TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Betriebswirtschaftslehre - Finanzmanagement und Kapitalmärkte

# Incorporating model uncertainty into the variable selection problem of expected return proxies

Christoph Klaus Jäckel

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

Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

genehmigten Dissertation.

| | |
|---|---|
| Vorsitzende: | Univ.-Prof. Dr. Alwine Mohnen |
| Prüfer der Dissertation: | 1. Univ.-Prof. Dr. Christoph Kaserer |
| | 2. Univ.-Prof. Dr. Markus Glaser |

Die Dissertation wurde am 25.02.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Wirtschaftswissenschaften am 15.07.2014 angenommen.

Dedicated to my late father.

# ABSTRACT

Expected stock returns are one of the most relevant variables in finance, both among practitioners and academics. Unfortunately, they are unobservable and thus suitable proxies have to be found to approximate them. Realized returns are by far the most common proxy, from which estimates are either inferred directly or indirectly via factor-based asset pricing models or predictive regressions. Their success stems from their wide availability, observability and asymptotic unbiasedness. However, they have one major shortcoming, namely the large noise induced by information surprises, or news, that clouds the true underlying process of expected returns that one is ultimately interested in. As a solution to this problem, alternative expected return proxies have been proposed with substantially lower variation than realized returns. They are forward looking in nature and therefore, at least in theory, unaffected by any news. They rely on earnings forecasts, CDS spreads, and corporate bond yields. Due to their substantially lower standard deviation as compared to realized returns, they allow a much sharper statistical inference and deliver very robust results. In other words, parameter uncertainty is greatly reduced, which has helped to identify seemingly robust relations between expected returns and factors that would otherwise be overshadowed by noise when realized returns are used.

In this thesis, I argue that the results of these alternative expected return estimates may be driven, at least partly, by the ignorance of model uncertainty. This uncertainty is introduced because there are many alternative specifications and it is difficult to ascertain which is the correct one. As a consequence, results may be due to the selection of a proxy for which the measurement error happens to be associated with the variable of interest, and not the true underlying expected return process. By contrast, there is only one specification for realized returns.

In the theoretical part of this thesis, I first show that model uncertainty is a relevant issue for the most prominent alternative proxy, the implied cost of capital (ICC). For this proxy, a multitude of different specifications exist. Although they all have in common that they are defined as the internal rate of return that equates a vector of all fu-

ture expected dividends with the current stock price, they differ in their assumptions about expected dividends. I introduce these specifications in a detailed literature review.

Next, I present at length studies that deal with the issue of selecting the best specification among the available expected return proxies. I argue that none of the current approaches is able to generate a confident selection. We simply do not have a reliable method to identify the best alternative proxy, due to the fact that all of these evaluation approaches rely on noisy realized returns. In particular, I criticize one approach that tries to solve the noisiness by adding additional news proxies. In this approach proxies are defined inconsistently. Following this discussion in my thesis, it is clear that one faces great uncertainty in the selection process.

To deal with this problem, I introduce a Bayesian model averaging approach that directly incorporates uncertainty into the statistical inference. This approach allows someone to condition on the information set including all expected return proxies while conducting inferences, as opposed to conditioning on a single proxy. It thereby compares all proxies simultaneously to the extent to which they are able to explain subsequent realized returns. Put differently, the weight of a proxy, which denotes the posterior belief one has in its quality, is higher, the better it is able to explain subsequent realized returns. This external validation prevents the problem of "measurement error optimization", in which someone is choosing a proxy, either intentionally or otherwise, because the measurement error of this proxy is associated with the variable of interest, and not true expected returns. However, this directly implies that the evaluation of any alternative proxy relies again on noisy realized returns, a circularity argument that I am able to work out.

The empirical part of this thesis starts with an extensive description of my US data set which implements eight commonly used specifications of the ICC approach. I present evidence that there are noteworthy differences between the different specifications, both on an aggregate and on a firm-level. After showing that one cannot reliably identify the best of these eight ICC specifications, I apply the Bayesian model averaging approach to three research questions to average the evidence across those specifications. Taken together, the evidence from this empirical exercise suggests that the incorporation of model uncertainty is important to prevent the reporting of biased and overconfident results, but it also shows that alternative proxies can still yield additional insight.

In summary, my thesis levels the playing field between realized returns, with large parameter uncertainty and no model uncertainty, and alternative proxies, with typically modest parameter uncertainty, but potentially large model uncertainty.

# PUBLICATIONS

This dissertation is based on papers that I wrote during my PhD, some of them together with coauthors. The Bayesian model averaging approach is based on Jäckel (2014). Also, the parts of the dissertation that deal with the evaluation method of Easton and Monahan (2005) are mostly taken from Jäckel (2013). From both of these papers, I adopt parts freely. Furthermore, similar ideas, arguments, tables and figures have appeared previously in Jäckel and Mühlhäuser (2011), Jäckel, Kaserer, and Mühlhäuser (2013), and Hanauer, Jäckel, and Kaserer (2013).

## CONTENTS

LIST OF FIGURES

# LIST OF TABLES

# NOMENCLATURE

| | |
|---|---|
| $\Delta d_{t+1}$ | log dividend growth between time $t$ and time $t+1$ |
| $\Delta$ | quantity or parameter of interest, such as a slope coefficient |
| $\epsilon_t$ | normally distributed mean zero shock at time $t$ |
| $\gamma$ | slope coefficient between $\mu_t$ and a variable of interest $x_t$ |
| $\kappa$ | parameter of linearization, defined as $-\log(\rho) - (1-\rho)\log(1/\rho - 1)$ |
| $\mu_{i,t}$ | short notation for the expected return next period, $E_t[r_{t+1}]$, for asset $i$ |
| $\omega_t$ | regression coefficient of a pooled regression of the $roe$ on its lagged value (Easton and Monahan (2005) analysis) |
| $\overline{x}$ | overline indicates a posterior value of a parameter $x$ |
| $\phi$ | shrinkage parameter to control informativeness of prior |
| $\rho$ | parameter of linearization, defined as $1/(1+\exp(\overline{d-p}))$ |
| $\sigma_x$ | standard deviation of random variable $x$ |
| $\underline{x}$ | underline indicates a prior value of a parameter $x$ |
| $\widehat{x}$ | hat indicates an empirically observable proxy of the latent variable $x$ or empirical estimator of fixed, but unknown parameter $x$ |
| $AGR_t$ | abnormal earnings growth in period $t$ |
| $b_{SS}$ | steady-state plowback rate |
| $b_t$ | plowback rate in period $t$ |
| $BPS_t$ | book value per share at time $t$ |
| $CF_{j,k}$ | part of capital gain returns $Retx_j$ that is explained by cash flow news, subject to ICC method $k$ |
| $CFN_{t+1}$ | log cash flow news arriving in period $t+1$ |
| $Cov(x,y)$ | covariance between random variables $x$ and $y$ |

| | |
|---|---|
| $D$ | data |
| $D_t$ | dividend, paid at time $t$ |
| $d_t$ | log dividend, paid at time $t$ |
| $DPS_t$ | dividends per share, paid at time $t$ |
| $DR_{j,k}$ | part of capital gain returns $Retx_j$ that is explained by discount rate news, subject to ICC method $k$ |
| $DRN_{t+1}$ | log discount rate news arriving in period $t+1$ |
| $E_t[\cdot]$ | expectation operator, where the expectations are conditional on information available at time $t$ |
| $EPS_t$ | earnings per share at time $t$ |
| $froe_{j,m}$ | forecasted log ROE for fiscal year $m$, computed with log earnings forecasts $eps_m$ made in December of year $j$ (Easton and Monahan (2005) analysis) |
| $g_{AGR}$ | long-term abnormal earnings growth rate |
| $g_{lt}$ | long-term earnings growth rate |
| $g_{SS}$ | steady-state earnings growth rate |
| $g_{st}$ | short-term earnings growth rate |
| $g_t$ | earnings growth rate in period $t$ |
| $h$ | error precision, i.e., $h \equiv \sigma^{-2}$ |
| $I_N$ | $N \times N$ identity matrix |
| $Ltg_t$ | empirical long-term earnings growth rate provided by IBES analyst forecasts at time $t$ |
| $M_k$ | model $k$ |
| $p(x)$ | marginal probability distribution of random variable $x$ |
| $P_t$ | price of an asset at time $t$ |
| $PO_t$ | payout ratio at time $t$ |
| $R_t^e$ | implied cost of equity capital at time $t$ |
| $R_t^k$ | implied cost of equity capital at time $t$, according to the ICC method $k$ |

| | |
|---|---|
| $R_t^{CT}$ | implied cost of equity capital at time $t$, according to the CT method |
| $R_t^{GLS}$ | implied cost of equity capital at time $t$, according to the GLS method |
| $R_t^{MPEG}$ | implied cost of equity capital at time $t$, according to the MPEG method |
| $R_t^{OJ}$ | implied cost of equity capital at time $t$, according to the OJ method |
| $R_t^{PEG}$ | implied cost of equity capital at time $t$, according to the PEG method |
| $R_t^{PE}$ | implied cost of equity capital at time $t$, according to the PE method |
| $R_t^{PSS}$ | implied cost of equity capital at time $t$, according to the PSS method |
| $R_{t+1}$ | realized return on an asset that is held from the end of period $t$ to the end of period $t+1$ |
| $r_{t+1}$ | log or continuously compounded realized return on an asset that is held from the end of period $t$ to the end of period $t+1$ |
| $Retx_j$ | capital gain returns between time $t+j$ and $t$ |
| $s$ | scaling factor |
| $SSE$ | sum of squared errors |
| $u_{t+1}$ | news arriving in period $t+1$, i.e., $u_{t+1} = CFN_{t+1} + DRN_{t+1}$ |
| $Var(x)$ | variance of random variable $x$ |
| $veps_k^t$ | vector of all future earnings, as assumed by ICC method $k$ and expected at time $t$ |
| $Vol_t$ | measure for stock market volatility, either the annualized variance or standard deviation of the daily value-weighted market returns from CRSP for period $t$ |
| $w_{i,t}$ | measurement error at time $t$ for the expected return of asset $i$ |

# ACRONYMS

| | |
|---|---|
| PE | implied cost of capital method based on the price-to-forward-earnings ratio |
| PEG | implied cost of capital method based on the PEG ratio, i.e., the price-earnings ratio divided by the short-term earnings growth rate |
| GLS | implied cost of capital method based on Gebhardt, Lee, and Swaminathan (2001) |
| OJ | implied cost of capital method based on Ohlson and Juettner-Nauroth (2005) and Gode and Mohanram (2003) |
| MPEG | implied cost of capital method based on the modified PEG ratio, proposed in Easton (2004) |
| CT | implied cost of capital method based on Claus and Thomas (2001) |
| PSS | implied cost of capital method based on Pástor, Sinha, and Swaminathan (2008) |
| CDZ | implied cost of capital method based on Chen, Da, and Zhao (2013) |
| Amex | American Stock Exchange |
| AIC | Akaike information criterion |
| AR | autoregressive |
| BMA | Bayesian model averaging |
| BIC | Bayesian information criterion |
| CAPM | capital asset pricing model |
| CDS | credit default swap |
| CF | cash flow |
| cf. | confer |
| CRSP | Center for Research in Security Prices |

| | |
|---|---|
| CS | Campbell and Shiller (1988) |
| DR | discount rate |
| DDM | dividend discount model |
| e.g. | exempli gratia |
| EM | Easton and Monahan (2005) |
| etc. | et cetera |
| f. | and the following |
| GDP | gross domestic product |
| GAAP | Generally Accepted Accounting Principles |
| ICC | implied cost of capital |
| IBES | Institutional Brokers' Estimate System |
| i.e. | id est |
| IPO | initial public offering |
| LHS | left-hand side |
| NASDAQ | National Association of Securities Dealers Automated Quotation |
| NYSE | New York Stock Exchange |
| OLS | ordinary least squares |
| PP&E | property, plant and equipment |
| FCFE | free cash flow to equity |
| RHS | right-hand side |
| ROE | return on equity |
| Std. dev. | standard deviation |
| US | United States of America |
| VAR | vector autoregression |
| WRDS | Wharton Research Data Services |
| YTM | yield to maturity |

# 1

# INTRODUCTION

## 1.1 MOTIVATION

According to Cochrane (2005), all asset pricing theory is founded on one simple concept: the price of an asset equals its expected discounted payoffs. However, the expected rate of return with which payoffs are discounted is unobservable and, as a consequence, a lot of effort has been put into answering the question of how to measure expected returns, both in the cross-section and the time series.

A satisfying answer to this question is of great importance to practitioners and academics alike. Chief Financial Officers have to know what rate of return their investors require to evaluate the gains from a project. Regulators need to understand what impact policy changes have on expected returns. Portfolio managers require expected returns as input parameters to compute the allocation weights for different asset classes and assets within these classes. Finally, financial economists want to understand what factors drive expected returns both over time and different assets.

To date, the most prominent approach to approximate expected returns for stocks, on which this thesis focuses, is the use of realized returns. Their popularity stems from the fact that they are an unbiased and observable proxy for expected returns next period. Over the last decades, some of the greatest breakthroughs and puzzles in financial economics have been reported based on evidence with realized returns. For example, Shiller (1981) showed that *realized* returns move too much to be reconcilable with the relatively smooth dividend process over time and constant *expected* discount rates, which was the standard assumption prior to his study.[1] This finding, for which Shiller was awarded the Nobel Prize in Economics in 2013, is the foundation of the literature on return predictability, which tries to measure aggregate time-varying return and cash flow expectations

1 Throughout this dissertation, I use the terms "returns", "cost of (equity) capital", and "discount rates" interchangeably.

via predictive regressions (cf. Kelly and Pruitt 2013).[2] Furthermore, empirical tests showed that the cross-section of *realized* returns cannot be explained by the capital asset pricing model (CAPM), which is a one-factor asset pricing model proposed by Sharpe (1964), Lintner (1965) and Mossin (1966) and meant to explain cross-sectional differences in *expected* returns. This has led to the development of many alternative multifactor models such as the three-factor model by Fama and French (1993). Moreover, the difference between the historical return on the US stock market and a riskless rate, the return for Treasury bills, was 6.9 percentage points over a 110 year period (cf. Mehra 2008). This large equity premium cannot be rationalized by standard economic models such as the consumption-based CAPM. This finding was dubbed as a "puzzle" by Mehra and Prescott (1985) and has sparked the development of many alternative theoretical asset pricing models such as the Campbell and Cochrane (1999) habit model and the long-run risks model by Bansal and Yaron (2004).

However, as Elton (1999) points out, realized returns as an estimate of expected returns only track latent expected returns with additional noise that is induced by information surprises in both expected dividends and returns. While these shocks are unbiased asymptotically, they can be biased in sample. As a consequence, the aforementioned findings could be spurious and could simply be due to the noise in realized returns, and not the true underlying process of expected returns. Take the example of the equity risk premium puzzle. If we take the average historical mean as the true mean of expected returns during this period, the difference between these returns and the risk-free rate is indeed hard to reconcile with standard economic theory. But there is also an alternative explanation that the expected returns were much lower and only the noise term, although zero asymptotically, was positive for this period. This is exactly the point made by Fama and French (2002), who argue that the very high stock returns of the latter half of the 20th century were due to an unexpected series of negative discount rate news.[3]

2 Cochrane (2005), Cochrane (2011), and Koijen and Van Nieuwerburgh (2011) are excellent reviews of the developments in the asset pricing literature in general and return predictability in particular over the last decades.

3 As shown later in Chapter 3, negative discount rate news translates to positive realized returns.

It is therefore not surprising that alternative measures of expected returns have been proposed.[4] All of these proxies for expected returns have one thing in common, that is, that they are all derived from an underlying theoretical model that links unobservable expected returns to observable data. Subject to further assumptions to make the theoretical model empirically tractable, the expected returns can then be backed out. One class of proxies links equity and credit markets to obtain conditional expected equity returns from either bond yield spreads or credit default swap (CDS) spreads. Recent examples of this approach are Campello, Chen, and Zhang (2008), Berg and Kaserer (2013), and Friewald, Wagner, and Zechner (2013). More established, however, is the implied cost of capital (ICC), which is defined as the discount rate that equates the current stock price of a firm with expected future dividends, for example approximated by analyst forecasts.[5] The ICC has found widespread use in both asset pricing and corporate finance applications.[6]

Because these alternative expected return measures are forward looking in nature and take expectations directly into account, they are not plagued with the large shocks induced by information surprises that drive realized returns. Hence, these measures are less volatile than realized returns and it is also argued that they are not affected by a possible in-sample correlation of the shocks. It is therefore commonly believed that these proxies offer fresh insights into questions concerning the economic drivers of expected returns (cf. Campello, Chen, and Zhang 2008).

Yet, these proxies have their own Achilles heel. While realized returns are directly observable without any further assumptions, which implies that they yield the correct results asymptotically, the same is not true for any other alternative proxy. Any error we make both in the assumptions of their underlying theoretical models and in the implementations to infer them empirically can bias the results. There

4 Throughout this dissertation, I use the terms "proxy" and "measure" interchangeably. Furthermore, I often refer to a "method" or "specification" from which the "proxy" or "measure" is derived from.

5 More correctly, it should be called the implied cost of equity capital, but the term equity is typically ignored in the literature. I follow this procedure throughout this dissertation as well.

6 For example, it has been applied in studies about the risk-return tradeoff (Pástor, Sinha, and Swaminathan 2008), the effect of cross-listings in the US (Hail and Leuz 2009), the relation between default risk and expected stock returns (Chava and Purnanandam 2010), and the impact of accounting quality on the cost of capital (Francis et al. 2004). Mühlhäuser (2013) gives a recent and extensive literature review in Chapter 2 of her dissertation. The empirical results in my dissertation are based exclusively on expected return estimates that are inferred from the ICC.

are now dozens of alternative implementations, all yielding different results, although they all want to measure the same underlying process. Of course, it directly follows that at best one of these proxies measures expected returns correctly and that at least all but one of them are measured with, potentially large, error. This measurement error could also lead to spurious results, just as in the case of realized returns, and therefore it is important to evaluate the quality of these alternative expected return measures.

First attempts in this direction have been made (cf., e.g., Easton and Monahan 2005 and Lee, So, and Wang 2011). In short, it can be shown that a proxy that explains subsequent realized returns better is tracking expected returns better, at least asymptotically. One part of this dissertation motivates these evaluation methods and criticizes certain assumptions made by them. In particular, I argue that the approach by Easton and Monahan (2005) should be regarded with suspicion.

Moreover, these approaches all have in common that they rely on realized returns again. Consequently, the evaluation tests are subject to the same points of criticism brought forward against realized returns. In particular, due to the large shocks to realized returns, an evaluation of any alternative proxy is notoriously difficult and a researcher faces great uncertainty about which proxy he should use.

This is a variable selection problem, which is part of a bigger problem in empirical research. Whenever a researcher tests many alternative specifications to derive his final model, but acts as if his final model is the only reasonable specification, he ignores the uncertainty inherent in this selection process. In his seminal work, Leamer (1978, p. 1) describes the problem for the classical linear regression model as follows:

> "Data mining," "fishing," "grubbing," "number crunching." These are the value-laden terms we use to disparage each other's empirical work with the linear regression model. A less provocative description would be "specification searching," and a catch-all definition is "the data-dependent process of selecting a statistical model." This definition encompasses both the estimation of different regression equations with different sets of explanatory variables and also the estimation of a single equation using different subsets of the data.

Later on, he describes the consequences of such specification searches:

> In general, the consequence of a specification search is what you might expect. There is greater uncertainty over parameters than is suggested by the final specification. [...] [T]he equation is estimated as if the specification were given, whereas the very fact that a search occurred reveals that there is uncertainty over the specification. Loosely speaking, the apparent statistical evidence implied by the final equation must be discounted; the greater the range of search, the greater must be the discount.[7]

Applied to the problem of choosing from a large set of expected return proxies, it could be that a researcher stumbles upon seemingly statistically significant results by simply trying many alternative proxies. In this case, the results would be driven by the association of the measurement error and the variable of interest, and not by the association of true, but latent expected returns and the variable of interest, which a researcher is eventually interested in. In brief, the results based on alternative proxies can be spurious as well, in this case even for large samples.

Fortunately, Leamer (1978) does not only describe the problem, but he also tackles it. Using a framework rooted in Bayesian statistics,[8] he proposes an intuitive solution: Bayesian model averaging (BMA). Instead of basing the inference on one model and acting as if this model is specified correctly, one can simply average the evidence across all models under consideration. Thereby, the model weights are positively related to the relative performance of each model. It has been shown that results obtained from such a BMA approach have better performance than the prevailing state-of-the-art method (cf. Raftery and Zheng 2003). Furthermore, BMA typically leads to better out-of-sample predictive power (cf., e.g., Raftery, Madigan, and Hoeting 1997), which is "particularly important because these are situations in which the model assumptions underlying BMA and other methods do not necessarily hold, and they provide a neutral criterion for comparing methods" (Raftery and Zheng 2003, p. 931).

In many research questions, the number of models under consideration is very large. For example, a linear regression model with 20 potential predictors has $2^{20} = 1,048,576$ different specifications. Until very recently, it was therefore not possible to implement a BMA

7 Cf. Leamer (1978, p. 12f.).
8 For an introduction to Bayesian statistics the reader is referred to Appendix A.4.

practically due to the computational restrictions. But recent theoretical and technological advancements have enabled researchers to overcome the difficulties related to implementing BMA (cf. Hoeting et al. 1999). Since then, BMA has found widespread use in empirical research.[9]

The main contribution of this dissertation is to map the variable selection problem that a researcher faces when dealing with alternative expected return proxies into a BMA setup. I treat each proxy as a separate model and show that many of the difficulties typically encountered with BMA are not an issue in this case because the number of models is manageable and the prior specification straightforward. The main difference to my implementation of BMA in contrast to other applications is that I compute the model weights based on the association between the expected return proxy and subsequent realized returns, and not based on the association between the proxy and the variable of interest. This requirement is necessary to prevent a researcher from finding spurious results due to "optimizing" the measurement error process, instead of reporting the true relation between his variable of interest and expected returns. I provide simulation-based evidence that the BMA outperforms current procedures to select from a set of expected return proxies. Furthermore, I apply the BMA approach to three research questions that have been previously examined with the help of the ICC, but only for one specific ICC method. Consequently, model uncertainty has been ignored in the original studies. I show that in some cases this ignorance can lead to uncertainty bands that are too narrow, which can result in misleading conclusions. In other cases, however, in which the results are similar across each ICC specification, model uncertainty is only of second-order importance and alternative proxies do indeed provide fresh insights due to their lower parameter uncertainty in comparison to realized returns. In summary, this dissertation helps in leveling the playing field between realized returns on the one hand and alternative proxies on the other hand, while the latter is unjustly favored in previous research that ignores the issue of model uncertainty.

9 Moral-Benito (2012) provides an overview over studies that apply BMA in economics. For studies in finance, see for example Pástor and Stambaugh (1999), Cremers (2002), Avramov (2002), and Binsbergen, Hueskes, et al. (2013).

## 1.2 OUTLINE

In Chapter 2, I introduce the implied cost of capital as the most prominent proxy class for expected returns. In the first part of this chapter, I present studies that examine the theoretical link between the true ICC on the one hand and true time-varying expected returns next period on the other hand. Next, I focus on the issue of inferring the ICC from the data. To do so, certain simplifying assumptions are necessary, which lead to measurement error in empirical ICC specifications. I introduce the most prominent methods to estimate the ICC. Finally, I briefly present empirical studies that use the ICC as their measure for expected returns and discuss how these studies deal with the problem of model uncertainty, i.e., the uncertainty of not knowing which of the many ICC specifications the correct one is.

Chapter 3 discusses previous studies that try to deal with this model uncertainty by identifying the best expected return measure from a set of proxies. It turns out that the only reasonable way to evaluate proxies is via a comparison of subsequent realized returns and the expected return proxies. Hence, we need realized returns to evaluate the proxies that are meant to replace these realized returns. I explain the approaches of Easton and Monahan (2005) and Lee, So, and Wang (2011) in detail and conclude the chapter by noting that all of the proposed approaches cannot solve the problem of identifying the best proxy with confidence. Therefore, a researcher faces great uncertainty in the selection process.

Chapter 4 introduces the BMA approach to deal with this uncertainty in the proxy selection problem. I link the computation of the proxy weights that are necessary to average across the evidence based on each proxy with the evaluation approaches discussed in Chapter 3. Equipped with these weights, I show how a researcher can conduct his empirical analysis unconditional of any specific model.

Chapter 5 marks the beginning of the empirical part of this thesis. I present the US sample ranging from 1985 to 2011 on which my empirical results are based. In a first step, I discuss the data sources for the input parameters necessary to compute the ICC and for other variables that I need. Afterwards, I show detailed summary statistics for the eight ICC methods that I choose to implement. Finally, I discuss issues such as coverage, analyst forecast bias, and time misalignment that could impact my findings.

Chapter 6 repeats the analysis of Easton and Monahan (2005), but with alternative proxies for cash flow and discount rate news that are motivated in Chapter 3. This chapter reinvigorates earlier claims that there is no way to reliably identify a single best proxy in the small samples that are available to the applied researcher.

In the first part of Chapter 7, I compute the model weights for the eight ICC methods that I implement. Because of the large shocks of realized returns, no clear winner among the proxies can be found and the weights, which denote the posterior confidence that a researcher should have in a proxy, are rather equally distributed. In the second part, I revisit three studies that use an aggregated ICC time series for their research question. Claus and Thomas (2001) measure the implied risk premium, Pástor, Sinha, and Swaminathan (2008) examine the risk-return tradeoff, and Chen, Da, and Zhao (2013) look at the relative impact of cash flow news and discount rate news, respectively, on stock price movements. While all of these studies base their inference on only one ICC method, I replicate their results with the BMA approach, which directly incorporates model uncertainty into the statistical inference.

Chapter 8 summarizes the major findings of this dissertation and discusses open questions for future research.

# 2

# LITERATURE REVIEW

The dividend discount model (DDM) states that the stock price, $P_t$, is the sum of expected future dividends per share, discounted to today (cf., e.g., Ang and Liu 2007):

$$P_t = E_t \left[ \sum_{j=1}^{\infty} \frac{DPS_{t+j}}{\prod_{k=1}^{j} 1 + R_{t+k}} \right], \tag{1}$$

where $E_t[DPS_{t+j}]$ are the dividends per share at time $t+j$ and $E_t[R_{t+j}]$ is the return from time $t+j-1$ to $t+j$, both expected at time $t$ by the investor. If we replace time-varying expected returns with a constant term $R_t^e$, then equation (1) simplifies to[10]

$$P_t = \sum_{j=1}^{\infty} \frac{DPS_{t+j}}{(1+R_t^e)^j}, \tag{2}$$

where $R_t^e$ is commonly referred to as the implied cost of capital. The ICC is therefore the counterpart for stocks to the yield to maturity (YTM) for bonds, with the additional complication that the cash flows are stochastic.[11]

In Section 2.1, I present studies that investigate the theoretical relation between the true ICC and time-varying expected returns. In other words, I discuss the transition from equation (1) to equation (2) which depends on the assumption that expected returns are constant. Because this assumption is unlikely to hold in reality, there are systematic differences between expected returns and the ICC. The ICC is therefore an imperfect proxy for expected returns, even if a researcher was able to estimate expected dividends without error. It is important to the applied researcher to understand the differences between the two concepts.

10 For ease of notation, I omit the expectation operator $E_t[\cdot]$ in the following. The reader should keep in mind that all future values are conditional on expectations based on information available at time $t$.

11 Cochrane (2005, p. 350) defines the YTM as the "fictional, constant, known, annual, interest rate that justifies the quoted price of a bond, assuming that the bond does not default."

Furthermore, expected dividends are just as unobservable as expected returns. Thus certain simplifying assumptions have to be made to approximate them. As a consequence, the true ICC is just as unobservable as true expected returns and empirically implementable specifications have to be found. They differ in the short-term cash flow forecasts and the approximation of the terminal value. In Section 2.2, I present these specifications in detail.

Finally, Section 2.3 gives an overview of empirical studies that rely on the ICC and discusses how these studies approximate the vector of unobservable expected dividends.

## 2.1 THEORETICAL RELATION BETWEEN THE ICC AND EXPECTED RETURNS

There are two reasons why the true ICC might differ from true expected returns, even if the vector of expected dividends is estimated correctly. First, the ICC is a geometric average of the term structure of equity returns (cf. Easton 2006), while in many studies one is interested in one particular expected return, typically the one for the next period. There is empirical evidence that the term structure of equity returns is downward sloping on average (cf. Binsbergen, Brandt, and Koijen 2012) and changes with business cycles (cf. Binsbergen, Hueskes, et al. 2013), which implies that the ICC is on average smaller than the expected return next period and that this difference is correlated with the business cycle.

In particular, this issue is relevant for the study of Pástor, Sinha, and Swaminathan (2008) who test the intertemporal CAPM by Merton (1973) that hypothesizes a positive linear relation between the conditional mean and variance of market returns. Pástor, Sinha, and Swaminathan (2008) approximate the conditional expected return for the next period with the ICC and therefore their results could be driven by systematic differences between the two. To alleviate such concerns, they theoretically show, based on a Campbell and Shiller (1988) (CS) loglinearization of returns, that the log ICC is perfectly correlated with log expected returns next period if both expected returns and expected dividend growth follow an AR(1) process.[12] Additionally, they show in simulation results that the ICC helps to detect a positive mean-variance relation that is overshadowed by the

12 I introduce the CS loglinearization in Appendix A.1 and the simple framework by Pástor, Sinha, and Swaminathan (2008) in Appendix A.2.

large noise when realized returns are used. While those analyses mitigate concerns that their results are driven by other factors in addition to an actual relation between time-varying conditional expected returns and its variance, they do not eliminate them. For example, they use a very simplistic AR(1) assumption for expected returns that conflicts with the empirical evidence. It would be interesting to calibrate the dynamics of expected returns in their simulations to the empirical results of Binsbergen, Brandt, and Koijen (2012) and Binsbergen, Hueskes, et al. (2013).

Furthermore, the difference between the ICC and the expected return next period is relevant for studies that evaluate the validity of different ICC methods with predictive regressions of the ICC on subsequent realized returns.[13] For example, Easton and Monahan (2005) claim that the criterion for the validation of an ICC method is how close the slope coefficient of such a regression is to one. However, the analytical analysis of Pástor, Sinha, and Swaminathan (2008) shows that this coefficient will diverge from one even in simple settings, a point I discuss in more detail in Appendix A.2.

Hughes, Liu, and Liu (2009) discuss a second reason why the ICC and expected returns can differ. They examine the effect of Jensen's inequality that arises because the price in equation (1) is a nonlinear function of stochastic expected returns.[14] This effect is best demonstrated with a small numerical example. Assume an asset is expected to pay 100 units next period, irrespective of the state of the economy. There are two states to be expected, boom and recession. In the boom state, investors expect a return on the asset of 5%; in the recession state, investors expect 15%. Both states are equally likely. Plugging these values into equation (1), we obtain

$$P_0 = 0.5 \times \frac{100}{1.15} + 0.5 \times \frac{100}{1.05} = 91.10. \tag{3}$$

Inferring the ICC from equation (2), we get $R^e = 100/91.10 - 1 = 9.77\%$. Hence, the ICC is in this case lower than the average expected return, which is 10%. It is also easy to see that the difference between the two is dependent on the variation in expected returns.

As Hughes, Liu, and Liu (2009) highlight, this is a well known result in the fixed income literature. The contribution of Hughes, Liu,

13 In Chapter 3 I describe these approaches in detail.

14 Note that the results in Pástor, Sinha, and Swaminathan (2008) are based on a log-linearized version of the present value identity and hence they ignore issues arising from the nonlinearity of the present value formula.

and Liu (2009) is the generalization of these results to the more complicated case of equities for which expected cash flows are stochastic as well. One of the main achievements of their study is that they are able to incorporate stochastic expected returns and cash flows as well as a correlation between the two in their model, while keeping it simple enough to allow for a closed form solution. They assume that true expected returns are only affected through the firm's beta. They go on to show that the implied cost of capital is also a function of volatilities of, and correlation between, expected returns and cash flows, growth in cash flows, and leverage. This is an important result for empirical studies. Whenever a study finds an association between the ICC and these characteristics, it is unclear if this association is driven by expected returns or merely by the misspecification of the ICC as a proxy for expected returns (cf. Lambert 2009).

As Tang, Wu, and Zhang (2013) point out, the model of Hughes, Liu, and Liu (2009) is tightly parameterized, which makes it hard to empirically calibrate. Furthermore, Lambert (2009) emphasizes that the assumption that cash flows grow with a constant rate over time, which is needed to make the model analytically tractable, is unrealistic. Therefore, this model helps in illustrating the differences between expected returns and the ICC in a setting where the former are stochastic. Since it abstracts from all other problems in estimating the latter, such as cash flows that grow with a time-varying growth rate, measurement error in cash flows, etc., and since the model parameters are not calibrated, it is an interesting open question if this effect is empirically relevant.

In summary, the study of Pástor, Sinha, and Swaminathan (2008) provides evidence that the ICC can be helpful in discovering an association between a variable of interest and expected returns that would otherwise be obscured by looking at realized returns due to their large random shocks. This result is obtained within a very simple framework and it would be an interesting venue for future research to check how this association works under more realistic assumptions. Furthermore, the study by Hughes, Liu, and Liu (2009) highlights that such an association can also be due to the peculiarities of the ICC, and not be driven by true, but latent expected returns.

Coming back to the original idea of inverting a dividend discount model to infer the expected rate of return, both studies assume that the vector of expected dividends in equation (1) and (2) is measured without error. They are only concerned with the assumption that is needed to transform equation (1) into equation (2). Put differently, these studies are dealing with the denominator of the dividend discount model. Next, I focus on the issue of approximating expected dividends, i.e., the numerator of the dividend discount model.

