldsymbol{S}}})$:

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{S}}} = \left(I_d - \mathbb{H}_s\right)^{-1} \boldsymbol{a}$$
$$vec\left(\Sigma_{\tilde{\boldsymbol{S}}}\right) = \left(I_d - \mathbb{H}_s\right)^{-1} vec\left(\Sigma_{\boldsymbol{e}}\right).$$

Proof:
For $\boldsymbol{\mu}_{\tilde{\boldsymbol{S}}}$, see Hamilton (1994, p. 258, last unnamed Equation before Eq. 10.1.8). For $vec(\Sigma_{\tilde{\boldsymbol{S}}})$, the same steps as in the proof of Lemma A.2.7 are applied. Similarly, we may transfer Remark A.2.8 to $\{\tilde{\boldsymbol{S}}_t\}$. □

Lemma A.2.18 (Mean and Variance of Shifted r_t Without Any Information)
Let $\{r_t\}$ be the covariance-stationary ARX$(\tilde{q},\tilde{p})$ with shift in time $s \geq 1$ in Definition 4.2.10. Then, the vector $\boldsymbol{r}_{t,s} = [r_t, \ldots, r_{t-s+1}]' \in \mathbb{R}^s$ is Gaussian, i.e., $\boldsymbol{r}_{t,s} \sim \mathcal{N}\left(\boldsymbol{\mu}_{r,s}, \Sigma_{r,s}\right)$, with parameters as follows:

$$\boldsymbol{\mu}_{r,s} = [I_s, O_{s\times d-s}]\,\boldsymbol{\mu}_{\tilde{\boldsymbol{S}}} \quad \text{and} \quad \Sigma_{r,s} = [I_s, O_{s\times d-s}]\,\Sigma_{\tilde{\boldsymbol{S}}}\,[I_s, O_{s\times d-s}]',$$

with $d = \max(1,\tilde{q}) + \max(1,p)\,K + (s-1)(K+1)$ and zero matrix $O_{s\times d-s} \in \mathbb{R}^{s\times d-s}$. The mean $\boldsymbol{\mu}_{\tilde{\boldsymbol{S}}}$ and covariance matrix $\Sigma_{\tilde{\boldsymbol{S}}}$ are given in Lemmata A.2.16 and A.2.17, respectively.

Proof:
The definition of $\tilde{\boldsymbol{S}}_t$ in Lemma 4.2.11 results in:

$$\begin{aligned}
\boldsymbol{\mu}_{r,s} &= \mathbb{E}_\Theta\left[\boldsymbol{r}_{t,s}\right] = \mathbb{E}_\Theta\left[[r_t,\ldots,r_{t-s+1}]'\right] = \mathbb{E}_\Theta\left[[I_s, O_{s\times d-s}]\,\tilde{\boldsymbol{S}}_t\right] = [I_s, O_{s\times d-s}]\,\boldsymbol{\mu}_{\tilde{\boldsymbol{S}}},\\
\Sigma_{r,s} &= \mathbb{V}\mathrm{ar}_\Theta\left[\boldsymbol{r}_{t,s}\right] = \mathbb{V}\mathrm{ar}_\Theta\left[[I_s, O_{s\times d-s}]\,\tilde{\boldsymbol{S}}_t\right] = [I_s, O_{s\times d-s}]\,\Sigma_{\tilde{\boldsymbol{S}}}\,[I_s, O_{s\times d-s}]',
\end{aligned}$$

which proves the assertion. □

Lemma A.2.19 (Mean and Variance of Shifted r_t Based on Full History)
Let $\boldsymbol{r} = [r_1,\ldots,r_T]$ and $F^c = [\boldsymbol{F}_1^c,\ldots,\boldsymbol{F}_T^c]$ be return and factor samples, respectively. For shift $s \geq 1$, we set $\tilde{m} = \max(\tilde{q},\tilde{p}) + s - 1$. The returns shall obey the shifted ARX in Definition 4.2.10 based on F^c, i.e.:

$$r_t = \alpha + \sum_{i=1}^{\tilde{q}}\left(\beta_i r_{t-s+1-i}\right) + \sum_{i=1}^{\tilde{p}}\left(\boldsymbol{\gamma}_i' \boldsymbol{F}_{t-s+1-i}^c\right) + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}\left(0,\sigma_\varepsilon^2\right) \text{ iid.}$$

Then, for all points in time $\tilde{m}+1 \leq t \leq T$, we have: $r_t | r_{t-1},\ldots,r_{t-s+1-\tilde{q}}, F^c \sim \mathcal{N}\left(\mu_{r_{t,s}|Full}, \sigma^2_{r_{t,s}|Full}\right)$ with the subsequent mean and variance:

$$\mu_{r_{t,s}|Full} = \alpha + \sum_{i=1}^{\tilde{q}}\left(\beta_i r_{t-s+1-i}\right) + \sum_{i=1}^{\tilde{p}}\left(\boldsymbol{\gamma}_i' \boldsymbol{F}_{t-s+1-i}^c\right) \quad \text{and} \quad \sigma^2_{r_{t,s}|Full} = \sigma_\varepsilon^2.$$

Proof:
The properties of the conditional expectation result in:

$$\begin{aligned}
\mu_{r_{t,s}|Full} &= \mathbb{E}_\Theta\left[r_t | r_{t-1},\ldots,r_{t-s+1-\tilde{q}}, F^c\right]\\
&= \mathbb{E}_\Theta\left[\alpha + \sum_{i=1}^{\tilde{q}}\left(\beta_i r_{t-s+1-i}\right) + \sum_{i=1}^{\tilde{p}}\left(\boldsymbol{\gamma}_i' \boldsymbol{F}_{t-s+1-i}^c\right) + \varepsilon_t \,\middle|\, r_{t-1},\ldots,r_{t+1-s-\tilde{q}}, F^c\right]\\
&= \alpha + \sum_{i=1}^{\tilde{q}}\left(\beta_i r_{t-s+1-i}\right) + \sum_{i=1}^{\tilde{p}}\left(\boldsymbol{\gamma}_i' \boldsymbol{F}_{t-s+1-i}^c\right),\\
\sigma^2_{r_{t,s}|Full} &= \mathbb{V}\mathrm{ar}_\Theta\left[r_t | r_{t-1},\ldots,r_{t-s+1\tilde{q}}, F^c\right]\\
&= \mathbb{V}\mathrm{ar}_\Theta\left[\alpha + \sum_{i=1}^{\tilde{q}}\left(\beta_i r_{t-s+1-i}\right) + \sum_{i=1}^{\tilde{p}}\left(\boldsymbol{\gamma}_i' \boldsymbol{F}_{t-s+1-i}^c\right) + \varepsilon_t \,\middle|\, r_{t-1},\ldots,r_{t-s+1-\tilde{q}}, F^c\right] = \sigma_\varepsilon^2,
\end{aligned}$$

which proves the claim. □

Lemma A.2.20 (Mean and Variance of Shifted r_t Based on Partial History)
Let $\boldsymbol{r} = [r_1, \ldots, r_T]$ and $F^c = [\boldsymbol{F}_1^c, \ldots, \boldsymbol{F}_T^c]$ be return and factor samples, respectively. For shift $s \geq 1$, we set $\tilde{m} = \max(\tilde{q}, \tilde{p}) + s - 1$. The returns shall obey the ARX in Definition 4.2.10 based on F^c, i.e.:

$$r_t = \alpha + \sum_{i=1}^{\tilde{q}} (\beta_i r_{t-s+1-i}) + \sum_{i=1}^{\tilde{p}} \left(\boldsymbol{\gamma}_i' \boldsymbol{F}_{t-s+1-i}^c\right) + \varepsilon_t, \ \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right) \ iid.$$

Then, for all $s + 1 \leq t \leq \tilde{m}$, we have: $r_t | r_{t-1}, \ldots, r_{\max(1, t-s+1-\tilde{q})}, F^c \sim \mathcal{N}\left(\mu_{r_{t,s}|Part}, \sigma^2_{r_{t,s}|Part}\right)$ with mean and variance as follows:

$$\begin{aligned}
\mu_{r_{t,s}|Part} &= \alpha + \sum_{i=1}^{\min(\tilde{q}, t-s)} (\beta_i r_{t-s+1-i}) + \sum_{i=1}^{\min(\tilde{p}, t-s)} \left(\boldsymbol{\gamma}_i' \boldsymbol{F}_{t-s+1-i}^c\right) + \boldsymbol{\kappa}'_{(t,\tilde{q},\tilde{p},s,\boldsymbol{\beta},\boldsymbol{\gamma})} \boldsymbol{\mu}_{\tilde{\boldsymbol{S}}}, \\
\sigma^2_{r_{t,s}|Part} &= \boldsymbol{\kappa}'_{(t,\tilde{q},\tilde{p},s,\boldsymbol{\beta},\boldsymbol{\gamma})} \Sigma_{\tilde{\boldsymbol{S}}} \boldsymbol{\kappa}_{(t,\tilde{q},\tilde{p},s,\boldsymbol{\beta},\boldsymbol{\gamma})} + \sigma_\varepsilon^2,
\end{aligned}$$

where mean $\boldsymbol{\mu}_{\tilde{\boldsymbol{S}}}$ and covariance matrix $\Sigma_{\tilde{\boldsymbol{S}}}$ are given in Lemmata A.2.16 or A.2.17 and we have for the vector $\boldsymbol{\kappa}_{(t,\tilde{q},\tilde{p},s,\boldsymbol{\beta},\boldsymbol{\gamma})} \in \mathbb{R}^{\max(1,\tilde{q}) + \max(1,p)K + (s-1)(K+1)}$:

$$\boldsymbol{\kappa}_{(t,\tilde{q},\tilde{p},s,\boldsymbol{\beta},\boldsymbol{\gamma})} = \left[\mathbb{1}_{\{t-s+1 \leq \tilde{q}\}} \left(\beta_{t-s+1}, \ldots, \beta_{\tilde{q}}, 0, \ldots 0\right), \mathbb{1}_{\{t-s+1 \leq \tilde{p}\}} \left(\boldsymbol{\gamma}'_{t-s+1}, \ldots, \boldsymbol{\gamma}'_{\tilde{p}}, 0 \ldots, 0\right)\right]'.$$

Proof:
The idea behind is to change to the stationary distribution as soon as there is no information in the form of return and factor samples available. Thereby, the properties of the conditional expectation provide:

$$\begin{aligned}
\mu_{r_{t,s}|Part} &= \mathbb{E}_\Theta \left[r_t | r_{t-1}, \ldots, r_{\max(1,t-s+1-\tilde{q})}, F^c\right] \\
&= \alpha + \sum_{i=1}^{\min(\tilde{q}, t-s)} (\beta_i r_{t-s+1-i}) + \mathbb{1}_{\{t-s+1 \leq \tilde{q}\}} \sum_{i=t-s+1}^{\tilde{q}} (\beta_i \mathbb{E}_\Theta \left[r_{t-s+1-i}\right]) \\
&\quad + \sum_{i=1}^{\min(\tilde{p}, t-s)} \left(\boldsymbol{\gamma}_i' \boldsymbol{F}_{t-s+1-i}^c\right) + \mathbb{1}_{\{t-s+1 \leq \tilde{p}\}} \sum_{i=t-s+1}^{\tilde{p}} (\boldsymbol{\gamma}_i' \mathbb{E}_\Theta \left[\boldsymbol{F}_{t-s+1-i}\right]) \\
&= \alpha + \sum_{i=1}^{\min(\tilde{q}, t-s)} (\beta_i r_{t-s+1-i}) + \sum_{i=1}^{\min(\tilde{p}, t-s)} \left(\boldsymbol{\gamma}_i' \boldsymbol{F}_{t-s+1-i}^c\right) + \boldsymbol{\kappa}'_{(t,\tilde{q},\tilde{p},s,\boldsymbol{\beta},\boldsymbol{\gamma})} \mathbb{E}_\Theta \left[\tilde{\boldsymbol{S}}_0\right], \\
\sigma^2_{r_{t,s}|Part} &= \mathbb{V}\mathrm{ar}_\Theta \left[r_t | r_{t-1}, \ldots, r_{\max(1,t-s+1-\tilde{q})}, F^c\right] \\
&= \mathbb{E}_\Theta \left[\left(\mathbb{1}_{\{t-s+1 \leq \tilde{q}\}} \sum_{i=t-s+1}^{\tilde{q}} \beta_i \left(r_{t-s+1-i} - \mathbb{E}_\Theta \left[r_{t-s+1-i}\right]\right)\right.\right. \\
&\quad \left.\left. + \mathbb{1}_{\{t-s+1 \leq \tilde{p}\}} \sum_{i=t-s+1}^{\tilde{p}} \boldsymbol{\gamma}_i' \left(\boldsymbol{F}_{t-s+1-i} - \mathbb{E}_\Theta \left[\boldsymbol{F}_{t-s+1-i}\right]\right) + \varepsilon_t\right)^2 \middle| F^c\right] \\
&= \boldsymbol{\kappa}'_{(t,\tilde{q},\tilde{p},s,\boldsymbol{\beta},\boldsymbol{\gamma})} \mathbb{V}\mathrm{ar}_\Theta \left[\tilde{\boldsymbol{S}}_0\right] \boldsymbol{\kappa}_{(t,\tilde{q},\tilde{p},s,\boldsymbol{\beta},\boldsymbol{\gamma})} + \sigma_\varepsilon^2.
\end{aligned}$$

Finally, we have $\mathbb{E}_\Theta \left[\tilde{\boldsymbol{S}}_0\right] = \boldsymbol{\mu}_{\tilde{\boldsymbol{S}}}$ and $\mathbb{V}\mathrm{ar}_\Theta \left[\tilde{\boldsymbol{S}}_0\right] = \Sigma_{\tilde{\boldsymbol{S}}}$, which completes the proof. □

A.3 Statistical Measures for Estimation Accuracy

In this section, we repeat the definitions of some common statistical measures, which we apply for assessing the quality of our out-of-sample point and interval forecasts in Chapters 3-5. In doing so, we follow Stock and Watson (2002a), Boivin and Ng (2006), Schumacher and Breitung (2008), Doz et al. (2012) as well as Bańbura and Modugno (2014), when we use the below trace R^2 for evaluating the goodness of estimated factors.

Definition A.3.1 (Trace R^2)
Let the matrices $F \in \mathbb{R}^{T\times K}$ and $\hat{F} \in \mathbb{R}^{T\times \hat{K}}$ collect the original and estimated factors, respectively, of the SFMs in Definition 2.1.3 or the DFMs in Definition 2.1.4. Then, the trace R^2 is given as follows:

$$\textit{trace } R^2 = \frac{tr\left(F'\hat{F}\left(\hat{F}'\hat{F}\right)^{-1}\hat{F}'F\right)}{tr\left(F'F\right)}.$$

Note, in empirical applications and some of our MC simulations we have to estimate the factor dimension. This is why we distinguish between the true and estimated factor dimensions, i.e., K and $\hat{K}$, in the above definition. As mentioned in previous sections, even for $K = \hat{K}$, the estimated factors are unique up to an invertible, linear transformation. Let the matrix $R^{\hat{K}\times K}$ satisfy the following equation: $\hat{F}R = F$. Then, an OLS provides as a solution: $R = (\hat{F}'\hat{F})^{-1}\hat{F}'F$, which justifies the definition of the trace R^2. Furthermore, it shows that the trace R^2 "is smaller than 1 and tends to 1 if the empirical canonical correlations between the true factors and their estimates tend to 1" (Doz et al., 2012, p. 1019).

To evaluate out-of-sample point forecasts for returns of future periods, we deploy the Root-Mean-Square Error, which can be found in standard textbooks and is defined as follows:

Definition A.3.2 (Root-Mean-Square Error)
Assume an out-of-sample period of length T. For any point in time $1 \leq t \leq T$, let $\hat{r}_t \in \mathbb{R}$ be the point estimate of the afterwards realized return $r_t \in \mathbb{R}$. Then, the Root-Mean-Square Error is given by:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{r}_t - r_t\right)^2}.$$

Finally, we focus on statistical measures for assessing the quality of interval forcasts. Since our prediction intervals in Chapters 3 and 4 rely on a prespecified level $\nu \in [0, 1]$, the Ratio of Interval Outliers measures the percentage of realized returns exceeding such ν-prediction intervals. Mathematically, this means:

Definition A.3.3 (Ratio of Interval Outliers)
Assume an out-of-sample period of length T. For any point in time $1 \leq t \leq T$, let $\hat{l}_t, \hat{u}_t \in \mathbb{R}$ with $\hat{l}_t \leq \hat{u}_t$ denote the lower and upper limits, respectively, of the ν-prediction interval with $\nu \in [0, 1]$ and let $r_t \in \mathbb{R}$ be the afterwards realized return. Then, the Ratio of Interval Outliers is defined as:

$$RIO = \frac{1}{T}\sum_{t=1}^{T}\left[\mathbb{1}_{\left\{r_t < \hat{l}_t\right\}} + \mathbb{1}_{\left\{r_t > \hat{u}_t\right\}}\right].$$

Hence, the closer RIO to level ν the more accurately the prediction intervals map the return behavior. For RIO $> \nu$, the prediction intervals seem too tight such that there are too many outliers. For RIO $< \nu$, the prediction intervals are too conservative, since the number of interval outliers is below our expectation.

Unfortunately, RIO does not assess the interval width. That means, for the same level ν tighter prediction intervals offering the same RIO should be preferred. As a solution, we use the Mean Interval Score, which balances the interval width and the number of interval outliers, as second measure for evaluating interval forecasts. Here, we define MIS as in Gneiting and Raftery (2007) or Brechmann and Czado (2015).

Definition A.3.4 (Mean Interval Score)
Assume an out-of-sample period of length T. For any point in time $1 \leq t \leq T$, let $(\hat{l}_t, \hat{u}_t)$ with $\hat{l}_t, \hat{u}_t \in \mathbb{R}$ denote the ν-prediction interval with $\nu \in [0,1]$ and let r_t be the afterwards realized return. Then, we have for the Mean Interval Score:

$$MIS = \frac{1}{T} \sum_{t=1}^{T} \left[\hat{u}_t - \hat{l}_t + \frac{2}{1-\nu} \left(\hat{l}_t - r_t \right) \mathbb{1}_{\left\{ r_t < \hat{l}_t \right\}} + \frac{2}{1-\nu} \left(r_t - \hat{u}_t \right) \mathbb{1}_{\{ r_t > \hat{u}_t \}} \right].$$

Appendix B

Data of Empirical Applications

We collect the data of the empirical analyses in Sections 3.2, 4.5 and 5.3. Thereby, we introduce for each time series an appropriate abbreviation, state its available period of time, repeat its updating frequency, list the assumed data type, indicate the performed preprocessing and provide a brief description including its source. In each section, we mention the time span of the overall data sample such that the individual time spans mark all time series starting at a later point in time or being discontinued. In case of updating frequencies we consider: daily (d), weekly (w), monthly (m) and quarterly (q). With Definition 2.2.2 and Lemma 2.2.3 in mind, we obtain for the assumed data types: stock (1), sum formulation of flow variable (2), average version of flow variable (3), sum formulation of change in flow variable (4) and average version of change in flow variable (5). Note, for complete time series the data type does not matter, since all yield an identity matrix for the matrix Q_i in Definition 2.2.1. Regarding data transformations we distinguish between: no transformation (1), first difference (2), second difference (3), logarithm (4) and first difference of logarithm (5). This classification is in accordance with Bernanke et al. (2005). Most of our data comes from the research database of the Federal Reserve Bank of St. Louis. For clarity reasons, we therefore shorten the Uniform Resource Locator (URL) “http://research.stlouisfed.org/fred2/series” by “fred”. As long as it makes sense and is possible, we add source abbreviations in subsequent sections. Finally, we comment in which manner publication delays are taken into account. Moreover, we shorten time series descriptions using the notations: Seasonally Adjusted (SA) and Not Seasonally Adjusted (NSA).

B.1 Mixed-Freq. Inform. Supporting Asset Allocation Decisions

The total data sample in Section 3.2 ranges from February 1, 1985 until November 11, 2016 and is weekly updated. For the sake of clarity, we separate the components of the return VARX in Definition 3.1.9 from the mixed-frequency panel data in Lemma 3.1.7. Besides data of the Federal Reserve Bank of St. Louis, we downloaded data from the URLs “http://de.global-rates.com/” and “http://finance.yahoo.com/”. In this context, the abbreviations “glrates” and “yahoo” stand for the URL prefixes “http://de.global-rates.com /zinssatze/libor/amerikanischer-dollar” and “http://finance.yahoo.com/quote”, respectively.