## 2.2 ISSUES WITH APPROXIMATING EXPECTED DIVIDENDS

While expected dividends are the ultimate driver of a firm's stock price, they are hard to approximate. For example, according to Fama and French (2001) only 20.8% of US firms paid dividends in 1999, which implies that stock returns are mostly driven by capital gains through retained earnings. Miller and Modigliani (1961) demonstrate that in the absence of taxes, dividend policy does not affect firm value. Also, it does not invalidate the dividend discount model from equation (1) since retained earnings today lead to larger dividends in the future and the effects cancel each other out in an efficient market. However, it matters for a researcher who does not know the vector of true expected dividends and has to approximate them. Entertaining a dividend discount model with direct estimates of expected dividends in the short term is just not practical, also because dividends are subject to arbitrary decisions about the payout ratio. Consequently, earnings are preferred as a performance measure both by practitioners and academics (cf., e.g., Bakshi and Zhiwu 2008). Penman (2007, p. 121f.) summarizes the shortcomings of the dividend discount model nicely in the following quote:

> The truth of the matter is that dividend payout over the foreseeable future doesn't mean much. Some firms pay a lot of dividends, others none. A firm that is very profitable and worth a lot can have zero payout and a firm that is marginally profitable can have high payout, at least in the short run. Dividends usually are not necessarily tied to value creation. Indeed, firms can borrow to pay dividends, and this has nothing to do with their investing and operating activities where value is created. Dividends are distributions of value, not the creation of value.

It is therefore not surprising that it is easier to obtain proxies for expected earnings than for expected dividends (cf. Daske 2006). In the next subsection, I introduce the most common data sources of these forecasts, analysts and regression-based predictions based on historical data, and discuss their pros and cons. A researcher then has to adjust the dividend discount model in such a way that he can represent a firm's stock price by expected earnings. I show in subsection 2.2.2 that he can either transform the earnings into dividends or transform the dividend discount model into alternative valuation models that accept earnings as an input parameter. Finally, because these valuation models rely on forecasts for an infinite horizon, but reasonable forecasts for expected earnings are only available for the next couple of years, researchers have to make further simplifying assumptions. In subsection 2.2.3, I present several studies that propose such empirically implementable versions of the theoretical valuation models.

### 2.2.1 *Data sources for expected earnings*

#### 2.2.1.1 *Analyst earnings forecasts*

The most widely used proxy for expected earnings are forecasts by sell-side analysts, i.e., analysts employed by brokerage houses, independent research institutes, and investment banks (cf. Beyer et al. 2010). These forecasts are provided by data vendors such as Institutional Brokers' Estimate System (IBES), Value Line, First Call, and Zacks.

As Bradshaw (2011) points out, both the forecasts provided by these analysts as well as the analysts themselves as economic agents within capital markets are interesting research topics in accounting and finance, which is why there are hundreds of studies that deal with analysts and their forecasts in one way or another. Excellent surveys on this research can be found in Ramnath, Rock, and Shane (2008) and Bradshaw (2011).

Here, I only want to discuss the relevant questions with respect to the computation of the implied cost of capital. The main question is whether analysts are able to measure investors' expectations correctly. One obvious way to test this question is to compare the forecasts provided by analysts with the actual earnings. Early studies established the superiority of direct measures of earnings expectations through analyst forecasts over the former de facto standard, univariate time series forecasts (cf. Brown and Rozeff 1978; Brown, Hagerman, et al. 1987).[15] In a recent study, however, Bradshaw, Drake, et al. (2012)

15 The reader is referred to Brown (1993) for a comprehensive review of the early literature on earnings forecasting research.

show evidence that a random walk time series forecast can be superior over longer horizons, for younger or smaller firms, and when analysts forecast positive or more extreme changes in earnings. Nevertheless, analyst forecasts are the most widely used surrogate for market's expectations on earnings nowadays.

A common result found in the literature is that analysts are on average too optimistic, that is, they provide forecasts that are higher than the actual earnings (cf., e.g., Dugar and Nathan 1995; Das, Levine, and Sivaramakrishnan 1998; Lim 2001; Abarbanell and Lehavy 2003; Hovakimian and Saenyasiri 2012). If this overoptimism is exclusive to analysts and not shared by investors, it would lead to ICCs that are biased upwards, which is particularly worrisome for studies that want to estimate the level of expected returns such as Claus and Thomas (2001).

More generally, several studies find predictable forecast errors in analyst estimates. So (2013) is a recent example that also shows that investors do not fully adjust for these errors. Hence, he is able to develop a trading strategy that sorts firms by predictable forecast errors and achieves abnormal returns. This illustrates a general problem of the evaluation of earnings forecasts. The fact that forecasts systematically deviate from subsequent actual earnings does not automatically imply that these forecasts do not measure investors' expectations correctly. Maybe the analysts are too optimistic on average because the investors are too optimistic as well? Of course, there are also reasons to believe that most of the bias is limited to analysts, and not shared by investors. This can happen if unbiasedness is not the most important goal of an analyst. If it is, however, for the researcher or the investor, it results in a misalignment between analysts and those end users of the earnings forecasts. For example, an analyst might be willing to be too optimistic towards a firm to build a better relation with the management and hence, get better information. As a consequence, his forecasts will be biased upwards, but will also be more accurate (cf. Lim 2001). In a recent study, Malmendier and Shanthikumar (2013) try to disentangle such strategic optimism from non-strategic optimism by analysts. For a review on the incentive structure of analysts, refer to Ramnath, Rock, and Shane (2008) or Beyer et al. (2010).

Analyst bias is an important issue for researchers that compute their ICC with analyst data. If this bias is limited to analysts and correlated with their variable of interest, a significant relation between a researcher's ICC proxy and his variable of interest could be the result of this bias, and not because of a relation with expected returns, which he is interested in. In other words, his results would be spurious, a point I discuss in more detail in Chapter 4.

There are several studies that propose methods to adjust the forecasts for these predictable errors to improve the ICC estimates. Easton and Sommers (2007) recommend using realized earnings as a proxy for expected earnings, which implicitly relies on the assumption that earnings follow a random walk. In light of the recent results of Bradshaw, Drake, et al. (2012) (see above) this may be a reasonable first-order approximation after all. Not surprisingly, they find that their implied cost of capital estimates are lower than those based on analyst earnings. One major shortcoming of this approach is that realized earnings do not change monthly and thus do not incorporate new information in a timely manner. Furthermore, in contrast to analysts who typically provide forecasts for the next couple of years there is no obvious way to extrapolate current earnings into the future. Leaving them constant implies the unreasonable assumption of no earnings growth. Easton and Sommers (2007) work around this problem by using valuation models that only require one earnings forecast and assume a constant growth rate afterwards. As I show in the next section, this is a very limited model that does not allow a researcher to incorporate different growth assumptions over different horizons.

Guay, Kothari, and Shu (2011) argue that one source of the forecast error is the sluggishness of analysts with respect to information in past stock returns. They show that stock prices adjust to new information more quickly than analysts, which induces a negative correlation between the ICC and recent stock price performance. This is best illustrated with a simple example. Suppose that a stock price declines due to lower expected earnings by *investors*. However, *analysts* do not adjust their forecasts simultaneously, but lag a few months behind. In this case, a researcher who relies on those analyst forecasts has to explain a lower stock price with unchanged earnings forecasts. To do so, he has to assume, incorrectly, that the discount rate has increased. Guay, Kothari, and Shu (2011) support this story by showing empirically that there is a strong negative relation between past realized returns and forecast errors. If a stock performed poorly in the last year, analysts overestimate future earnings and vice versa. This effect translates also into higher ICC estimates for stocks with poor recent performance. They propose two alternative estimation procedures for the ICC to adjust for the sluggishness of analysts. In the first procedure, they adjust the earnings forecasts for each firm by the expected forecast error. This forecast error is either predicted from a regression or estimated as the time series median of a portfolio of firms, where the portfolios are determined by the past stock performance. That is, for portfolios of stocks with poor recent performance, they reduce the analyst forecasts accordingly, and vice versa. In the second procedure, they simply lag the stock prices by roughly five months.

The approach of Mohanram and Gode (2013) also uses a regression to estimate predictable forecast errors and adjusts for them.[16] In contrast to the approach of Guay, Kothari, and Shu (2011), their regression model is specified more generally and not focused on the sluggishness problem of analysts. In particular, they use a regression specification similar to the one proposed by Hughes, Liu, and Su (2008). This specification explains forecast errors with factors related to analysts' overreaction (accruals, sales, growth, long-term growth estimates, growth in PP&E) and underreaction (recent returns, recent revisions in forecasts). Note that this model includes recent returns, which are the main explanatory variable in Guay, Kothari, and Shu (2011). Both studies show that their adjusted ICC estimates are better able to explain subsequent realized returns than the unadjusted ones, which is taken as evidence that the proxies have been improved.

One point of criticism towards the procedure of Guay, Kothari, and Shu (2011) is that it relies on the whole time series of earnings forecasts to infer the predicted forecast error. It would have been interesting to repeat their analysis using a rolling window to only rely on information that was publicly available up to the date of the adjustment. In contrast, Mohanram and Gode (2013) only use publicly available information for their adjustments. Thus they do not have a look-ahead bias.

Another relevant question about analyst earnings is which data vendor to use. The most prominent one is IBES, which aggregates earnings forecasts from many analysts to one mean or median estimate. Ramnath, Rock, and Shane (2005) show that the data from IBES is superior to the one provided by Value Line in terms of accuracy. Two factors explain this finding. First, Value Line provides single forecasts, while IBES aggregates over many analysts, which mitigates concerns of idiosyncratic analyst error. Second, Value Line only up-

16 For a similar approach, see also Larocque (2013). Because her approach is much more parsimonious and quite similar to Guay, Kothari, and Shu (2011), I present only the study of Mohanram and Gode (2013) in detail here.

dates its forecast for each firm on a quarterly cycle, in contrast to the monthly cycle of IBES. Thus IBES has a timing advantage.

Additionally, there is another caveat against data sets based on analyst forecasts: those forecasts are only available for larger firms with considerable institutional following and more extensive disclosures (cf. Bradshaw, Drake, et al. 2012). Also, the coverage of firms for which analysts provide forecasts has increased over time (cf. Bradshaw 2011), which could induce systematic biases. For example, an implied risk premium could rise over time simply because more and more small firms are added to the sample and smaller firms require higher expected returns on average.[17] Ecker et al. (2013) generalize this problem and show that ICC samples are a non-random sample from the population distribution and that this problem can bias empirical results.

In the empirical part of this dissertation, I rely on IBES forecasts as well. While the previous paragraphs have shown that there are some potential biases with analyst earnings forecasts in general and while there are also some issues with IBES in particular,[18] they are still the most widely used surrogate of earnings forecasts both by practitioners and academics. Next, I discuss an alternative estimate for expected earnings that is free of the specific problems that plague analyst forecasts, but one that is subject to its own shortcomings.

#### 2.2.1.2 *Regression based earnings forecasts*

Recently, Hou, Dijk, and Zhang (2012) have proposed the use of earnings forecasts from a cross-sectional regression model instead of analyst forecasts. That is, in each year they predict the firm-level earnings for the next five years based on a pooled cross-sectional regression using the previous ten years of data. As predictors, they use lagged earnings, total assets, dividend payments, a dummy variable indicating dividend payers, a dummy variable indicating negative earnings, and accruals.

They show that such forecasts are superior to analyst forecasts in terms of coverage, forecast bias, and earnings response coefficient. Furthermore, implied costs of capital computed from their earnings forecasts are better able to explain subsequent realized returns, which they consider as additional evidence for the superiority of their proxy.

17 Dijk (2011) is a recent survey on the size effect.
18 For example, Ljungqvist, Malloy, and Marston (2009) document widespread ex post revisions between the years 2000 and 2007 to the IBES database, which resulted in significant changes in the results of empirical research.

This is a promising approach and it will be interesting to see how it is adopted by other researchers. One shortcoming, however, is the periodicity. They employ regressions on an annual basis and therefore, their approach cannot be used to infer ICC estimates on a monthly basis. This reduces the sample period substantially and is not fully compensated by the longer time period of available forecasts.[19]

### 2.2.2 *Theoretical transformations of the dividend discount model to alternative valuation models*

To make the earnings forecasts compatible with the dividend discount model, one has to adjust either the former or the latter to use earnings forecasts to infer an ICC estimate.

The most obvious approach is the assumption of a payout ratio. Then, expected dividends can be computed as the product of expected earnings and the assumed payout ratio. This approach is followed by Pástor, Sinha, and Swaminathan (2008) and subsequent studies that have built upon their approach (e.g., Lee, Ng, and Swaminathan 2009, Chava and Purnanandam 2010, Li, Ng, and Swaminathan 2013, and Chen, Da, and Zhao 2013).[20]

Alternatively, the dividend discount model in equation (2) can be transformed into a residual income model and abnormal earnings growth model, respectively. To derive them, start with a zero-sum equality:[21]

$$0 = x_t + \frac{x_{t+1} - (1 + R_t^e) \times x_t}{(1 + R_t^e)} + \frac{x_{t+2} - (1 + R_t^e) \times x_{t+1}}{(1 + R_t^e)^2} + \dots, \quad (4)$$

where $x$ can be any variable for which the transversality condition $x_{t+T}/(1 + R_t^e)^T \to 0$ as $T \to \infty$ holds.

19 Hou, Dijk, and Zhang (2012) are able to compute ICCs from 1968 on, while earnings forecasts from IBES are only available at the end of the 70s. Since the number of firms tracked by IBES was very low in the first few years, Claus and Thomas (2001) start their sample in 1985. I follow this procedure in my empirical analysis.

20 Note that Pástor, Sinha, and Swaminathan (2008) interpret the term "dividends" quite generally to describe the free cash flow to equity (FCFE) since they also adjust for stock repurchases and new equity issues, at least for their US sample. Contrarily, I use a simpler approach later on in which I use the dividend payout ratio, which is in line with most of the other studies.

21 Easton (2007) gives an excellent introduction into the various models, their derivations, and empirical implementations. This part draws heavily on it.

To convert the dividend discount model to the residual income model, we further require that the following relation, known as "clean surplus" relation, holds:[22]

$$BPS_{t+1} = BPS_t + EPS_{t+1} - DPS_{t+1}, \tag{5}$$

where $BPS_t$ is the book value per share and $EPS_t$ are the earnings per share at the end of period $t$. This relation requires that all items affecting the book value of equity are included in earnings (cf. Claus and Thomas 2001). If we set $x_j$ to the expected book value per share, $BPS_j$, add equation (4) to equation (2) while considering equation (5), and rearrange, we obtain the residual income model:

$$P_t = BPS_t + \sum_{j=1}^{\infty} \frac{EPS_{t+j} - R_t^e \times BPS_{t+j-1}}{(1 + R_t^e)^j}. \tag{6}$$

As Claus and Thomas (2001) point out, equation (6) is a simple algebraic restatement of the dividend discount model that only needs the additional assumption that forecasted earnings satisfy the clean surplus relation. If this assumption holds, both models theoretically yield identical results. However, Claus and Thomas (2001) argue that the residual income model has two empirical advantages: First, a large part of the stock price $P_t$ is explained by the book value per share, $BPS_t$, which is directly observable. Second, it is easier to make reasonable assumptions for the growth rate in residual incomes or abnormal earnings than for dividends. If book values measure input costs correctly, residual incomes measure the additional rent over the fair compensation an investor expects for those input costs.[23] Due to reasons such as global competition and antitrust actions, it is a commonly made assumption that the abnormal rents cannot grow in the long run.

Ohlson (2005), however, criticizes the clean surplus assumption, which is violated by the GAAP's earnings construct.[24] More impor-

22 The residual income model is sometimes also referred to as the abnormal earnings model.

23 Penman (2007, p. 162) defines residual earnings as "the return on common equity, expressed as a dollar excess return rather than a ratio."

24 Ohlson (2005) does not discuss the empirical relevancy of this issue, but there is evidence that dirty surplus accounting, i.e., violations of the clean surplus relation, have a minor effect on the correctness of the residual income model. For an international sample, Isidro, O'Hanlon, and Young (2004, 2006) identify a range of dirty surplus practices and evaluate the effect of those practices on the residual income model. They find little evidence on systematic valuation errors. Furthermore, Frankel and Lee (1999) highlight that the residual income model does not impose that the clean surplus relation holds in the past. It is merely required that it holds for the future,

tantly, though, he shows that capital transactions such as share issuances and repurchases break the relation in equation (5) on a per share basis. As a solution, he advocates the use of the abnormal earnings growth model, which does not rely on the clean surplus assumption and is developed in Ohlson and Juettner-Nauroth (2005).[25]

The derivation is identical to the one of the residual income model, except that $x_j$ is set to $EPS_{j+1}/R_t^e$. Then, the dividend discount model can be rewritten as

$$P_t = \frac{EPS_{t+1}}{R_t^e} + \sum_{j=2}^{\infty} \frac{EPS_{t+j} + R_t^e \times DPS_{t+j-1} - (1 + R_t^e) \times EPS_{t+j-1}}{R_t^e \times (1 + R_t^e)^{j-1}}. \tag{7}$$

Abnormal earnings growth refers to the numerator in the infinite sum expression. According to Ohlson (2005, p. 331) one should interpret abnormal earnings growth "as the expected eps-increment in excess of what should be expected due to earnings retained in the prior period."

In summary, a researcher can choose between three different valuation models. Theoretically, all models yield identical results, although the residual income model needs an additional assumption for this to hold. Empirically, all models can be implemented with proxies for expected earnings and additional assumptions about the payout ratio to infer expected dividends. However, while expected dividends are directly needed for both the dividend discount model and the abnormal earnings growth model, they are only needed to compute future expected book values per share in the case of the residual income model. Therefore, this model is more robust to misspecifications with respect to the payout ratio of a firm.

---

which can be ensured mechanically by the researcher. The only additional requirement is that the earnings forecasts are also consistent with clean surplus accounting. They go on to show that the majority of dirty surplus accounting practices, at least in the US, are unpredictable and their expected value is zero. Therefore, the residual income model should not be systematically biased by dirty surplus accounting.

25 In his discussion of the abnormal earnings growth model, Penman (2005) acknowledges the points raised by Ohlson (2005) that the residual income model is inconsistent with the dividend discount model on a per share basis, but he also brings forward points of criticism towards the abnormal earnings growth model. For example, he argues that the dismissal of book values comes at the cost of losing information inherent in the balance sheet of a firm. Also, he points out that the "anchor" of the residual income model is the current book value, which is observable. In contrast, the anchor in the abnormal earnings growth model, $EPS_{t+1}/R_t^e$, is not a number in the financial statement, but already a forecast that is subject to potentially large measurement error.

Because proxies for expected earnings are only available for a few years ahead, all models need to make an assumption about the years after. This is one of the major differentiators between the various empirical studies and I discuss the most prominent approaches in the next subsection.

### 2.2.3 *Empirically implementable valuation models*

Before I introduce the main ICC approaches, I want to emphasize that I am deliberately vague on what empirical proxies the respective studies use in their application. I only mention them when I am of the opinion that it helps the reader to better understand the specific method. I do so because I want to separate the methods, which are commonly applied by many follow-up studies, as much as possible from the specific input parameters used in the studies that introduce the methods. For example, the approach by Pástor, Sinha, and Swaminathan (2008) needs a steady-state earnings growth assumption, which these authors compute as the rolling average of nominal gross domestic product (GDP) growth rates starting from 1930. In contrast, Chen, Da, and Zhao (2013) use the year 1947 as a start, and Chava and Purnanandam (2010) use the GDP growth rate of the previous year. This is just one of many examples in which the original empirical specification is slightly changed by follow-up studies, quite often without any motivation. Nevertheless, the elemental components of the methods, which I describe here, are left unchanged.

#### 2.2.3.1 *Residual income models*

Claus and Thomas (2001) and Gebhardt, Lee, and Swaminathan (2001) were among the first to apply the ICC for a large US sample. Both of them use a derivative of the residual income model to approximate the implied cost of capital.

Because Claus and Thomas (2001) want to estimate the equity risk premium for the aggregate US stock market, they do not have to worry about differences in long-term growth rates across firms. Instead, as already mentioned above, they argue that aggregated residual income for the US stock market should not grow in the long run due to reasons such as global competition and antitrust actions. Because they want to establish an upper bound on the equity risk premium, they assume that residual incomes grow at the expected inflation rate $g_{lt}$, which they approximate as the yield on a nominal 10-

year government bond minus an assumed real risk-free rate of three percent. They require estimates for expected earnings up to five years ahead. In their study, they rely on analyst forecasts from IBES, which are only available for two to three years ahead for many firms. But IBES also provides an expected growth rate for earnings, which can be used to extrapolate available earnings forecasts when missing.

With these assumptions, the residual income model from equation (6) can be rewritten into an empirically tractable version:

$$P_t = BPS_t + \sum_{j=1}^{5} \frac{EPS_{t+j} - R_t^{CT} \times BPS_{t+j-1}}{(1 + R_t^{CT})^j} + \frac{(EPS_5 - R_t^{CT} \times BPS_4) \times (1 + g_{lt})}{(R_t^{CT} - g_{lt}) \times (1 + R_t^{CT})^5}, \tag{8}$$

where $R_t^{CT}$ is the ICC of the CT method.

The focus of the study by Gebhardt, Lee, and Swaminathan (2001) is on the drivers of the cross-sectional variation in implied risk premiums. For such a research question, the growth assumption made by Claus and Thomas (2001) seems too simplistic because it would imply that all firms grow with the expected inflation rate after five years. Instead, Gebhardt, Lee, and Swaminathan (2001) propose a two-stage approach. In the first stage, they forecast earnings explicitly for the next three years. In the second stage, they linearly mean revert the firm's expected three-year ahead return on equity (ROE) to the median industry ROE by period $t + 12$. According to them, this captures the long-term erosion of abnormal ROEs over time that an individual firm can earn over its peers. After this second stage, it is assumed that residual incomes do not grow anymore, which means that any growth in earnings or cash flows is value neutral.

In summary, the ICC of the GLS method, $R_t^{GLS}$, can be obtained by solving the following equation numerically:

$$P_t = BPS_t + \sum_{j=1}^{11} \frac{(ROE_{t+j} - R_t^{GLS}) \times BPS_{t+j-1}}{(1 + R_t^{GLS})^j} + \frac{(ROE_{12} - R_t^{GLS}) \times BPS_{11}}{R_t^{GLS} \times (1 + R_t^{GLS})^{11}}. \tag{9}$$

#### 2.2.3.2 *Dividend discount models*

Pástor, Sinha, and Swaminathan (2008) build upon the GLS method in that they also break the intrinsic stock price into three parts: an explicit forecast period for the next three years, an intermediate forecast period in which they interpolate the three-year ahead earnings to a long-term growth rate, and a terminal value for which they assume, just as in the GLS case, that economic profits are value irrelevant. However, the major difference between the two is that Pástor, Sinha, and Swaminathan (2008) directly infer the dividends from forecasted earnings and estimated payout ratios and do not rely on the residual income transformation.[26,27]

They implement the following empirically tractable finite-horizon model:

$$P_t = \sum_{j=1}^{15} \frac{EPS_{t+j} \times (1 - b_{t+j})}{(1 + R_t^{PSS})^j} + \frac{EPS_{t+16}}{R_t^{PSS} \times (1 + R_t^{PSS})^{15}}, \tag{10}$$

where $b_j$ is the expected plowback rate in period j, which is defined as one minus the payout ratio. Pástor, Sinha, and Swaminathan (2008) use an intermediate forecast horizon of 15 years as their base case, instead of the 12 years in the GLS method. Earnings from year 4 on are inferred from the previous earnings forecast, $EPS_{j-1}$, and an estimated earnings growth rate $g_j$:

$$EPS_{t+j} = EPS_{t+j-1} \times (1 + g_{t+j}). \tag{11}$$

They impose an exponential rate of decline to mean revert the year $t + 3$ growth rate to a long-term growth rate, $g_{lt}$:

$$g_{t+j} = g_{t+j-1} \times \exp\left[\frac{\log(g_{lt}/g_{t+3})}{T - 1}\right]. \tag{12}$$

Finally, they have to forecast the plowback rates, which they do in two stages. For the first two years, they explicitly forecast the plowback rates based on historical values. After that, they mean revert the plowback rates between $t + 3$ to $t + T + 1$ linearly to a steady-state value computed from the sustainable growth rate formula. This for-

26 As mentioned in footnote 20, Pástor, Sinha, and Swaminathan (2008) use the term "dividends" quite generally.

27 Lee, Ng, and Swaminathan (2009) provide a motivation for this modification. This approach imposes less stringent data requirements, which is particularly important in an international context in which the number of firms per country is substantially lower compared to a US sample.

mula assumes that, in the steady-state, the product of the return on new investments, ROI, and the plowback rate, $b_{SS}$, is equal to the steady-state growth rate in earnings, $g_{SS}$. Due to competition driving returns on investment down to the cost of equity, ROI can be set to $R^{PSS}$. In summary, the steady-state plowback rate, $b_{SS}$, is set to the ratio of the steady-state earnings growth rate, $g_{SS}$, and the implied cost of capital, $R^{PSS}$: $b_{SS} = \frac{g_{SS}}{R^{PSS}}$. Pástor, Sinha, and Swaminathan (2008) assume for each firm the same steady-state growth rate. Specifically, they set $g_{SS}$ to a rolling average of nominal GDP growth rates. Now, the terminal value is easily computed as

$$
\begin{aligned}
TV_{t+T} &= \sum_{i=1}^{\infty} \frac{EPS_{t+T+1} \times (1+g_{SS})^{i-1} \times (1-b_{SS})}{(1+R_t^{PSS})^i} \\
&= EPS_{t+T+1} \times \frac{1-b_{SS}}{1+R_t^{PSS}} \sum_{i=0}^{\infty} \frac{(1+g_{SS})^i}{(1+R_t^{PSS})^i} \\
&= EPS_{t+T+1} \times \frac{1-\frac{g_{SS}}{R_t^{PSS}}}{1+R_t^{PSS}} \times \frac{1}{1-\frac{1+g_{SS}}{1+R_t^{PSS}}} \\
&= EPS_{t+T+1} \times \frac{\frac{R_t^{PSS}-g_{SS}}{R_t^{PSS}}}{1+R_t^{PSS}} \times \frac{1+R_t^{PSS}}{R_t^{PSS}-g_{SS}} \\
&= \frac{EPS_{t+T+1}}{R_t^{PSS}}.
\end{aligned}
$$

The last line is identical to the terminal value part in equation (10). The terminal value is therefore identical to the PE method, which I introduce in the next section (see equation 14). This method assumes that there is no abnormal earnings growth and the derivations here illustrate the point nicely. Note that there might still be growth in earnings in steady-state and this growth might be different for different firms. However, this earnings growth does not add any value. For example, the larger the part of the earnings that are retained, the larger the earnings growth, but this is value-irrelevant because had those retained earnings been paid out instead, an investor could have just invested these earnings for the same rate of return, $R^{PSS}$, elsewhere.

Because it is assumed that the plowback rates revert linearly to $b_{SS}$, they can be inferred recursively as

$$b_{t+j} = b_{t+j-1} - \frac{b_{t+2} - b_{SS}}{T-1}. \qquad (13)$$

Given equation (11), (12), and (13) as well as explicit forecasts for earnings, earnings growth, and the plowback rates for the first few years, equation (10) can be solved for $R_t^{PSS}$.

To the best of my knowledge, all studies but one set $g_{lt}$ equal to $g_{SS}$. That is, they interpolate the three-year ahead earnings growth rate to the steady-state growth rate. On the contrary, the study of Chen, Da, and Zhao (2013), which I replicate in Chapter 7, uses a different assumption. Like Pástor, Sinha, and Swaminathan (2008), they use a rolling average of nominal GDP growth rates to approximate $g_{SS}$, but they set $g_{lt}$ to the mean long-term analyst industry growth forecast. As I show in the empirical part of this thesis, this seemingly minor modification has a large impact on the results and, therefore, I give this method its own abbreviation and label it the CDZ method, in contrast to the original approach by Pástor, Sinha, and Swaminathan (2008), which I abbreviate as the PSS method.

#### 2.2.3.3 *Abnormal earnings growth models*

Finally, I now turn to modifications of the abnormal earnings growth model from equation (7). If one is willing to make the very simplistic assumption that the abnormal earnings growth for all future periods is zero, the PE method follows directly. This method is simply the price-to-forward-earnings ratio:

$$R_t^{PE} = \frac{EPS_{t+1}}{P_t}. \tag{14}$$

I only use this method as a naïve benchmark because it completely ignores earnings growth.

More realistically, it is a common assumption that the expected abnormal earnings growth in equation (7), abbreviated with $AGR_j$

here, grows by a constant rate $g_{AGR}$. Then, the abnormal earnings growth model simplifies to[28]

$$\begin{aligned}
P_t &= \frac{EPS_{t+1}}{R_t^e} \\
&\quad + \sum_{j=2}^{\infty} \frac{EPS_{t+j} + R_t^e \times DPS_{t+j-1} - (1 + R_t^e) \times EPS_{t+j-1}}{R_t^e \times (1 + R_t^e)^{j-1}} \\
&= \frac{EPS_{t+1}}{R_t^e} + \sum_{j=2}^{\infty} \frac{AGR_{t+j}}{R_t^e \times (1 + R_t^e)^{j-1}} \\
&= \frac{EPS_{t+1}}{R_t^e} + \frac{AGR_{t+2}}{R_t^e} \sum_{j=1}^{\infty} \frac{(1 + g_{AGR})^{j-1}}{(1 + R_t^e)^j} \\
&= \frac{EPS_{t+1}}{R_t^e} + \frac{AGR_{t+2}}{R_t^e \times (R_t^e - g_{AGR})} \\
&= \frac{EPS_{t+1}}{R_t^e} + \frac{EPS_{t+2} + R_t^e \times DPS_{t+1} - (1 + R_t^e) \times EPS_{t+1}}{R_t^e \times (R_t^e - g_{AGR})}.
\end{aligned} \tag{15}$$

Equation (15) can be further simplified by assuming that both the dividends in the next year and the growth rate $g_{AGR}$ are zero. In this case, we can analytically solve for the ICC. As Easton (2004) shows, this ICC proxy is equal to the square root of the inverse of the price-earnings-growth ratio, which is the price-earnings ratio divided by a growth rate between the two-year and one-year ahead earnings. Therefore, I label this ICC proxy the PEG method:

$$R_t^{PEG} = \sqrt{\frac{EPS_{t+2} - EPS_{t+1}}{P_t}}. \tag{16}$$

Easton (2004) also proposes a modification to the PEG method in which he does not assume that the next year's dividends are zero. In this case, equation (15) can be rearranged to the MPEG method:

$$P_t = \frac{EPS_{t+2} + R_t^{MPEG} \times DPS_{t+1} - EPS_{t+1}}{(R_t^{MPEG})^2}. \tag{17}$$

Finally, Gode and Mohanram (2003) make further adjustments to the abnormal earnings growth model by Ohlson and Juettner-Nauroth (2005). First, they assume that $g_{AGR}$ is equal to the expected inflation rate, which they compute as in Claus and Thomas (2001) as the nominal risk-free rate minus three percent. Second, they replace the short-term earnings growth rate, $g_{st}$, between the earnings two years

28 For this derivation to hold, $g_{AGR}$ has to be smaller than $R_t^e$.

and one year ahead that is implicit in equation (15) with an average of this growth rate and the long-term growth rate provided by analysts. This decision is made so that the information in the long-term growth rate provided by analysts is not discarded. I abbreviate this method as the OJ method in the following. The ICC, $R_t^{OJ}$, based on this method is obtained by solving the following equation numerically:

$$P_t = \frac{EPS_{t+1}}{R_t^{OJ}} + \frac{g_{st} \times EPS_{t+1} + R_t^{OJ} \times (DPS_{t+1} - EPS_{t+1})}{R_t^{OJ} \times (R_t^{OJ} - g_{AGR})}. \quad (18)$$

It is instructive to compare the different versions of the abnormal earnings growth model with their empirical counterparts based on the residual income model and the dividend discount model. The PE, the PEG, the MPEG, and the OJ method rely only on earnings forecasts for the next two years. Forecasts after this very short forecasting period are ignored. Only the OJ method allows for the consideration of longer term growth assumptions through $g_{AGR}$ and $g_{st}$. It is at least questionable to assume that forecasts for the next two earnings are sufficient for valuation, which implies that the abnormal earnings growth methods may be too simplistic.

This issue is particularly troublesome if the short-term earnings forecasts are negative. For example, this is likely to happen during a financial crisis in which many analysts, and probably investors as well, expect negative short-term earnings for many firms. This is not a problem for the ICC approach per se as long as those negative earnings in the short term are offset by a positive long-term view. And this is exactly what the other methods do that rely on further-ahead earnings forecasts and long-term growth rates that are assumed to be positive. Such a long-term view is not possible for the abnormal earnings growth models presented here and negative earnings in the short term directly imply nonsensical results or no results at all.[29] Therefore, it is a requirement for those methods that the expected earnings are always positive, which can induce a systematic bias both in the cross-section and in the time series. This bias is hard to evaluate, but one can at least check how many observations are lost due to this requirement. I perform this check in Chapter 5.

29 The PEG method has the additional requirement that the earnings forecast two years ahead has to be larger than the forecast one year ahead.

#### 2.2.3.4 *Target price model*

So far, all models make certain simplifying assumptions about the terminal value, which are needed to make the models empirically implementable. There is, however, an alternative solution to the terminal value problem. The data provider Value Line provides a four-year ahead target price, which exempts a researcher from the estimation of a terminal value. All that is left to do is to estimate the intermediate dividend payments. Of course, the quality of this approach is dependent on the terminal value assumptions made by the analyst instead of the assumptions made by the researcher. This is probably one reason why the target price model is rarely used in empirical studies because the terminal value assumptions of the analysts are opaque.

One exemption is the study of Brav, Lehavy, and Michaely (2005). They argue that their data could be less affected by analyst optimism and conflict of interest issues because Value Line is an independent research service with no affiliation to investment banking activity. However, as mentioned in Section 2.2.1.1, Ramnath, Rock, and Shane (2005) provide evidence that IBES is superior to Value Line in terms of accuracy due to an aggregation and a timing advantage.

#### 2.2.3.5 *Estimating earnings growth rates*

With the exception of the target price model, all methods introduced above rely on *assumed* growth rates after the detailed forecast period, supported by economic reasoning. Easton, Taylor, et al. (2002) propose a regression-based approach to simultaneously *estimate* the growth rate and the ICC for a residual income model. Easton (2004) expands this approach to an abnormal earnings growth model.[30] Basically, these studies replace a possible specification error in the researcher's assumption by a specification error in the regression approach. One problem of these approaches is that they only allow the estimation of average ICC and growth rates estimates across all firms and are therefore not suitable for cross-sectional analyses. Nekrasov and Ogneva (2011) eliminate this shortcoming by adding firm-level explanatory variables to the regression.

Tang, Wu, and Zhang (2013) show that results based on assumed growth rates on the one hand and estimated growth rates on the other hand can differ substantially and argue that the specification error in the regression approach is more pronounced. This opinion seems to

30 Ashton and Wang (2013) propose a similar approach which has fewer data requirements and makes fewer assumptions.

be shared by most researchers that apply the ICC because regression-based approaches to simultaneously estimate the implied growth rate are rarely used in empirical studies.

## 2.3 implementation details of empirical studies

As the previous sections have illustrated, a researcher who wants to measure expected returns using the ICC faces a multitude of options with respect to the empirical implementation.[31] The impact of these various options on the results are thereby unclear and naturally the question arises on how robust the results are to the quite often arbitrary decisions made by the researcher. Furthermore, the multitude of options also increases the probability of finding results by chance and allows researchers to practice data fishing or data snooping. In summary, there are many degrees of freedom for a researcher. Because there is only one true expected return process, these degrees of freedom are only affecting the measurement error part of the expected return proxy. Focusing on only one or a few specifications ignores the large uncertainty that the researcher has about the correct specification and can lead to overconfident and biased results. In Chapter 4, I introduce an approach that allows the incorporation of model uncertainty into the statistical inference. In this section, however, I shortly want to discuss how current studies using the ICC in their empirical analysis deal with the issue of model uncertainty.[32]

31 The decision to focus on the ICC already implies that other options, mentioned in the introduction, are not considered.

32 I only focus on the implementation details with respect to the ICC computation. In Chapter 2 of her dissertation, Mühlhäuser (2013) gives a recent summary of the research questions and findings of many empirical studies that apply the ICC as an estimate for expected returns.

Table 1: **Implementation details of published articles.** This table reports the source of the short-term earnings forecasts, the specific ICC methods implemented, relevant implementation details, and the checks performed with respect to the robustness of the ICC approach for several published studies that use the ICC as a proxy for expected returns. Methods used in the robustness section of an article are in brackets. The implementation details are somewhat arbitrary chosen and incomplete. For a complete summary of all assumptions refer to the original article.

| Study | Earnings data | Methods | Implementation details | Robustness checks |
|---|---|---|---|---|
| Pástor, Sinha, and Swaminathan (2008) | IBES | PSS | Short-term forecasts (first three years) from IBES; intermediate forecasts until year 15 based on earnings growth that is exponentially interpolated to the historical GDP growth average; long-term assumption: value-neutral earnings growth | Replacement of analysts' ex ante forecasts by ex post realized earnings; different intermediate horizons: $T = 10, 20$; ICC and growth rate are simultaneously computed as in Easton, Taylor, et al. (2002) |
| Li, Ng, and Swaminathan (2013) | IBES | PSS | As in Pástor, Sinha, and Swaminathan (2008) | Different intermediate horizons: $T = 10, 20$; incorporation of repurchases and new equity issues in the plowback rate computation; ICC and growth rate are simultaneously computed as in Easton (2004) |
| Lee, Ng, and Swaminathan (2009) | IBES | PSS (MPEG, OJ, CT) | Short-term forecasts (first three years) from IBES; intermediate forecasts until year 15 based on earnings growth that is exponentially interpolated to the world average GDP growth over the last 10 years; long-term assumption: value-neutral earnings growth; additionally, sluggish earnings forecasts are adjusted with the approach by Guay, Kothari, and Shu (2011) | Different intermediate horizons: $T = 10, 20$; explicit adjustment for ex post country-level forecast bias |

| Study | Earnings data | Methods | Implementation details | Robustness checks |
|---|---|---|---|---|
| Chen, Da, and Zhao (2013) | IBES | CDZ | Short-term forecasts (first three years) from IBES; intermediate forecasts until year 15 based on earnings growth that is exponentially interpolated to the average long-term growth rate assumed by analysts; long-term assumption: value-neutral earnings growth | Steady-state growth assumptions as functions of firm characteristics instead of same growth for all firms; several adjustments for forecast errors |
| Chava and Purnanandam (2010) | IBES | PSS | Short-term forecasts (first three years) from IBES; intermediate forecasts until year 15 based on earnings growth that is exponentially interpolated to the GDP growth rate of the previous year; long-term assumption: value-neutral earnings growth | Several robustness checks on analyst forecasts; one robustness check on the intermediate period (T=10 instead of T=15); one test adjusts the future cash flows based on a simulated time to bankruptcy |
| Claus and Thomas (2001) | IBES | CT | Earnings forecasts for the first five years are inferred from IBES; after this, earnings are assumed to grow at the expected inflation rate | Three percentage points more and less for the long-term growth rate, i.e., the expected inflation rate |
| Gebhardt, Lee, and Swaminathan (2001) | IBES | GLS | Short-term forecasts (first three years) from IBES; intermediate forecasts until year 12 based on a ROE that is linearly interpolated to the rolling 10-year median ROE of all firms within the same industry that have a positive ROE; long-term assumption: value-neutral earnings growth | Different horizons: $T = 6, 9, 15, 18, 21$ |
| Hann, Ogneva, and Ozbas (2013) | IBES | GLS (PEG, CT) | As in Gebhardt, Lee, and Swaminathan (2001) | Account for unexpected and expected analyst forecast errors |

| Study | Earnings data | Methods | Implementation details | Robustness checks |
|---|---|---|---|---|
| Ortiz-Molina and Phillips (2013) | IBES | GLS | As in Gebhardt, Lee, and Swaminathan (2001) | Adjustment of sluggish revisions in analyst forecasts; sample restriction to observations with small analyst earnings forecast errors; deletion of ICC estimates below the risk-free rate |
| Hribar and Jenkins (2004) | IBES | GLS, OJ, adjusted PEG | As in the original studies | Use of ICCs in later months to reduce the impact of stale earnings forecasts; ICC and growth rate are simultaneously computed as in Easton, Taylor, et al. (2002) |
| Hail and Leuz (2006) | IBES | Average of GLS, CT, OJ, MPEG | Refer to their appendix | Replication of results for each valuation method separately (untabulated); ICC and growth rate are simultaneously computed as in Easton, Taylor, et al. (2002) and Easton (2004); lagging of stock prices by three months, as proposed by Guay, Kothari, and Shu (2011) |
| Hail and Leuz (2009) | IBES | Average of GLS, CT, OJ, MPEG | Refer to their appendix | Replication of results for each valuation method separately; adjusted long-run growth expectations that vary by firm, cross-listing status (their variable of interest), country, and/or year; control for differences in analyst forecast bias across firms and countries; robustness check of impact of missing and negative earnings forecasts |

| Study | Earnings data | Methods | Implementation details | Robustness checks |
|---|---|---|---|---|
| Dhaliwal, Heitzman, and Zhen Li (2006) | IBES | Average of GLS, CT, OJ, MPEG | Refer to their appendix | Replication of results for each valuation method separately |
| Attig, Guedhami, and Mishra (2008) | IBES | Average of GLS, CT, OJ, MPEG | Refer to their appendix | Replication of results for each valuation method separately (untabulated) |
| Daske (2006) | IBES | GLS, OJ | GLS method is adjusted to take all five-year ahead earnings from IBES into account; for further details, refer to Section 3 in the article | ICC and growth rate are simultaneously computed as in Easton, Taylor, et al. (2002) and Easton (2004) |
| Chen, Chen, and John Wei (2011) | IBES | Median of GLS, CT, OJ, MPEG | Refer to their appendix | Several checks with firm-specific terminal value assumptions |
| Brav, Lehavy, and Michaely (2005) | Value Line | DDM | The ICC is backed out from the forecast for the next period dividends, a dividend growth rate, and the four-year target price | Forecasts from First Call instead of Value Line |
| Francis et al. (2004) | Value Line | DDM (MPEG, PEG) | As in Brav, Lehavy, and Michaely (2005) | Alternative methods are computed with Zacks data |

For a non-exhaustive selection of published articles that use the ICC, Table 1 presents an overview of the implementation details and robustness checks. In column "Implementation details", I focus on the main differentiators between these studies, which is somewhat subjective and arbitrary. Therefore, this table is not meant to be a complete list of all assumptions and decisions made by the cited studies. Furthermore, I want to stress that the robustness checks only refer to control checks with respect to the ICC computation. Other checks are therefore excluded, even if they are related to the quality of the ICC. This mostly happens when studies do not directly adjust the ICC for forecast errors by analysts, but include a control variable of the forecast error in their regression analysis.

With respect to the source for earnings forecasts for the next few years, earlier statements about the preference towards IBES are confirmed. With the exception of the studies by Francis et al. (2004) and Brav, Lehavy, and Michaely (2005), all studies rely on IBES data. This might change in the future if the recently proposed regression-based approach by Hou, Dijk, and Zhang (2012) is adopted by other researchers.

The known shortcomings of analyst forecasts, discussed in Section 2.2.1.1, are addressed somewhat arbitrarily. Some studies ignore this issue, some studies adjust the earnings for the optimism bias, some studies deal with this bias by using realized earnings, and some studies account for the sluggishness. The same is true for the assumption on future growth rates. Several studies implement the approach by Easton, Taylor, et al. (2002) and Easton (2004) to simultaneously estimate the growth rate. However, since this approach is only applicable to portfolios of firms and many studies focus on differences across firms, they prefer to only use this approach in their robustness section. It is interesting that these portfolio approaches are the standard robustness check with respect to the growth assumption. In my opinion, it is at least doubtful that investors do not assume different growth rates within industries or countries, but across firms. In case of a misspecification, this can induce a systematic bias in the expected return estimates. This is less of an issue for studies that only look at the aggregated ICC such as Pástor, Sinha, and Swaminathan (2008) and Li, Ng, and Swaminathan (2013). But, even studies that focus on cross-sectional differences mostly focus on industry or country growth rates. An exception is the study of Hail and Leuz (2009), which acknowledges the substantial impact long-run growth assumptions can have on the ICC. Therefore, they rerun their analysis with a variety of different growth assumptions, some of which are also firm-specific.

Take as an example the study of Gebhardt, Lee, and Swaminathan (2001). One of their results is that the industry implied mean risk premium from the prior year is an important variable to explain the cross-sectional variation of ICCs. However, this result is at least partly driven because of a mechanical relation between the independent and dependent variable. The GLS method mean reverts three-year ahead ROEs of each firm to an industry-wide ROE that is based on the median of all profitable firms within each industry over the last ten years. Hence, the assumed growth rate by Gebhardt, Lee, and Swaminathan (2001) is constant within industries and highly persistent. A firm that had a higher than average ICC last year because of a higher than average growth rate is more likely to have a higher than average ICC this year. So it is obvious that a higher than average industry implied risk premium last year is correlated with a firm's current ICC. It is easy to think about similar arguments even on an aggregated level. The more general point here is that one of the most crucial input parameters, long-run growth rates, is also one of the hardest to estimate. Therefore, it would not be surprising if a systematic error is introduced.

In terms of methods, there is a clear separation between asset pricing studies on the one hand and corporate finance as well as accounting studies on the other hand. Brav, Lehavy, and Michaely (2005), Pástor, Sinha, and Swaminathan (2008), Lee, Ng, and Swaminathan (2009), Chava and Purnanandam (2010), Li, Ng, and Swaminathan (2013), and Chen, Da, and Zhao (2013) belong to the former field and, except for Brav, Lehavy, and Michaely (2005), all of these studies rely on derivatives of the PSS method. As mentioned above, Lee, Ng, and Swaminathan (2009) motivate the use of this method with less stringent data requirements, which is particularly important for international samples for which fewer firms per country are available. Yet, the studies by Chava and Purnanandam (2010), Li, Ng, and Swaminathan (2013) and Chen, Da, and Zhao (2013) only use a US sample, which makes this argument irrelevant to those studies. Also, it is interesting that these studies hardly make any robustness checks of the ICC method. Brav, Lehavy, and Michaely (2005) and Chen, Da, and Zhao (2013) only make robustness checks within their respective ICC method. Pástor, Sinha, and Swaminathan (2008) and Li, Ng, and Swaminathan (2013) use only one alternative ICC estimate, based on the regression approach that simultaneously estimates the growth rate (see Section 2.2.3.5). Only Lee, Ng, and Swaminathan (2009) implement several different methods for robustness (MPEG, OJ, CT).

All other studies focus on residual income and abnormal earnings growth models. The most common procedure here is to focus on the mean or median of several ICC methods. In these cases the studies also show the evidence based on every method separately.

Table 1 nicely illustrates the many degrees of freedom a researcher has in approximating expected returns. It also shows that the decisions about the selection of a specific method are rather arbitrary and ad-hoc. This is best illustrated in the following quote by Li, Ng, and Swaminathan (2013, p. 11):

> In addition to the methodology used in this paper, several other procedures are used in the literature to compute the ICC. Instead of going through all of them, we pick a procedure recommended by Easton (2004) [...].

There seems to be a common belief that it does not really matter which method is chosen in particular, as long as the method is sensible. As I show in Chapter 4, this belief is misguided. In the next chapter, however, I focus first on the question of how to identify sensible proxies in the first place, that is, how to select the best proxy from a multitude of different options.

# 3

# EVALUATING EXPECTED RETURN PROXIES

Because of the many different specifications available to the applied researcher in computing the ICC, it was only a matter of time before comparative studies of these specifications emerged. This chapter summarizes existing efforts in this area. Broadly, there are two options to evaluate an expected return proxy.[33]

First, earlier studies checked the relation between the proxy and common risk factors. If the proxy under consideration is associated in the hypothesized direction with such risk factors, it is taken as confirmatory evidence that the proxy is measuring latent expected returns. However, this approach is subject to a joint hypothesis problem similar to efficient market tests (cf., e.g., Fama 1970, 1991). I discuss this approach in Section 3.1.

Therefore, this approach has mostly been replaced by regressions of an expected return proxy on subsequent realized returns. It turns out that any reasonable proxy of *expected* returns has to explain *realized* returns eventually. Since my BMA approach builds upon it, I motivate this approach in detail in Section 3.2.

Both Lee, So, and Wang (2011) and Easton and Monahan (2005) extend the second approach. The former study distinguishes between the evaluation in the time series and in the cross-section, while the latter study argues that the empirical validation of alternative proxies can be improved by controlling ex post for ex ante unexpected cash flow and discount rate news. I introduce and criticize these approaches in Section 3.3 and Section 3.4.

Finally, in Section 3.5 I summarize the findings of this chapter and conclude that an evaluation method that tries to pick the best proxy is unsatisfactory because it ignores the uncertainty inherent in this decision. Instead, a researcher should incorporate his uncertainty into the statistical inference. This insight is the main motivation for the BMA approach, which I introduce in detail in Chapter 4.

33 I am only aware of studies within the ICC literature that try to evaluate different ICC specifications. For this reason, this chapter focuses again on the ICC. Despite my focus on the ICC, the reader should keep in mind that in general these evaluation methods can also be applied to different proxies, such as proxies inferred from CDS spreads or bond yields.

### 3.1 ASSOCIATION WITH RISK FACTORS

This evaluation approach is advocated by Botosan and Plumlee (2005). They argue that estimates of the true cost of equity capital have to be associated with firm-specific risk factors in a consistent and predictable manner. Therefore, they identify five such risk factors that are supported by economic theory, approximate these theoretical proxies by empirical measures, and regress the expected return or cost of capital proxies on the risk factors.[34] For their US sample based on Value Line forecasts, they find that only the cross-sectional variation in ICC estimates based on target prices and the PEG method are consistently and predictably related to the risk factors. Some of the associations of other methods (GLS method, OJ method, and a simple finite horizon Gordon growth model) with the examined risk factors run counter to the theory. For example, the GLS method has a negative relation with the unlevered beta, which implies that investors require a higher return for stocks with lower market risk. This is inconsistent with the predictions of the CAPM. Botosan and Plumlee (2005, p. 51) therefore conclude that the target price method and PEG method "dominate the alternatives, and recommend that individuals requiring firm-specific estimates of expected cost of equity capital rely on either of these two methods as opposed to the alternatives we examine."

The main point of criticism brought forward against this approach is that it introduces a circularity argument. One of the main motivations for the ICC approach, in contrast to approaches that extract expected returns from models such as the CAPM or the Fama-French three-factor model (cf. Fama and French 1993), is that it does not rely on any asset pricing model. These asset pricing models identify risk factors based on theoretical arguments, estimate the premiums of those risk factors as well as the exposure of a stock to these risk factors with appropriate proxies, and back out the expected return from this information. Of course, this approach is dependent on the correctness of the model. Consequently, the apparent unexpected relation between an expected return proxy on the one hand and a risk factor on the other hand can be due to a bad proxy or the wrong risk factor. Thus, a researcher does not know if his proxy or his risk factor is misspecified.

34 They approximate five *theoretical* risk factors (market risk, leverage, information risk, firm size, growth) with the following *empirical* proxies (same ordering): unlevered beta, ratio of long-term debt to market value of common equity, dispersion in analyst forecasts, market value of equity, and expected growth in earnings.

This circularity is nicely illustrated with a short example. As already mentioned, Botosan and Plumlee (2005) deem the GLS method to be inferior to other proxies because it is, among other things, weakly (in their case even negatively) related to market risk. They assume that their risk factor is correctly specified and because they do not find the expected relation between this risk factor and the proxy based on the GLS method, they take this as evidence against the GLS method. Gebhardt, Lee, and Swaminathan (2001) also find a weak relation between their ICCs and the beta of a firm. However, they interpret this finding as evidence that beta, or market risk, is not priced, not that their proxy is incorrectly specified. In brief, we get two interpretations from the same finding and cannot identify which interpretation is the correct one.

Because of this shortcoming, Easton and Monahan (2010) label this approach "logically inconsistent" and recommend an alternative test, proposed in an earlier study (cf. Easton and Monahan 2005), that is based on realized returns. I present the underlying idea of such tests with subsequent realized returns in the next section.

## 3.2 ASSOCIATION WITH SUBSEQUENT REALIZED RETURNS

### 3.2.1 *Motivation*

In Appendix A.1, I introduce the Campbell and Shiller (1988) (CS) loglinearization of returns. CS develop a useful approximate identity for the log realized return $r_{t+1}$ on an asset $i$ that is held from the end of period $t$ to the end of period $t+1$:[35]

$$\begin{aligned} r_{t+1} \approx\ & E_t[r_{t+1}] \\ & + (E_{t+1} - E_t)\left[\sum_{j=0}^{\infty} \rho^j \Delta d_{t+j+1}\right] \\ & - (E_{t+1} - E_t)\left[\sum_{j=1}^{\infty} \rho^j r_{t+j+1}\right], \end{aligned} \tag{19}$$

where $\Delta d_{t+1} \equiv d_{t+1} - d_t$ is the log dividend growth during period $t+1$ and $\rho$ is a parameter of linearization defined by $\rho = 1/(1 + \exp(\overline{d-p}))$, where $\overline{d-p}$ is the average log dividend-price ratio.[36]

35 For ease of notation, I suppress the identifier $i$ in the following if it is not needed.
36 If not defined otherwise, lowercase letters of already used symbols denote log variables, e.g., $\log D_t \equiv d_t$.

Equation (19) shows that unexpected returns, $r_{t+1} - E_t[r_{t+1}]$, can only occur for two reasons: changing expectations about future dividends and changing expectations about future returns (or a combination of both). An increase in expected future dividends results in capital gains today, while an increase in expected future returns results in a capital loss today. As an example, suppose that expected future dividends do not change over a period, but expected future returns rise. These higher returns can only be generated by future capital gains from a lower current price (cf. Campbell, Lo, and MacKinlay 1997).

In the literature, the last equation is often abbreviated as follows:[37]

$$r_{t+1} = \mu_t + CFN_{t+1} - DRN_{t+1}, \tag{20}$$

where $\mu_t \equiv E_t[r_{t+1}]$ and $CFN_{t+1}$ is defined as the cash flow (CF) news and $DRN_{t+1}$ is defined as the discount rate (DR) news that arrive between time t and $t + 1$.

Equation (20) motivates the use of realized returns, $r_{t+1}$, as a proxy for expected returns next period, $\mu_t$. It is a commonly made assumption that CF news and DR news are zero on average, which is equal to the statement that there are no systematic biases in investors' expectations. In this case, the mean of historical realized returns will be an unbiased estimator of expected returns. Furthermore, news are by definition information surprises that are unknown at time t and therefore, they have to be uncorrelated with $\mu_t$. The justification behind these assumptions is that investors are assumed to be rational, which implies that they take all relevant information available at time t into account. As a direct consequence, all new information that arrives between time t and $t + 1$ is unexpected and hence orthogonal to the forecast from time t.[38]

### 3.2.2 *Predictive regressions*

The discussion in the previous paragraph has shown that realized returns over period $t + 1$ are an unbiased estimate of returns for this period, expected at time t. This relation also implies that *any* predic-

37 Note that this relation is only an approximation. However, the approximation error is mostly ignored in the literature and an equal sign is used (cf., e.g., Pástor, Sinha, and Swaminathan 2008; Vuolteenaho 2002).

38 Pástor and Stambaugh (2009) and Sadka and Sadka (2009) are recent studies that exploit the unbiasedness and lack of correlation between expected returns on the one hand and the news term on the other hand.

tor that is able to explain subsequent realized returns asymptotically is tracking expected returns. This is the motivation for *predictive regressions*, which are the most common approach to extract cash flow and expected return expectations (cf. Kelly and Pruitt 2013).[39] If the slope coefficient on the predictors differ from zero in these regressions, that is, if returns are predictable, one can also estimate the time series of expected returns as the fitted regression line.

Predictive regressions can also be used to evaluate the quality of expected return proxies. A proxy that is of any value in tracking expected returns has to explain subsequent realized returns eventually. The only additional requirement of the proxy in comparison to any predictor is that the proxy should measure expected returns theoretically, not only track them. Common examples for predictors, such as the dividend-price ratio, stock variance, inflation, or the long-term government bond yield, do not fulfill this requirement because they are only meant to be related to expected returns, not to measure them directly. By contrast, ICC and proxies derived from structural models try to measure expected returns.

To get a better understanding for the predictive regression framework, I summarize the two unexpected news parts in equation (20) into one unanticipated news error term, $u_{t+1}$:

$$r_{t+1} = \mu_t + u_{t+1}. \tag{21}$$

Furthermore, I assume that a proxy $\widehat{\mu}_t$, e.g. the log implied cost of capital $r^e_t$, tracks expected returns, but is scaled by a factor $s$ and also affected by additional measurement error $w_t$:

$$\widehat{\mu}_t = s\mu_t + w_t. \tag{22}$$

The additional measurement error in equation (22) can arise due to misspecifications in the underlying proxy. In the case of the ICC, misspecifications such as incorrect analyst forecasts or terminal value assumptions have been discussed in Chapter 2.

The scaling factor $s$ is motivated by the simple framework of Pástor, Sinha, and Swaminathan (2008) that shows that even in the case of no

39 The literature on return predictability is too voluminous to cover here. Recent important contributions, some of which also deal with statistical issues that arise from this approach, are Stambaugh (1999), Lewellen (2004), Campbell and Yogo (2006), Cochrane (2008), Pástor and Stambaugh (2009), and Binsbergen and Koijen (2010).

measurement error in cash flows, the ICC does not track expected returns one by one. I show this in detail in Appendix A.2.[40]

A researcher can now run a predictive regression by regressing subsequent realized returns $r_{t+1}$ on a proxy $\widehat{\mu}_t$. Ideally, if a researcher could observe true expected returns $\mu_t$, the intercept $\alpha$ of this regression would be 0 and the slope coefficient $\beta$ would be 1 asymptotically. In the more realistic case in which a researcher has to use a proxy for expected returns, $\beta$ is given as

$$\begin{aligned}\beta &= \frac{Cov\left(\widehat{\mu}_t, r_{t+1}\right)}{Var\left(\widehat{\mu}_t\right)} \\ &= \frac{Cov\left(s\mu_t + w_t, \mu_t + u_{t+1}\right)}{Var\left(s\mu_t + w_t\right)} \\ &= \frac{sVar\left(\mu_t\right) + Cov\left(w_t, \mu_t\right)}{Var\left(s\mu_t + w_t\right)}.\end{aligned} \tag{23}$$

If we assume that the proxy is measured without any error, equation (23) simplifies to $\beta = 1/s$. This result is also obtained in Appendix A.2, with the only difference being that the scaling factor is calibrated based on the input parameters in the Pástor, Sinha, and Swaminathan (2008) framework. The main takeaway here is that the most prominent expected return proxy, the ICC, will not yield slope coefficients of 1 in predictive regressions, even if the cash flows are measured without any error. This happens because the ICC is a geometric average of the yield curve of equity returns, which is not identical to expected returns next period.

As another example, suppose that the scaling factor is 1, but that the proxy is measured with additional white noise, so that equation (23) simplifies to $\beta = Var(\mu_t)/(Var(\mu_t) + Var(w_t))$. In this case, $\beta$ will be lower than 1 and biased towards 0. This result is known as the attenuation bias in the literature (cf., e.g., Wansbeek and Meijer 2000). However, this is an extreme assumption and is not likely to hold in the case of expected return proxies. For example, there is ample evidence, as shown in Chapter 2, that analyst forecasts are systematically biased. It is also hard to imagine that the assumptions

40 One could argue that the introduction of a scaling factor is redundant. As I show in Appendix A.3, it is possible to split up the measurement error into a part that is driven by expected returns and a part not driven by expected returns. Since I do not assume that the measurement error is uncorrelated with expected returns, one could also model $\widehat{\mu}_t$ as $\mu_t + w_t$ and assume that one part of $w_t$ is given as $(s-1)\mu_t$. I deliberately introduce the scaling factor here to emphasize that the proxies can differ from expected return next period even in cases in which the proxy is otherwise correctly measured. As I argue further below, I think that this issue is not of great relevance to empirical studies.

about terminal values for the different ICC methods do not introduce a systematic error.

If the scaling factor is less of an issue for a researcher and he is mostly interested in minimizing $w_t$, a good evaluation criterion is the $R^2$ of the regression, which is defined as the squared correlation between $r_{t+1}$ and $\widehat{\mu}_t$ in this simple setup. In this case, we get

$$\begin{aligned} R^2 &= \left( \frac{Cov\left(\widehat{\mu}_t, r_{t+1}\right)}{\sqrt{Var\left(\widehat{\mu}_t\right)}\sqrt{Var\left(r_{t+1}\right)}} \right)^2 \\ &= \frac{\left(Cov\left(s\mu_t + w_t, \mu_t + u_{t+1}\right)\right)^2}{Var\left(s\mu_t + w_t\right) Var\left(\mu_t + u_{t+1}\right)} \\ &= \frac{\left(sVar\left(\mu_t\right) + Cov\left(w_t, \mu_t\right)\right)^2}{Var\left(s\mu_t + w_t\right)\left(Var\left(\mu_t\right) + Var\left(u_{t+1}\right)\right)} \end{aligned} \tag{24}$$

In Appendix A.3 I show that the $R^2$ is maximized if the variance of $w_t$ is zero, irrespective of the scaling factor $s$. This is intuitive: Because the best predictor of subsequent realized returns are true, but unobservable expected returns today, any other predictor that tracks expected returns with an additional error term that is at least partly uncorrelated with expected returns must do worse.

With respect to the evaluation of an expected return proxy, the above discussion shows that predictive regressions allow a researcher to evaluate his proxy. In my opinion, it is not necessary to make the additional assumption that $\beta$ is equal to 1 in the regressions because a researcher should argue on prior grounds that his proxy tracks expected returns. In the case of the ICC, we actually know that $\beta$ will not be equal to 1. Empirically, studies are mostly satisfied that the ICC tracks expected returns and do not require a 1:1 relation between expected returns next period and the ICC. Actually, the requirement that the regression coefficient should be 1 can actually lead to nonsensical results. For example, Li, Ng, and Swaminathan (2013) find betas larger than 1 for their time series predictive regressions of market returns on the aggregated ICC. This can be explained by the fact that the scaling factor $s$ is lower than 1 in the case of the ICC. It directly follows that one could "improve" an ICC proxy by adding additional white noise to it because this will ceteris paribus lower the regression coefficient and bias the coefficient towards 1. This, of course is nonsensical, but it is the evaluation criterion for the evaluation method by Easton and Monahan (2005). I introduce their approach in detail in Section 3.4, but in their defense I want to mention that they look at the cross-sectional predictive power, for which the slope coefficients are typically far below 1 and therefore this particular problem is less of an issue.

### 3.2.3 *Issues*

The previous section showed that expected return proxies can be evaluated with predictive regressions. Unfortunately, this approach relies on realized returns again and inherits all of the problems associated with that. For example, I argued above that realized returns are an unbiased estimator of expected returns, which is a positive characteristic. Yet, this characteristic only pays off for large samples. In small samples that are available to an econometrician, inferences are imprecise due to the very large variation of both the CF and DR news part. This is the main point of Elton (1999, p. 1199) when he writes:

> The use of average realized returns as a proxy for expected returns relies on a belief that information surprises tend to cancel out over the period of a study and realized returns are therefore an unbiased estimate of expected returns. However, I believe that there is ample evidence that this belief is misplaced. There are periods longer than 10 years during which stock market realized returns are on average less than the risk-free rate (1973 to 1984). There are periods longer than 50 years in which risky long-term bonds on average underperform the risk free rate (1927 to 1981).

Fama and French (2002) further support Elton's argument. They show evidence that the large capital gains in the second half of the 20th century were largely unexpected, driven by a decline in expected discount rates. In terms of equation (20), this means that the high average of realized returns is not due to high expected returns, $\mu_t$, but persistent negative shocks to the discount rate part $DRN_{t+1}$.

Unfortunately, any evaluation method based on the correlation between a proxy and realized returns is subject to this large variation. Furthermore, since true expected returns are unobservable, we do not know how much of the variation in realized returns is driven by variation in expected returns and how much is driven by variation in the news part. In terms of equation (24), we do not know the $R^2$ between realized returns and true expected returns. Therefore, without further assumptions any evaluation of alternative proxies can only be *relative*. That is, we can compare two proxies with each other, but we cannot tell how close one proxy is to the benchmark, $\mu_t$. In the next chapter, I discuss these issues in more detail and also argue why a researcher would want to rely on alternative proxies, despite these shortcomings.

Before doing so, I introduce two extensions of the simple predictive regression approach to evaluate expected return proxies.

## 3.3 THE EVALUATION APPROACH BY LEE, SO, AND WANG (2011)

The main contribution of Lee, So, and Wang (2011) is that they look at the time series and cross-sectional performance simultaneously. They start with the assumption that the proxies under evaluation are tracking true expected returns for asset i, $\mu_{i,t} \equiv E_t[r_{i,t+1}]$, with an additional additive firm-specific error term, $w_{i,t}$:

$$\widehat{\mu}_{i,t} = \mu_t + w_{i,t}. \tag{25}$$

In terms of equation (22), they set $s = 1$ or, equivalently, they model scaling issues into the error term (see footnote 40). Ideally, a researcher wants $w_{i,t}$ to be zero for each period and each firm. However, they acknowledge the problem that as soon as there is any measurement error, the criterion to evaluate a proxy is not one-dimensional anymore because the measurement error term has both a time series and a cross-sectional component. As a consequence, they argue that a researcher should evaluate a proxy based on its performance in both dimensions. It should have a measurement error that is stable over time so that it tracks expected returns more closely in the time series. And it should have a measurement error structure in the cross-section that preserves the ranks of true expected returns.

They provide a simple example that illustrates that these two requirements do not imply each other. Suppose there are two stocks A and B with constant expected returns of 10% and 2%. Suppose further that one proxy produces estimates of 2% and 10% for stock A and B, respectively. Another proxy produces time-varying estimates that are either 15% and 5% for some periods or 10% and 2% for the other periods. Lee, So, and Wang (2011) argue that the first proxy is to be preferred in terms of the time series property because it tracks expected returns for both stocks with a constant measurement error. For cross-sectional analyses the second proxy should be chosen instead because it preserves the ranking of expected returns across stocks.