Return Vector

No.	**Series ID**	**Time span**	**Freq.**	**Type**	**Trans.**	**Series description**
1.	GOLD	1985/02/01-2016/11/11	d	1	5	Gold fixing price in London Bullion Market at 10.30 am (London time), USD per troy ounce, NSA, delay of 0 days, fred/GOLDAMGBD228NLBM

2.	NASDAQ	1985/02/01-2016/11/11	*d*	1	5	Nasdaq Composite, price index, NSA, delay of 0 days, yahoo/%5EIXIC/history?period1=34556400&period2=1479164400&interval=1d&filter=history&frequency=1d
3.	S&P500	1985/02/01-2016/11/11	*d*	1	5	Standard & Poor's 500, price index, NSA, delay of 0 days, yahoo/%5EGSPC/history?period1=-630982800&period2=1479164400&interval=1d&filter=history&frequency=1d
4.	DJIA	1985/02/01-2016/11/11	*d*	1	5	Dow Jones Industrial Average, price index, NSA, delay of 0 days, yahoo/%5EDJI/history?period1=475801200&period2=1479164400&interval=1d&filter=history&frequency=1d

Mixed-Frequency Panel Data

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
5.	DGS3-1	1985/02/01-2016/11/11	*d*	1	1	Spread between 3-Year and 1-Year Treasury Constant Maturity Rates, NSA, delay of 0 days, fred/DGS3, fred/DGS1
6.	DGS10-1	1985/02/01-2016/11/11	*d*	1	1	Spread between 10-Year and 1-Year Treasury Constant Maturity Rates, NSA, delay of 0 days, fred/DGS10, fred/DGS1
7.	LIBOR6-3	1986/01/02-2016/11/11	*d*	1	1	Spread between 6-Month and 3-Month LIBOR, NSA, delay of 1 day, fred/USD6MTD156N, fred/USD3MTD156N (history), glrates/usd-libor-zinssatz-6-monate.aspx, glrates/usd-libor-zinssatz-3-monate.aspx (latest values)
8.	LIBOR12-3	1986/01/02-2016/11/11	*d*	1	1	Spread between 12-Month and 3-Month LIBOR, NSA, delay of 1 day, fred/USD12MD156N, fred/USD3MTD156N (history), glrates/usd-libor-zinssatz-12-monate.aspx, glrates/usd-libor-zinssatz-3-monate.aspx (latest values)
9.	DB-AAA	1986/01/02-2016/10/07	*d*	1	1	Spread between Moody's Seasoned Baa and Aaa Corporate Bond Yield, percent, NSA, delay of 1 day, fred/DBAA, fred/DAAA
10.	BBB-AAA	1996/12/31-2016/11/11	*d*	1	1	Spread between BofA Merrill Lynch US Corporate BBB and AAA Option-Adjusted Spread, percent, NSA, delay of 1 day, fred/BAMLC0A4CBBB, fred/BAMLC0A1CAAA
11.	H0A0HYM2	1996/12/31-2016/11/11	*d*	1	2	BofA Merrill Lynch US High Yield Option-Adjusted Spread, percent, NSA, delay of 1 day, fred/BAMLH0A0HYM2
12.	C0A0CM	1996/12/31-2016/11/11	*d*	1	2	BofA Merrill Lynch US Corporate Master Option-Adjusted Spread, percent, NSA, delay of 1 day, fred/BAMLC0A0CM
13.	EMCBPIOAS	1998/12/31-2016/11/11	*d*	1	2	BofA Merrill Lynch Emerging Markets Corporate Plus Index Option-Adjusted Spread, percent, NSA, delay of 1 day, fred/BAMLEMCBPIOAS
14.	VIXCLS	1990/01/02-2016/11/11	*d*	1	5	CBOE Volatility Index: VIX, NSA, delay of 1 day, fred/VIXCLS
15.	VXOCLS	1986/01/02-2016/11/11	*d*	1	5	CBOE S&P 100 Volatility Index: VXO, NSA, delay of 1 day, fred/VXOCLS
16.	VXDCLS	1997/10/07-2016/11/11	*d*	1	5	CBOE DJIA Volatility Index, NSA, delay of 1 day, fred/VXDCLS
17.	VXNCLS	2001/02/02-2016/11/11	*d*	1	5	CBOE NASDAQ 100 Volatility Index, NSA, delay of 1 day, fred/VXNCLS
18.	DCOILWTICO	1986/01/02-2016/11/11	*d*	1	5	Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma, USD per Barrel, NSA, delay of 4 days, fred/DCOILWTICO
19.	PPIACO	1985/02/01-2016/11/11	*m*	2	5	Producer Price Index for All Commodities, Index 1982=100, NSA, delay of 45 days after 1st of respective month, fred/PPIACO
20.	CPIAUCSL	1985/02/01-2016/11/11	*m*	2	5	Consumer Price Index for All Urban Consumers: All Items, Index 1982-1984=100, NSA, delay of 49 days after 1st of respective month, fred/CPIAUCSL

Mixed-Frequency Panel Data

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
21.	GDP	1985/02/01-2016/11/11	q	2	5	Gross Domestic Product, billions of USD, SA annual rate, delay of 126 days after 1st of respective quarter, fred/GDP
22.	INDPRO	1985/02/01-2016/11/11	m	2	5	Industrial Production Index, Index 2012=100, SA, delay of 49 days after 1st of respective month, fred/INDPRO
23.	IPB50001N	1985/02/01-2016/11/11	m	2	5	Industrial Production: Total index, Index 2012=100, NSA, delay of 49 days after 1st of respective month, fred/IPB50001N
24.	PAYEMS	1985/02/01-2016/11/11	m	2	5	All Employees: Total Nonfarm Payrolls, thousands of persons, SA, delay of 40 days after 1st of respective month, fred/PAYEMS
25.	UNRATE	1985/02/01-2016/11/11	m	1	1	Civilian Unemployment Rate, percent, SA, delay of 40 days after 1st of respective month, fred/UNRATE
26.	PCE	1985/02/01-2016/11/11	m	2	5	Personal Consumption Expenditures, billions of USD, SA annual rate, delay of 62 days after 1st of respective month, fred/PCE
27.	PSAVERT	1985/02/01-2016/11/11	m	1	1	Personal Saving Rate, percent, SA annual rate, delay of 62 days after 1st of respective month, fred/PSAVERT

B.2 Estimation of Approximate Dynamic Factor Models

This section lists the panel data of the empirical study in Section 4.5. The overall sample ranges from January 8, 1999 until February 5, 2016 and is weekly updated. As before, most of our panel data comes from the research database of the Federal Reserve Bank of St. Louis. In addition, we downloaded data from the URL “http://www.global-rates.com/”. This is why “global” refers to the URL prefix “http://www.global-rates.com/interest-rates/libor/american-dollar” in the sequel. For clarity reasons, we group the variables as described in Section 4.5.

US Treasuries

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
1.	DGS3MO	1999/01/08-2016/02/05	d	1	2	3-Month Treasury Constant Maturity Rate, percent, NSA, delay of 1 day, fred/DGS3MO
2.	DTB3	1999/01/08-2016/02/05	d	1	2	3-Month Treasury Bill: Secondary Market Rate, percent, NSA, delay of 1 day, fred/DTB3
3.	DGS1	1999/01/08-2016/02/05	d	1	2	1-Year Treasury Constant Maturity Rate, percent, NSA, delay of 1 day, fred/DGS1
4.	DGS2	1999/01/08-2016/02/05	d	1	2	2-Year Treasury Constant Maturity Rate, percent, NSA, delay of 1 day, fred/DGS2
5.	DGS3	1999/01/08-2016/02/05	d	1	2	3-Year Treasury Constant Maturity Rate, percent, NSA, delay of 1 day, fred/DGS3
6.	DGS5	1999/01/08-2016/02/05	d	1	2	5-Year Treasury Constant Maturity Rate, percent, NSA, delay of 1 day, fred/DGS5
7.	DGS7	1999/01/08-2016/02/05	d	1	2	7-Year Treasury Constant Maturity Rate, percent, NSA, delay of 1 day, fred/DGS7
8.	DGS10	1999/01/08-2016/02/05	d	1	2	10-Year Treasury Constant Maturity Rate, percent, NSA, delay of 1 day, fred/DGS10

US Corporates

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
9.	DAAA	1999/01/08-2016/02/05	d	1	2	Moody’s Seasoned Aaa Corporate Bond Yield, percent, NSA, delay of 1 day, fred/DAAA
10.	DBAA	1999/01/08-2016/02/05	d	1	2	Moody’s Seasoned Baa Corporate Bond Yield, percent, NSA, delay of 1 day, fred/DBAA
11.	C0A0CM	1999/01/08-2016/02/05	d	1	2	BofA Merrill Lynch US Corporate Master Option-Adjusted Spread, percent, NSA, delay of 1 day, fred/BAMLC0A0CM
12.	C0A4CBBB	1999/01/08-2016/02/05	d	1	2	BofA Merrill Lynch US Corporate BBB Option-Adjusted Spread, percent, NSA, delay of 1 day, fred/BAMLC0A4CBBB

US LIBOR

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
13.	LIBOR1	1999/01/08-2016/02/05	d	1	2	1-Month LIBOR, based on USD, percent, NSA, delay of 1 day, fred/USD1MTD156N (history), global/usd-libor-interest-rate-1-month.aspx (latest values)
14.	LIBOR2	1999/01/08-2016/02/05	d	1	2	2-Month LIBOR, based on USD, percent, NSA, delay of 1 day, fred/USD2MTD156N (history), global/usd-libor-interest-rate-2-months.aspx (latest values)
15.	LIBOR3	1999/01/08-2016/02/05	d	1	2	3-Month LIBOR, based on USD, percent, NSA, delay of 1 day, fred/USD3MTD156N (history), global/usd-libor-interest-rate-3-months.aspx (latest values)
16.	LIBOR6	1999/01/08-2016/02/05	d	1	2	6-Month LIBOR, based on USD, percent, NSA, delay of 1 day, fred/USD6MTD156N (history), global/usd-libor-interest-rate-6-months.aspx (latest values)
17.	LIBOR12	1999/01/08-2016/02/05	d	1	2	12-Month LIBOR, based on USD, percent, NSA, delay of 1 day, fred/USD12MD156N (history), global/usd-libor-interest-rate-12-months.aspx (latest values)