To assess the stability of the measurement error for a specific proxy *over time*, they propose the use of a *modified variance measure*. Taking the time series variance on both sides of equation (25) and rearranging, firm's $i$ variance in the measurement error can be expressed as

$$Var_i(w_{i,t}) = Var_i(\widehat{\mu}_{i,t}) - Var_i(\mu_{i,t}) - 2Cov_i(\mu_{i,t}, w_{i,t}), \quad (26)$$

which can be reorganized as

$$Var_i(w_{i,t}) = Var_i(\widehat{\mu}_{i,t}) - 2[Var_i(\mu_{i,t}) - Cov_i(\mu_{i,t}, w_{i,t})] + Var_i(\mu_{i,t}). \quad (27)$$

Lee, So, and Wang (2011) are interested in a relative measure of the validity of expected return proxies. Therefore, the third term in equation (27) can be omitted because it does not depend on the specific proxy:

$$MVar_i(w_{i,t}) = Var_i(\widehat{\mu}_{i,t}) - 2[Var_i(\mu_{i,t}) - Cov_i(\mu_{i,t}, w_{i,t})], \quad (28)$$

where $MVar_i(w_{i,t})$ is called the modified variance for firm $i$ and a specific expected return proxy.

Next, the second term in equation (28) has to be rearranged because it is dependent on the unobservable expected return process. Fortunately, it is easily shown that this term is equal to the covariance between subsequent realized returns and the expected return proxy:

$$\begin{aligned} Cov_i(r_{i,t+1}, \widehat{\mu}_{i,t}) &= Cov_i(\mu_{i,t} + u_{i,t+1}, \mu_{i,t} + w_{i,t}) \\ &= Var_i(\mu_{i,t}) + Cov_i(w_{i,t}, \mu_{i,t}). \end{aligned} \quad (29)$$

This derivation holds because $Cov_i(u_{i,t+1}, \mu_{i,t})$ and $Cov_i(u_{i,t+1}, w_{i,t})$ are equal to 0.

Plugging equation (29) into (28), we get

$$MVar_i(w_{i,t}) = Var_i(\widehat{\mu}_{i,t}) - 2Cov_i(r_{i,t+1}, \widehat{\mu}_{i,t}). \quad (30)$$

In equation (30), the modified variance $MVar$ is only a function of observable variables. Obviously, the lower the modified variance, the better a proxy is considered to be. Furthermore, note that the lower bound for the modified variance measure is not zero. Since we omitted a positive term, it can actually be below zero. The lower bound is $-Var_i(\mu_{i,t})$, which would be obtained if the proxy is equal to the true expected returns next period. Therefore, just like in the simple predictive regressions, $\mu_t$ is the benchmark against which every other proxy is compared to. Nonetheless, we also have the same problem as before, that is, we do not know this benchmark and therefore, we cannot infer how good our proxy is compared to $\mu_t$. Here, we do not know what $-Var_i(\mu_{i,t})$ is.[41]

It is instructive to compare this evaluation criterion against the simpler $R^2$ criterion. The modified variance measure implicitly punishes scaled proxies of expected returns next period, such as the ICC. To see this, suppose we have two proxies. Proxy A has a scaling factor $s_A$ that is unequal to one, but no additional measurement error $w_{i,t}$. Proxy B has a scaling factor of one, but additional measurement error that is white noise. If we take the $R^2$ criterion, the first proxy will win because $R^2$ is unaffected by any scale transformations of the independent variable.

For proxy A, we get the following modified variance measure:

$$\begin{aligned} MVar_i^A(w_{i,t}^A) &= Var_i(\widehat{\mu}_{i,t}^A) - 2Cov_i(r_{i,t+1}, \widehat{\mu}_{i,t}^A) \\ &= s_A^2 Var_i(\mu_{i,t}) - 2Cov_i(\mu_{i,t} + u_{i,t+1}, s_A \mu_{i,t}) \\ &= s_A^2 Var_i(\mu_{i,t}) - 2s_A Var_i(\mu_{i,t}) \\ &= (s_A^2 - 2s_A) Var_i(\mu_{i,t}). \end{aligned} \tag{31}$$

For proxy B, the modified variance measure can be computed as

$$\begin{aligned} MVar_i^B(w_{i,t}^B) &= Var_i(\widehat{\mu}_{i,t}^B) - 2Cov_i(r_{i,t+1}, \widehat{\mu}_{i,t}^B) \\ &= Var_i(\mu_{i,t}) + Var_i(w_{i,t}) \\ &\quad - 2Cov_i(\mu_{i,t} + u_{i,t+1}, \mu_{i,t} + w_{i,t}) \\ &= Var_i(\mu_{i,t}) + Var_i(w_{i,t}) - 2Var_i(\mu_{i,t}) \\ &= Var_i(w_{i,t}) - Var_i(\mu_{i,t}). \end{aligned} \tag{32}$$

The modified variance measure for proxy A is a function of the scaling factor $s_A$ and therefore this criterion is not scale invariant. Con-

41 In an additional analysis, they make further stringent assumptions which make it possible to approximate the variance of $\mu_{i,t}$. This, in turn, allows them to compute the total variance of the measurement error, for which a simple benchmark exists, 0. However, this approach is more of a theoretical exercise due to the additional assumptions made that introduce specification errors that cannot be evaluated easily.

sequently, proxies that do not measure expected returns next period, such as the ICC, get punished by this approach, while the same is not true for the $R^2$ criterion. By contrast, the modified variance measure for proxy $B$ is only dependent on the measurement error and the variance of true expected returns, which can be regarded as a constant because it is equal for all proxies. Therefore, it could be that $B$ gets a lower modified variance measure than $A$ although the latter actually tracks expected returns perfectly.

The question remains how relevant the scaling issue is in practice. While this question has not been addressed in the current literature and might be a fruitful problem for future research, I think that there is reason to believe that it is not of first-order importance. First, as I mentioned above, a researcher has to make sure a priori that his proxy actually measures expected returns in an economically reasonable way, and does not just track it. For example, the ICC, although it does not measure expected returns next period, does measure a discount rate and therefore, it fulfills this criterion. A researcher that would propose an "improved" ICC measure by just scaling it would likely have a difficult time convincing his colleagues that this is a reasonable approach. In the end, one would lose the economic interpretability of the ICC. Second, one is mostly interested in comparisons within one proxy class of expected returns and within this proxy class, the scaling factor should be identical to all proxies under consideration.

For the cross-section, Lee, So, and Wang (2011) evaluate methods depending on how well they explain the cross-sectional ranking in terms of their true expected returns. Since these are not observable, they have to rely again on the correlation with subsequent realized returns as the criterion, which is identical to the $R^2$ criterion.

In summary, the evaluation approach of Lee, So, and Wang (2011) is very similar to simple predictive regressions. The main point of their study is that a two-dimensional evaluation (time series and cross-section) is necessary because the two dimensions are not redundant. Just because one proxy is good in one dimension does not necessarily imply that it is good in the other dimension. There are in my opinion two objections to this argument. The first one is empirical. As their Figure 1 shows, there is actually an almost perfect relation between the two dimensions. Proxies with lower modified variance (time series dimension) also predict the cross-sectional variation in realized returns better. These proxies are two proxies based on the Gordon-growth formula and the GLS method.

The second objection is one of relevancy. For me, it is unclear why a researcher should require both criteria to hold simultaneously. If he is interested in the time series properties, why should he also consider the cross-sectional explanatory power of his proxy, and vice versa? Lee, So, and Wang (2011) do not discuss this issue and they also do not offer a solution to the question of how a researcher should trade off the time series and the cross-sectional performance of a proxy.

## 3.4 THE EVALUATION APPROACH BY EASTON AND MONAHAN (2005)

So far, I showed that the evaluation of any alternative expected return proxy is notoriously difficult. We cannot rely on the risk factor approach because this introduces a circularity argument. Unfortunately, if we rely on realized returns instead we reintroduce problems which avoidance was our main motivation for alternative proxies in the first place. This is just another circularity problem. Again, we have to deal with the large noise in realized returns that makes statistical inference very imprecise and, as a consequence, the evaluation very uncertain. Easton and Monahan (2010) (EM hereafter in this chapter) try to tackle this problem. Their idea is to control for the ex ante unexpected news terms $CFN_{t+1}$ and $DRN_{t+1}$ in equation (20) with proxies that are observable ex post. Their hope is that this will improve the power of the regression because the large variation due to information surprises is controlled for.

I do not believe in this argument. In brief, even with the advantage of hindsight the measurement error in the CF and DR news part is large. A large body of literature in asset pricing is dedicated to answer the question of how much of the variation in realized returns is due to CF news and how much is due to DR news.[42] Since the variation in those parts is an order of magnitude larger than the variation of true expected returns, any measurement error has a far more severe impact here. While EM propose a complicated method to evaluate the measurement error, this method is subject to further assumptions.

42 Examples are Campbell (1991), Campbell and Ammer (1993), Vuolteenaho (2002), Chen and Zhao (2009), Engsted, Pedersen, and Tanggaard (2012), and Chen, Da, and Zhao (2013). The study of Chen, Da, and Zhao (2013) is discussed in detail in Chapter 7.

Also, it is unclear how results from this measurement error analysis should be incorporated back into the evaluation approach.

However, my main objection is that EM throw away the most relevant part of knowledge that they gain by taking information over period $t+1$ into account. Because we know the realization of subsequent returns at time $t+1$, we know that the sum of expected returns and the two news parts has to be identical to the realized returns over period $t+1$ (see equation 20). Instead, EM use a CF news proxy that is inconsistent with this relation, which directly shows that their proxies are mismeasured.

Since this approach has found widespread use especially in the accounting literature (cf. Botosan, Plumlee, and Wen 2011; Nekrasov and Ogneva 2011; Mohanram and Gode 2013; Larocque 2013), I want to formalize this critique in detail in the next section. In the empirical part of this dissertation, I support my theoretical arguments empirically.[43]

### 3.4.1 *Recapitulation of the Easton and Monahan (2005) approach*

EM begin their analysis by replacing the three parts on the right-hand side (RHS) of equation (20) with empirical implementable proxies for each firm $i$ and by estimating the following cross-sectional regression:

$$r_{t+1}^{i} = \alpha + \beta_e \hat{r}_t^{k,i} + \beta_{CFN} \widehat{CFN}_{t+1}^{k,i} + \beta_{DRN} \widehat{DRN}_{t+1}^{k,i} + \epsilon_{t+1}^{i}, \quad (33)$$

where $\hat{r}_t^{k,i}$ is the log ICC of firm $i$ for a specific method $k$ that should be evaluated as a proxy for $E_t[r_{t+1}^{i}]$.[44] $\widehat{CFN}_{t+1}^{k,i}$ and $\widehat{DRN}_{t+1}^{k,i}$ are the CF news and DR news proxy, respectively.[45] As EM point out, if the proxies are measured without error, the betas will all be 1 in absolute terms ($\beta_{DRN}$ should be $-1$) and the alpha will be 0. This relation would hold without error because it is derived from a tautology. Realized returns *have to* be explained as the sum of the three parts. Hence, they argue that an ICC proxy is the better, the closer the coefficient is

43 This part is based on Jäckel (2013).

44 In general, this approach could also be applied to proxies other than the ICC. However, EM derive a DR proxy subject to the characteristics of the ICC and it is unclear how such a proxy could be estimated with other expected return measures.

45 EM multiply $\widehat{DRN}_{t+1}^{k,i}$ by $-1$, so that $\beta_{DRN}$ should be positive. I do not follow this procedure because in the variance decomposition literature, $\widehat{DRN}_{t+1}^{k,i}$ is also defined in such a way that a negative DR news shock implies a positive realized return. Later on, this allows me to compare my results with Vuolteenaho (2002).

to 1 in a univariate regression. However, a deviation from 1 is not necessarily proof of a bad expected return proxy, but could also be due to CF and DR news components that are biased in sample. Therefore, they include proxies for those parts as well and derive an estimate of the measurement error variance of the expected return proxy in a second step.

As a return news proxy $\widehat{DRN}^k_{t+1}$, they use the following formula:[46]

$$\widehat{DRN}^k_{t+1} = \frac{\rho}{1-\rho}(\hat{r}^k_{t+1} - \hat{r}^k_t), \tag{34}$$

where $\rho$ is a parameter of linearization that arises in the CS loglinearization of returns (see Appendix A.1). I motivate this formula further below.

For the CF news proxy $\widehat{CFN}_{t+1}$, they use the following formula:

$$\begin{aligned}\widehat{CFN}_{t+1} = {} & (roe_t - froe_{t,t}) + (froe_{t+1,t+1} - froe_{t,t+1}) \\ & + \frac{\rho}{1-\rho\omega_t}(froe_{t+1,t+2} - froe_{t,t+2}).\end{aligned} \tag{35}$$

In equation (35), $froe_{j,m}$ denotes the forecasted log ROE for fiscal year $m$ and is computed with log earnings forecasts $eps_m$ made in December of year $j$. $\omega_t$ is the regression coefficient of a pooled regression of the $roe$ on its lagged value. It only depends on realized and forecasted (up to two years into the future) return on equity values. Easton and Monahan (2005, p. 511) write that "our cash flow news proxy embeds the assumption that $roe$ follows a first order autoregressive process after year $t+1$", an assumption which they claim is supported by empirical evidence. Note that this proxy is *independent* of the specific ICC method $k$, i.e., the CF news proxy proposed by EM is constant across methods.

### 3.4.2 *An analytical derivation of the consistent cash flow and discount rate news proxies*

In this section, I first derive an analytical expression of the log ICC as well as an ICC proxy that is plagued with measurement error. In a next step, I show that the DR news proxy is just the scaled difference between two subsequent ICC proxy values, consistent with equation

46 I suppress the firm index $i$ in the following.

(34) and results recently derived by Chen, Da, and Zhao (2013).[47] Finally, I define the CF news as the part of the log price that is left unexplained by the log ICC and the DR news proxy.

The basic equation of the loglinearization that allows us to derive equation (19) is the following approximate identity:[48]

$$r_{t+1} = \kappa + \rho p_{t+1} + (1-\rho)d_{t+1} - p_t, \tag{36}$$

where $\kappa$ is a parameter of linearization and defined as $\kappa = -\log(\rho) - (1-\rho)\log(1/\rho - 1)$. This formula says that log realized returns $r_{t+1}$ in period $t+1$ can be approximated by a weighted average of log capital gains and log dividends for this period. Putting the log price $p_t$ on the left-hand side (LHS) and solving this equation iteratively yields[49]

$$p_t = \frac{\kappa}{1-\rho} + (1-\rho)\sum_{j=0}^{\infty} \rho^j E_t(d_{t+j+1}) - \sum_{j=0}^{\infty} \rho^j E_t(r_{t+1+j}). \tag{37}$$

This formula is just a loglinearized approximation of the classical present value formula.

The ICC is defined as the constant discount rate that equates prices with the sum of discounted cash flows. Due to the loglinearization, we are now able to solve for the ICC analytically. As Pástor, Sinha, and Swaminathan (2008) highlight, the ICC is then defined as the value of $r_t^e$ that solves[50]

$$p_t = \frac{k}{1-\rho} + (1-\rho)\sum_{j=0}^{\infty} \rho^j E_t(d_{t+j+1}) - r_t^e \sum_{j=0}^{\infty} \rho^j. \tag{38}$$

Solving for $r_t^e$ yields

$$r_t^e = \kappa - (1-\rho)p_t + (1-\rho)^2 \sum_{j=0}^{\infty} \rho^j E_t(d_{t+j+1}). \tag{39}$$

Equation (39) gives the theoretical correct ICC, that is, the ICC that equates the current stock price with true, but unobservable future expected dividends. Unfortunately, $r_t^e$ is just as unobservable as $E_t[r_{t+1}]$

47 They show that $E_t\left[\sum_{j=1}^{\infty} \rho^j r_{t+j+1}\right]$ and the ICC contain similar information. Their approach, however, does not rely on a loglinear return decomposition and they only work with capital returns instead of total returns.
48 Refer to Appendix A.1 for more details.
49 For this derivation the transversality condition, $\lim_{j\to\infty} \rho^j p_{t+j} = 0$, has to hold.
50 Refer to Appendix A.2 for more information.

because the vector of expected dividends is unobservable. Any empirically implementable ICC method makes certain assumptions about the expected dividend vector and hence, one can think of the estimate of a specific ICC method, $\hat{r}_t^k$, as the sum of two parts: the true ICC, $r_t^e$, and an additional measurement error $FE^k$ for each period j:

$$\begin{aligned} \hat{r}_t^k &= \kappa - (1-\rho)p_t + (1-\rho)^2 \sum_{j=0}^{\infty} \rho^j \left[E_t(d_{t+j+1}) + FE_{t+j+1}^k\right] \\ &= r_t^e + (1-\rho)^2 \sum_{j=0}^{\infty} \rho^j FE_{t+j+1}^k. \end{aligned} \tag{40}$$

The actual form of $FE_{t+j}^k$ does not matter for the following analysis.[51] Here, I abbreviate the sum with $VFE_t^k$, which simplifies equation (40) further:

$$\hat{r}_t^k = r_t^e + VFE_t^k. \tag{41}$$

Next, we can compute the true, but unobservable discount rate part, $DRN_{t+1}$, as

$$DRN_{t+1} = (E_{t+1} - E_t)\left[\sum_{j=1}^{\infty} \rho^j r_{t+j+1}^i\right] = \frac{\rho}{1-\rho}(r_{t+1}^e - r_t^e). \tag{42}$$

Equation (42) directly follows from the assumption of the ICC that expected returns for all future periods are constant, an assumption inherent in the ICC approach. A researcher, however, does not observe $DRN_{t+1}$ and hence, he has to find a proxy. Given that the researcher already has a proxy for the ICC, he can just substitute his proxy from equation (41) into equation (42). This results in

$$\begin{aligned} \widehat{DRN}_{t+1}^k &= \frac{\rho}{1-\rho}(\hat{r}_{t+1}^k - \hat{r}_t^k) \\ &= \frac{\rho}{1-\rho}(r_{t+1}^e - r_t^e + VFE_{t+1}^k - VFE_t^k). \end{aligned} \tag{43}$$

Equation (43) is identical with the proxy from equation (34) that EM use in their study. Despite some criticism,[52] this is the correct proxy for DR news *given a specific ICC method*. If one believes that a particular ICC method is the correct proxy for the expected return next period, it

51 Reasons on why the cash flows are most likely measured with error are given in Chapter 2.

52 Easton and Monahan (2010) have a summary of the points of criticism brought forward against this DR proxy. They also defend it there.

directly follows from this belief that the DR news component is just a scaled difference of two subsequent ICC values, whereby the scaling factor is $\rho/(1-\rho)$. Of course, a researcher is well aware of the fact that his ICC proxy is only an approximation of the true underlying expected return for next period because he knows that it is plagued by measurement error $\mathsf{VFE}_t^k$. Yet he does not know how to obtain a better proxy, i.e., a proxy with better characteristics of the measurement error part $\mathsf{VFE}^k$.[53] Because if he did, his ICC proxy would not be his best guess to begin with and an evaluation of such a method would be unnecessary.

Therefore, I completely agree with EM on their DR news proxy. However, I show next that the inherent assumption that the evaluated ICC method is the best proxy a researcher has also determines the CF news proxy.[54]

To do so, we can replace the LHS of equation (20) with equation (36) and the RHS of equation (20) with $\hat{r}_t^k$ for $E_t[r_{t+1}]$ and the expression in equation (43) for the DR news part. Solving for the CF news part and acknowledging that we use proxies now, not the true unobservable variables, we get

$$\widehat{CFN}_{t+1}^k = \kappa + \frac{\rho}{1-\rho}\hat{r}_{t+1}^k - \frac{1}{1-\rho}\hat{r}_t^k + \rho p_{t+1} + (1-\rho)d_{t+1} - p_t. \quad (44)$$

Equation (44) defines $\widehat{CFN}_{t+1}^k$ as a function of the ICC of a specific method $k$ at the beginning and end of period $t+1$, respectively, the price changes within the period, and the dividend at the end of period $t+1$. I want to emphasize that this definition does not need any additional assumptions. Except for the well-accepted loglinearization, I make no further assumptions, for instance about the dividend or measurement error process. Setting the ICC as the constant discount rate for all future periods, as done in equation (38), is not an assump-

53 I am deliberately imprecise here because there is not a one-dimensional criterion to assess the measurement error, a point made by Lee, So, and Wang (2011) and discussed previously. For instance, a researcher might prefer an ICC method that has a larger (in absolute values), but constant measurement error over a method that has a smaller, but time-varying measurement error because the former would keep the ordering of expected returns in the cross-section.

54 Ang and Liu (2007) make a similar point. They show that given a dividend process, any one of the three variables – return volatility, expected returns, and price-dividend ratios – completely determine the other two. Here, I argue that given a dividend process assumed by a researcher and a market price, the expected returns, which are assumed to be constant, are completely determined by the CS return decomposition.

tion I need for my derivations, but the standard assumption within the ICC literature.

By putting equation (43) and (44) into equation (20), we get the following relation between realized returns $r_{t+1}$ in period $t+1$ on the one hand and the expected return proxy $\hat{r}_t^k$, the proxy for CF news $\widehat{CFN}_{t+1}^k$, and the proxy for DR news, $\widehat{DRN}_{t+1}^k$, on the other hand:

$$\begin{aligned} r_{t+1} =& \hat{r}_t^k \\ &+ \kappa + \frac{\rho}{1-\rho}\hat{r}_{t+1}^k - \frac{1}{1-\rho}\hat{r}_t^k + \rho p_{t+1} + (1-\rho)d_{t+1} - p_t \\ &- \frac{\rho}{1-\rho}\left(\hat{r}_{t+1}^k - \hat{r}_t^k\right). \end{aligned} \tag{45}$$

It is instructive to replace the three proxies with their true, but unobservable counterparts plus their measurement errors:

$$\begin{aligned} r_{t+1} =& r_t^e + VFE_t^k \\ &+ CFN_{t+1} + \frac{\rho}{1-\rho}VFE_{t+1}^k - \frac{1}{1-\rho}VFE_t^k \\ &- DRN_{t+1} - \frac{\rho}{1-\rho}VFE_{t+1}^k + \frac{\rho}{1-\rho}VFE_t^k. \end{aligned} \tag{46}$$

Equation (46) shows that the tautological relation between realized returns and its three parts can be maintained in the case in which possibly time-varying future expected returns are replaced by a constant ICC. The measurement error terms in this equation cancel out.[55]

Equation (45) proves that this tautological relation is also achievable empirically, i.e., in the case in which the ICC is measured with error due to the unobservable vector of expected dividends. It is obvious from this equation that the CF and DR news proxies are just as plagued with measurement error as is the ICC proxy. Yet, my point is that if one uses the tautological relation from equation (36), as EM do, then one also has to define the parts in a consistent way. And if one does so, the multivariate regression from equation (33) will always return betas of 1, no matter which ICC method is used. However, EM only define the DR news part as a function of the ICC proxy under evaluation. As I showed, this is inconsistent.

55 Note that this argument does not hold if one uses the proxy proposed by EM (see equation 35). In this case, the measurement error terms in the second and third line of equation (46) are not identical, which breaks the tautological relation. This is my main motivation to label the CF proxy by EM inconsistent.

But why is it inconsistent to use an independent CF news proxy? This is probably best answered with a short example: Suppose that the log realized return in one period is 20% and, for simplicity, no dividends are paid in that period and the expected return is 0%. That is, the returns are completely due to capital gains induced by information surprises. Suppose further that this positive return is driven equally by positive CF news and negative DR news. Moreover, a researcher does not have access to the true vector of future expected cash flows and returns, and therefore, does not know that CF and DR news were equally important for this period. Instead, he estimates with his specific methodology that the sum of the expected return proxy next period and the proxy for DR news is 15%. Now, what is his best estimate of the CF news proxy? Obviously, the answer is 5%. By contrast, if he applied a different method and obtained a value of 8% for the sum of the expected return next period and the DR news, his best guess would be 12% for the CF news part. This directly follows from the tautological relation between realized returns on the one hand and the three proxies on the other hand. If the researcher regresses the realized returns on his proxies, he will be able to perfectly explain the 20% for both methods. Of course, this does not mean that these two methods are perfect, it simply means that one can always offset any measurement error in one part by introducing an opposite measurement error in the other part.[56] By contrast, EM do not require that expected returns, CF news, and DR news sum up to realized returns. Therefore, they would not object to match observed realized returns of 20% with an ICC estimate of 10%, positive CF news of 10%, and negative DR news of 10%. I consider this to be inconsistent.

One could argue that it is better to use an independent CF news proxy instead of a consistent proxy plagued with added measurement error. In the example above, if the researcher could estimate a CF news proxy of 10% (i.e., the true CF news for this period) he should add this proxy to his regression. I totally agree with this ar-

56 This is a well known property of the CS return decomposition in the literature that estimates CF and DR news not with an ICC, but with a vector autoregression (VAR) approach. For example, Chen and Zhao (2009) highlight that every misspecification error to predict future expected discount rates necessarily shows up in the CF news term. Recently, Chen, Da, and Zhao (2013) entertain the ICC to separate capital gain returns into a CF and a DR news part. The argument made here can also be seen from their equations (3) to (5). Capital gain returns can be split up with any ICC method into the two parts. Only the percentage of each part that drive the capital gain returns differ from method to method. I introduce their approach in detail in Chapter 7.

gument, but I argue that it is not possible to find such a proxy. CF news are just as unobservable as the other two parts in the CS return decomposition approach, both ex ante and ex post. So before using such a CF news proxy, the researcher would want to evaluate its quality. To do so, the approach of EM recommends running a multivariate regression with proxies for the other two parts, thereby introducing a circularity argument. But what if the researcher has a prior belief that his CF news proxy is better than the consistent one that is dependent on the specific ICC method and hence, he does not need an empirical validation for it? In this case, he should not empirically evaluate inferior ICC methods, but instead propose a better ICC method based on his superior CF news proxy. Put differently, the CF news proxy on the one hand and the expected return proxy as well as the DR news proxy on the other hand are linked by a tautological relation, so the test of the quality of one is always a joint test of the quality of the other.[57]

Take the method proposed by Claus and Thomas (2001). They transform the first five IBES earnings forecasts into a residual income and assume from period 5 on that the residual income grows with the expected inflation rate. However, by using the CF news proxy as proposed by EM, a researcher inherits on the one hand the assumptions from the Claus and Thomas (2001) method and assumes on the other hand that an investor only uses analyst forecasts up to two years into the future – to compute the CF news (see equation 35). In terms of equation (46), a researcher uses two different sets of measurement errors $\mathsf{VFE}_t^k$.

It is obvious that both assumptions cannot hold simultaneously. One might object that the assumption made by Claus and Thomas (2001) is a necessary simplifying assumption, and not the true expectation generating process of an investor. That is, $\mathsf{VFE}_t^{CT}$ deviates from zero at least for some periods $t$. This is most certainly true, and the very reason one wants to evaluate ICC methods in the first place. But then, why use a CF news proxy that is just as incorrect as **and** inconsistent with the ICC method? The counter argument that the

57 This argument only holds, as shown before, in the case of the ICC, i.e., under the assumption that future expected returns are constant. In this case, the expected return proxy for next period and the DR news proxy are inherently linked, a point made before by Pástor, Sinha, and Swaminathan (2008) and Chen, Da, and Zhao (2013). In the more general case in which a researcher has only a proxy for next period's expected return and no such proxy for DR news, there are now two unknowns and an expected return proxy does not pin down the CF news proxy, but in this case the approach of EM is not applicable either.

proxy might be better than the one derived to be consistent with the ICC introduces a circularity argument: if we knew that it would be a better proxy, we could then extract a better ICC method in the first place. But since we do not know it and we have no way of evaluating that proxy, we should not use it to evaluate another proxy. Easton and Monahan (2010, p. 9) bring this argument up themselves, just in a different context in which they discredit the risk factor evaluation approach introduced in Section 3.1 by writing:

> Stated another way, it is illogical to evaluate the reliability of one proxy by comparing it to another set of proxies that may also be unreliable.

In summary, my analysis shows that measurement error per se does not break the tautological relation between realized returns on the one hand and any ICC proxy as well as CF and DR news proxies on the other hand, as long as the latter two are defined in a consistent way. EM, however, only define DR news in a consistent way. If one uses consistent proxies, the relation between *any* ICC proxy and subsequent realized returns will *always* erroneously indicate a perfect ICC method. As a side note, this is exactly what Chen, Da, and Zhao (2013) do, abstracting from minor methodological differences. They estimate the CF and DR news part, thereby assuming that their ICC method is measured without error. I describe their approach in detail in Chapter 7.

## 3.5 DISCUSSION

In this chapter, I discussed several empirical methods that have been proposed to select the best proxy among the many specifications available to the applied researcher. If one uses realized returns as a proxy for expected returns, there is no leeway for an econometrician in measuring this proxy because it is observable. Hence, the question of selecting the correct proxy is irrelevant.

By contrast, as the previous chapter has shown, there are a multitude of alternative proxies to choose from. In the case of the ICC alone, there are now several dozen combinations of methods and input parameters that are entertained in the literature. Furthermore, if anything, this number will increase in the future because the literature continues to expand, as can be seen from recent modifications by Hou, Dijk, and Zhang (2012), Chen, Da, and Zhao (2013), and Mohanram and Gode (2013).

This is a severe problem, as it increases the likelihood of finding a relation between a variable of interest and a proxy by chance. This problem is nicely illustrated by Freedman's paradox.[58] In a simple Monte Carlo simulation, he shows that predictors that are uncorrelated with the dependent variable can appear artificially important. This happens if there are many such predictors to choose from and the sample size is comparatively small. Of course, this is an extreme case of "number crunching" in which a researcher can choose from a large set of predictors at will. In the case of expected return proxies, a researcher is constrained by economic theory that requires the proxy to be meaningful in the first place. Nevertheless, the previous chapter has shown that even in the boundaries of reasonable proxy specifications there are many different proxies with different measurement error processes to choose from. Furthermore, I show in the empirical part of this thesis that the use of different proxies leads to noteworthy differences in the results. There is also the additional problem that these processes might not be white noise, as Freedman (1983) assumes to show his paradox, but are actually related to the variable of interest in the research question at hand. Thus the issue of "data fishing", "data mining" or "number crunching", as Leamer (1978) called it, is a relevant one in the case of expected return proxies.[59]

Easton and Monahan (2005) and Lee, So, and Wang (2011) acknowledge this problem and propose a first step of a solution. They argue that an external validation of any proxy is necessary to prevent such data fishing. In my opinion, this is an important insight and contribution to the literature, but it has been mostly ignored by empirical research.[60]

58 See Freedman (1983) and Lukacs, Burnham, and Anderson (2010) for a recent discussion.

59 Manski (2007, p. 7) gives a reason why such data fishing might occur: "Forthright acknowledgement of ambiguity should be the norm, but it is distressingly rare. The scientific community rewards those who produce strong novel findings. The public, impatient for solutions to its pressing concerns, rewards those who offer simple analyses leading to unequivocal policy recommendations. These incentives make it tempting for researchers to maintain assumptions far stronger than they can persuasively defend, in order to draw strong conclusions." Often, this phenomenon is called the "publication bias" (cf., e.g., Harvey, Liu, and Zhu 2013).

60 To be more precise, it is ignored in the literature that simply wants to apply alternative expected return proxies to answer questions about the relation between expected returns and a variable of interest. Studies that try to improve upon expected return proxies have to justify that their proposals actually work. To do so, these studies use the evaluation methods presented in this chapter.

The ignorance is most probably due to the fact that this evaluation again relies on realized returns. Alternative proxies are meant to replace realized returns, so econometricians have an objection on relying on those proxies again to evaluate their alternative proxies. In the end, what could one learn from such alternative proxies that could not be learned from realized returns in the first place? Furthermore, the reason we want to replace realized returns, i.e., their large unsystematic shocks, also leads to weak power of those evaluation tests. In brief, just as we are unable to get precise estimates of expected returns with realized returns in short samples, we are also unable to identify the best alternative proxy with realized returns in such short samples.

Currently, there is no way to incorporate this uncertainty in the selection of the best proxy into the statistical inference. It is therefore not surprising that researchers prefer to select a proxy they deem more reasonable personally rather than to rely on such a weak test. However, this introduces the problem again that a researcher might not have chosen a proxy he deems more reasonable, but simply one that better supports his research question.

Therefore, the current procedure is unsatisfactory because it can lead to overconfident and biased results. In the next chapter, I introduce a method that incorporates the information of *all* proxies simultaneously and automatically, thereby alleviating concerns of data fishing. It is an extension of the evaluation approaches introduced here that does not just select a single best proxy, but weighs the evidence across all proxies based on how well they explain subsequent realized returns. Therefore, the uncertainty in the selection process is acknowledged.[61] Because the problems with realized returns are a small-sample problem, in large enough samples we can identify the best proxy with a probability that converges to 1. Thus a model selection approach and my model averaging approach are identical asymptotically. The same is not true for small samples, in which a model selection approach leads to incorrect results. Additionally, I am able to work out a caveat that applies to all alternative proxies. We are only able to make relative comparisons between different proxies,

61 This idea is related to the study of Pástor and Stambaugh (1999). They acknowledge that an investor who wants to estimate expected returns from a factor-based asset pricing model such as the CAPM or the Fama-French three-factor model faces great uncertainty about the correct parameter values within the models and about the correctness of the models themselves. They apply a Bayesian framework as well to incorporate this uncertainty into the statistical inference, although their focus is on the parameter uncertainty within each model.

but not absolute comparisons between any proxy and true expected returns. Hence, results on alternative proxies might be biased even asymptotically, a shortcoming a researcher has to live with if he wants to replace realized returns.

# 4 BAYESIAN MODEL AVERAGING AND EXPECTED RETURN PROXIES

In the last chapters, I argued that the evidence from the performance evaluation of expected return proxies is mostly ignored in applied research due to the large uncertainty inherent in this evaluation. Instead, researchers typically choose one or a few proxies in an ad-hoc manner.