Foreign Exchange Rates and GOLD

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
18.	USAL	1999/01/08-2016/02/05	d	1	5	US / Australia Foreign Exchange Rate, NSA, delay of 0 days, http://www.rba.gov.au/statistics /historical-data.html#exchange-rates (history), http://www.rba.gov.au/statistics/frequency/ exchange-rates.html (latest values)
19.	CAUS	1999/01/08-2016/02/05	d	1	5	Canada / US Foreign Exchange Rate, NSA, delay of 0 days, fred/DEXCAUS (history), http://www.bankofcanada.ca/rates/exchange/ noon-rates-5-day/ (latest values)
20.	USUK	1999/01/08-2016/02/05	d	1	5	US / UK Foreign Exchange Rate, NSA, delay of 1 day, http://www.bankofengland.co.uk/boeapps/ iadb/index.asp?Travel=NIxIRx&levels=1&XNotes =Y&C=C8P&G0Xtop.x=28&G0Xtop.y=10&X Notes2=Y&Nodes=X3790X3791X3836&Section Required=I&HideNums=-1&ExtraInfo=true #BM
21.	USEU	1999/01/08-2016/02/05	d	1	5	US / Euro Foreign Exchange Rate, NSA, delay of 1 day, https://www.ecb.europa.eu/stats/exchang e/eurofxref/html/index.en.html
22.	GOLD	1999/01/08-2016/02/05	d	1	5	Gold fixing price in London Bullion Market at 10.30 am (London time), USD per troy ounce, NSA, delay of 0 days, fred/GOLDAMGBD228NLBM

Demand

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
23.	UNRATENSA	1999/01/08-2016/02/05	m	5	2	Civilian Unemployment Rate, percent, NSA, delay of 35 days after 1st of respective month, fred/UNRATENSA
24.	PSAVERT	1999/01/08-2016/02/05	m	5	2	Personal Saving Rate, percent, SA annual rate, delay of 61 days after 1st of respective month, fred/PSAVERT
25.	PI	1999/01/08-2016/02/05	m	3	5	Personal Income, billions of USD, SA annual rate, delay of 61 days after 1st of respective month, fred/PI
26.	PCE	1999/01/08-2016/02/05	m	3	5	Personal Consumption Expenditures, billions of USD, SA annual rate, delay of 61 days after 1st of respective month, fred/PCE
27.	GOVEXP	1999/01/08-2016/02/05	q	3	5	Government total expenditures, billions of USD, SA annual rate, delay of 140 days after 1st of respective quarter, fred/W068RCQ027SBEA

Supply

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
28.	GDP	1999/01/08-2016/02/05	q	3	5	Gross Domestic Product, billions of USD, SA annual rate, delay of 140 days after 1st of respective quarter, fred/GDP
29.	INDPRO	1999/01/08-2016/02/05	m	3	5	Industrial Production Index, Index 2007=100, SA, delay of 45 days after 1st of respective month, fred/INDPRO
30.	EXPGSC1	1999/01/08-2016/02/05	q	3	5	Real Exports of Goods & Services, billions of chained 2009 USD, SA annual rate, delay of 140 days after 1st of respective quarter, fred/EXPGSC1
31.	IMPGSC1	1999/01/08-2016/02/05	q	3	5	Real Imports of Goods & Services, billions of chained 2009 USD, SA annual rate, delay of 140 days after 1st of respective quarter, fred/IMPGSC1

Inflation

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
32.	CPIAUCNS	1999/01/08-2016/02/05	m	3	5	Consumer Price Index for All Urban Consumers: All Items, Index 1982-1984=100, NSA, delay of 45 days after 1st of respective month, fred/CPIAUCNS
33.	PPIACO	1999/01/08-2016/02/05	m	3	5	Producer Price Index for All Commodities, Index 1982=100, NSA, delay of 43 days after 1st of respective month, fred/PPIACO

B.3 FAVARs for Incomplete Panel Data

This data is for the most part an updated version of the one in Bernanke et al. (2005). A few of their time series are missing, since we did not find them anymore, while new times series, in particular, incomplete ones were added. For clarity reasons, we group the data as follows: real output and income; employment and hours; consumption; housing starts and sales; real inventories, orders, and unfilled orders; stock prices; foreign exchange rates; interest rates; money and credit quantity aggregates; price indices; average hourly earnings; miscellaneous; mixed-frequency time series; observed variables $\boldsymbol{Y}_t$. The total data sample ranges from January 1959 to October 2015 and is monthly updated. In addition to the series number, the first K variables of the sorted data provide their position number in brackets (Bork, 2009). An asterix * next to an abbreviation marks the respective variable as slow-moving. The distinction between slow- and fast-moving variables arises from Bernanke et al. (2005). In doing so, they assume slow-moving variables "not to respond contemporaneously to unanticipated changes in monetary policy", however, they allow fast-moving variables "to respond contemporaneously to policy shocks". Again, the notation "fred" refers to the research database of the Federal Reserve Bank of St. Louis.

Real output and income

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
1.	IPFINAL*	1959:01-2015:10	m	1	5	Industrial Production: Final Products (Market Group), Index 2012=100, SA, delay of 0 months, fred/IPFINAL
2.[6]	IPCONGD*	1959:01-2015:10	m	1	5	Industrial Production: Consumer Goods, Index 2012=100, SA, delay of 0 months, fred/IPCONGD
3.	IPDCONGD*	1959:01-2015:10	m	1	5	Industrial Production: Durable Consumer Goods, Index 2012=100, SA, delay of 0 months, fred/IPDCONGD
4.	IPNCONGD*	1959:01-2015:10	m	1	5	Industrial Production: Nondurable Consumer Goods, Index 2012=100, SA, delay of 0 months, fred/IPNCONGD
5.	IPBUSEQ*	1959:01-2015:10	m	1	5	Industrial Production: Business Equipment, Index 2012=100, SA, delay of 0 months, fred/IPBUSEQ
6.	IPMAT*	1959:01-2015:10	m	1	5	Industrial Production: Materials, Index 2012=100, SA, delay of 0 months, fred/IPMAT

7.	IPB53100N*	1959:01-2015:10	m	1	5	Industrial Production: Durable goods materials, Index 2012=100, NSA, delay of 0 months, fred/IPB53100N
8.	IPB53200N*	1959:01-2015:10	m	1	5	Industrial Production: Nondurable Goods Materials, Index 2012=100, NSA, delay of 0 months, fred/IPB53200N
9.	IPMANSICS*	1959:01-2015:10	m	1	5	Industrial Production: Manufacturing (SIC), Index 2012=100, SA, delay of 0 months, fred/IPMANSICS
10.	INDPRO*	1959:01-2015:10	m	1	5	Industrial Production Index, Index 2012=100, SA, delay of 0 months, fred/INDPRO
11.	CUMFNS*	1959:01-2015:10	m	1	1	Capacity Utilization: Manufacturing (SIC), Percent of Capacity, SA, delay of 0 months, fred/CUMFNS
12.	NAPM*	1959:01-2015:10	m	1	1	ISM Manufacturing: PMI Composite Index, Index, SA, delay of 0 months, fred/NAPM
13.	NAPMPI*	1959:01-2015:10	m	1	1	ISM Manufacturing: Production Index, Index, SA, delay of 0 months, fred/NAPMPI
14.	RPI*	1959:01-2015:10	m	1	5	Real Personal Income, billions of chained 2009 USD, SA Annual Rate, delay of 0 months, fred/RPI
15.	W875RX1*	1959:01-2015:10	m	1	5	Real Personal Income Excluding Current Transfer Receipts, billions of chained 2009 USD, SA annual rate, delay of 0 months, fred/W875RX1

Employment and hours

No.	**Series ID**	**Time span**	**Freq.**	**Type**	**Trans.**	**Series description**
16.	CE16OV*	1959:01-2015:10	m	1	5	Civilian Employment, thousands of persons, SA, delay of 0 months, fred/CE16OV
17.[4]	UNRATE*	1959:01-2015:10	m	1	1	Civilian Unemployment Rate, percent, SA, delay of 0 months, fred/UNRATE
18.	UEMPMEAN*	1959:01-2015:10	m	1	5	Average (Mean) Duration of Unemployment, Weeks, SA, delay of 0 months, fred/UEMPMEAN
19.	UEMPLT5*	1959:01-2015:10	m	1	5	Number of Civilians Unemployed for Less Than 5 Weeks, thousands of persons, SA, delay of 0 months, fred/UEMPLT5
20.	UEMP5TO14*	1959:01-2015:10	m	1	5	Number of Civilians Unemployed for 5 to 14 Weeks, thousands of persons, SA, delay of 0 months, fred/UEMP5TO14
21.	UEMP15OV*	1959:01-2015:10	m	1	5	Number of Civilians Unemployed for 15 Weeks and Over, thousands of persons, SA, delay of 0 months, fred/UEMP15OV
22.	UEMP15T26*	1959:01-2015:10	m	1	5	Number of Civilians Unemployed for 15 to 26 Weeks, thousands of persons, SA, delay of 0 months, fred/UEMP15T26
23.[1]	PAYEMS*	1959:01-2015:10	m	1	5	All Employees: Total Nonfarm Payrolls, thousands of persons, SA, delay of 0 months, fred/PAYEMS
24.	USPRIV*	1959:01-2015:10	m	1	5	All Employees: Total Private Industries, thousands of persons, SA, delay of 0 months, fred/USPRIV
25.	USGOOD*	1959:01-2015:10	m	1	5	All Employees: Goods-Producing Industries, Thousands of Persons, SA, delay of 0 months, fred/USGOOD
26.	CES1021000001*	1959:01-2015:10	m	1	5	All Employees: Mining and Logging: Mining, thousands of persons, SA, delay of 0 months, fred/CES1021000001
27.	USCONS*	1959:01-2015:10	m	1	5	All Employees: Construction, thousands of persons, SA, delay of 0 months, fred/USCONS
28.	MANEMP*	1959:01-2015:10	m	1	5	All Employees: Manufacturing, thousands of persons, SA, delay of 0 months, fred/MANEMP
29.	DMANEMP*	1959:01-2015:10	m	1	5	All Employees: Durable Goods, thousands of persons, SA, delay of 0 months, fred/DMANEMP
30.	NDMANEMP*	1959:01-2015:10	m	1	5	All Employees: Nondurable Goods, thousands of persons, SA, delay of 0 months, fred/NDMANEMP
31.	CES0800000001*	1959:01-2015:10	m	1	5	All Employees: Private Service-Providing, thousands of persons, SA, delay of 0 months, fred/CES0800000001
32.	USTPU*	1959:01-2015:10	m	1	5	All Employees: Trade, Transportation and Utilities, thousands of persons, SA, delay of 0 months, fred/USTPU
33.	USWTRADE*	1959:01-2015:10	m	1	5	All Employees: Wholesale Trade, thousands of persons, SA, delay of 0 months, fred/USWTRADE

Employment and hours

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
34.[5]	USFIRE*	1959:01-2015:10	m	1	5	All Employees: Financial Activities, thousands of persons, SA, delay of 0 months, fred/USFIRE
35.	USPBS*	1959:01-2015:10	m	1	5	All Employees: Professional and Business Services, thousands of persons, SA, delay of 0 months, fred/USPBS
36.	USGOVT*	1959:01-2015:10	m	1	5	All Employees: Government, thousands of persons, SA, delay of 0 months, fred/USGOVT
37.	AWHMAN*	1959:01-2015:10	m	1	1	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing, Hours, SA, delay of 0 months, fred/AWHMAN
38.[7]	AWOTMAN*	1959:01-2015:10	m	1	1	Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing, Hours, SA, delay of 0 months, fred/AWOTMAN
39.	NAPMEI*	1959:01-2015:10	m	1	1	ISM Manufacturing: Employment Index, Index, SA, delay of 0 months, fred/NAPMEI

Consumption

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
40.[8]	PCE*	1959:01-2015:10	m	1	5	Personal Consumption Expenditures, billions of USD, SA annual rate, delay of 0 months, fred/PCE
41.	PCEDG*	1959:01-2015:10	m	1	5	Personal Consumption Expenditures: Durable Goods, billions of USD, SA annual rate, delay of 0 months, fred/PCEDG
42.	PCEND*	1959:01-2015:10	m	1	5	Personal Consumption Expenditures: Nondurable Goods, billions of USD, SA annual rate, delay of 0 months, fred/PCEND
43.	PCES*	1959:01-2015:10	m	1	5	Personal Consumption Expenditures: Services, billions of USD, SA annual rate, delay of 0 months, fred/PCES

Housing starts and sales

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
44.	HOUST	1959:01-2015:10	m	1	4	Housing Starts: Total: New Privately Owned Housing Units Started, thousands of units, SA annual rate, delay of 0 months, fred/HOUST
45.	HOUSTNE	1959:01-2015:10	m	1	4	Housing Starts in Northeast Census Region, thousands of units, SA annual rate, delay of 0 months, fred/HOUSTNE
46.	HOUSTMW	1959:01-2015:10	m	1	4	Housing Starts in Midwest Census Region, thousands of units, SA annual Rate, delay of 0 months, fred/HOUSTMW
47.	HOUSTS	1959:01-2015:10	m	1	4	Housing Starts in South Census Region, thousands of units, SA annual rate, delay of 0 months, fred/HOUSTS
48.	HOUSTW	1959:01-2015:10	m	1	4	Housing Starts in West Census Region, thousands of units, SA annual rate, delay of 0 months, fred/HOUSTW
49.	PERMITNSA	1959:01-2015:10	m	1	4	New Private Housing Units Authorized by Building Permits, thousands of units, NSA, delay of 0 months, fred/PERMITNSA

Real inventories, orders, and unfilled orders

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
50.	NAPMII	1959:01-2015:10	m	1	1	ISM Manufacturing: Inventories Index, Index, NSA, delay of 0 months, fred/NAPMII
51.	NAPMNOI	1959:01-2015:10	m	1	1	ISM Manufacturing: New Orders Index, Index, SA, delay of 0 months, fred/NAPMNOI
52.	NAPMSDI	1959:01-2015:10	m	1	1	ISM Manufacturing: Supplier Deliveries Index, Index, SA, delay of 0 months, fred/NAPMSDI

Stock prices

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
53.	FSPCOM	1959:01-2015:10	m	1	5	S&P's Common Stock Price Index: Composite, delay of 0 months, http://www.econ.yale.edu/~shiller/data/ie_data.xls
54.	FSDXP	1959:01-2015:10	m	1	1	S&P's Composite Common Stock: Dividend Yield, delay of 0 months, http://www.econ.yale.edu/~shiller/data/ie_data.xls
55.	FSPXE	1959:01-2015:10	m	1	1	S&P's Composite Common Stock: Price-Earnings Ratio, delay of 0 months, http://www.econ.yale.edu/~shiller/data/ie_data.xls

Foreign exchange rates

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
56.	EXSZUS	1959:01-2015:10	m	1	5	Switzerland / US Foreign Exchange Rate, Swiss Francs to One USD, NSA, delay of 0 months, fred/EXSZUS
57.	EXJPUS	1959:01-2015:10	m	1	5	Japan / US Foreign Exchange Rate, Japanese Yen to One USD, NSA, delay of 0 months, fred/EXJPUS
58.	EXUSUK	1959:01-2015:10	m	1	5	US / UK Foreign Exchange Rate, USDs to One British Pound, NSA, delay of 0 months, fred/EXUSUK
59.	EXCAUS	1959:01-2015:10	m	1	5	Canada / US Foreign Exchange Rate, Canadian Dollars to One USD, NSA, delay of 0 months, fred/EXCAUS

Interest rates

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
60.	TB3MS	1959:01-2015:10	m	1	1	3-Month Treasury Bill: Secondary Market Rate, percent, NSA, delay of 0 months, fred/TB3MS
61.	TB6MS	1959:01-2015:10	m	1	1	6-Month Treasury Bill: Secondary Market Rate, percent, NSA, delay of 0 months, fred/TB6MS
62.	GS1	1959:01-2015:10	m	1	1	1-Year Treasury Constant Maturity Rate, percent, NSA, delay of 0 months, fred/GS1
63.	GS5	1959:01-2015:10	m	1	1	5-Year Treasury Constant Maturity Rate, percent, NSA, delay of 0 months, fred/GS5
64.	GS10	1959:01-2015:10	m	1	1	10-Year Treasury Constant Maturity Rate, percent, NSA, delay of 0 months, fred/GS10
65.	AAA	1959:01-2015:10	m	1	1	Moody's Seasoned Aaa Corporate Bond Yield, percent, NSA, delay of 0 months, fred/AAA
66.	BAA	1959:01-2015:10	m	1	1	Moody's Seasoned Baa Corporate Bond Yield, percent, NSA, delay of 0 months, fred/BAA
67.	TB3SMFFM	1959:01-2015:10	m	1	1	3-Month Treasury Bill Minus Federal Funds Rate, percent, NSA, delay of 0 months, fred/TB3SMFFM
68.	TB6SMFFM	1959:01-2015:10	m	1	1	6-Month Treasury Bill Minus Federal Funds Rate, percent, NSA, delay of 0 months, fred/TB6SMFFM
69.	T1YFFM	1959:01-2015:10	m	1	1	1-Year Treasury Constant Maturity Minus Federal Funds Rate, percent, NSA, delay of 0 months, fred/T1YFFM
70.	T5YFFM	1959:01-2015:10	m	1	1	5-Year Treasury Constant Maturity Minus Federal Funds Rate, percent, NSA, delay of 0 months, fred/T5YFFM
71.	T10YFFM	1959:01-2015:10	m	1	1	10-Year Treasury Constant Maturity Minus Federal Funds Rate, percent, NSA, delay of 0 months, fred/T10YFFM
72.	AAAFFM	1959:01-2015:10	m	1	1	Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate, percent, NSA, delay of 0 months, fred/AAAFFM
73.	BAAFFM	1959:01-2015:10	m	1	1	Moody's Seasoned Baa Corporate Bond Minus Federal Funds Rate, percent, NSA, delay of 0 months, fred/BAAFFM

Money and credit quantity aggregates

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
74.	M1SL	1959:01-2015:10	m	1	5	M1 Money Stock, billions of USD, SA, delay of 0 months, fred/M1SL
75.	M2SL	1959:01-2015:10	m	1	5	M2 Money Stock, billions of USD, SA, delay of 0 months, fred/M2SL
76.	TOTRESNS	1959:01-2015:10	m	1	5	Total Reserves of Depository Institutions, billions of USD, NSA, delay of 0 months, fred/TOTRESNS
77.	BUSLOANS	1959:01-2015:10	m	1	5	Commercial and Industrial Loans, All Commercial Banks, billions of USD, SA, delay of 0 months, fred/BUSLOANS
78.	NONREVSL	1959:01-2015:10	m	1	5	Total Nonrevolving Credit Owned and Securitized, Outstanding, billions of USD, SA, delay of 0 months, fred/NONREVSL

Price indices

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
79.	NAPMPRI	1959:01-2015:10	m	1	1	ISM Manufacturing: Prices Index, Index, NSA, delay of 0 months, fred/NAPMPRI
80.	PPIFGS*	1959:01-2015:10	m	1	5	Producer Price Index by Commodity for Finished Goods, Index 1982=100, SA, delay of 0 months, fred/PPIFGS
81.[3]	PPIFCG*	1959:01-2015:10	m	1	5	Producer Price Index by Commodity for Finished Consumer Goods, Index 1982=100, SA, delay of 0 months, fred/PPIFCG