In Section 4.1 of this chapter,[62] I show in a simple setup that such a course of action can severely bias the empirical results because a researcher ignores the large uncertainty he has about the correctness of his proxies. As the following quote illustrates, this issue is well known in the model selection and model averaging literature:

> However, even when we do objective, data-based model selection [...], the selection process is expected to introduce an added component of sampling uncertainty into any estimated parameter; hence classical theoretical sampling variances are too small: They are conditional on the model and do not reflect model selection uncertainty. One result is that conditional confidence intervals can be expected to have less than nominal coverage.[63]

Fortunately, model averaging is a solution to this problem that allows the incorporation of such uncertainty into the statistical inference. In Section 4.2, I introduce the most prominent averaging approach, BMA, and apply it to the proxy selection problem. The basic idea of the BMA approach is to average across the evidence of each proxy, conditional on the relative weight this proxy should have in comparison to its competitors. In Section 4.3, I discuss the computation of these model weights in detail. Section 4.4 summarizes this chapter.

Model uncertainty and issues arising from it have recently gained attention in the finance literature and approaches similar to mine that also rely on Bayesian statistics have been used to address these issues.

62 This chapter is based on Jäckel (2014).
63 Cf. Burnham and Anderson (2002, p. 156).

For example, Pástor and Stambaugh (1999), Pástor (2000), and Pástor and Stambaugh (2000) use a Bayesian framework to consider the uncertainty an investor has in choosing the correct factor-based asset pricing model and issues that arise out of this uncertainty, such as estimating the cost of equity capital or selecting the portfolio weights. Cremers (2002) and Avramov (2002) use BMA to evaluate the evidence of return predictability of many predictors simultaneously. Pástor and Stambaugh (2012) show that stocks are more volatile in the long run from an investor's perspective than commonly believed due to various uncertainties that the investor faces.[64] I contribute to this growing literature by introducing the issue of uncertainty into the proxy variable selection problem.

## 4.1 MOTIVATION

Suppose we are interested if the time series of variable $x_t$ is related to expected returns, $E_t[r_{t+1}] \equiv \mu_t$. For illustration purposes, suppose the following linear relation:[65]

$$\mu_t = \gamma x_t + \epsilon_t, \qquad t = 1, \ldots, T. \tag{47}$$

Using a univariate, classical regression setting with the assumption of normally distributed mean zero errors $\epsilon_t$, we can regress $\mu_t$ on $x_t$ and test if the estimator $\widehat{\gamma}$ is significantly different from zero. Because $\mu_t$ is unobservable, the researcher has to identify observable proxies, one of which is realized returns $r_{t+1}$.

As discussed in the previous chapter, it is easily shown via equation (21) that the unconditional mean of $r_{t+1}$ is equal to the unconditional mean of $\mu_t$. In other words, realized returns are an unbiased estimator of expected returns. Despite this advantageous characteristic, realized returns have come under criticism due to the fact that the variation in $u_{t+1}$ is an order of magnitude larger than the variation in $\mu_t$. This makes statistical inference notoriously difficult, as highlighted

64 Pástor and Veronesi (2009) give an excellent introduction into the literature on learning in financial markets. This literature deals with the uncertainty an economic agent has about the correct parameters and models, how they have to learn about them over time and how this affects the results. My research deals with one type of uncertainty in particular, namely the uncertainty about the correct specification of alternative expected return proxies.

65 For simplicity, I assume that the intercept is zero.

earlier. Plugging equation (21) into (47), the sampling variance of the estimator based on realized returns, $\widehat{\gamma}_{rr}$, is given by

$$Var(\widehat{\gamma}_{rr}) = \frac{Var(\epsilon_t) + Var(u_{t+1})}{\sum_{t=1}^{T}(x_t - E[x])^2}. \tag{48}$$

In small samples, the large variance of $u_{t+1}$ results in a large sampling variance for $\widehat{\gamma}_{rr}$ that can hinder the detection of an existing relation between $\mu_t$ and $x_t$. This insight has led to a "proxy variable search" with the hope of identifying alternative proxies for expected returns that are not plagued with the large noise inherent in realized returns. One particularly fruitful proxy class has been the ICC that was introduced in Chapter 2.

Unfortunately, these proxies are subject to measurement error. I ignore the additional scaling factor $s$ here and assume that a proxy $\widehat{\mu}_{t,k}$, measured at time $t$, is tracking $\mu_t$ with an additional, additive proxy-specific error term $w_{t,k}$:

$$\widehat{\mu}_{t,k} = \mu_t + w_{t,k}, \qquad t = 1, \dots, T. \tag{49}$$

If we run regression (47) with a proxy as defined in equation (49), the resulting regression coefficient $\widehat{\gamma}_k$ converges asymptotically to the true coefficient, $\gamma$, and an additional bias term:

$$\begin{aligned}
\gamma_k &= \frac{Cov(\widehat{\mu}_{t,k}, x_t)}{Var(x_t)} \\
&= \frac{Cov(\mu_t + w_{t,k}, x_t)}{Var(x_t)} \\
&= \frac{Cov(\mu_t, x_t) + Cov(w_{t,k}, x_t)}{Var(x_t)} \\
&= \frac{Cov(\gamma x_t + \epsilon_t, x_t)}{Var(x_t)} + \frac{Cov(w_{t,k}, x_t)}{Var(x_t)} \\
&= \gamma\frac{Var(x_t)}{Var(x_t)} + \frac{Cov(w_{t,k}, x_t)}{Var(x_t)} \\
&= \gamma + \frac{Cov(w_{t,k}, x_t)}{Var(x_t)} \\
&= \gamma + Bias_k.
\end{aligned} \tag{50}$$

Obviously, the hope of a researcher who applies proxy $k$ is that the mean and variance of $w_{t,k}$ are close to zero. In this case, the researcher is able to detect a relation between $x_t$ and $\mu_t$ much more precisely, compared to the analysis that employs realized returns. The danger, however, is quite obvious as well: if $w_{t,k}$ is systematically correlated with $x_t$, or if by chance the two comove in sample, then one might incorrectly deduce a relation between $\mu_t$ and $x_t$ from the data, although this relation is solely due to the specific measurement error of the proxy under consideration.

This simple example provides a nice motivation for the evaluation methods introduced in Chapter 3. Studies such as Easton and Monahan (2005) and Lee, So, and Wang (2011) realize that we need an *external* validation of the quality of alternative proxies. Furthermore, these studies show that such an external validation exists in the form of predictive regressions.

However, I argued before that predicting subsequent realized returns is notoriously difficult. Until now, there has been a heated debate if *any* predictor, not just expected return proxies, can actually predict realized returns.[66] We therefore have a large number of proxies to choose from and we are unable to precisely determine which of these proxies is best. In short, we face large model uncertainty. As a solution to deal with this uncertainty, I introduce model averaging to the expected return proxy literature in the next section.

## 4.2 MODEL AVERAGING APPROACH

Within a Bayesian framework, Leamer (1978) shows that the posterior distribution of a quantity of interest $\Delta$ can be computed, given data D, as the average of the posterior distributions under each model, weighted by their posterior model probability.[67] If we interpret each proxy as a separate model, we therefore get

$$p(\Delta|D) = \sum_{k=1}^{K} p(M_k|D)p(\Delta|D, M_k), \tag{51}$$

where $M_k = M_1, \ldots, M_K$ are the models considered. Equation (51) implies that the marginal distribution of the parameter of inter-

66 For recent surveys of the literature, see Koijen and Van Nieuwerburgh (2011) and Cochrane (2011).

67 The most popular derivative of model averaging is the one based on Bayes' theorem, which is also the foundation in Leamer's work. This approach subsumes under the name of Bayesian model averaging. There are also frequentist alternatives based on information-theoretical criteria such as Akaike information criterion (AIC) or Bayesian information criterion (BIC) (cf. Claeskens and Hjort 2008). As I show later on, due to the simplicity of my setup both approaches yield identical results. Therefore, the debate about differences between the two approaches is not an issue for the purpose of my approach. Still, in line with most studies that focus on model averaging techniques, I use a Bayesian setting to motivate my approach. Appendix A.4 gives an introduction into Bayesian statistics.

est is a mixture distribution. The mixture probabilities are the posterior model weights, $p(M_k|D)$, and the individual distributions are the distributions of the parameter of interest, conditional on a specific model.

In this simple framework, the only difference between two models $M_k$ and $M_j$ is that the proxy $\widehat{\mu}_k$ is replaced with $\widehat{\mu}_j$.[68] Additionally, D is split into two parts here: the part $D_{RQ}$ is the part of the data needed to answer the research question at hand. In the simple setup introduced in the previous section, $D_{RQ}$ consists of the matrix of expected return proxies and $x_t$. $D_P$ is the part of the data needed to compute the posterior probability of each proxy measuring true expected returns, given that one of the proxies is indeed correct. As argued before, this data consists of the set of proxies under consideration and subsequent realized returns. The separation of the data set is a direct consequence of the previous discussion: if we evaluate the quality of a proxy in terms of how well it explains the research question at hand, we run the risk of finding spurious relations driven by measurement error, and not by true expected returns. This is the main differentiation between my approach and other studies that apply BMA: Since measurement error is not such an obvious problem in other studies, a model is considered to be superior if it is better able to explain the research question.[69] In contrast, I use subsequent realized returns to infer the posterior model weights $p(M_k|D_P)$, which is a measure of the quality of the proxy.[70] This computation is independent of the specific research question. Results for each proxy are then obtained for the research question and averaged across the proxies based on the model weights.

68 This is important for the reader to keep in mind because I use the words proxies and models somewhat interchangeably. The latter is often used to conform with the language of the model selection and averaging literature.

69 For example, Fernandez, Ley, and Steel (2001) and Sala-I-Martin, Doppelhofer, and Miller (2004) employ BMA to test the robustness of explanatory variables in cross-country economic growth regressions. Since the literature has come up with a multitude of possible regressors, the question arises which combinations of those regressors help in explaining cross-country economic growth and how to take issues of model uncertainty into consideration. In this application problems of measurement error are ignored. As a consequence, the better a model explains the dependent variable, the higher its posterior model probability will be.

70 The computation of the weights is discussed further below.

To emphasize the separation between the posterior model weight computation and the subsequent statistical inference, equation (51) can be rewritten as

$$p(\Delta|D_{RQ}, D_P) = \sum_{k=1}^{K} p(M_k|D_P) p(\Delta|D_{RQ}, M_k). \tag{52}$$

The posterior distribution of the parameter of interest, $p(\Delta|D_{RQ}, M_k)$, can be derived from a Bayesian perspective. That is, one would specify the prior for this parameter, $p(\Delta|M_k)$, and the likelihood conditional on the parameter, $p(D_{RQ}|\Delta, M_k)$. The posterior is proportional to the product of likelihood and prior. While this seems to be a daunting task at first glance, it is greatly simplified by the fact that each model has the same interpretation in the case of expected return proxies. Since each proxy wants to measure the same, all parameters have the same interpretation. In this thesis, I choose a simpler approach that is more in line with current practice that mostly applies frequentist approaches: For the posterior distribution of the parameter of interest I use the sampling distribution from the frequentist approach. For example, if we run a time series regression as specified in (47) and are interested in the slope coefficient, $p(\Delta|D_{RQ}, M_k)$ is just the sampling distribution of the slope coefficient from a regression of a specific proxy on $x_t$. These distributions are easily adjusted to incorporate heteroskedastic or autocorrelated error structures, as will be shown in the empirical examples in Chapter 7.

Coming back to the research question from the previous section, the first two moments of $\widehat{\gamma}$ can then be calculated as[71]

$$E[\widehat{\gamma}_{BMA}|D_{RQ}, D_P] = \sum_{k=1}^{K} p(M_k|D_P) \widehat{\gamma}_k, \tag{53}$$

and

$$\begin{aligned} Var(\widehat{\gamma}_{BMA}|D_{RQ}, D_P) = & \sum_{k=1}^{K} p(M_k|D_P) Var(\widehat{\gamma}_k|D_{RQ}, M_k) \\ & + \sum_{k=1}^{K} p(M_k|D_P)(\widehat{\gamma}_k - E[\widehat{\gamma}_{BMA}|D_{RQ}, D_P])^2. \end{aligned} \tag{54}$$

While the mean estimate across all models is simply a weighted average across the estimate of each model, the variance of the com-

71 Leamer (1978) provides a derivation of these results.

bined estimate $\widehat{\gamma}_{BMA}$ exceeds a weighted average of the variances of the estimates *within each model* by an amount that depends on the variability of the estimates *across models*. Consider a case in which $Var(\widehat{\gamma}_k|D_{RQ}, M_k)$ is quite small for all models, i.e., conditional on a certain proxy the regression coefficient is measured accurately, but across proxies, the coefficients vary widely. In such a case a researcher would severely underestimate the variability of the parameter of interest if he was only to focus on one proxy.

In the case of the proxy literature, it is an apparent advantage (see, for instance, Lee, Ng, and Swaminathan 2009) that the statistical inference is much sharper due to considerably lower standard errors. This statement, however, is often based on evidence for one proxy. Equation (54) shows that the variance could indeed be much larger if the coefficients between different proxies differ substantially.

By plugging equation (50) into equation (53) and (54) and rearranging, we can express $E[\widehat{\gamma}_{BMA}|D_{RQ}, D_P]$ and $Var(\widehat{\gamma}_{BMA}|D_{RQ}, D_P)$ as follows:

$$E[\widehat{\gamma}_{BMA}|D_{RQ}, D_P] = \widehat{\gamma} + Bias_{BMA}, \tag{55}$$

and

$$\begin{aligned} Var(\widehat{\gamma}_{BMA}|D_{RQ}, D_P) = \sum_{k=1}^{K} p(M_k|D_P) Var(\widehat{\gamma} + Bias_k) \\ + \sum_{k=1}^{K} p(M_k|D_P)(Bias_k - Bias_{BMA})^2, \end{aligned} \tag{56}$$

where $Bias_{BMA} = \sum_{k=1}^{K} p(M_k|D_P) Bias_k$. Equation (55) and (56) are instructive representations of the discussion above. First, if one of the proxies is measured without error, we want its posterior model probability to approach unity. In this case $E[\widehat{\gamma}_{BMA}|D_{RQ}, D_P]$ and $Var(\widehat{\gamma}_{BMA}|D_{RQ}, D_P)$ will converge to $\widehat{\gamma}$ and $Var(\widehat{\gamma})$. Second, if the bias over all models varies randomly around zero and all proxies get equal support in the data, the average estimate across the models, $\widehat{\gamma}_{BMA}$, will be unbiased, but there is considerable model uncertainty that is automatically incorporated into the BMA analysis. Conversely, if an econometrician only examines a subset of the proxies, one might end up with biased estimates. Third, if all proxies under consideration are systematically biased, BMA will fail.[72] Finally, all approaches

72 Therefore, it is a commonly made assumption in the BMA literature that the true model is part of the set of models considered (cf., e.g., Lunn et al. 2012, Chapter 8).

that base their results on only one proxy, whether this proxy is chosen ad-hoc or by its ability to predict subsequent realized returns, ignore the variability of the parameters from different models. This leads to overoptimistic decisions and can result in the false identification of seemingly robust relations.[73]

## 4.3 COMPUTATION OF POSTERIOR MODEL WEIGHTS

I follow Avramov (2002), Cremers (2002), and Binsbergen, Hueskes, et al. (2013), which are all studies that run predictive regressions in a BMA framework to compute posterior model weights. Consider a set of $k$ linear univariate models $M_k = M_1, \ldots, M_K$ with the $k$th model be given by

$$r_{t+1} = \beta_0 + \beta_1 \widehat{\mu}_{t,k} + \varepsilon_{t+1}, \qquad t = 1, \ldots, T, \tag{57}$$

where $\varepsilon_{t+1}$ is assumed to be identically, independently, and normally distributed with mean zero and unknown variance $\sigma^2$. Furthermore, I assume that the regressors are strictly exogenous. This assumption is clearly false because both the ICC at time $t$ and realized returns in period $t+1$ are dependent on asset prices at time $t$ (cf. Stambaugh 1999). Nevertheless, it is a commonly made assumption in the literature (cf., e.g., Cremers 2002 and Wright 2008) and previous studies have argued that they expect the introduced bias to be small. Moreover, it is an issue that is typically ignored in the model selection approaches introduced in Chapter 3 as well.

73 In Appendix A.5, I introduce a Monte-Carlo exercise that compares the performance of the BMA approach with current selection procedures. I simulate the relations between a variable of interest, latent expected returns, and expected return proxies for four different specifications in a simple setup. The simulation results confirm the statements here. In short samples, BMA can severely decrease the bias in estimates and increase the coverage, i.e., BMA results in coverage regions that include the true underlying parameter more often. Alternative approaches, such as using the proxy that shows the strongest relation with the variable of interest or averaging across the proxies before the analysis, perform worse for those four specifications. In untabulated results, I repeat the simulation for many more specifications and find that alternative selection procedures perform mostly worse, sometimes equally well, and almost never better. The cases in which they perform better are cases of random measurement error added to the proxies and in which there is an actual relation between the variable of interest and true expected returns. In these cases, the attenuation bias is compensated by the overestimation due to selecting the best proxy. Hence, in these cases the BMA approach is more conservative.

In general, the posterior model probability for model $k$ is computed, given data $D_P$, via Bayes' theorem as

$$p(M_k|D_P) = \frac{p(D_P|M_k)p(M_k)}{\sum_k p(D_P|M_k)p(M_k)}, \tag{58}$$

where

$$p(D_P|M_k) = \int\int p(D_P|\beta_k, \sigma^2, M_k)p(\beta_k|\sigma^2, M_k)p(\sigma^2)d\beta_k d\sigma^2. \tag{59}$$

Therefore, we have to specify two priors. First, a prior about the probability of each model, $p(M_k)$. Second, priors about the parameters $\beta = (\beta_0, \beta_1)$ and $\sigma^2$. Both cases can be tricky if the number of explanatory variables differs between models and if the parameters' interpretation changes from model to model.[74] In my case, however, this is not an issue because each model has only one explanatory variable and the interpretation in each model is the same. The default assumption about $p(M_k)$ is to give each model the same weight a priori:

$$p(M_k) = \frac{1}{K}. \tag{60}$$

I use the same priors as Wright (2008) and Binsbergen, Hueskes, et al. (2013). That is, I make the assumption that $\beta$ takes the natural conjugate g-prior specification proposed by Zellner (1986). The prior on $\beta$ conditional on the variance of the error term $\sigma^2$ is therefore given as $N(0, \phi\sigma^2(X_k'X_k)^{-1})$, where $\phi$ is a shrinkage parameter that controls the informativeness of the prior and $X_k$ is the $T \times 2$ matrix of a $T$ vector of ones and the $T$ vector $\widehat{\mu}_k$. Since $\sigma^2$ is identical across models, we can use an improper prior of an inverse gamma (0,0) that is proportional to $1/\sigma$. Then, the posterior model weights can be computed from[75]

$$p(M_k|D_P) \propto \left(r'r - \left(\frac{\phi}{1+\phi}\right) r'X_k(X_k'X_k)^{-1}X_k'r\right)^{-T/2}, \tag{61}$$

where $r \equiv (r_2, \ldots, r_{T+1})$ denotes the vector of subsequent realized returns. Finally, we just have to normalize equation (61) so that all model weights sum to one.

74 For a discussion of these issues, see, for instance, Ley and Steel (2009).

75 I can ignore all terms here that are constant across models because they cancel out in equation (58).

The parameter $\phi$ governs the informativeness of the researcher's prior information. The lower $\phi$, the more weight is put on prior information. In the limit, if $\phi = 0$, $p(M_k|D_p)$ is equal for all models, i.e., the posterior probabilities are identical to the prior probabilities $p(M_k)$.

To provide a link with frequentist approaches and to get rid of the subjective aspects of the prior assumptions, we can increase $\phi$ to reduce the impact of the priors. In the limiting case, i.e., $\phi \to \infty$, the posterior model weights in equation (61) become proportional to $(SSE)^{-T/2}$, where $SSE$ is the sum of squared errors from an ordinary least squares (OLS). This result is also derived by Leamer (1978) who is in search of a reasonable diffuse prior. Because the sum of squared errors is proportional to the negative of the $R$ squared, the weights can be computed from the latter instead as well.

Furthermore, it is easy to show that in this case the weights computed from equation (61) are identical to the weights that would be obtained from information-theoretical approaches that use AIC or BIC based on the following formula:[76]

$$p_{AIC}(k) = \frac{\exp(0.5\Delta_{AIC,k})}{\sum_i \exp(0.5\Delta_{AIC,i})}, \tag{62}$$

where $\Delta_{AIC,k} = AIC_k - \max(AIC_1, \ldots, AIC_K)$. This subtraction is made merely for computational reasons. In equation (62), we can replace AIC with BIC; since the model sizes are identical across models, the penalty term that normally differs between AIC and BIC does not matter.

To summarize, both a noninformative Bayesian approach as well as a frequentist approach yield identical results due to the simplicity of the setup (univariate linear regression for each model). Consequently, debates about which approach is superior are not relevant here and model weights, given the data, are easily computed. The better a proxy is able to explain subsequent realized returns, i.e., the lower the sum of squared errors, the more credible this proxy is in comparison to its competitors and the more weight a researcher should assign to it.

76 Claeskens and Hjort (2008) give a good introduction into frequentist approaches of model selection and averaging and also motivate the formula given here. George and Foster (2000) examine in detail under which assumptions the ordering of the posterior probability corresponds exactly to the ordering based on approaches such as AIC or BIC.

Due to the high level of noise inherent in realized returns, alternative proxies have been proposed. That is, the main motivation for these proxies is the replacement of realized returns. However, the previous analysis to infer the quality of these proxies has to rely again on the very same realized returns it wants to replace. In my opinion, this is a severe shortcoming of *any* alternative expected return proxy.[77]

A main contribution of this study is that the introduction of model averaging techniques allows me to shed light on this issue. For the sake of simplicity, let's focus only on two alternative proxies with equal prior probability. In this case, the Bayes factor can be interpreted as a summary of the evidence provided by the data in favor of one proxy, in comparison to another proxy.[78] In our case, the Bayes factor $BF$ is given by

$$BF_{12} = \left(\frac{SSE_2}{SSE_1}\right)^{T/2} = \left(\frac{Var(w_{t,2}) + Var(u_{t+1})}{Var(w_{t,1}) + Var(u_{t+1})}\right)^{T/2}. \tag{63}$$

As argued above, the main motivation of any alternative proxy is the large variation in realized returns induced by $u_{t+1}$. Thus, equation (63) will be dominated by the term $Var(u_{t+1})$, which means that in small samples it will be notoriously hard to separate proxies with low measurement error from proxies with large measurement error. This means that the weights will not converge quickly to the best proxies in small samples. Only if sample size increases, even SSE ratios that are close to one will eventually become large and reveal the superiority of one proxy over the other. Furthermore, it is often argued that $u_{t+1}$ can be correlated with other variables in sample and therefore,

77 This issue has been recognized before, but only in a qualitative way. For example, Guay, Kothari, and Shu (2011, p. 129) write: "Like Easton and Monahan (2005) and a large literature in finance, we use realized returns as a metric to assess the cost of capital estimates and the effectiveness of our proposed remedies. Although our returns-based tests are consistent with a large asset-pricing literature, we acknowledge that realized returns are a noisy proxy for expected returns, and that this is, in fact, an important motivation behind implied cost of capital measures. However, despite the limitations, we are unaware of a superior benchmark to *validate* [emphasis in the original] cost of capital measures that does not rely on realized returns."

78 Kass and Raftery (1995, p. 773) define the Bayes factor as "the posterior odds of the null hypothesis when the prior probability on the null is one-half." For a detailed discussion of the Bayes factor, refer to their article.

inferences based on it can be misleading.[79] Consequently, it might happen that inferior proxies get more weight in small samples.

## 4.4 DISCUSSION

This chapter has shown how it is possible to compute posterior model weights for each proxy. These weights denote the posterior belief a researcher has in the quality of a proxy, based on his prior beliefs and the evidence in the data. This evidence is solely based on how well the proxy is able to explain subsequent realized returns. Taking evidence based on how well a proxy would explain the variable of interest $x$ would erroneously favor proxies for which the measurement error is related to $x$, and not true expected returns.

This chapter has also shown that the proxy weights are affected by the large noise of realized returns. As I show in Chapter 7, this results in rather equally distributed weights across the ICC specifications. In other words, we cannot reliably differentiate between good or bad proxies because of the small samples that are available to us. This, however, directly implies that it is important to average across the evidence based on each proxy, instead of just relying on the evidence of a single proxy.

79 Fama and French (2002), for instance, argue that the high realized returns in the US stock market in the second part of the 20th century were a result of a series of negative discount rate news that generated positive shocks $u_{t+1}$ and not a result of high $\mu_t$ for this period. Additionally, Campello, Chen, and Zhang (2008) also run the predictive regressions of their expected return proxy and do not find a large explanatory power of their proxy. Instead of attributing this result as a negative sign for the quality of their proxy, they blame the shock structure in the sample for this result and conclude that it is therefore worthwhile to explore alternative proxies. Yet another indication of the sensitivity of the specific sample is the evidence presented in Koijen and Van Nieuwerburgh (2011) that the regression coefficients in predictive regressions are unstable over time.

# 5

# DATA AND DESCRIPTIVE STATISTICS

In this chapter, I discuss empirical issues when computing the ICC and show extensive descriptive statistics for a US sample ranging from 1985 to 2011. Section 5.1 describes the data sources of my data set. Section 5.2 presents descriptive statistics, both for the aggregate time series of the ICC and the cross-sectional variation in firm-level ICCs. In Section 5.3 I perform several sensitivity checks.

## 5.1 DATA SOURCES

In this dissertation, I focus on two derivatives of the residual income model (CT, GLS), two derivatives of the dividend discount model (PSS, CDZ), and three derivatives of the abnormal earnings growth model (PEG, MPEG, OJ). I also implement the forward PE ratio as a naïve benchmark. These methods are the most widely used ICC methods in empirical studies, as has been shown in Table 1. While the models were discussed in detail in Chapter 2, Table 2 summarizes the assumptions I make that are necessary to implement the models empirically. I emphasize that I do not exactly replicate the original studies that proposed the methods. For example, the GDP data of Pástor, Sinha, and Swaminathan (2008) already begins in 1930, while I use data that starts in 1947 in accordance with Chen, Da, and Zhao (2013). In brief, I try to use identical inputs across methods whenever possible. Nevertheless, my implementations should be reasonably close to the original studies because I am not aware of any major assumptions that I change. Furthermore, I compare my data with other studies below and the results are reasonably close.

Table 2: **Summary of empirical ICC proxies and data sources.** This table shows the underlying valuation model for each of the implemented ICC proxies. For each model, the equation referred to in column "Eq." and repeated in column "$P_t$ equals" is solved numerically for $R$. The share price of firm $i$ is obtained from IBES (the firm index is suppressed throughout this table); if available, $EPS_{t+k}$ is the consensus mean earnings per share forecast for year $t+k$ from IBES. To be included in the sample, I require at least two analyst forecasts. Additional forecasts, if not available, are inferred by multiplying the forecast for the previous year with the analysts' long-term earnings growth rate: $EPS_{t+k+1} = EPS_{t+k} \times (1 + Ltg_t)$. Expected dividends $DPS_{t+k}$ are computed as the product of $EPS_{t+k}$ and a payout ratio $PO_t$. If not defined differently, future payout ratios, expected at time $t$, are inferred from Compustat data as common dividends (DVC) divided by income before extraordinary items (IBCOM) for profitable firms. For firms with a negative income, 6% of total assets is used as the denominator instead. $BPS_t$ is the equity book value per share at the end of period $t$, computed as shareholder's equity (SEQ) divided by common shares outstanding from IBES. In cases after the fiscal-year end, but before the book value was reported, I infer $BPS_t$ from shareholder's equity of the previous year and the clean-surplus relation. The not-yet reported earnings per share are taken from IBES.

| Method | Eq. | $P_t$ equals | Implementation details |
|---|---|---|---|
| PE | (14) | $\frac{EPS_{t+1}}{R}$ | |
| PEG | (16) | $\frac{EPS_{t+2}-EPS_{t+1}}{R^2}$ | |
| MPEG | (17) | $\frac{EPS_{t+2}+R\times DPS_{t+1}-EPS_{t+1}}{R^2}$ | |
| OJ | (18) | $\frac{EPS_{t+1}}{R} + \frac{g_{st}\times EPS_{t+1}-R\times(EPS_{t+1}-DPS_{t+1})}{R\times(R-g_{AGR})}$ | $g_{st}$ is the average of the growth rate between $EPS_{t+1}$ and $EPS_{t+2}$ on the one hand and $Ltg_t$ on the other hand. $g_{AGR}$ is the yield on a 10-year government bond minus three percent. If this number is negative, I set $g_{AGR}$ to zero. |

| Method | Eq. | $P_t$ equals | Implementation details |
|---|---|---|---|
| CT | (8) | $BPS_t + \sum_{j=1}^{5} \frac{EPS_{t+j} - R \times BPS_{t+j-1}}{(1+R)^j} + \frac{(EPS_5 - R \times BPS_4) \times (1+g_{lt})}{(R-g_{lt}) \times (1+R)^5}$ | $g_{lt}$ is the yield on a 10-year government bond minus three percent. If this number is negative, I set $g_{lt}$ to zero. Book values per share of year $t+j$ are recursively inferred from book values of year $t+j-1$, the earnings per share of year $t+j$ and a constant payout ratio $PO_t$: $BPS_{t+j} = BPS_{t+j-1} + EPS_{t+j} \times (1 - PO_t)$. |
| GLS | (9) | $BPS_t + \sum_{j=1}^{11} \frac{(ROE_{t+j} - R) \times BPS_{t+j-1}}{(1+R)^j} + \frac{(ROE_{12} - R) \times BPS_{11}}{R \times (1+R)^{11}}$ | $ROE_{t+j} = EPS_{t+j}/BPS_{t+j-1}$ for $j = 1, 2, 3$. From $j = 4$ on, $ROE_{t+j}$ is linearly interpolated to $ROE_{12}$, which is defined as the median industry ROE (IBCOM/SEQ of previous year) over the last ten years for all profitable firms. I use the 48 Fama and French (1997) industry classification. $BPS_{t+j}$ is computed as in the CT method. |
| PSS | (10) | $\sum_{j=1}^{15} \frac{EPS_{t+j} \times PO_{t+j}}{(1+R)^j} + \frac{EPS_{t+16}}{R \times (1+R)^{15}}$ | From $k = 4$ on, earnings are computed as $EPS_{t+j} = EPS_{t+j-1} \times (1 + g_{t+j})$. The earnings growth rate in year $t+3$, $g_{t+3}$, is set to $Ltg_t$ and mean-reverted to a long-term growth value by year $t+17$, $g_{lt}$, with formula $g_{t+k} = g_{t+k-1} \times \exp(\log(g_{lt}/g_{t+3})/14)$. After the first two years the plowback rate, defined as $b_t = 1 - PO_t$, is recursively computed as $b_{t+k} = b_{t+k-1} - (b_{t+2} - b_{SS})/14$. $g_{lt}$ and the steady-state earnings growth, $g_{SS}$, are both set to the average nominal GDP growth rate estimated using an expanding rolling window starting from 1947. The steady-state plowback rate is given as $b_{SS} = g_{SS}/R$. |
| CDZ | (10) | As PSS | This method is identical to PSS except for $g_{lt}$, which is set to the industry-wide mean of long-term growth rates $Ltg$ by the analysts. From $t+17$ on, in the period in which $g_{lt}$ would be reached, it is instead assumed that earnings grow with $g_{SS}$ that are computed identical to the PSS method. I use the same industry classification as in the GLS case. |

### 5.1.1 *IBES*

Throughout this dissertation, I follow the majority of studies that implement the ICC and approximate expected earnings with analyst forecasts from IBES. This data source provides mean earnings forecasts up to five years ahead as well as a long-term earnings growth rate estimated by the analysts. While there is no exact definition of this growth rate, it is typically assumed to refer to earnings growth rates for the next five years (cf. Claus and Thomas 2001). With the exception of the OJ, the PSS, and the CDZ method it is only used to infer missing earnings forecasts from the previous forecasts and this earnings growth rate: $EPS_{i,t+j+1} = EPS_{i,t+j} \times (1 + Ltg_i)$. I fill missing earnings forecasts according to Chen, Da, and Zhao (2013), i.e., I compute missing $EPS_3$, $EPS_4$, and $EPS_5$ with the help of the long-term earnings growth rate $Ltg$, but I do not recover missing $Ltg$ values from the earnings forecasts, as done in other papers.

To get a better understanding of the coverage of the earnings forecasts, Table 3 shows the average number of monthly one-year ahead earnings forecasts available for each year as well as the percentage of two-to-five years ahead earnings forecasts and long-term earnings growth forecasts, in relation to the one-year ahead earnings forecasts. This table is produced from the raw IBES summary file without any additional filters that I set below.

The number of earnings forecasts for the next year indicates that IBES, at least in the US, covers many firms. Additionally, since analysts typically focus on larger firms, IBES represents 90% or more of the total US market capitalization (cf. Claus and Thomas 2001). The peak in firms covered was around the turn of the millennium, which could be driven by the Dot-com bubble that was accompanied by a large increase in publicly traded firms.[80] Moreover, Table 3 also shows that the coverage for further ahead forecasts declines dramatically. While two-year ahead earnings forecasts are available for at least 80% of the firms in each year, they are almost never available for forecasts for year 5. The coverage for long-term earnings growth rates is far better, which is why this growth rate is commonly used to fill in missing data. It can also be seen from this table that the coverage for further ahead forecasts has increased over time. For example, in 1985 five-year ahead forecasts were only available for 3% of the obser-

80 For example, Ritter and Welch (2002) show that the number of IPOs was on a very high level from 1994 to 2000 with an average of 420 IPOs per year, before the number falls abruptly to 80 IPOs in 2001.

Table 3: **Coverage of earnings forecasts in IBES.** This table shows the number of average monthly observations per year for which an earnings forecast for the next year is available. Additionally, it shows the percentage of available forecasts for the two-to-five years ahead forecasts as well as the long-term earnings growth rate (Ltg), in relation to the number of available one-year ahead earnings forecasts.

| | Number of | % of forecasts in relation to $EPS_1$ forecasts | | | | |
|---|---|---|---|---|---|---|
| Year | $EPS_1$ forecasts | $EPS_2$ | $EPS_3$ | $EPS_4$ | $EPS_5$ | Ltg |
| 1985 | 2978 | 81.25 | 14.31 | 7.58 | 3.26 | 76.32 |
| 1986 | 3050 | 82.24 | 10.56 | 2.59 | 1.72 | 76.16 |
| 1987 | 3315 | 82.50 | 9.91 | 2.81 | 1.33 | 74.43 |
| 1988 | 3448 | 82.28 | 12.12 | 4.31 | 2.15 | 69.34 |
| 1989 | 3517 | 83.83 | 20.95 | 11.33 | 5.05 | 68.00 |
| 1990 | 3430 | 83.87 | 16.26 | 5.37 | 2.53 | 69.30 |
| 1991 | 3361 | 85.97 | 19.10 | 6.08 | 2.99 | 69.91 |
| 1992 | 3526 | 88.72 | 25.90 | 12.62 | 9.50 | 72.14 |
| 1993 | 3943 | 90.67 | 30.77 | 19.08 | 15.07 | 74.21 |
| 1994 | 4432 | 91.60 | 31.41 | 19.09 | 14.40 | 74.81 |
| 1995 | 4857 | 89.62 | 24.36 | 10.36 | 5.23 | 72.24 |
| 1996 | 5407 | 89.15 | 23.51 | 6.04 | 1.45 | 74.16 |
| 1997 | 5842 | 89.88 | 24.20 | 3.82 | 1.30 | 76.88 |
| 1998 | 5923 | 89.62 | 25.64 | 4.18 | 1.81 | 77.25 |
| 1999 | 5755 | 88.98 | 25.32 | 4.40 | 1.76 | 75.88 |
| 2000 | 5442 | 87.13 | 26.14 | 5.16 | 1.50 | 75.52 |
| 2001 | 4806 | 87.47 | 28.80 | 5.49 | 1.55 | 74.97 |
| 2002 | 4155 | 94.56 | 32.28 | 7.40 | 0.88 | 80.28 |
| 2003 | 4084 | 96.12 | 40.57 | 14.10 | 9.44 | 82.12 |
| 2004 | 4201 | 98.22 | 51.30 | 21.02 | 15.39 | 78.61 |
| 2005 | 4399 | 98.18 | 56.04 | 21.31 | 14.16 | 75.98 |
| 2006 | 4538 | 97.95 | 60.01 | 22.13 | 15.91 | 73.07 |
| 2007 | 4656 | 97.07 | 62.42 | 26.48 | 16.87 | 70.26 |
| 2008 | 4443 | 97.53 | 62.32 | 27.06 | 20.40 | 70.07 |
| 2009 | 4171 | 97.86 | 66.05 | 39.17 | 26.86 | 66.69 |
| 2010 | 4227 | 98.48 | 73.48 | 40.57 | 25.70 | 65.89 |
| 2011 | 4141 | 99.69 | 73.09 | 32.05 | 22.11 | 69.05 |

vations with one-year ahead forecasts. In 2011, this ratio increased to 22%. Interestingly, the coverage ratio for long-term earnings growth remained roughly constant over time and even declined after 2003.

It is an interesting question if the systematic changes in coverage systematically bias the results. For example, if the earnings growth rates deviate from the earnings growth rate an analyst expects from year 2 to 3, but methods like the GLS, PSS, and CDZ method use this growth rate to infer the three-year ahead forecasts, then the ICC estimate can simply change due to a changing coverage of three-year ahead forecasts. As can be seen from Table 3, there was a large trend for this coverage ratio in particular.

Next to earnings forecasts, I also obtain the earnings announcement date, the stock price as well as the shares outstanding from IBES. The product of the latter two values serves as the market capitalization throughout this dissertation and the stock price is also needed as an input parameter to infer the ICC. It is important to use the same data vendor here to make sure that any adjustments, such as stock splits or share repurchases, are handled consistently. This also ensures that the different values refer to the same point in time. IBES does not release its forecasts at the end of the month, but on the Thursday before the third Friday of every month. This is important to keep in mind, particularly for research questions that relate the ICC to other monthly updated data. For example, in the analysis of Pástor, Sinha, and Swaminathan (2008), which I replicate in Chapter 7, the aggregate ICC is regressed on the market volatility each month. It is important to check the robustness of the results with respect to the exact point in time for which the data was computed.

### 5.1.2 *CRSP/Compustat*

I obtain all US firms from the CRSP/Compustat merged file from WRDS. CRSP provides me with the total monthly returns for each firm that are necessary for my cross-sectional analysis in the next chapter.[81]

81 Note that I also obtain the aggregated continuously compounded with-dividend returns on the value-weighted portfolio of all NYSE, Amex, and NASDAQ stocks from CRSP as a benchmark for market-wide realized returns. In Chapter 7, this benchmark is needed for the aggregated predictive regressions that I run and to compute the stock market volatility to replicate the results of Pástor, Sinha, and Swaminathan (2008).

For the computation of the ICC, additional data from Compustat is needed. I obtain shareholder's equity (item SEQ) to infer the book value per share. The number of shares is taken from IBES to make sure that split-adjustments are consistently applied to both the price and the book value per share. This is in line with the studies of Hail and Leuz (2009) and Gebhardt, Lee, and Swaminathan (2001) that obtain the stock prices and shares outstanding from one source as well.[82] Also, the payout ratios for most methods are assumed to be constant and equal to the historical ratios, as such, I divide common dividends (item DVC) by income before extraordinary items available to common shareholders (item IBCOM).[83] For firms with a negative or missing income, 6% of total assets (item AT) is used as the denominator instead. I winsorize the payout ratios at 0 and 1, respectively. The historical ROE is calculated as IBCOM divided by SEQ of the previous year. It is needed to infer the 12-year ahead ROE in the GLS method, which is defined as the median industry ROE (IBCOM/SEQ of previous year) over the last ten years for all profitable firms.[84]

### 5.1.3 *Additional data*

For the PSS and CDZ method, I also need GDP growth data starting from 1947. I obtain quarterly updated, annualized growth rates from the website of the Bureau of Economic Analysis. Also, as a proxy for interest rates, I use the nominal yield on a 10-year government bond. This data is available in the H15 report of the Board of Governors of the Federal Reserve System.

### 5.1.4 *Matching and computation issues*

For the residual income model I need book values in addition to earnings forecasts. As highlighted by Easton (2007, Chapter 13), certain

82 Hail and Leuz (2009) use data from IBES, while Gebhardt, Lee, and Swaminathan (2001) get their prices and shares outstanding from CRSP. As a sensitivity check that my results for the residual income models are not driven by a mismatch between book values from Compustat and prices from IBES, I repeat the empirical analysis of this thesis with an updated data set. In this data set, I delete the observations for which their price-book ratio lies within the top and bottom 1% from the cross-sectional distribution in each year. The results remain virtually unchanged.

83 The only exceptions are the PSS and CDZ method, for which the plowback rates, defined as 1 minus the payout ratio, are linearly interpolated to a steady-state plowback rate after year $t+2$.

84 Throughout this dissertation, I use the 48 Fama and French (1997) industry classification. This classification can be found on Kenneth French's website: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/Industry_Definitions.zip`.

data issues are encountered when matching prices, book values and future earnings on the same date. One such issue arises from the delay between the fiscal year-end and the reporting of actual earnings and book values. Because basing the computation on actual not-yet-reported book values would contradict the principle of using only publicly available information on the estimation date, I create synthetic book values until book values are publicly released. To approximate this release date, I use the earnings announcement date from IBES. If this date is not available, I set it to 120 days after the fiscal-year end instead.

Similar to Gebhardt, Lee, and Swaminathan (2001) and consistent with clean-surplus accounting, I generate synthetic book values based on previous book values, earnings and an estimated payout ratio. This ensures that I do not base my calculation on stale book values. However, the actual earnings are only available after the earnings announcement date. I therefore use the IBES earnings forecasts for the next period to again ensure that only publicly available information is used. In cases in which the IBES release date is before the earnings announcement date, the first forecast refers to the *previous* fiscal-year end.

Another issue is the misalignment of prices, book values, and earnings forecasts. Equation (2) in Chapter 2 is only valid if the price is measured on the same date that all expected dividends are discounted to, which happens to be the fiscal-year end for the firm. In the more common case where the price is taken from any day during the year, Daske, Gebhardt, and Klein (2006) propose a technique for the residual income model that computes the book value on the estimation date and adjusts the residual incomes accordingly.

It turns out that there is a simpler solution to this problem. Instead of adjusting the RHS of equation (2) and its derivatives, such as the abnormal earnings growth model and the dividend discount model, one can also adjust the LHS, i.e., the price (cf., e.g., Easton 2007). Suppose that the price is not measured at the fiscal-year end date $t$, but at any date $t + x$, where $x$ is a number between 0 and 365. For example, if one wants to compute the ICC for firm $i$ on January 20 and the fiscal-year end for this firm is at the end of December, $x$ would be 20. Because the dividends are discounted to date $t$, and not to date $t + x$, the price, which is measured at this date, has to be discounted to date t too. The correct discount rate in this model is just the ICC, which yields an updated version of the dividend discount model:

$$\frac{P_t}{(1+R_t^e)^{\frac{x}{365}}} = \sum_{j=1}^{\infty} \frac{DPS_{t+j}}{(1+R_t^e)^j}. \tag{64}$$

In the interest of remaining consistent with the literature, I ignore this issue in my main empirical analysis. In Section 5.3.2, I show that this decision has a minor effect on the statistics.

In summary, I match analyst forecasts and price data from IBES with accounting data from Compustat. Thereby, I ensure that I only use publicly available information. Now I have all of the information that is necessary to compute the ICC based on the different methods in Table 2.

Prior to computing the ICC, I filter all observations for which at least one of the following items is not available:

- IBES: price, shares outstanding, one-and-two-year ahead earnings forecasts, either three-year ahead earnings forecast or long-term earnings growth rate.
- Compustat: shareholder's equity (item SEQ), common dividends (item DVC), income before extraordinary items available to common shareholders (item IBCOM).[85]

As an additional filter, I delete all observations with a negative book value from Compustat. All other filters are subject to the specific ICC method. Because the abnormal earnings growth models anchor their valuation on capitalized expected one-year ahead earnings, I set the ICC for those methods to missing if the one-year ahead earnings forecast is negative. Other requirements, such as larger two-year than one-year ahead forecast for the PEG, result automatically in missing values because the ICC cannot be numerically solved for. This leads to a different number of observations for different ICC methods, which I show the statistics for further below.

I solve for all ICC methods numerically, even for those where an analytical solution is available.[86] I solve for the value of the ICC that sets the difference of the current price and discounted dividends to

85 If item IBCOM is negative or missing, the item total assets (item AT) has to be available to compute the plowback rate.

86 Note that equation (64), which I implement in Section 5.3.2, requires a numerical solution even for the methods for which an analytical solution would be available without the consideration of timing issues.

Table 4: **Summary statistics for aggregated ICC.** This table contains the summary statistics for the monthly time series of the ICC for different methods. Only observations are considered for which an ICC can be computed for each method. The firm-level ICCs are value-weighted (Panel A) or equal-weighted (Panel B). All numbers are reported in percent. The time period ranges from 1985 to 2011.

| | PE | PEG | MPEG | OJ | CT | GLS | PSS | CDZ |
|---|---|---|---|---|---|---|---|---|
| Panel A: Value-weighted implied cost of capital | | | | | | | | |
| Mean | 6.49 | 9.71 | 11.07 | 11.42 | 9.66 | 8.82 | 9.85 | 12.66 |
| Median | 6.40 | 9.74 | 11.02 | 11.02 | 9.24 | 8.75 | 9.82 | 12.68 |
| Std. dev. | 1.45 | 1.08 | 1.54 | 1.66 | 1.81 | 1.36 | 1.39 | 1.12 |
| Minimum | 3.73 | 7.61 | 8.21 | 9.14 | 7.11 | 6.03 | 7.29 | 10.52 |
| Maximum | 11.43 | 14.14 | 16.84 | 17.42 | 15.45 | 13.29 | 14.88 | 17.43 |
| Panel B: Equal-weighted implied cost of capital | | | | | | | | |
| Mean | 6.85 | 11.94 | 12.84 | 12.69 | 10.38 | 10.08 | 10.85 | 13.94 |
| Median | 6.73 | 12.08 | 12.91 | 12.68 | 10.47 | 9.96 | 10.93 | 14.10 |
| Std. dev. | 1.29 | 1.10 | 1.36 | 1.62 | 1.73 | 1.03 | 1.32 | 1.56 |
| Minimum | 4.78 | 9.68 | 10.24 | 10.00 | 7.52 | 8.37 | 8.46 | 10.81 |
| Maximum | 11.07 | 15.53 | 16.86 | 17.41 | 15.21 | 13.13 | 14.90 | 18.05 |

zero. I set the lower bound on the ICC estimate to 0.00001% for all methods but for the OJ and CT method. For those methods the lower bound is set to be 0.00001% larger than the long-term growth rate. I abort the root search as soon as the change in the ICC is less than 0.00001 percentage points for one step.

## 5.2 DESCRIPTIVE STATISTICS

This section presents summary statistics for the ICC methods that I implement. While I focus on the aggregate time series for each ICC method in subsection 5.2.1, I look at the cross-sectional variation for each method in subsection 5.2.2.

### 5.2.1 *Properties of the aggregate ICC*

Table 4 shows summary statistics for both the value-weighted and the equal-weighted time series of the ICC for each method. In this table as well as in the discussion to follow and also in the empirical analysis of Chapter 6 and 7, I require an observation to be complete, i.e., to have a numeric ICC value for each method. If one or more methods have a missing value, this observation gets filtered.[87]

87 Further below, I discuss differences in the number of observations across methods.

The mean of the value-weighted ICC varies from 6.5% for the PE method to 12.7% for the CDZ method. The low mean for the PE is due to the ignorance of any earnings growth after the next year. As mentioned before, this method is only meant to be a naïve benchmark and thus the true ICC during that period was probably substantially larger. At the other extreme, the high average for the CDZ method is driven by the high expected earnings growth rate that Chen, Da, and Zhao (2013) assume. The other methods all lie within a much closer interval between 8.8% and 11.4%. The ICCs based on the residual income models (CT, GLS) as well as the PSS method are on the lower end of this range, while the abnormal earnings growth methods have typically larger values. The only exception is the PEG method which assumes that dividends in the next year and the growth rate thereafter are zero. This finding is in line with results from Daske, Van Halteren, and Maug (2010) who provide evidence based on simulation results that derivatives of the abnormal earnings growth model lead to higher estimates for the ICC than derivatives of the residual income model. Empirically, Hail and Leuz (2009) find the same pattern for their international sample.

There are also noteworthy differences in the standard deviation of the different methods. The CT has the highest standard deviation with a value of 1.8%. A possible explanation can be given based on the CS loglinearization of prices, which decomposes the log price into a cash flow and a discount rate part (see equation 37). This relation is tautological, which means that one of the two parts has to explain a stock price. The approach for any ICC method is to directly model the CF part and back out the DR part as the residual. As a consequence, if a specific ICC method allows less variation in the CF part, it has to explain the remainder of the price changes by changes in the ICC. The CT anchors the valuation on a very persistent book value and estimates only the earnings for the next five years. After this explicit forecast period, it assumes that residual incomes grow with the expected inflation rate. This is a conservative assumption. Thus most of the variation in prices has to be explained by changes in the ICC and this drives the larger ICC variation in the CT method. By contrast, the standard deviation of the CDZ method is one of the lowest, which is explained by the assumption made by Chen, Da, and Zhao (2013) that earnings growth converges to the current industry long-term earnings growth rate assumed by analysts. Since this growth rate varies together with changes in prices over time, the majority of these price changes is already explained by the CF part for this method. Therefore, the ICC has to explain a smaller unexplained part and thus varies less.

Furthermore, the variation for each ICC method is an order of magnitude smaller than the variation in realized returns.[88] This is a standard finding in the literature and advertised as one of the ICC's main selling points (cf., e.g., Lee, Ng, and Swaminathan 2009).

With respect to the equally weighted time series, Panel B of Table 4 shows that these are typically larger than their value-weighted counterparts. Just as in the case of realized returns as a proxy for expected returns, investors seem to require a premium for small firms.[89] Nevertheless, the differences across methods are very similar. For example, the PE method results in the lowest average ICC and the CDZ method leads to the highest average for both weighting procedures. The standard deviations of the equally weighted time series are similar to their value-weighted counterparts as well.

Figure 1 plots the monthly value-weighted time series of the ICC for the eight methods. Not surprisingly, the CDZ/PE method has the highest/lowest ICC throughout the sample. Also, this plot nicely illustrates the above discussion about the differences in standard deviations between the methods. Most of the methods constrain future cash flow growth. This becomes especially apparent before and during the Dot-com bubble. The ever increasing prices during that period had to be explained by something. But because most methods shut down an exclusive explanation through updated beliefs in future dividends, they have to explain a part of the rising stock prices with lower discount rates. Therefore, all methods except the CDZ method are relatively low during the Dot-com period. Conversely, the CDZ method remains rather constant throughout this period since most of the price changes are explained by higher expected growth rates.

Figure 1 also shows that the correlation between the methods is quite high. In particular, all methods except the PE and CDZ method mostly move in unison. Also, all methods show a strong reaction to the financial crisis with a large rise in expected returns.

Table 5 presents the correlation between the value-weighted and equal-weighted time series of the eight methods under considera-

88 The annualized standard deviation of monthly realized returns on the value-weighted portfolio of all NYSE, Amex, and NASDAQ stocks was 16.1% for the same period.

89 Hanauer, Jäckel, and Kaserer (2013) examine the size premium in detail for an international ICC sample.

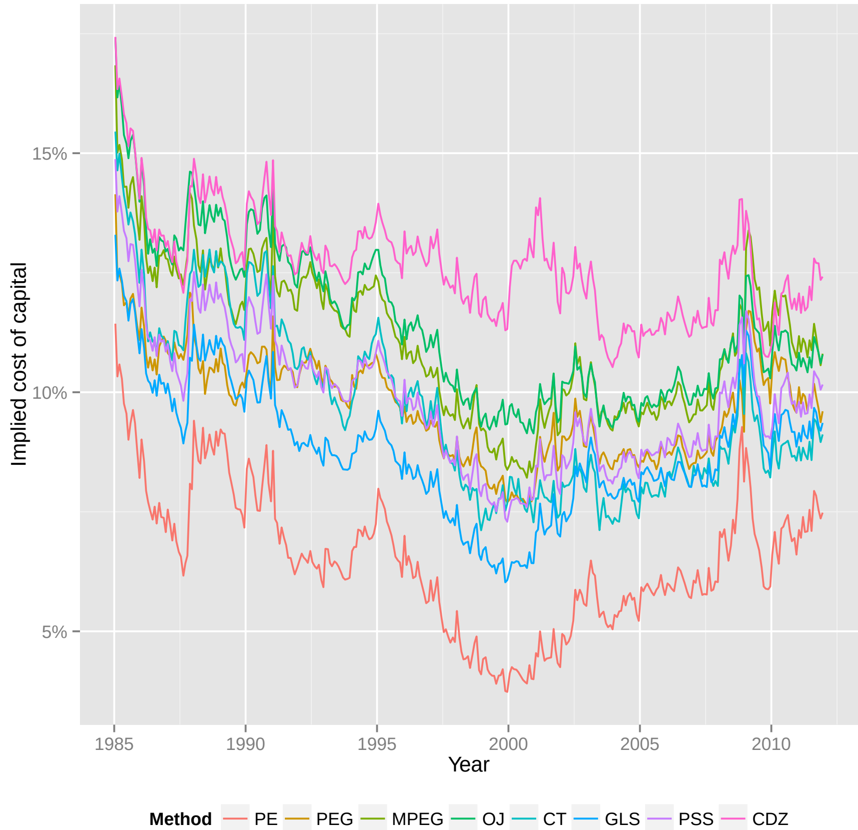


Figure 1: **Value-weighted ICC for different methods over time.** This figure shows the monthly time series of the ICC for eight different methods. The ICC is computed as a value-weighted average of ICCs across all US firms for which an ICC is available for all methods.

Table 5: **Correlations for aggregated ICC.** This table shows the correlation between the monthly time series of the ICC for different methods. Only observations are considered for which an ICC can be computed for each method. The firm-level ICCs are value-weighted (Panel A) or equal-weighted (Panel B). All correlations are reported in percent. The time period ranges from 1985 to 2011.

| | PE | PEG | MPEG | OJ | CT | GLS | PSS | CDZ |
|---|---|---|---|---|---|---|---|---|
| Panel A: Value-weighted implied cost of capital | | | | | | | | |
| PE | 100.00 | 85.39 | 89.22 | 86.81 | 87.55 | 98.12 | 96.21 | 71.92 |
| PEG | 85.39 | 100.00 | 98.93 | 91.90 | 86.25 | 89.25 | 91.74 | 69.57 |
| MPEG | 89.22 | 98.93 | 100.00 | 95.96 | 91.77 | 92.00 | 95.71 | 74.41 |
| OJ | 86.81 | 91.90 | 95.96 | 100.00 | 98.60 | 87.76 | 96.03 | 84.38 |
| CT | 87.55 | 86.25 | 91.77 | 98.60 | 100.00 | 86.87 | 95.55 | 86.20 |
| GLS | 98.12 | 89.25 | 92.00 | 87.76 | 86.87 | 100.00 | 95.85 | 69.08 |
| PSS | 96.21 | 91.74 | 95.71 | 96.03 | 95.55 | 95.85 | 100.00 | 83.22 |
| CDZ | 71.92 | 69.57 | 74.41 | 84.38 | 86.20 | 69.08 | 83.22 | 100.00 |
| Panel B: Equal-weighted implied cost of capital | | | | | | | | |
| PE | 100.00 | 86.18 | 89.53 | 88.68 | 91.07 | 97.42 | 94.24 | 74.53 |
| PEG | 86.18 | 100.00 | 98.88 | 93.65 | 90.68 | 90.01 | 93.14 | 79.59 |
| MPEG | 89.53 | 98.88 | 100.00 | 96.39 | 93.65 | 92.86 | 94.77 | 76.07 |
| OJ | 88.68 | 93.65 | 96.39 | 100.00 | 99.01 | 91.12 | 96.96 | 80.95 |
| CT | 91.07 | 90.68 | 93.65 | 99.01 | 100.00 | 91.79 | 98.03 | 83.55 |
| GLS | 97.42 | 90.01 | 92.86 | 91.12 | 91.79 | 100.00 | 94.66 | 74.76 |
| PSS | 94.24 | 93.14 | 94.77 | 96.96 | 98.03 | 94.66 | 100.00 | 88.12 |
| CDZ | 74.53 | 79.59 | 76.07 | 80.95 | 83.55 | 74.76 | 88.12 | 100.00 |

tion. The correlation for almost all methods is above 90% for both the value-weighted and equal-weighted time series. Even the correlation between the PE method and the other methods is almost perfect, although the level of this method is substantially lower. In short, all methods seem to be driven by the same factor, but differ in their level and how they are scaled by this factor. The only outlier is again the CDZ method. Although there is still a large positive correlation with all other methods, it is substantially lower with correlations as low as 69%.

Finally, Figure 2 shows the value-weighted implied risk premium over time for each of the eight methods, i.e., the ICC minus the yield of a 10-year government bond. In many research applications, the focus is on this implied risk premium instead of the ICC. Because the yield is the same for each method, the relative difference across methods is identical between Figure 1 and Figure 2. All methods except for the PE method result in positive risk premiums for the whole period.[90] Because a negative *expected* risk premium is irreconcilable with finance theory, this is supporting evidence for the validity of the ICC

90 For the GLS method, it is below zero for one month only.

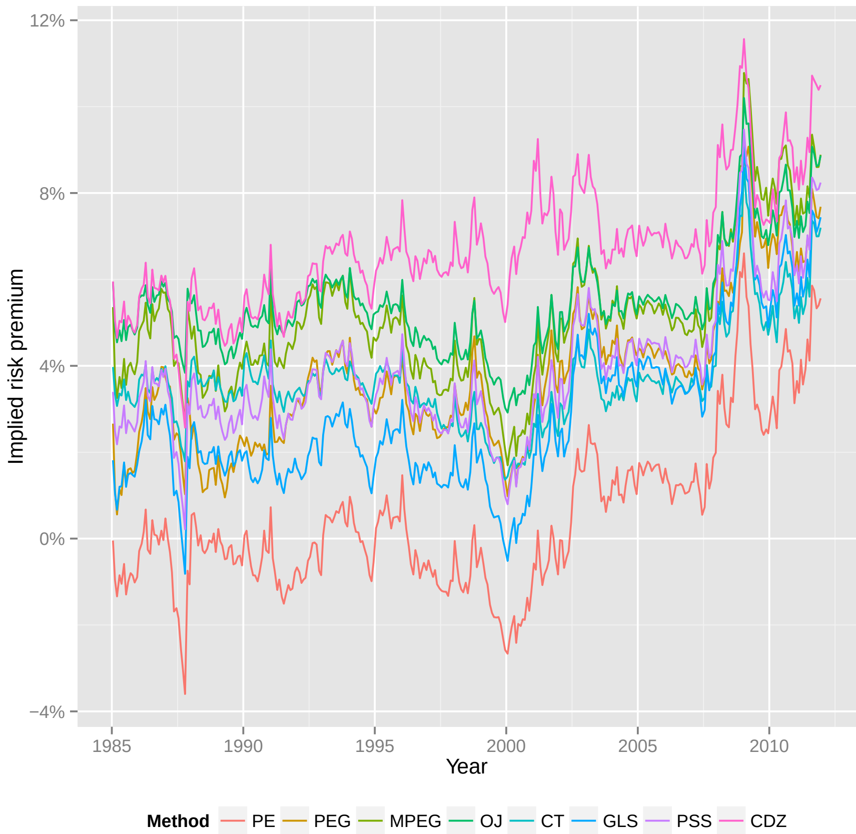


Figure 2: **Value-weighted implied risk premium for different methods over time.** This plot shows the monthly time series of the implied risk premium for eight different methods, computed as the difference between the ICC and the yield of a 10-year government bond. The ICC is computed as a value-weighted average of ICCs across all US firms for which an ICC is available for all methods.

Table 6: **Summary statistics for firm-level ICCs.** This table contains the summary statistics for the pooled cross-section of the ICC for different methods. Only observations are considered for which an ICC can be computed for each method. The columns Mean, Std. dev., 5th, 25th, 50th, 75th, and 95th refer to the mean, the standard deviation, and the specific percentile, respectively, over the pooled sample. All numbers are reported in percent. The time period ranges from 1985 to 2011.

| | Mean | Std. dev. | 5th | 25th | 50th | 75th | 95th |
|---|---|---|---|---|---|---|---|
| PE | 6.72 | 3.94 | 1.88 | 4.45 | 6.28 | 8.33 | 12.68 |
| PEG | 11.87 | 5.19 | 6.00 | 8.75 | 10.72 | 13.72 | 21.41 |
| MPEG | 12.72 | 5.13 | 7.19 | 9.57 | 11.55 | 14.54 | 22.15 |
| OJ | 12.56 | 3.88 | 8.05 | 10.06 | 11.84 | 14.18 | 19.46 |
| CT | 10.27 | 3.45 | 5.82 | 8.10 | 9.85 | 11.89 | 15.93 |
| GLS | 9.97 | 2.96 | 5.72 | 8.11 | 9.76 | 11.59 | 14.74 |
| PSS | 10.77 | 3.39 | 6.56 | 8.67 | 10.25 | 12.24 | 16.54 |
| CDZ | 13.98 | 4.45 | 8.19 | 11.14 | 13.33 | 16.07 | 21.80 |

approach. It also shows that the ignorance of growth opportunities in the case of the PE method results in unreasonable risk premiums. In brief, investors take earnings growth into consideration.

In summary, there are noteworthy differences between the first and second moments of the eight ICC methods under consideration, although there is only one true, but latent ICC. Also, the proxies are not perfectly correlated. It becomes clear from the previous discussion that a researcher would severely underestimate the true uncertainty in the statistical inference if he would base the results solely on one method. In this case he would focus on parameter uncertainty only, thereby completely ignoring model uncertainty.

### 5.2.2 *Properties of the firm-level ICC*

Table 6 shows summary statistics for the cross-sectional variation in the ICC for each method.

Not surprisingly, the cross-sectional average across all firms is similar to the time series average for each method. With respect to the cross-sectional variation, Table 6 shows that the PEG and MPEG methods have the largest standard deviations. A possible explanation for this finding is that these are dependent on the growth between the two-year and one-year ahead earnings and that this growth rate does not vary a lot in the cross-section. Consequently, a large part of the cross-sectional variation has to be explained through the residual, the ICC. By contrast, all other methods allow for much larger differences in the expected cash flows across firms through industry or firm spe-

Table 7: **Correlations for cross-sectional ICC.** This table shows the average of the annual cross-sectional correlations between the ICC of different methods. Only observations are considered for which an ICC can be computed for each method. The firm-level ICCs are value-weighted (Panel A) or equal-weighted (Panel B). All correlations are reported in percent. The time period ranges from 1985 to 2011.

| Method | PE | PEG | MPEG | OJ | CT | GLS | PSS | CDZ |
|---|---|---|---|---|---|---|---|---|
| PE | 100.00 | 9.66 | 17.78 | 29.62 | 64.10 | 61.22 | 52.96 | 30.02 |
| PEG | 9.66 | 100.00 | 96.12 | 89.16 | 51.46 | 51.81 | 52.35 | 50.51 |
| MPEG | 17.78 | 96.12 | 100.00 | 94.91 | 57.69 | 51.96 | 59.36 | 50.22 |
| OJ | 29.62 | 89.16 | 94.91 | 100.00 | 73.32 | 55.49 | 77.40 | 67.67 |
| CT | 64.10 | 51.46 | 57.69 | 73.32 | 100.00 | 60.83 | 91.88 | 77.61 |
| GLS | 61.22 | 51.81 | 51.96 | 55.49 | 60.83 | 100.00 | 59.68 | 49.59 |
| PSS | 52.96 | 52.35 | 59.36 | 77.40 | 91.88 | 59.68 | 100.00 | 89.52 |
| CDZ | 30.02 | 50.51 | 50.22 | 67.67 | 77.61 | 49.59 | 89.52 | 100.00 |

cific growth rates. Therefore, the variation in the ICC is lower. Hail and Leuz (2009) find a similar ranking of standard deviations for four implemented methods (MPEG, OJ, CT, GLS) and an international sample.

Table 7 displays the average of the annual cross-sectional correlations between the different ICC methods. These correlations are substantially lower than the time series correlations reported in Table 5. Also, the differences between the cross-sectional correlations for different valuation models are more pronounced. The correlation within the abnormal earnings growth models is very high and the correlation between the dividend discount models is also high, but the correlation between methods from different valuation models is only around 50%. Furthermore, the simple PE method is clearly an outlier with correlations that never exceed 65% and that can be as low as 10%. Easton and Monahan (2005) report similar cross-sectional correlations for their sample.

In summary, a comparison between the statistics of the time series and the cross-sectional variation in the ICC methods implies that a researcher who focuses on the cross-sectional variation faces much more uncertainty than a researcher who focuses on the time series characteristics. In particular, the cross-sectional correlation between the methods is substantially lower than the correlation over time, which implies that the firm-level ICCs are either affected by random noise or driven by different, but systematic measurement errors.