Price indices

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
82.	PPIITM*	1959:01-2015:10	m	1	5	Producer Price Index by Commodity Intermediate Materials: Supplies and Components, Index 1982=100, SA, delay of 0 months, fred/PPIITM
83.[9]	PPICRM*	1959:01-2015:10	m	1	5	Producer Price Index by Commodity for Crude Materials for Further Processing, Index 1982=100, SA, delay of 0 months, fred/PPICRM
84.	CPIAUCSL*	1959:01-2015:10	m	1	5	Consumer Price Index for All Urban Consumers: All Items, Index 1982-1984=100, SA, delay of 0 months, fred/CPIAUCSL
85.	CPIAPPSL*	1959:01-2015:10	m	1	5	Consumer Price Index for All Urban Consumers: Apparel, Index 1982-1984=100, SA, delay of 0 months, fred/CPIAPPSL
86.	CPITRNSL*	1959:01-2015:10	m	1	5	Consumer Price Index for All Urban Consumers: Transportation, Index 1982-1984=100, SA, delay of 0 months, fred/CPITRNSL
87.	CPIMEDSL*	1959:01-2015:10	m	1	5	Consumer Price Index for All Urban Consumers: Medical Care, Index 1982-1984=100, SA, delay of 0 months, fred/CPIMEDSL
88.	CUSR0000SAC*	1959:01-2015:10	m	1	5	Consumer Price Index for All Urban Consumers: Commodities, Index 1982-1984=100, SA, delay of 0 months, fred/CUSR0000SAC
89.	CUSR0000SAD*	1959:01-2015:10	m	1	5	Consumer Price Index for All Urban Consumers: Durables, Index 1982-1984=100, SA, delay of 0 months, fred/CUSR0000SAD
90.	CUSR0000SAS*	1959:01-2015:10	m	1	5	Consumer Price Index for All Urban Consumers: Services, Index 1982-1984=100, SA, delay of 0 months, fred/CUSR0000SAS
91.[2]	CPILFESL*	1959:01-2015:10	m	1	5	Consumer Price Index for All Urban Consumers: All Items Less Food and Energy, Index 1982-1984=100, SA, delay of 0 months, fred/CPILFESL
92.	CUSR0000SA0L2*	1959:01-2015:10	m	1	5	Consumer Price Index for All Urban Consumers: All items less shelter, Index 1982-1984=100, SA, delay of 0 months, fred/CUSR0000SA0L2
93.	CUSR0000SA0L5*	1959:01-2015:10	m	1	5	Consumer Price Index for All Urban Consumers: All items less medical care, Index 1982-1984=100, SA, delay of 0 months, fred/CUSR0000SA0L5

Average hourly earnings

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
94.	CES2000000008*	1959:01-2015:10	m	1	5	Average Hourly Earnings of Production and Nonsupervisory Employees: Construction, USD per Hour, SA, delay of 0 months, fred/CES2000000008
95.	CES3000000008*	1959:01-2015:10	m	1	5	Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing, USD per Hour, SA, delay of 0 months, fred/CES3000000008

Miscellaneous

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
96.	MEI	1959:01-2015:10	m	1	1	Composite Leading Indicators, Amplitude Adjusted, delay of 0 months, http://stats.oecd.org/Index.aspx?DataSetCode=MEI_CLI

Mixed-frequency time series

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
97.	EXGEUS	1971:01-2001:12	m	1	5	Germany / US Foreign Exchange Rate, German Deutsche Marks to One USD, NSA, delay of 0 months, fred/EXGEUS
98.	EXFRUS	1971:01-2001:12	m	1	5	France / US Foreign Exchange Rate, French Francs to One USD, NSA, delay of 0 months, fred/EXFRUS

Mixed-frequency time series

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
99.	EXITUS	1971:01-2001:12	m	1	5	Italy / US Foreign Exchange Rate, Italian Lire to One USD, NSA, delay of 0 months, fred/EXITUS
100.	EXUSEU	1999:01-2015:10	m	1	5	US / Euro Foreign Exchange Rate, USDs to One Euro, NSA, delay of 0 months, fred/EXUSEU
101.	GDP	1959:01-2015:10	q	2	5	Gross Domestic Product, billions of USD, SA annual rate, delay of 0 months, fred/GDP
102.	W068RCQ027SBEA	1960:01-2015:10	q	2	5	Government Total Expenditures, billions of USD, SA annual rate, delay of 0 months, fred/W068RCQ027SBEA
103.	IMPGSC1	1959:01-2015:10	q	2	5	Real Imports of Goods and Services, billions of Chained 2009 USD, SA annual rate, delay of 0 months, fred/IMPGSC1
104.	EXPGSC1	1959:01-2015:10	q	2	5	Real Exports of Goods and Services, billions of Chained 2009 USD, SA annual rate, delay of 0 months, fred/EXPGSC1
105.	WALCL	2002:12-2015:10	m	1	5	All Federal Reserve Banks - Total Assets, Eliminations from Consolidation, millions of USD, NSA, delay of 0 months, fred/WALCL
106.	MBST	2002:12-2015:10	m	1	5	Mortgage-backed securities held by the Federal Reserve: All Maturities, millions of USD, NSA, delay of 0 months, fred/MBST
107.	TREAST	2002:12-2015:10	m	1	5	US Treasury securities held by the Federal Reserve: All Maturities, millions of USD, NSA, delay of 0 months, fred/TREAST
108.	WRESBAL	1984:01-2015:10	m	1	5	Reserve Balances with Federal Reserve Banks, billions of USD, NSA, delay of 0 months, fred/WRESBAL

Observed variables $\boldsymbol{Y}_t$

No.	Series ID	Time span	Freq.	Type	Trans.	Series description
109.	CURRCIR	1959:01-2015:10	m	1	5	Currency in Circulation, billions of USD, NSA, delay of 0 months, fred/CURRCIR
110.	AMBSL	1959:01-2015:10	m	1	5	St. Louis Adjusted Monetary Base, billions of USD, SA, delay of 0 months, fred/AMBSL
111.	FEDFUNDS	1959:01-2015:10	m	1	1	Effective Federal Funds Rate, percent, NSA, delay of 0 months, fred/FEDFUNDS

Acronyms

ADFM	Approximate Dynamic Factor Model
AIC	Akaike Information Criterion
AR	Autoregressive Model
ARIMA	Autoregressive Integrated Moving Average Model
ARX	Autoregressive Extended Model
ASFM	Approximate Static Factor Model
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BofA	Bank of America
bp	basis point
B&H	Buy&Hold
CBOE	Chicago Board Options Exchange
CPPI	Constant Proportion Portfolio Insurance
DFM	Dynamic Factor Model
DJIA	Dow Jones Industrial Average
EDFM	Exact Dynamic Factor Model
EM	Expectation-Maximization Algorithm
ESFM	Exact Static Factor Model
FA	Factor Analysis
FAVAR	Factor-Augmented Vector Autoregression Model
FEVD	Forecast Error Variance Decomposition
FEDFUNDS	Effective Federal Funds Rate
FM	Factor Model
FX	Foreign Exchange
GDFM	Generalized Dynamic Factor Model
GDP	Gross Domestic Product
iid	identically and independently distributed
IRF	Impulse Response Function
LIBOR	London Interbank Offered Rate
L&S	Leverage & Short Sales
KF	Kalman Filter
KS	Kalman Smoother
MC	Monte Carlo
MIDAS	Mixed-Data Sampling
MIS	Mean Interval Score
MLE	Maximum-Likelihood Estimation

NASDAQ	Nasdaq Composite
NSA	Not Seasonally Adjusted
OLS	Ordinary Least Squares Regression
PCA	Principal Component Analysis
PL	Prediction Level
PPCA	Probabilistic Principal Component Analysis
RE	Ranking Error
RIO	Ratio of Interval Outliers
RMSE	Root-Mean-Square Error
SA	Seasonally Adjusted
SFM	Static Factor Model
S&P500	Standard & Poor's 500
UK	United Kingdom
URL	Uniform Resource Locator
US	United States
USD	United States Dollar
VAR	Vector Autoregression Model
VARX	Vector Autoregression Model with Exogenous Variables

Nomenclature

$abs\left(\cdot\right)$	Absolute value
$\arg\max$	Argument of maximum
$\lceil\cdot\rceil$	Ceiling function
$\hat{P}^F_{t\mid t-1} \in \mathbb{R}^{K\times K}$	Covariance matrix of factor $\boldsymbol{F}_t$ conditioned on Ω_{t-1}
$\hat{\Sigma} \in \mathbb{R}^{K\times K}$	Estimator of matrix $\Sigma \in \mathbb{R}^{K\times K}$
$\hat{\boldsymbol{\mu}} \in \mathbb{R}^N$	Estimator of mean $\boldsymbol{\mu} \in \mathbb{R}^N$
$\lvert\lvert \cdot \rvert\rvert_2$	Euclidean norm / 2-norm
$\mathbb{E}_\Theta\left[\cdot\right]$	Expectation based on model parameters Θ
$\hat{\boldsymbol{F}}_{t\mid t-1} \in \mathbb{R}^K$	Expectation of factor $\boldsymbol{F}_t$ conditioned on Ω_{t-1}
$exp\left(\cdot\right)$	Exponential function
$\lfloor\cdot\rfloor$	Floor function
$\perp$	Independence symbol
Ω_t	Information up to point in time $t \geq 0$
Γ^{KF}_t	Kalman Filter Gain at time $t \geq 0$
Γ^{KS}_t	Kalman Smoother Gain at time $t \geq 0$
$\otimes$	Kronecker product
$\boldsymbol{F}_t \in \mathbb{R}^K$	K-dimensional vector of hidden factors
$\mathbb{1}_K \in \mathbb{R}^K$	K-dimensional vector of ones
$\mathbf{0}_K \in \mathbb{R}^K$	K-dimensional zero vector
$\operatorname{diag}\left(\boldsymbol{z}\right) \in \mathbb{R}^{K\times K}$	$K \times K$-dimensional diagonal matrix with elements $\boldsymbol{z} \in \mathbb{R}^K$
$I_K \in \mathbb{R}^{K\times K}$	$K \times K$-dimensional identity matrix
$\Sigma^{-1} \in \mathbb{R}^{K\times K}$	$K \times K$-dimensional matrix inverse of non-singular matrix $\Sigma \in \mathbb{R}^{K\times K}$
$\Sigma^{1/2} \in \mathbb{R}^{K\times K}$	$K \times K$-dimensional matrix square root of matrix $\Sigma \in \mathbb{R}^{K\times K}$
$\Sigma' \in \mathbb{R}^{K\times K}$	$K \times K$-dimensional transpose of matrix $\Sigma \in \mathbb{R}^{K\times K}$
$O_K \in \mathbb{R}^{K\times K}$	$K \times K$-dimensional zero matrix
$\mathcal{L}\left(\Theta \mid X\right)$	Log-likelihood function for parameters Θ and sample data X
$\lvert \cdot \rvert$	Matrix determinant
$tr\left(\cdot\right)$	Matrix trace
$\mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right)$	Multivariate normal distribution with parameters $\boldsymbol{\mu}$ and Σ
$ln\left(\cdot\right)$	Natural logarithm
$\boldsymbol{X}_t \in \mathbb{R}^N$	N-dimensional vector of panel data
$f_\Theta\left(\cdot\right)$	Probability density function based on model parameters Θ
$\Phi^{-1}\left(\cdot\right)$	Quantile function of standard normal distribution
Θ	Set of model parameters
$\mathbb{R}_+$	Set of positive real numbers

$\{\boldsymbol{X}_t\}$	Time series of vectors $\boldsymbol{X}_t$
$\mathcal{U}(a, b)$	Univariate uniform distribution on the interval $[a, b]$
$\mathbb{V}\mathrm{ar}_{\Theta}[\cdot]$	Variance / covariance matrix based on model parameters Θ
$vec(\cdot)$	Vectorization operator

List of Figures

List of Tables

Publications and Working Papers

Some results of this thesis have been prepared for publication in peer-reviewed journals:

M. Defend, A. Min, L. Portelli, **F. Ramsauer**, F. Sandrini and R. Zagst: Mixed-frequency information supporting asset allocation decisions. *Working Paper, 2017.*
→ this paper covers the findings of Chapter 3.