Table 8: **Average monthly number of observations for different ICC methods per year.** For each year of the sample, this table shows the average monthly number of observations for which the specific ICC method returns a value. In the column "All", the requirement is that all methods return a value. This is the sample selection criterion throughout this thesis.

| Year | PE | PEG | MPEG | OJ | CT | GLS | PSS | CDZ | All |
|---|---|---|---|---|---|---|---|---|---|
| 1985 | 1686 | 1617 | 1628 | 1612 | 1659 | 1760 | 1729 | 1729 | 1543 |
| 1986 | 1672 | 1600 | 1612 | 1609 | 1694 | 1776 | 1725 | 1725 | 1555 |
| 1987 | 1734 | 1662 | 1673 | 1660 | 1734 | 1826 | 1770 | 1770 | 1603 |
| 1988 | 1726 | 1618 | 1632 | 1621 | 1710 | 1798 | 1746 | 1746 | 1557 |
| 1989 | 1769 | 1630 | 1643 | 1640 | 1756 | 1842 | 1792 | 1792 | 1568 |
| 1990 | 1766 | 1684 | 1695 | 1680 | 1756 | 1841 | 1799 | 1799 | 1624 |
| 1991 | 1754 | 1704 | 1714 | 1705 | 1766 | 1850 | 1804 | 1804 | 1654 |
| 1992 | 1896 | 1857 | 1865 | 1855 | 1928 | 1994 | 1944 | 1944 | 1816 |
| 1993 | 2151 | 2105 | 2111 | 2093 | 2203 | 2270 | 2192 | 2192 | 2059 |
| 1994 | 2514 | 2461 | 2468 | 2440 | 2562 | 2646 | 2557 | 2557 | 2402 |
| 1995 | 2663 | 2608 | 2614 | 2580 | 2684 | 2780 | 2690 | 2690 | 2544 |
| 1996 | 2911 | 2818 | 2827 | 2812 | 2952 | 3086 | 2959 | 2959 | 2758 |
| 1997 | 3147 | 3064 | 3072 | 3053 | 3251 | 3415 | 3261 | 3261 | 2998 |
| 1998 | 3134 | 3050 | 3056 | 3046 | 3286 | 3437 | 3288 | 3288 | 2991 |
| 1999 | 2909 | 2833 | 2840 | 2838 | 3070 | 3222 | 3084 | 3084 | 2781 |
| 2000 | 2525 | 2440 | 2445 | 2450 | 2694 | 2900 | 2716 | 2716 | 2392 |
| 2001 | 2130 | 1990 | 1994 | 2007 | 2340 | 2651 | 2353 | 2353 | 1938 |
| 2002 | 2126 | 2065 | 2067 | 2067 | 2348 | 2598 | 2360 | 2360 | 2024 |
| 2003 | 2260 | 2120 | 2128 | 2135 | 2475 | 2613 | 2434 | 2434 | 2073 |
| 2004 | 2416 | 2284 | 2289 | 2269 | 2536 | 2641 | 2469 | 2469 | 2209 |
| 2005 | 2513 | 2368 | 2374 | 2338 | 2610 | 2730 | 2531 | 2531 | 2273 |
| 2006 | 2532 | 2350 | 2358 | 2296 | 2576 | 2735 | 2485 | 2485 | 2212 |
| 2007 | 2470 | 2306 | 2312 | 2224 | 2491 | 2690 | 2414 | 2414 | 2148 |
| 2008 | 2318 | 2054 | 2063 | 2017 | 2396 | 2603 | 2288 | 2288 | 1896 |
| 2009 | 2033 | 1780 | 1789 | 1661 | 2116 | 2508 | 1957 | 1956 | 1532 |
| 2010 | 2254 | 2077 | 2084 | 1797 | 2148 | 2596 | 1966 | 1966 | 1703 |
| 2011 | 2209 | 2033 | 2040 | 1842 | 2155 | 2436 | 2005 | 2005 | 1758 |

## 5.3 Sensitivity analyses

### 5.3.1 *Number of observations*

In Table 8, I show the average monthly number of observations per year for which an ICC is available. The table reveals a similar trend over time as Table 3, which showed the IBES forecasts from the raw data file. Starting from 1985, the number of observations increases until the turn of the millennium, where it peaks and decreases afterwards for a couple of years. This pattern is probably driven by the rise and fall of many technological firms during the Dot-com bubble. Since then, the number of observations has remained fairly constant until the end of the sample in 2011. The column "All" shows the number of complete observations for each year. If not stated differently, this is the sample throughout this dissertation. As can be seen, this number is lower than the number for each method, which implies that reasons for missing values are somewhat idiosyncratic.

More interestingly, however, are the systematic differences between the number of observations across methods. The GLS method typically has the most observations, while the other methods have a lower number of observations throughout the sample. In case of the abnormal earnings growth models, this difference is partly due to the requirement of positive one-year ahead forecasts. A similar argument can be made in the case of the PSS and CDZ method. If the long-term earnings growth rates are missing or the earnings three years ahead are negative, the cash flow forecasts assumed by these methods are not reconcilable with the price reported by IBES. Therefore, there is a possibility that the above statistics are biased towards better performing firms because I require complete observations for my main empirical analyses.

As a first indication of how large this effect could be, I plot the GLS method for two cases. The first case is based on all observations for which the GLS ICC is available. The second case is based on observations for which all ICC values are available.

The results can be seen in Figure 3. The differences of the aggregate ICC between the two samples is negligible, both for a value-weighted and equal-weighted aggregation.

### 5.3.2 *Time misalignment*

In this section, I check the robustness of the results with respect to the misalignment of prices, book values, and earnings forecasts discussed above.

Panel A and B of Table 9 are identical to Table 4 except that the ICC for each method is computed with discounted prices as shown in equation (64). While such an adjustment of prices is economically correct, it is mostly ignored in empirical studies. It is therefore interesting to check the importance of the ignorance of this timing issue.

A comparison between the two tables reveals that the more correct ICCs computed with adjusted prices are always higher than those computed with unadjusted prices. This directly follows from the fact that prices which are discounted with a positive ICC are lower than unadjusted prices and a lower price is only reconcilable with the same cash flows if the discount rate is higher. This effect is on average roughly 0.32 percentage points across methods, both for the value-

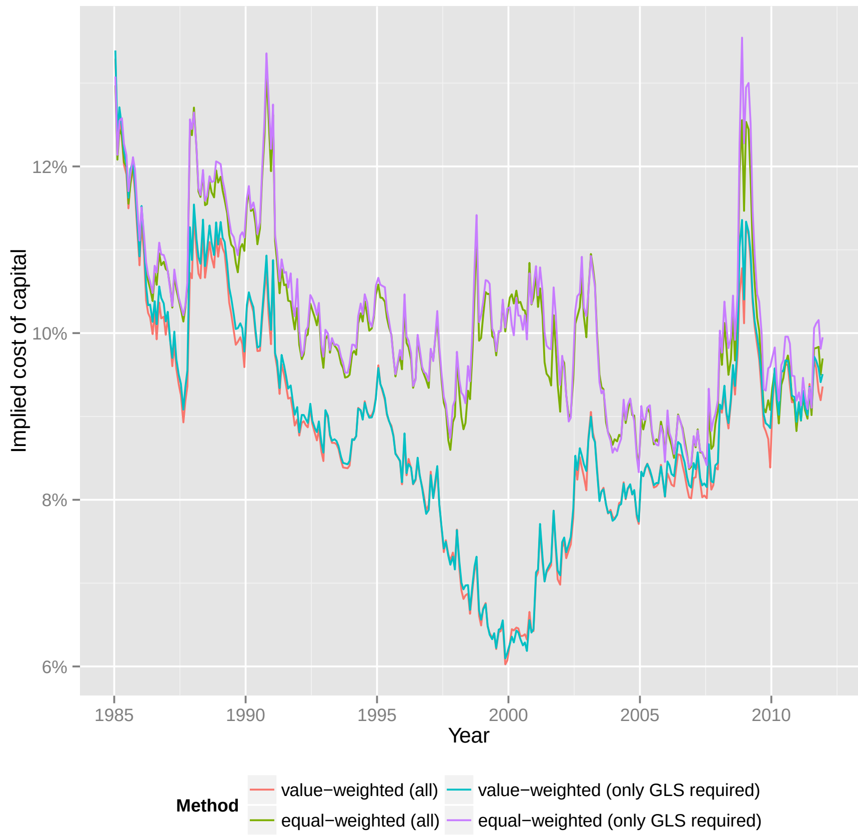


Figure 3: **Impact of omitted observations.** The plot shows the monthly value-weighted and equal-weighted time series of the implied cost of capital for the GLS method computed for two different samples: (1) all firm-month observations for which an ICC value from the GLS method is available; (2) all firm-month observations for which an ICC value for all methods is available.

Table 9: **Summary statistics for the aggregated ICC and discounted prices.** Panel A and B of this table replicate Table 4. The only difference between the two tables is that in this table the prices that enter the ICC computation are adjusted according to equation (64). For each method, Panel C shows the correlation between the value-weighted or equal-weighted ICC computed with adjusted and unadjusted prices. All numbers are reported in percent. The time period ranges from 1985 to 2011.

| | PE | PEG | MPEG | OJ | CT | GLS | PSS | CDZ |
|---|---|---|---|---|---|---|---|---|
| Panel A: Value-weighted implied cost of capital with adjusted prices | | | | | | | | |
| Mean | 6.74 | 9.96 | 11.44 | 11.74 | 9.95 | 9.08 | 10.18 | 13.10 |
| Median | 6.62 | 10.08 | 11.40 | 11.42 | 9.66 | 8.94 | 10.13 | 13.09 |
| Std. dev. | 1.54 | 1.13 | 1.63 | 1.73 | 1.86 | 1.43 | 1.47 | 1.17 |
| Minimum | 3.92 | 7.81 | 8.49 | 9.36 | 7.33 | 6.25 | 7.63 | 10.76 |
| Maximum | 11.62 | 14.29 | 17.06 | 17.59 | 15.58 | 13.45 | 15.07 | 17.67 |
| Panel B: Equal-weighted implied cost of capital with adjusted prices | | | | | | | | |
| Mean | 7.14 | 12.37 | 13.35 | 13.09 | 10.72 | 10.37 | 11.21 | 14.43 |
| Median | 6.95 | 12.46 | 13.44 | 13.13 | 10.79 | 10.24 | 11.25 | 14.57 |
| Std. dev. | 1.40 | 1.20 | 1.49 | 1.70 | 1.77 | 1.09 | 1.39 | 1.63 |
| Minimum | 5.01 | 9.86 | 10.48 | 10.13 | 7.70 | 8.65 | 8.69 | 11.28 |
| Maximum | 12.19 | 16.59 | 18.19 | 17.68 | 15.42 | 13.83 | 15.32 | 18.41 |
| Panel C: Correlation between ICC time series with adjusted and unadjusted prices | | | | | | | | |
| Value-weighted | 99.50 | 99.34 | 99.20 | 99.48 | 99.64 | 99.49 | 99.24 | 98.13 |
| Equal-weighted | 99.28 | 99.04 | 98.98 | 99.49 | 99.64 | 99.30 | 99.33 | 99.22 |

and equally weighted time series. Hence, studies that are interested in the absolute level of the ICC should account for the misalignment.

However, Panel C of Table 9 shows that the dynamics are almost identical between the two approaches (price adjusted vs. unadjusted) because the correlation between them is almost perfect for all methods.

### 5.3.3 *Analyst forecast bias*

So far, I only looked at firm-level ICCs that were computed with the consensus analyst forecasts. IBES also provides the lowest (most pessimistic) and the highest (most optimistic) forecasts across the analysts that cover a specific firm. In this section, I show statistics for the ICC that are computed with these alternative forecasts. This exercise serves three purposes. First, it evaluates the impact of reasonable alternative estimates of expected earnings and is a reminder that even the forecasts for the next couple of years are subject to considerable uncertainty. Second, it allows an initial, albeit simplistic analysis of the impact of analyst forecast bias because the lowest earnings forecasts might be less affected by it. Third, it nicely illustrates differences between the various ICC methods.

In this section, I require that all observations have an ICC available for all methods *and* for all forecasts (lowest, mean, highest). Although the raw IBES data set has as many observations for the lowest and the highest forecasts as it has for the mean forecast, this requirement still results in a lower number of observations.[91] This is because some ICCs cannot be computed for the lowest or highest forecasts, which are by definition more extreme than the mean forecasts. For example, it is a requirement for some methods that the forecasts are always positive. Obviously, there are more observations with negative forecasts in the case in which the most pessimistic forecast is used, but this effect should be negligible. For example, the correlation between the equal-weighted ICC time series for the GLS method computed with the filter that only requires all methods to have an ICC in the mean forecast case and the filter that requires all methods to have an ICC for all forecasts is 99.8%. For the MPEG method, it is 99.5%.

Figure 4 shows the value-weighted monthly time series for each ICC method and the mean forecasts and is therefore identical to Figure 1 except for the different number of observations used. Additionally, it also plots a shaded area around each method which is limited by the ICC computed with the most pessimistic and most optimistic forecasts. Figure 5 is the equivalent to Figure 4 with an equal-weighted time series.

Table 10 shows summary statistics for the spread between the ICC computed with the highest forecasts on the one hand and the lowest forecasts on the other hand. Furthermore, it shows the correlation between the two time series for each method.

As the figures and the table illustrate, there are substantial differences between the methods. On one extreme, the GLS and PE method are hardly affected by the different forecasts. For the former, the ICC based on the most optimistic analyst forecasts and the ICC based on the most pessimistic forecasts are only separated by a little more than half a percentage point on average. For the PE method, the difference is only about one percentage point. At the other extreme, the differ-

91 Note that the lowest and highest forecasts are identical to the mean forecast in cases in which there is only one analyst forecast available. This case applies to 31% of the sample and results in an ICC estimate that is the same for all forecast types. This introduces a systematic bias in the analysis because the dispersion for smaller firms is arbitrarily smaller, simply because it is covered by fewer analysts. This should be kept in mind when interpreting the statistics.

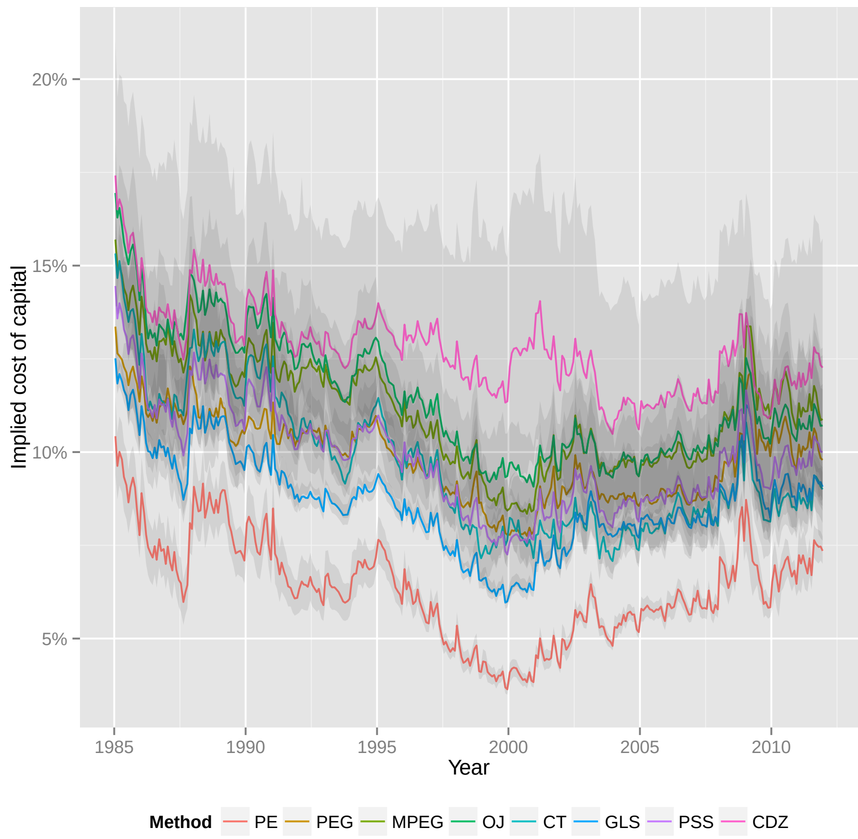


Figure 4: **Spread between the ICC computed with most optimistic and most pessimistic analyst forecasts (value-weighted).** This figure shows the monthly time series of the ICC for eight different methods. The ICC is computed as a value-weighted average of ICCs across all US firms for which an ICC is available for all methods. Only observations are considered for which an ICC can be computed for all methods and for all forecast types (pessimistic, mean, optimistic). The upper and lower end of the shaded area around each line represent the ICC computed with the most optimistic and most pessimistic analyst forecast, respectively.

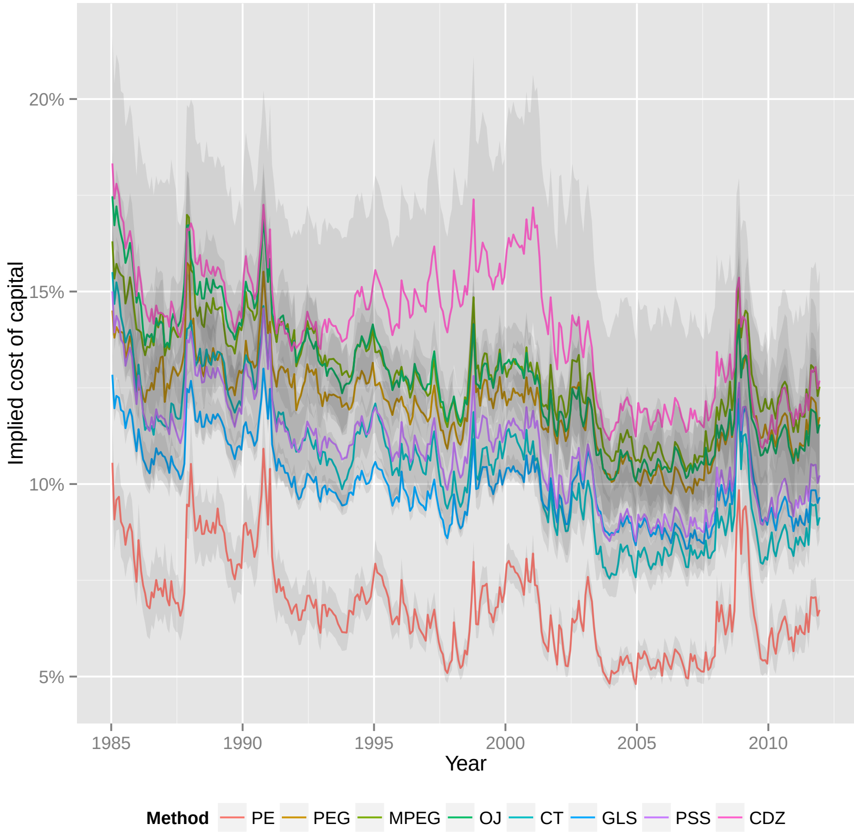


Figure 5: **Spread between the ICC computed with most optimistic and most pessimistic analyst forecasts (equal-weighted).** This figure shows the monthly time series of the ICC for eight different methods. The ICC is computed as an equal-weighted average of ICCs across all US firms for which an ICC is available for all methods. Only observations are considered for which an ICC can be computed for all methods and for all forecast types (pessimistic, mean, optimistic). The upper and lower end of the shaded area around each line represent the ICC computed with the most optimistic and most pessimistic analyst forecast, respectively.

Table 10: **Summary statistics for aggregated ICC computed with most optimistic and pessimistic analyst forecasts.** Panel A and B of this table show summary statistics of the difference between the monthly ICC time series computed with the most optimistic analyst forecast and the monthly ICC time series computed with the most pessimistic forecast. The firm-level ICCs are either value-weighted (Panel A) or equal-weighted (Panel B). For each method, Panel C shows the correlation between the value-weighted or equal-weighted ICC computed with the most optimistic and most pessimistic forecasts, respectively. Only observations are considered for which an ICC can be computed for all methods and for all forecast types (pessimistic, mean, optimistic). Numbers in Panel A and B are reported in percentage points and in Panel C in percent. The time period ranges from 1985 to 2011.

| | PE | PEG | MPEG | OJ | CT | GLS | PSS | CDZ |
|---|---|---|---|---|---|---|---|---|
| Panel A: Difference between value-weighted ICC with highest and lowest forecasts | | | | | | | | |
| Mean | 0.95 | 2.03 | 2.22 | 3.02 | 1.46 | 0.62 | 3.39 | 6.33 |
| Median | 0.87 | 2.08 | 2.27 | 2.98 | 1.31 | 0.55 | 3.30 | 6.29 |
| Std. dev. | 0.48 | 0.99 | 0.94 | 0.59 | 0.55 | 0.20 | 0.50 | 0.71 |
| Minimum | 0.25 | -1.73 | -1.11 | 1.24 | 0.68 | 0.34 | 2.54 | 5.17 |
| Maximum | 3.06 | 4.54 | 4.65 | 4.50 | 3.13 | 1.48 | 4.66 | 8.29 |
| Panel B: Difference between equal-weighted ICC with highest and lowest forecasts | | | | | | | | |
| Mean | 1.07 | 1.51 | 1.65 | 2.19 | 1.69 | 0.63 | 2.70 | 5.39 |
| Median | 0.96 | 1.47 | 1.59 | 2.17 | 1.61 | 0.58 | 2.61 | 5.21 |
| Std. dev. | 0.46 | 0.93 | 0.90 | 0.48 | 0.48 | 0.18 | 0.41 | 0.67 |
| Minimum | 0.30 | -0.59 | -0.34 | 1.22 | 0.94 | 0.35 | 2.06 | 4.24 |
| Maximum | 2.90 | 4.16 | 4.27 | 3.67 | 2.87 | 1.14 | 3.74 | 7.16 |
| Panel C: Correlation between ICC time series with highest and lowest forecasts | | | | | | | | |
| Value-weighted | 98.22 | 67.16 | 82.60 | 94.86 | 97.42 | 99.47 | 96.97 | 86.51 |
| Equal-weighted | 97.80 | 72.96 | 81.46 | 96.15 | 98.76 | 99.50 | 98.01 | 94.55 |

ence for the CDZ method is above five percentage points for both the value-weighted and equal-weighted time series.

These extreme differences can be attributed to the different assumptions made by the methods. The PE method only relies on the forecast for the next year. This forecast is the easiest to estimate. In particular, at the end of the fiscal year an analyst already has much of the information that affects the next earnings. The GLS method is also insensitive to changes of the earnings forecasts. Because it is largely driven by the current book value per share and the historical industry median ROE, the difference between the ICC based on the highest and lowest forecasts, respectively, is small. In contrast, the CDZ method relies heavily on the accuracy of the earnings forecasts, especially the long-term earnings growth rate. Not surprisingly then, variation in these forecasts has a large impact on the time series of the CDZ method.

Another interesting finding is that the minimum spread in Panel A and B is *negative* for the PEG and MPEG method in Table 10. That is, there are months in which the aggregated ICC based on the most optimistic forecasts is actually lower than the ICC based on the most pessimistic forecasts. This result emphasizes the meaning of the word *growth* in the abnormal earnings growth models. As can be seen from equation (16), the PEG method is only driven by the difference or growth, respectively, between the two-year ahead and one-year ahead forecast. The level of these earnings forecasts is irrelevant. To a lesser extent, this is also true for the MPEG method, which is also affected by next year's dividends. Therefore, the ICC for these methods can be larger if the difference between the two forecasts is larger, even if both forecasts are lower. This explains why the ICC can be higher even if the forecasts are more pessimistic.

The correlations in Panel C confirm the previous findings. They are almost perfect for all methods that rely on forecasts that are easy to estimate (PE method) or that are mostly influenced by input parameters other than earnings forecasts (CT, GLS, and PSS method). They are below 85% for the PEG and MPEG method that rely mostly on earnings growth instead of the earnings level. And they are in between these values for the other methods that rely both on earnings forecasts and other input parameters (OJ and CDZ method).

The analysis in this section is only meant to illustrate the impact that errors in the estimated earnings forecasts can have. As has been seen, this impact is the larger, the more a method relies on these forecasts. In the literature, there are alternative approaches to approximate this impact. Some studies use realized earnings instead of analyst forecasts, some studies try to adjust the analyst forecasts, and Hou, Dijk, and Zhang (2012) propose the use of regression-based earnings forecasts. However, all of these approaches are subject to their own shortcomings, some of which I discussed in Chapter 2. The main takeaway is that there is no perfect or superior way to control for systematic errors in expected earnings. Therefore, concerns of measurement error cannot be eliminated.

## 5.4 DISCUSSION

The findings in this chapter can be summarized as follows. First, there are noteworthy differences in the levels and dynamics of the ICCs estimated with different methods. These differences are more pronounced in the cross-section, but are also apparent in the aggregated time series. Second, reasonable changes to the earnings forecasts have a large impact on the results. Third, at least in the time series, all methods seem to be driven by the same underlying factor or factors. However, it is unclear if the only factor is indeed the true expected return process, or if there are additional systematic biases that affect all methods simultaneously. All in all, the evidence in this chapter suggests that a researcher faces considerable model uncertainty when he approximates true expected returns or implied costs of capital with an observable proxy. Basing his inference on only one or two such proxies can severely underestimate the uncertainty he should have in his results. For example, the low implied risk premiums found by Claus and Thomas (2001) with the CT method are on the lower end of the spectrum. Because we do not know the true expected earnings from investors, drawing inferences about the equity risk premium from evidence based on the CT method alone leads to overconfident results.

While the sensitivity analyses in this chapter tried to cover a wide range of reasonable specifications, they are far from exhaustive. I did not use alternative earnings forecasts, for instance from different data vendors for analyst forecasts or from predictive regressions. I did not implement approaches that simultaneously estimate the earnings growth rates. I did not adjust the earnings forecasts with methods proposed in the literature. I did not examine alternative long-term growth assumptions for specific ICC methods. Therefore, the evidence in this chapter is a *lower* bound on the true uncertainty a researcher should have about the true level of the unobservable ICC.

# 6

# EMPIRICAL APPLICATION OF THE EVALUATION APPROACH BY EASTON AND MONAHAN (2005)

In this chapter,[92] I apply the evaluation approach by Easton and Monahan (2005) to my data set and illustrate the points of criticism I brought forward against this approach in Chapter 3. To allow for an easy comparison between my results and the results by EM, I have to make additional transformations to my data set. They are described in Section 6.1. In Section 6.2, I show summary statistics for this specific data set. Section 6.3 contains the results for the evaluation approach by EM, both for the CF news proxy proposed by EM and for a consistent CF news proxy. Finally, in Section 6.4 I show evidence that the inconsistent proxy defined by EM is severely misspecified.

## 6.1 ADDITIONAL DATA RESTRICTIONS

To empirically confirm my analytical derivations of Chapter 3, I replicate the analysis of EM with an updated data set and consistent cash flow proxies. In accordance with EM, I limit my sample to firms that have their fiscal-year end in December and I only focus on the ICC each December. Consequently, all book values refer to the prior fiscal-year end almost a year ago and it is thus a safe bet that the information is already publicly available. I also need realized total returns as well as dividends paid to investors for the subsequent year, which I obtain from the CRSP monthly return file.

To be included in the sample, a firm-year observation must be complete, i.e., 12 subsequent monthly returns from CRSP, a value for each ICC proxy, and all CF and DR news proxies must be available. Also, if a company does not pay dividends at all, its log dividends are set to zero,[93] and as a consistency check, I require that the stock price from

92 This chapter is based on Jäckel (2013).

93 The log dividends are needed to compute the CF news part, as defined in equation (44). Unfortunately, the log transformation causes problems in cases in which a variable is zero or close to it, as noted before, for instance, by Vuolteenaho (2002). A solution to this problem is to apply the return decomposition approach by Chen, Da, and Zhao (2013), which does not rely on a loglinearization. Because this issue is not the purpose of this chapter, I rely on the loglinearization as well. My concluding remark is that one should not use the approach proposed by EM. The fact that one cannot use it for some firms without further assumptions because log transforma-

IBES is larger than the dividends from CRSP. Finally, in line with EM I eliminate the top and bottom 1/2 percentile of realized subsequent returns, the ICC estimates, and the CF and DR news proxies for the cross-sectional distribution each year. The final sample consists of 22,267 firm-year observations. Because the computation of the news proxies requires earnings forecasts by analysts for the next year, my sample ends in December 2010.

Each year, EM sort a firm into one of five groups, subject to their price-dividend ratio, and assign a group specific, time-constant $\rho$ for each group. I follow their procedure and use the $\rho$ estimates reported in EM.[94] $\omega_t$, which is needed to compute the CF news proxy proposed by EM (see equation 35), is estimated through a rolling pooled time series regression for each of the 48 Fama and French (1997) industries of log return on equity $roe_{i,t}$ on its lagged value. That is, I run $roe_{i,t-\tau} = \omega_{0,t} + \omega_t roe_{i,t-(\tau-1)} + \varepsilon_{i,t}$, where $\tau$ is a number between zero and nine. My median (mean) value of $\omega_t$ is slightly lower, but similar, to EM (median: 0.48 vs. 0.52; mean: 0.49 vs. 0.55).

## 6.2 SUMMARY STATISTICS

Table 11 gives the summary statistics for the relevant variables and the specific data set used in this chapter. The main differences to the data set used in the previous chapter are the filtering discussed above and the logging of the variables. In particular, the lower level of the ICCs as well as a smaller standard deviation in comparison to Table 6 is mainly explained by the latter. Nevertheless, the statistics in Panel A of Table 11 are similar to the statistics shown in Table 6.

In the sample period, the realized continuously compounded return $r_{i,t+1}$ is 9.1% on average. Also, realized returns have a standard deviation roughly tenfold as large as any expected return measure, a common result in the literature that I previously discussed. This finding is a direct consequence of the fact that realized returns are driven

tions are not available is just an additional, but solvable, problem of the method. My critique, on the other hand, is more fundamental. Nevertheless, I run two robustness checks. First, I ignore dividends altogether in the computations, even for firms that pay dividends. The results are mostly unchanged. Second, I only focus on firms that pay dividends. While this changes the statistics quite a bit (the sample size is roughly cut in half), the results based on consistent CF news remain very similar.

94 Easton and Monahan (2005) report the following values: (1) non-dividend-paying stocks: $\rho = 0.988$, (2) fourth quartile of price-dividend ratio for dividend paying stocks (i.e., the quartile with the highest price-dividend ratio): $\rho = 0.957$, (3) third quartile: $\rho = 0.921$, (4) second quartile: $\rho = 0.927$, and (5) first quartile: $\rho = 0.924$.

Table 11: **Summary statistics for the data set used in the EM analysis.** In Panel A, this table provides summary statistics for the relevant input parameters of the EM analysis. RR is the realized, continuously compounded return for year $t+1$, as obtained from CRSP. All ICCs represent continuously compounded returns. In Panel B, the CF news proxy as used in EM is computed from equation (35) and the ICC specific CF news proxies are computed from equation (44). In Panel C, the ICC specific DR news proxies are computed from equation (43). The columns Mean, Std. dev., 5th, 25th, 50th, 75th, and 95th refer to the mean, the standard deviation, and the specific percentile, respectively, over the pooled sample. Only observations are considered for which an ICC, a CF news proxy, and a DR news proxy can be computed for each method. Furthermore, only ICCs measured in December are considered for which the underlying firm has its fiscal-year end in December. All numbers are reported in percent. The time period ranges from 1985 to 2010.

| | Mean | Std. dev. | 5th | 25th | 50th | 75th | 95th |
|---|---|---|---|---|---|---|---|
| Panel A: Realized returns and expected return proxies | | | | | | | |
| RR | 9.07 | 33.94 | -49.42 | -8.70 | 11.01 | 29.39 | 60.75 |
| PE | 6.06 | 2.66 | 2.11 | 4.27 | 5.84 | 7.56 | 10.72 |
| PEG | 10.06 | 3.42 | 5.52 | 7.83 | 9.45 | 11.65 | 16.67 |
| MPEG | 10.96 | 3.33 | 6.78 | 8.72 | 10.32 | 12.53 | 17.42 |
| OJ | 10.98 | 2.53 | 7.59 | 9.17 | 10.59 | 12.35 | 15.69 |
| CT | 9.10 | 2.28 | 5.70 | 7.49 | 8.91 | 10.50 | 13.06 |
| GLS | 8.83 | 2.21 | 5.37 | 7.34 | 8.73 | 10.24 | 12.65 |
| PSS | 9.50 | 2.18 | 6.44 | 7.99 | 9.23 | 10.71 | 13.44 |
| CDZ | 11.89 | 2.87 | 7.69 | 9.93 | 11.59 | 13.53 | 17.11 |
| Panel B: CF news proxies (as defined by EM and ICC-specific) | | | | | | | |
| EM | -4.13 | 23.55 | -30.68 | -9.42 | -2.51 | 2.45 | 17.14 |
| PE | 14.91 | 111.30 | -119.08 | -9.25 | 7.44 | 31.07 | 181.60 |
| PEG | -3.55 | 156.69 | -252.85 | -40.74 | -0.19 | 35.18 | 230.11 |
| MPEG | -4.93 | 156.43 | -255.72 | -39.88 | -1.18 | 32.18 | 228.15 |
| OJ | -4.24 | 98.22 | -159.53 | -29.45 | -1.20 | 24.04 | 143.99 |
| CT | -0.25 | 91.21 | -142.50 | -20.64 | 1.09 | 21.55 | 140.85 |
| GLS | 8.12 | 66.62 | -80.97 | -12.64 | 2.45 | 19.30 | 124.98 |
| PSS | -0.17 | 91.48 | -145.70 | -21.44 | 1.81 | 22.59 | 144.00 |
| CDZ | -4.13 | 114.14 | -191.33 | -26.48 | 0.09 | 23.00 | 173.71 |
| Panel C: ICC-specific DR news proxies | | | | | | | |
| PE | 11.52 | 122.23 | -133.54 | -18.91 | 1.69 | 28.53 | 203.44 |
| PEG | -2.94 | 163.02 | -256.90 | -37.88 | -3.15 | 30.33 | 250.15 |
| MPEG | -3.42 | 163.14 | -262.64 | -37.99 | -2.96 | 31.00 | 248.44 |
| OJ | -2.71 | 107.25 | -174.45 | -27.26 | -2.28 | 22.99 | 164.13 |
| CT | -0.61 | 101.69 | -159.60 | -22.76 | -2.05 | 20.56 | 159.52 |
| GLS | 7.51 | 84.69 | -110.38 | -18.19 | -0.16 | 22.27 | 158.32 |
| PSS | -0.12 | 101.27 | -160.38 | -22.59 | -0.63 | 22.34 | 161.98 |
| CDZ | -1.69 | 121.69 | -195.44 | -25.06 | -0.61 | 24.43 | 192.94 |

by changes in expectations about *all* future expected dividends and returns.

The different ICC measures have similar properties as in the EM study, even though the time period is quite different (1981 to 1998 vs. 1985 to 2010): The standard deviations are roughly the same and the PE method yields the lowest mean ICC, due to an expected earnings growth of zero. Ignoring the CDZ method, which is not evaluated by EM, the mean and median ICC is the highest for the abnormal earnings growth valuation methods in both samples.