M. Defend, A. Min, L. Portelli, **F. Ramsauer**, F. Sandrini and R. Zagst: Estimation of approximate dynamic factor models with incomplete data. *Working Paper, 2017.*
→ this paper considers the content of Chapter 4.

M. Lingauer and **F. Ramsauer**: Estimation of FAVAR models for incomplete data with a modified Kalman Filter. *Working paper, 2017.*
→ this paper summarizes the results of Chapter 5.

Besides my own studies, I contributed to and worked on the following projects:

E. Ivanov, A. Min and F. Ramsauer: Copula-based factor models for multivariate asset returns. *Econometrics 5(2):20, 2017, pp. 1-24.*
→ this paper arose from the master's thesis of E. Ivanov, which was supervised by A. Min and advised by A. Min and me.

M. Escobar, M. Krayzler, **F. Ramsauer**, D. Saunders and R. Zagst: Incorporation of stochastic policyholder behavior in analytical pricing of GMABs and GMDBs. *Risks 4(4):41, 2016, pp. 1-36.*
→ this paper resulted from my master's thesis, which was supervised by M. Escobar, D. Saunders and R. Zagst and advised by M. Krayzler.

J. Hauptmann, A. Hoppenkamps, **A. Min**, F. Ramsauer and R. Zagst: Forecasting market turbulence using regime-switching models. *Financial Markets and Portfolio Management 28(2), 2014, pp. 139-164.*
→ this article is based on the diploma thesis of A. Hoppenkamps, which was supervised by R. Zagst, advised by A. Min and followed a joint project with J. Hauptmann. Here, my contributions are restricted to its revisions by providing alternative model implementations in MATLAB for comparison reasons.

Note, in the above overview all corresponding authors are highlighted in **bold**.

Bibliography

K. Aastveit and T. Trovik. Nowcasting Norwegian GDP: The role of asset prices in a small open economy. *Empirical Economics*, 42(1):95 – 119, 2012.

K. Aastveit, K. Gerdrup, A. Jore, and L. Thorsrud. Nowcasting GDP in real time: A density combination approach. *Journal of Business & Economic Statistics*, 32(1):48 – 68, 2014.

H. Abdi and L. Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433 – 459, 2010.

T. Abeysinghe. Forecasting Singapore's quarterly GDP with monthly external trade. *International Journal of Forecasting*, 14(4):505 – 513, 1998.

H. Akaike. Factor analysis and AIC. *Psychometrika*, 52(3):317 – 332, 1987.

H. Alt. *Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung*. Springer Berlin Heidelberg, 2012.

F. Altissimo, R. Cristadoro, M. Forni, M. Lippi, and G. Veronese. New Eurocoin: Tracking economic growth in real time. *Review of Economics & Statistics*, 92(4):1024 – 1034, 2010.

D. Amengual and M. Watson. Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business & Economic Statistics*, 25(1):91 – 96, 2007.

E. Andreou, E. Ghysels, and A. Kourtellos. Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2):246 – 261, 2010.

E. Andreou, E. Ghysels, and A. Kourtellos. Forecasting with mixed-frequency data. In M. Clements and D. Henry, editors, *The Oxford Handbook on Economic Forecasting. Part II. Data Issues*, pages 225 – 246. 2011.

E. Andreou, E. Ghysels, and A. Kourtellos. Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, 31(2):240 – 251, 2013.

E. Angelini, J. Henry, and M. Marcellino. Interpolation and backdating with a large information set. *Journal of Economic Dynamics and Control*, 30(12):2693 – 2724, 2006.

E. Angelini, M. Bańbura, and G. Rünstler. Estimating and forecasting the Euro area monthly national accounts from a dynamic factor model. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2010(1), 2010.

E. Angelini, G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler. Short-term forecasts of Euro area GDP growth. *The Econometrics Journal*, 14(1):25 – 44, 2011.

M. Armesto, K. Engemann, and M. Owyang. Forecasting with mixed frequencies. *Federal Reserve Bank of St. Louis Review*, 96(6):521 – 536, 2010.

M. Artis, A. Banerjee, and M. Marcellino. Factor forecasts for the UK. *Journal of Forecasting*, 24(4):279 – 298, 2005.

S. Aruoba and F. Diebold. Real-time macroeconomic monitoring: Real activity, inflation, and interactions. *National Bureau of Economic Research. Working Paper No. 15657*, 2010.

S. Aruoba, F. Diebold, and C. Scotti. Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, 27(4):417 – 427, 2009.

J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70 (1):191 – 221, 2002.

J. Bai and S. Ng. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133 – 1150, 2006.

J. Bai and S. Ng. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304 – 317, 2008a.

J. Bai and S. Ng. Large dimensional factor analysis. *Foundations and Trends® in Econometrics*, 3(2): 89 – 163, 2008b.

J. Bai and S. Ng. Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1):18 – 29, 2013.

J. Bai, E. Ghysels, and J. Wright. State space models and MIDAS regressions. *Econometric Reviews*, 32 (7):779 – 813, 2013.

J. Bai, K. Li, and L. Lu. Estimation and inference of FAVAR models. *Journal of Business & Economic Statistics*, 2015.

M. Bańbura and M. Modugno. Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133 – 160, 2014.

M. Bańbura and G. Rünstler. A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, 27(2):333 – 346, 2011.

M. Bańbura, D. Giannone, and L. Reichlin. Nowcasting. In M. Clements and D. Hendry, editors, *The Oxford Handbook on Economic Forecasting. Part II. Data Issues*, pages 193 – 224. 2011.

M. Bańbura, D. Giannone, M. Modugno, and L. Reichlin. Now-casting and the real-time data flow. In G. Elliott and A. Timmermann, editors, *Handbook of Economic Forecasting*, volume 2, Part A, pages 195 – 237. 2013.

M. Bańbura, D. Giannone, and M. Lenza. Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *ECB Working Paper No. 1733. http://ssrn.com/abstract=2491561*, 2014.

K. Barhoumi, O. Darné, and L. Ferrara. Are disaggregate data useful for factor analysis in forecasting French GDP? *Journal of Forecasting*, 29(1-2):132 – 144, 2010.

M. Barigozzi, M. Lippi, and M. Luciani. Dynamic factor models, cointegration, and error correction mechanisms. *http://ssrn.com/abstract=2402185*, 2014.

A. Benkwitz, H. Lütkepohl, and J. Wolters. Comparison of bootstrap confidence intervals for impulse responses of German monetary systems. *Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes. Discussion Paper No. 1999/29*, 1999.

B. Bernanke and J. Boivin. Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3):525 – 546, 2003.

B. Bernanke, J. Boivin, and P. Eliasz. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, 120(1):387 – 422, 2005.

F. Black and A. Perold. Theory of constant proportion portfolio insurance. *Journal of Economic Dynamics and Control*, 16(3):403 – 426, 1992.

J. Boivin and M. Giannoni. Global forces and monetary policy effectiveness. *National Bureau of Economic Research. Working Paper No. 13736*, 2008.

J. Boivin and S. Ng. Understanding and comparing factor-based forecasts. *International Journal of Central Banking*, 1(3):117 – 152, 2005.

J. Boivin and S. Ng. Are more data always better for factor analysis? *Journal of Econometrics*, 132(1): 169 – 194, 2006.

J. Boivin, M. Giannoni, and D. Stevanovic. Monetary transmission in a small open economy: More data, fewer puzzles. *Columbia University. Working Paper*, 2010.

L. Bork. Estimating US monetary policy shocks using a factor-augmented vector autoregression: An EM algorithm approach. *http://ssrn.com/abstract=1358876*, 2009.

L. Bork. A large-dimensional factor analysis of the Federal Reserve's large-scale asset purchases. *Aalborg University. Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2618378*, 2015.

L. Bork, H. Dewachter, and R. Houssa. Identification of macroeconomic factors in large panels. *Center for Research in the Economics of Development. Working Paper No. 2010/10*, 2010.

E. Brechmann and C. Czado. COPAR-multivariate time series modeling using the copula autoregressive model. *Applied Stochastic Models in Business and Industry*, 31(4):495 – 514, 2015.

J. Breitung and S. Eickmeier. Dynamic factor models. *Allgemeines Statistisches Archiv*, 90(1):27 – 42, 2006.

P. Brockwell and R. Davis. *Introduction to time series and forecasting.* Springer, 2002.

G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281 – 1304, 1983.

C. Chiu, B. Eraker, A. Foerster, T. Kim, and H. Seoane. Estimating VAR's sampled at mixed or irregular spaced frequencies: A Bayesian approach. *The Federal Reserve Bank of Kansas City. Research Working Papers*, 2012.

M. Clements and A. Galvão. Macroeconomic forecasting with mixed-frequency data. *Journal of Business & Economic Statistics*, 26(4):546 – 554, 2008.

T. Crone and A. Clayton-Matthews. Consistent economic indexes for the 50 states. *The Review of Economics and Statistics*, 87(4):593 – 603, 2005.

D. Croushore. Chapter 17 forecasting with real-time macroeconomic data. volume 1 of *Handbook of Economic Forecasting*, pages 961 – 982. 2006.

P. De Jong. Smoothing and interpolation with the state-space model. *Journal of the American Statistical Association*, 84(408):1085 – 1088, 1989.

P. De Jong and M. Mackinnon. Covariances for smoothed estimates in state space models. *Biometrika*, 75(3):601 – 602, 1988.

C. De Mol, D. Giannone, and L. Reichlin. Forecasting using a large number of predictors: Is Bayesian regression a valid alternative to principal components? *ECB Working Paper No. 700. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=947129*, 2006.

C. De Mol, D. Giannone, and L. Reichlin. Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2):318 – 328, 2008.

A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1 – 38, 1977.

C. Doz, D. Giannone, and L. Reichlin. A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164(1):188 – 205, 2011.

C. Doz, D. Giannone, and L. Reichlin. A quasi-maximum likelihood approach for large, approximate dynamic factor models. *The Review of Economics and Statistics*, 94(4):1014 – 1024, 2012.

R. Durrett. *Probability: Theory and examples.* Cambridge University Press, 2010.

M. Evans. Where are we now? Real-time estimates of the macro economy. *National Bureau of Economic Research. Working Paper No. 11064*, 2005.

M. Forni and M. Lippi. The generalized dynamic factor model: Representation theory. *Econometric Theory*, 17(6):1113 – 1141, 2001.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, 82(4):540 – 554, 2000.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model consistency and rates. *Journal of Econometrics*, 119(2):231 – 255, 2004.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, 100(471):830 – 840, 2005.

M. Forni, D. Giannone, M. Lippi, and L. Reichlin. Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, 25(5):1319 – 1347, 2009.

N. Francis, E. Ghysels, and M. Owyang. The low-frequency impact of daily monetary policy shocks. *Federal Reserve Bank of St. Louis. Working Paper Series 2011-009B. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1793358*, 2011.

A. Galvão. Changes in predictive ability with mixed frequency data. *International Journal of Forecasting*, 29(3):395 – 410, 2013.

Z. Ghahramani and G. Hinton. Parameter estimation for linear dynamical systems. *University of Toronto. Working paper. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.55.5997&rep=rep1 &type=pdf*, 1996.

E. Ghysels. Macroeconomics and the reality of mixed frequency data. *http://ssrn.com/abstract =2069998*, 2015.

E. Ghysels and J. Wright. Forecasting professional forecasters. *Journal of Business & Economic Statistics*, 27(4):504 – 516, 2009.

E. Ghysels, P. Santa-Clara, and R. Valkanov. The MIDAS touch: Mixed data sampling regression models. *http://escholarship.org/uc/item/9mf223rs*, 2004.

E. Ghysels, P. Santa-Clara, and R. Valkanov. There is a risk-return trade-off after all. *Journal of Financial Economics*, 76(3):509 – 548, 2005.

E. Ghysels, P. Santa-Clara, and R. Valkanov. Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1-2):59 – 95, 2006.

E. Ghysels, A. Sinko, and R. Valkanov. MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1):53 – 90, 2007.

D. Giannone, L. Reichlin, and L. Sala. Monetary policy in real time. In M. Gertler and K. Rogoff, editors, *NBER Macroeconomics Annual*, pages 161 – 200. 2004.

D. Giannone, L. Reichlin, and D. Small. Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665 – 676, 2008.

D. Giannone, L. Reichlin, and S. Simonelli. Nowcasting Euro area economic activity in real time: The role of confidence indicators. *National Institute Economic Review*, 210(1):90 – 97, 2009.

T. Gneiting and A. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359 – 378, 2007.

G. Golub and C. Van Loan. *Matrix computations*. Johns Hopkins University Press, 1996.

W. Greene. *Econometric analysis*. Prentice Hall, 2003.

M. Hallin and R. Liška. Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603 – 617, 2007.

J. Hamilton. *Time series analysis*. Princeton University Press, 1994.

A. Harvey and R. Pierse. Estimating missing observations in economic time series. *Journal of the American Statistical Association*, 79(385):125 – 131, 1984.

D. Harville. *Matrix algebra from a statistician's perspective*. Springer New York, 1997.

J. Hauptmann, A. Hoppenkamps, A. Min, F. Ramsauer, and R. Zagst. Forecasting market turbulence using regime-switching models. *Financial Markets and Portfolio Management*, 28(2):139 – 164, 2014.

S. Haykin. Kalman filters. In S. Haykin, editor, *Kalman Filtering and Neural Networks*, pages 1 – 21. 2002.

J. Hogrefe. Forecasting data revisions of GDP: A mixed frequency approach. *AStA Advances in Statistical Analysis*, 92(3):271 – 296, 2008.

N. Hyung and C. Granger. Linking series generated at different frequencies. *Journal of Forecasting*, 27 (2):95 – 108, 2008.

J. Jobson and B. Korkie. Performance hypothesis testing with the Sharpe and Treynor measures. *The Journal of Finance*, 36(4):889 – 908, 1981.

I.T. Jolliffe. *Principal component analysis*. Springer, 2002.

B. Jungbacker and S. Koopman. Likelihood-based analysis for dynamic factor models. *Tinbergen Institute. Discussion Paper No. 08-007/4*, 2008.

B. Jungbacker and S. Koopman. Likelihood-based dynamic factor analysis for measurement and forecasting. *The Econometrics Journal*, 18(2):C1 – C21, 2015.

B. Jungbacker, S. Koopman, and M. van der Wel. Dynamic factor analysis in the presence of missing data. *Tinbergen Institute. Discussion Paper No. 09-010/4*, 2009.

B. Jungbacker, S. Koopman, and M. van der Wel. Maximum likelihood estimation for dynamic factor models with missing data. *Journal of Economic Dynamics and Control*, 35(8):1358 – 1368, 2011.

R. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35 – 45, 1960.

L. Kilian. Small-sample confidence intervals for impulse response functions. *Review of Economics and Statistics*, 80(2):218–230, 1998.

V. Kuzin, M. Marcellino, and C. Schumacher. MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro area. *International Journal of Forecasting*, 27(2):529 – 542, 2011.

V. Kuzin, M. Marcellino, and C. Schumacher. Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries. *Journal of Applied Econometrics*, 28(3):392 – 411, 2013.

K. Lahiri and G. Monokroussos. Nowcasting US GDP: The role of ISM business surveys. *International Journal of Forecasting*, 29(4):644 – 658, 2013.

H. Liu and S. Hall. Creating high-frequency national accounts with state-space modelling: A Monte Carlo experiment. *Journal of Forecasting*, 20(6):441 – 449, 2001.

M. Luciani. Forecasting with approximate dynamic factor models: The role of non-pervasive shocks. *http://ssrn.com/abstract=1925807*, 2011.

M. Luciani. Large-dimensional dynamic factor models in real-time: A survey. *http://ssrn.com/abstract =2511872*, 2014.

M. Luciani and L. Ricci. Nowcasting Norway. *International Journal of Central Banking*, 10(4):215 – 248, 2014.

H. Lütkepohl. *New introduction to multiple time series analysis.* Springer Science & Business Media, 2005.

M. Marcellino and C. Schumacher. Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP*. *Oxford Bulletin of Economics and Statistics*, 72(4):518 – 550, 2010.

M. Marcellino and V. Sivec. Monetary, fiscal and oil shocks: Evidence based on mixed frequency structural FAVARs. *Journal of Econometrics*, 193(2):335 – 348, 2016.

M. Marcellino, M. Porqueddu, and F. Venditti. Short-term GDP forecasting with a mixed frequency dynamic factor model with stochastic volatility. *Bank of Italy Temi di Discussione. Working Paper No. 896. http://ssrn.com/abstract=2221327*, 2013.

R. Mariano and Y. Murasawa. A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18(4):427 – 443, 2003.

R. Mariano and Y. Murasawa. A coincident index, common factors, and monthly real GDP*. *Oxford Bulletin of Economics and Statistics*, 72(1):27 – 46, 2010.

M. Modugno. Nowcasting inflation using high frequency data. *ECB Working Paper No. 1324. http://ssrn.com/abstract=1797802*, 2011.

L. Nunes. Nowcasting quarterly GDP growth in a monthly coincident indicator model. *Journal of Forecasting*, 24(8):575 – 592, 2005.

T. Proietti. Temporal disaggregation by state space methods: Dynamic regression methods revisited. *Econometrics Journal*, 9(3):357–372, 2006.

T. Proietti and F. Moauro. Dynamic factor analysis with non-linear temporal aggregation constraints. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(2):281 – 300, 2006.

C. Rao and H. Toutenburg. *Linear models: Least squares and alternatives.* Springer, 1999.

R. Reis and M. Watson. Relative goods' prices, pure inflation, and the Phillips correlation. *National Bureau of Economic Research. Working Paper No. 13615*, 2007.

T. Roncalli. *Introduction to risk parity and budgeting.* Taylor & Francis, 2013.

D. Rubin and D. Thayer. EM algorithms for ML factor analysis. *Psychometrika*, 47(1):69 – 76, 1982.

D. Ruppert. *Statistics and data analysis for financial engineering.* Springer, 2011.

S. Särkkä, A. Vehtari, and J. Lampinen. Time series prediction by Kalman smoother with cross-validated noise density. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 2, pages 1653 – 1657, 2004.

F. Schmid and R. Schmidt. Statistical inference for performance measure Omega. *Working Paper, https://www.researchgate.net/profile/Friedrich_Schmid/publication/228466525_Statistical_Inference_for_Performance_Measure_Omega/links/0deec528d077026795000000.pdf*, 2008.

H. Schneeweiss and H. Mathes. Factor analysis and principal components. *Journal of Multivariate Analysis*, 55(1):105 – 124, 1995.

F. Schorfheide and D. Song. Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics*, 33(3):366 – 380, 2015.

C. Schumacher and J. Breitung. Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, 24(3):386 – 398, 2008.

R. Shumway and D. Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):253 – 264, 1982.

J. Stock and M. Watson. Diffusion indices. *National Bureau of Economic Research. Working Paper No. 6702. Revised Version.*, 1999a.

J. Stock and M. Watson. Forecasting inflation. *Journal of Monetary Economics*, 44(2):293 – 335, 1999b.

J. Stock and M. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167 – 1179, 2002a.

J. Stock and M. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147 – 162, 2002b.

J. Stock and M. Watson. Implications of dynamic factor models for VAR analysis. *National Bureau of Economic Research. Working Paper No. 11467*, 2005.

J. Stock and M. Watson. Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, pages 173 – 205, 2009.

J. Stock and M. Watson. Dynamic factor models. *Oxford Handbook of Economic Forecasting. Oxford University Press, USA*, pages 35 – 59, 2011.

M. Tipping and C. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611 – 622, 1999.

M. Watson and R. Engle. Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23(3):385 – 400, 1983.

K. Wohlrabe. Forecasting with mixed-frequency time series models. *Ludwig-Maximilians-Universität München. Dissertation*, 2008.

C. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95 – 103, 1983.

Y. Yamamoto. Bootstrap inference for impulse response functions in factor-augmented vector autoregressions. *Hi-Stat Discussion Paper*, 2012.

Index