Both CF and DR news proxies have similar means and standard deviations in my and EM's data set. The means are all close to zero, which is in line with the argument that there should be no systematic bias in the news part. Also, the standard deviations are substantially larger than those of realized returns. This is a counter-intuitive result, something that has gone unnoticed in previous studies. The return decomposition also implies a variance decomposition.[95] That is, the variation of unexpected realized returns has to be explained by the variation in CF news, the variation in DR news, or a combination of both. If the two are positively correlated, they offset each other (because positive DR news has a negative effect on realized returns) and the sum of the two variance terms can be larger than the variance of unexpected returns. Nevertheless, the DR news components implied by the different ICC methods are quite extreme. EM, Mohanram and Gode (2013), and myself report standard deviations of more than 100% for the DR news proxies. Since the standard deviation of unexpected returns is only around 34%, a very large, offsetting CF news part is needed to fulfill the tautological relation. This is again best demonstrated by a simple example. Suppose that the realized unexpected returns this period are 20%, but a researcher estimates DR news of -100%.[96] Hence, he has to assume an almost equally large, but offsetting CF news part of -80% proxy to reconcile the lower variation in realized returns with the very large variation in DR news. For that reason I find a very high positive correlation between the CF and DR news proxies within one ICC method, as detailed further below.

But why is the standard deviation of the DR news proxies so large in the first place? This is most likely due to the specific procedure of calculating the DR news first as a function of changes in the ICC

95 In accordance with the VAR literature, I focus on unexpected returns, i.e., $r_{t+1} - E_t[r_{t+1}]$, here. For those returns, the expected return for next period is already considered and only CF and DR news are left to explain these unexpected returns.

96 Again, a negative DR news shock implies a positive return.

and then backing out CF news as the residual. Any specification error that is made within a specific ICC method directly translates into a specification error in the DR news part. Due to the fact that the latter is affected by changes in every forecasted period, changes in measurement error are scaled by a large number, as can be seen from equation (43). For example, EM use a $\rho$ of 0.988 for their non-dividend paying stocks. This implies, all other things being equal, that a change in measurement error of only $0.1$ percentage points results in a change of the DR news part of $0.988/(1-0.988) \times 0.1 = 8.23$ percentage points.[97]

A VAR approach is typically used in the literature to estimate CF and DR news. This approach takes the opposite route in computing the two news parts. Instead of obtaining a vector of expected future cash flows, e.g. from analysts, it predicts a vector of future expected returns by basically running several regressions in which state variables including realized returns are explained by their predecessors. The CF news part can then be backed out as the residual.[98] This approach guarantees, just as my consistent proxies, that the tautological relation always holds. However, due to the statistical nature of this approach, any extreme relations are only possible by construction: with economic reasonable time series (normally, the VAR consists of realized returns, the dividend yield, and one or more additional economic variables) it is not possible to generate extreme variations. For example, Vuolteenaho (2002) reports a standard deviation of 28% for the CF news and of 13% for the DR news, which is much lower than the standard deviations I report for proxies based on the ICC.

Since the ICC-independent CF news proxy from EM is not burdened with the task to offset the very large variation in the DR news proxies, its standard deviation is substantially lower (23.6% in my

97 The return decomposition proposed by Chen, Da, and Zhao (2013) is more robust to such specification errors. First, it does not rely on a loglinearization of returns. Second, Chen, Da, and Zhao (2013) compute the CF and DR news part for each firm individually. In contrast, the parameter of linearization $\rho$ that is used here is constant for each firm within one group. I introduce their approach in detail in the next chapter and show that the results are more reasonable. To further support the argument that the issue of measurement error is more relevant for specific groups of firms, I repeat the analysis after excluding all firms that do not pay dividends, as mentioned in footnote 93. These firms are assigned the value of $\rho$ that is closest to one and for which measurement errors have the largest impact. Consistent with the argument here, I find lower variances in the CF and DR news parts and lower correlations. Because the latter are still very high, the problem is only alleviated, not solved.

98 For a detailed description, as well as a critique of this approach, see Chen and Zhao (2009). For a response to the critique, refer to Campbell, Polk, and Vuolteenaho (2010) or Engsted, Pedersen, and Tanggaard (2012).

sample and 38.4% in EM's sample). Obviously, this directly implies that the three parts do not add up to realized returns. For example, the sum of the ICC, the EM CF news proxy, and the method-specific DR proxy is 10.26% for the MPEG method and -2.8% for the GLS method. This is a first indication that the news part is substantially misspecified by EM. By contrast, it is 9.45% for all methods if the consistent CF news proxies are used, which is almost identical to the realized returns of 9.1%. The reason for the minor difference is that the empirical implementation by EM, which I replicate here, breaks the tautological relation between realized returns and the three proxy parts. The realized returns that are used in this study are obtained from CRSP and are defined as the total return index over the calendar year. The ICCs are computed with data that comes from IBES and Compustat and there are certain mismatches between the two databases. The most important one is probably due to a different time period: While the realized returns from CRSP correspond to the calendar year, the ICCs are computed with IBES prices from around mid-December. Hence, the tautological relation is broken. Further reasons could be data errors or inconsistencies. Also, the derivations are based on the CS return decomposition, which is an approximation, while the returns are directly computed and therefore free of any approximation error.

My results are also roughly in line with Mohanram and Gode (2013), who present summary statistics for a sample that ranges from 1983 to 2007. The only noteworthy difference is that they report CF news with a lower standard deviation (9.8%).

Table 12 shows the average of the yearly cross-sectional correlations among the key variables. Except for the CDZ method, all ICCs are positively correlated with subsequent returns in the cross-section. Also, the correlation between the different ICC proxies is positive except for the PE/PEG relation, which is preliminary evidence that they are all driven by a common factor.

The most interesting part of Table 12 is the correlation between the different CF and DR news proxies. The first thing to notice is the very high positive correlation between the CF and DR news proxy within one ICC method. I already gave an explanation for this almost linear relation between the CF and DR news proxy above.

But with the exception of the very simplistic PE method, all correlations of CF and DR news between methods are also positive. Therefore, this can be seen as evidence that on a firm-level CF and DR news

Table 12: **Correlations between variables in the EM analysis.** This table shows the average of the annual cross-sectional correlations for the relevant input parameters of the EM analysis. In Panel A, the correlations with subsequent realized returns and the expected return proxies are given. Panel B and C show the correlation with the CF and DR news part, respectively. RR is the realized, continuously compounded return for year $t+1$, as obtained from CRSP. All ICCs represent continuously compounded returns. In Panel B, the CF news proxy as used in EM is computed from equation (35) and the ICC specific CF news proxies are computed from equation (44). In Panel C, the ICC specific DR news proxies are computed from equation (43). Only observations are considered for which an ICC, a CF news proxy, and a DR news proxy can be computed for each method. Furthermore, only ICCs measured in December are considered for which the underlying firm has its fiscal-year end in December. All correlations are reported in percent. The time period ranges from 1985 to 2010.

| | | Expected return proxy | | | | | | | | CF proxy | | | | | | | | | DR proxy | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RR | PE | PEG | MPEG | OJ | CT | GLS | PSS | CDZ | EM | PE | PEG | MPEG | OJ | CT | GLS | PSS | CDZ | PE | PEG | MPEG | OJ | CT | GLS | PSS | CDZ |
| Panel A: Correlation with realized returns and expected return proxies | | | | | | | | | | | | | | | | | | | | | | | | | | |
| RR | 100 | 7 | 1 | 3 | 3 | 5 | 6 | 4 | -1 | 33 | -5 | -10 | -11 | -5 | -1 | -25 | -1 | 1 | -33 | -27 | -28 | -32 | -32 | -60 | -32 | -25 |
| PE | 7 | 100 | -10 | 2 | 14 | 64 | 58 | 49 | 13 | 3 | -29 | 7 | 5 | -2 | -17 | -18 | -17 | -14 | -27 | 5 | 4 | -3 | -16 | -16 | -17 | -15 |
| PEG | 1 | -10 | 100 | 94 | 86 | 37 | 40 | 44 | 43 | -9 | 13 | -36 | -35 | -30 | -7 | -9 | -7 | -7 | 12 | -33 | -32 | -27 | -7 | -7 | -6 | -6 |
| MPEG | 3 | 2 | 94 | 100 | 94 | 45 | 38 | 52 | 38 | -7 | 11 | -37 | -37 | -33 | -9 | -10 | -9 | -8 | 9 | -34 | -34 | -30 | -8 | -8 | -8 | -7 |
| OJ | 3 | 14 | 86 | 94 | 100 | 62 | 41 | 73 | 57 | -11 | 6 | -31 | -32 | -34 | -14 | -10 | -18 | -17 | 5 | -28 | -29 | -30 | -14 | -8 | -16 | -15 |
| CT | 5 | 64 | 37 | 45 | 62 | 100 | 52 | 88 | 62 | -14 | -14 | -8 | -9 | -17 | -31 | -16 | -30 | -27 | -13 | -7 | -9 | -16 | -29 | -14 | -28 | -26 |
| GLS | 6 | 58 | 40 | 38 | 41 | 52 | 100 | 51 | 38 | 4 | -11 | -7 | -6 | -9 | -13 | -25 | -14 | -14 | -11 | -7 | -6 | -9 | -14 | -20 | -14 | -14 |
| PSS | 4 | 49 | 44 | 52 | 73 | 88 | 51 | 100 | 80 | -16 | -9 | -8 | -11 | -19 | -25 | -10 | -33 | -32 | -9 | -8 | -10 | -18 | -24 | -9 | -31 | -31 |
| CDZ | -1 | 13 | 43 | 38 | 57 | 62 | 38 | 80 | 100 | -20 | -1 | -7 | -8 | -16 | -20 | -2 | -28 | -32 | -1 | -6 | -8 | -14 | -19 | -1 | -25 | -28 |
| Panel B: Correlation with CF news proxies (as defined by EM and ICC-specific) | | | | | | | | | | | | | | | | | | | | | | | | | | |
| EM | 33 | 3 | -9 | -7 | -11 | -14 | 4 | -16 | -20 | 100 | 16 | -1 | 0 | 9 | 24 | -4 | 22 | 21 | 5 | -7 | -6 | -1 | 12 | -17 | 9 | 11 |
| PE | -5 | -29 | 13 | 11 | 6 | -14 | -11 | -9 | -1 | 16 | 100 | -21 | -20 | 4 | 62 | 60 | 56 | 49 | 95 | -20 | -19 | 4 | 59 | 50 | 54 | 48 |
| PEG | -10 | 7 | -36 | -37 | -31 | -8 | -7 | -8 | -7 | -1 | -21 | 100 | 99 | 89 | 24 | 28 | 24 | 23 | -17 | 98 | 97 | 86 | 26 | 27 | 26 | 25 |
| MPEG | -11 | 5 | -35 | -37 | -32 | -9 | -6 | -11 | -8 | 0 | -20 | 99 | 100 | 91 | 26 | 26 | 27 | 25 | -16 | 97 | 98 | 89 | 28 | 26 | 29 | 27 |
| OJ | -5 | -2 | -30 | -33 | -34 | -17 | -9 | -19 | -16 | 9 | 4 | 89 | 91 | 100 | 52 | 39 | 57 | 54 | 4 | 86 | 88 | 95 | 50 | 34 | 55 | 54 |
| CT | -1 | -17 | -7 | -9 | -14 | -31 | -13 | -25 | -20 | 24 | 62 | 24 | 26 | 52 | 100 | 63 | 88 | 82 | 58 | 22 | 24 | 48 | 94 | 50 | 82 | 78 |
| GLS | -25 | -18 | -9 | -10 | -10 | -16 | -25 | -10 | -2 | -4 | 60 | 28 | 26 | 39 | 63 | 100 | 60 | 53 | 63 | 31 | 29 | 44 | 67 | 91 | 64 | 58 |
| PSS | -1 | -17 | -7 | -9 | -18 | -30 | -14 | -33 | -28 | 22 | 56 | 24 | 27 | 57 | 88 | 60 | 100 | 96 | 52 | 23 | 26 | 53 | 83 | 48 | 94 | 93 |
| CDZ | 1 | -14 | -7 | -8 | -17 | -27 | -14 | -32 | -32 | 21 | 49 | 23 | 25 | 54 | 82 | 53 | 96 | 100 | 45 | 21 | 23 | 50 | 77 | 42 | 90 | 96 |
| Panel C: Correlation with ICC-specific DR news proxies | | | | | | | | | | | | | | | | | | | | | | | | | | |
| PE | -33 | -27 | 12 | 9 | 5 | -13 | -11 | -9 | -1 | 5 | 95 | -17 | -16 | 4 | 58 | 63 | 52 | 45 | 100 | -10 | -10 | 14 | 65 | 65 | 60 | 53 |
| PEG | -27 | 5 | -33 | -34 | -28 | -7 | -7 | -8 | -6 | -7 | -20 | 98 | 97 | 86 | 22 | 31 | 23 | 21 | -10 | 100 | 99 | 90 | 31 | 37 | 31 | 29 |
| MPEG | -28 | 4 | -32 | -34 | -29 | -9 | -6 | -10 | -8 | -6 | -19 | 97 | 98 | 88 | 24 | 29 | 26 | 23 | -10 | 99 | 100 | 92 | 33 | 37 | 34 | 31 |
| OJ | -32 | -3 | -27 | -30 | -30 | -16 | -9 | -18 | -14 | -1 | 4 | 86 | 89 | 95 | 48 | 44 | 53 | 50 | 14 | 90 | 92 | 100 | 57 | 50 | 62 | 58 |
| CT | -32 | -16 | -7 | -8 | -14 | -29 | -14 | -24 | -19 | 12 | 59 | 26 | 28 | 50 | 94 | 67 | 83 | 77 | 65 | 31 | 33 | 57 | 100 | 68 | 89 | 83 |
| GLS | -60 | -16 | -7 | -8 | -8 | -14 | -20 | -9 | -1 | -17 | 50 | 27 | 26 | 34 | 50 | 91 | 48 | 42 | 65 | 37 | 37 | 50 | 68 | 100 | 66 | 58 |
| PSS | -32 | -17 | -6 | -8 | -16 | -28 | -14 | -31 | -25 | 9 | 54 | 26 | 29 | 55 | 82 | 64 | 94 | 90 | 60 | 31 | 34 | 62 | 89 | 66 | 100 | 96 |
| CDZ | -25 | -15 | -6 | -7 | -15 | -26 | -14 | -31 | -28 | 11 | 48 | 25 | 27 | 54 | 78 | 58 | 93 | 96 | 53 | 29 | 31 | 58 | 83 | 58 | 96 | 100 |

Table 13: **Univariate cross-sectional regressions of subsequent continuously compounded realized returns on the ICC for each method.** For each ICC method, the following cross-sectional regression is run each year:

$$r_{i,t+1} = \alpha + \beta_e \hat{r}^k_{i,t} + \epsilon_{i,t+1}.$$

The presented regression coefficients and adjusted $R^2$ are the averages over those annual regressions. The t-statistics are computed as the average regression coefficients divided by the standard error of the annual regression coefficients. $r_{i,t+1}$ is the realized, continuously compounded return for year $t+1$, as obtained from CRSP. $\hat{r}^k_{i,t}$ is the log ICC value of the respective method. Only observations are considered for which an ICC, a CF news proxy, and a DR news proxy can be computed for each method. Furthermore, only ICCs measured in December are considered for which the underlying firm has its fiscal-year end in December. The time period ranges from 1985 to 2010.

| Method | Stat. | $\alpha$ | $\beta_e$ | Adj. $R^2$ |
|---|---|---|---|---|
| PE | Coef. | 0.0418 | 0.708 | 0.0292 |
| | t-value | 1.13 | 1.61 | |
| PEG | Coef. | 0.0882 | -0.0205 | 0.0229 |
| | t-value | 3.24 | -0.0694 | |
| MPEG | Coef. | 0.0755 | 0.0941 | 0.0173 |
| | t-value | 2.76 | 0.352 | |
| OJ | Coef. | 0.072 | 0.139 | 0.0194 |
| | t-value | 2 | 0.346 | |
| CT | Coef. | 0.0405 | 0.497 | 0.02 |
| | t-value | 1.08 | 1.11 | |
| GLS | Coef. | 0.0223 | 0.683 | 0.0256 |
| | t-value | 0.629 | 1.65 | |
| PSS | Coef. | 0.0536 | 0.319 | 0.0187 |
| | t-value | 1.61 | 0.785 | |
| CDZ | Coef. | 0.118 | -0.287 | 0.0246 |
| | t-value | 3.14 | -0.764 | |

are positively correlated, a result previously obtained by Vuolteenaho (2002) who uses a VAR approach to compute the news parts. This is a good consistency check and shows that the derived CF news proxy are driven by common underlying economic factors.

## 6.3 EMPIRICAL RESULTS

Before I run the multivariate regressions proposed by EM, I first show results for univariate regressions in Table 13. The advantage of a univariate regression is that we know that the residual contains all of the information surprises that affected realized returns over period $t+1$. We further know that the residual has a mean of zero and is uncorrelated with every variable conditional on time $t$ asymptotically.

For each year, I run a cross-sectional regression of subsequent realized returns on the ICC estimated in December of the previous year. In Table 13, I report the average regression coefficients of those annual regressions. The t-statistics are computed via the approach described in Fama and MacBeth (1973).

The coefficient on the ICC proxy across methods varies widely. While it is negative for the PEG and CDZ method, it is close to one for the PE and GLS method. In line with these results, the $R^2$ are also the highest for the PE and GLS method. Interestingly, the $R^2$ for the CDZ method is also quite high.

A possible reason for the negative relation of the CDZ method in explaining the cross-sectional variation in subsequent realized returns was already given previously. It could be that the very aggressive earnings growth assumption made by Chen, Da, and Zhao (2013) are at odds with investors' expectations.

In summary, based on the results of Table 13, the GLS and the PE method are doing a good job in explaining the cross-sectional variation in subsequent realized returns and can therefore be considered reasonable proxies for true expected returns. Moreover, all other methods except for the PEG and CDZ method at least have a positive coefficient. Of course, due to the large standard deviation of the news shocks, the power of these regressions is low, which can be seen from the low adjusted $R^2$. Therefore, the results could also be rationalized by a sequence of positive or negative news shocks, and not by the superiority of the GLS or PE method. As a solution to this problem, EM propose to control for the news shocks. I turn to their approach next.

Table 14 replicates the analysis in EM with their CF news proxy. In line with their study and the study of Mohanram and Gode (2013), I find regression coefficients that are all far below one. Only the sign for CF and DR news is correct, while the betas of all ICC methods are negative. This is an interesting result because the univariate regression analysis implied a positive association between most of the ICC methods and subsequent realized returns. Taking those results for face value, none of the ICC methods are of any use. However, as discussed in Section 6.2, these results are most likely driven by a news proxies plagued with extreme measurement error.[99]

99 This argument is supported by their measurement error variance analysis. In Table 5 of EM the proxies have large measurement errors. Note that I do not focus on the measurement error variance analysis of EM here because it is irrelevant to my argument and would only complicate the discussion.

Table 14: **Multivariate cross-sectional regressions of subsequent continuously compounded realized returns on the ICC for each method with CF news proxy as defined in EM.** For each ICC method, the following cross-sectional regression is run each year:

$$r_{i,t+1} = \alpha + \beta_e \hat{r}^k_{i,t} + \beta_{CFN,EM} \widehat{CFN}^{EM}_{i,t+1} + \beta_{DRN} \widehat{DRN}^{k}_{i,t+1} + \epsilon_{i,t+1}.$$

The presented regression coefficients and adjusted $R^2$ are the averages over those annual regressions. The t-statistics are computed as the average regression coefficients divided by the standard error of the annual regression coefficients. $r_{i,t+1}$ is the realized, continuously compounded return for year $t+1$, as obtained from CRSP. $\hat{r}^k_{i,t}$ is the log ICC value of the respective method. The CF news proxy independent of a specific ICC method is computed from equation (35). The ICC specific DR news proxies are computed from equation (43). Only observations are considered for which an ICC, a CF news proxy, and a DR news proxy can be computed for each method. Furthermore, only ICCs measured in December are considered for which the underlying firm has its fiscal-year end in December. The time period ranges from 1985 to 2010.

| Method | Stat. | $\alpha$ | $\beta_e$ | $\beta_{CFN,EM}$ | $\beta_{DRN}$ | Adj. $R^2$ |
|---|---|---|---|---|---|---|
| PE | Coef. | 0.161 | -0.555 | 0.65 | -0.104 | 0.259 |
| | t-value | 5.51 | -1.68 | 11.3 | -15.3 | |
| PEG | Coef. | 0.155 | -0.425 | 0.578 | -0.0509 | 0.206 |
| | t-value | 5.71 | -1.48 | 11.6 | -9.34 | |
| MPEG | Coef. | 0.162 | -0.452 | 0.583 | -0.0534 | 0.205 |
| | t-value | 5.89 | -1.74 | 11.5 | -9.53 | |
| OJ | Coef. | 0.178 | -0.581 | 0.603 | -0.0967 | 0.239 |
| | t-value | 5.39 | -1.7 | 11.7 | -11.8 | |
| CT | Coef. | 0.126 | -0.122 | 0.693 | -0.117 | 0.269 |
| | t-value | 4.34 | -0.374 | 11 | -13.4 | |
| GLS | Coef. | 0.207 | -0.978 | 0.44 | -0.239 | 0.435 |
| | t-value | 7.69 | -4.37 | 9.63 | -31.1 | |
| PSS | Coef. | 0.151 | -0.336 | 0.676 | -0.116 | 0.258 |
| | t-value | 5.31 | -1.01 | 11.8 | -13.8 | |
| CDZ | Coef. | 0.145 | -0.23 | 0.678 | -0.0803 | 0.225 |
| | t-value | 4.51 | -0.764 | 11.7 | -11.4 | |

Table 15: **Multivariate cross-sectional regressions of subsequent continuously compounded realized returns on the ICC for each method with consistent CF news proxies.** For each ICC method, the following cross-sectional regression is run each year:

$$r_{i,t+1} = \alpha + \beta_e \hat{r}^k_{i,t} + \beta_{CFN} \widehat{CFN}^k_{i,t+1} + \beta_{DRN} \widehat{DRN}^k_{i,t+1} + \epsilon_{i,t+1}.$$

The presented regression coefficients and adjusted $R^2$ are the averages over those annual regressions. The t-statistics are computed as the average regression coefficients divided by the standard error of the annual regression coefficients. $r_{i,t+1}$ is the realized, continuously compounded return for year $t+1$, as obtained from CRSP. $\hat{r}^k_{i,t}$ is the log ICC value of the respective method. The ICC specific CF news proxies are computed from equation (44). The ICC specific DR news proxies are computed from equation (43). Only observations are considered for which an ICC, a CF news proxy, and a DR news proxy can be computed for each method. Furthermore, only ICCs measured in December are considered for which the underlying firm has its fiscal-year end in December. The time period ranges from 1985 to 2010.

| Method | Stat. | $\alpha$ | $\beta_e$ | $\beta_{CFN}$ | $\beta_{DRN}$ | Adj. $R^2$ |
|---|---|---|---|---|---|---|
| PE | Coef. | -0.00353 | 1.08 | 0.939 | -0.942 | 0.881 |
| | t-value | -0.472 | 9.85 | 104 | -104 | |
| PEG | Coef. | 0.0152 | 0.799 | 0.939 | -0.941 | 0.879 |
| | t-value | 1.96 | 13.7 | 95.3 | -96.5 | |
| MPEG | Coef. | 0.011 | 0.85 | 0.94 | -0.941 | 0.879 |
| | t-value | 1.44 | 15.6 | 94.4 | -96.1 | |
| OJ | Coef. | 0.016 | 0.811 | 0.94 | -0.942 | 0.879 |
| | t-value | 1.83 | 10.6 | 94.6 | -97.2 | |
| CT | Coef. | 0.0137 | 0.859 | 0.942 | -0.943 | 0.88 |
| | t-value | 1.52 | 7.65 | 97.2 | -99.2 | |
| GLS | Coef. | 0.0203 | 0.753 | 0.919 | -0.936 | 0.881 |
| | t-value | 2.44 | 8.73 | 84.6 | -102 | |
| PSS | Coef. | 0.0245 | 0.725 | 0.94 | -0.943 | 0.88 |
| | t-value | 3.39 | 8.33 | 99.9 | -99.5 | |
| CDZ | Coef. | 0.0437 | 0.597 | 0.939 | -0.942 | 0.881 |
| | t-value | 4.08 | 6.22 | 98 | -97.3 | |

Note that the inclusion of news proxies into the regression does not only change the coefficients quite a bit, but also the ranking across methods based on the slope coefficients. The GLS method is now one of the worst methods. Furthermore, the CDZ method now has one of the lowest regression coefficients in absolute values. The ranking also changes if we look at the $R^2$ instead.

Table 15 repeats the multivariate regression of EM with updated, consistent CF news proxies. The previous results are completely reversed. Not only do both news proxies have an almost perfect 1:1 relation with realized returns of the following year for every ICC method, but also $\beta_e$ is above 0.5 for all methods and very close to one for most of them. According to the EM evaluation method, this means that all proxies track expected returns for next period almost perfectly. Also, the ranking across methods changes again. For example, the coefficient on the CDZ method is the lowest again, in contrast to Table 14.

As I argued before, I do not consider these empirical results as evidence for the quality of particular proxies, but for the inappropriateness of the evaluation approach. By definition, the coefficients have to be close to one. As pointed out earlier, they actually should be one theoretically. As noted before, the reason for the minor differences are due to a mismatch of CRSP and IBES data.

In untabulated results, I repeat the multivariate regression with an updated realized return. Instead of using returns provided by CRSP, I compute an approximated loglinearized return as defined in equation (36). The correlation between the two realized returns is almost perfect, with differences driven by mismatching reasons as discussed before. Not surprisingly, every proxy now measures expected returns perfectly, according to the evaluation approach by EM.

In summary, this section showed that it is possible for *every* proxy to define matching CF and DR proxies so that the evaluation method by EM implies a perfect relation between the proxy and the expected return next period. Therefore, the EM approach, applied with consistent proxies, cannot be used to separate between different methods.

In the next section, I provide evidence that using an inconsistent proxy introduces additional, large measurement error and is at odds with the tautological relation that defines realized returns.

## 6.4 EXAMINATION OF THE RESIDUAL IN THE MULTIVARIATE REGRESSIONS WITH THE CF NEWS PROXY AS DEFINED BY EM

It is easy to check whether the specification proposed by EM and replicated in Table 14 is correct. Because realized returns are just the sum of expected returns next period and information surprises, the residuals in equation (33) have to be zero if one measures the parts correctly or at least consistently. In contrast, EM use an inconsistent CF news proxy, so an examination of the residuals reveals how good the specification is. If the two news proxies are not too far off, the residuals should be close to zero. In Section 6.2, I already showed that the mean of the news part is not reconcilable with a specific ICC method and realized returns. Here, I focus on the residuals instead.

The average of the cross-sectional standard deviations of the residuals for the regressions run in Table 14 ranges from 21% to 26% across methods. In comparison, the variation in the dependent variable, i.e., the total variation that has to be explained by the independent variables, is only 29%.[100] In other words, almost all of the variation of realized returns in the specification proposed by EM has to be explained by the residuals. I therefore conclude that the regressions by EM are subject to a large misspecification error.

This can also be seen from the adjusted $R^2$ column in Table 14. If the proxies would be specified correctly or consistently, the adjusted $R^2$ would be one. On the contrary, the values are all well below 30% except for the GLS method. And even in the case of the GLS method, less than half of the cross-sectional variation in realized returns can be explained by the three proxies.

Figure 6 is an illustration of this point. It plots the residuals for the multivariate regression with the inconsistent proxy by EM and the ICC and DR proxy according to the PSS method. This histogram is overlaid with the realized returns. If the proxies by EM did a good job in approximating these realized returns, the residuals would be small because there would not be much variation in realized returns left to explain. By contrast, this figure shows that the variation of the residuals is almost as large as the variation of the dependent variable, realized returns. In other words, the proxies are severely misspecified and are rather useless in explaining the variation of realized returns. Almost all of this variation has to be explained by the residuals. In theory, we know that the residuals should be zero for every observation, i.e., there is no left-out factor. Therefore, the large residuals indicate severe measurement errors in the CF and DR news part.

To summarize, this section has shown that even in hindsight a researcher cannot estimate CF and DR news in a reliable manner. Eventually, this result also explains why the methodology proposed by Chen, Da, and Zhao (2013) is not robust to the specific ICC method. I introduce their approach in detail in the next chapter and show how to apply a BMA approach to check the sensitivity of this approach to the specific ICC methods.

100 Note that I report a cross-sectional standard deviation of 34% in Table 11. This number is based on a pooled sample. Here, I report the average across the annual cross-sectional standard deviations.

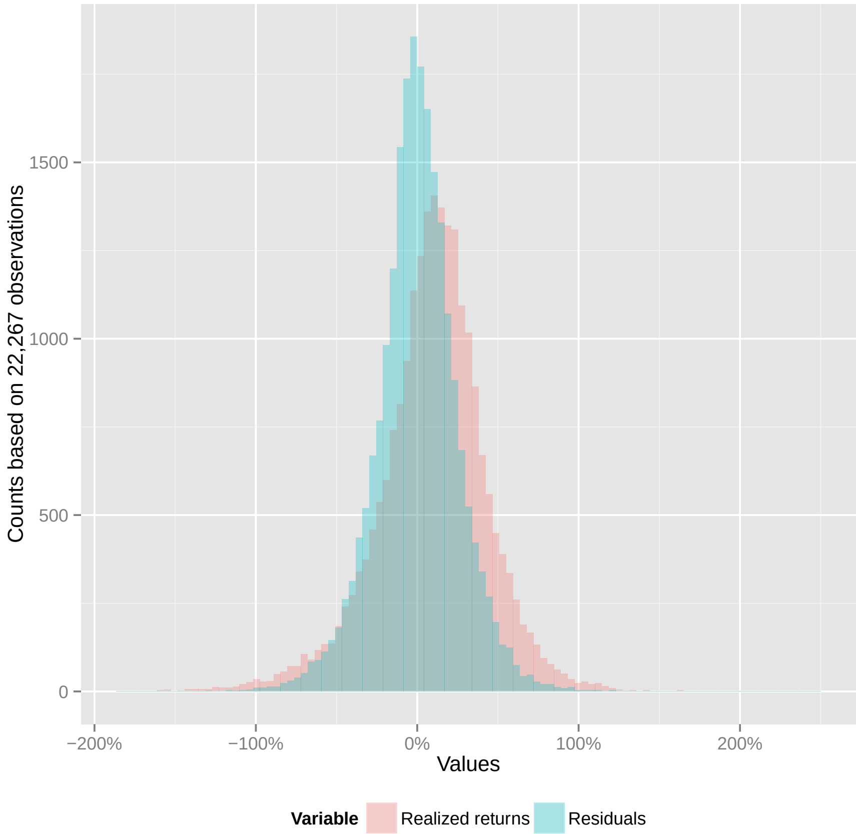


Figure 6: **Histogram of log realized returns and residuals of a multivariate cross-sectional regression of subsequent log realized returns on the ICC with the PSS method and a CF news proxy as defined by EM.** For the PSS method, the following cross-sectional regression is run each year:

$$r_{i,t+1} = \alpha + \beta_e \hat{r}_{i,t}^{PSS} + \beta_{CFN,EM} \widehat{CFN}_{i,t+1}^{EM} + \beta_{DRN} \widehat{DRN}_{i,t+1}^{PSS} + \epsilon_{i,t+1}.$$

From this regression, the residuals are stacked in a vector. This figure overlays the histograms of this vector of residuals and the vector of log realized returns, $r_{i,t+1}$. The sample ranges from 1985 to 2010.

# 7

# EMPIRICAL APPLICATION OF THE BAYESIAN MODEL AVERAGING APPROACH

In this chapter,[101] the Bayesian model averaging (BMA) approach presented in Chapter 4 is applied to three research questions that have been previously answered with the help of the ICC, but only conditional on one specification. In Section 7.1, I compute the posterior model weights that are necessary for the application of the BMA approach. Because these weights are independent of the specific research question at hand, I can discuss their computations in a preceding section. Section 7.2 revisits the study of Claus and Thomas (2001) that infers an implied risk premium for the aggregated US market. In Section 7.3, I replicate the study of Pástor, Sinha, and Swaminathan (2008) that looks at the intertemporal risk-return tradeoff. And Section 7.4 applies the BMA approach to the study of Chen, Da, and Zhao (2013) that wants to explain whether cash flow or discount rate news drive stock price movements.

## 7.1 WEIGHTS

Table 16 shows the posterior model weights that are obtained from applying equation (61) with different shrinkage parameters $\phi$.[102] As has been argued in Chapter 4, a shrinkage parameter close to zero puts almost all weight on prior information and leaves little room for the data to change the researcher's view on his priors. Since the priors are equally weighted across models, so are the posteriors in the case of $\phi = 0.01$.

101 This chapter is based on Jäckel (2014).

102 In line with Li, Ng, and Swaminathan (2013), I use continuously compounded realized returns, but I do not log the implied cost of capital estimates. Because this is a somewhat inconsistent procedure, I repeat the analysis with simple returns in Table 26 in the appendix. The results are very similar. Also, in the return predictability literature, excess returns are often used. As another robustness check, I compute the posterior model weights with such excess log realized returns and implied risk premiums instead. The posterior weights, which can be found in Table 27, are more evenly distributed than in Table 16, but the results are similar. In particular, the MPEG method is the best performing method in all cases, while the CDZ method is the worst one.

Table 16: **Posterior model weights for different shrinkage parameters.** This table shows the posterior model weights of the ICC methods for different shrinkage parameters $\phi$. The weights are based on predictive regressions of subsequent continuously compounded realized returns for the next month on the ICCs. The following priors are specified: Equal prior model probabilities $p(M_k)$ across ICC methods, an improper prior on $\sigma^2$, and the natural conjugate g-prior specification for $\beta$: $N(0, \phi\sigma^2(X_k'X_k)^{-1})$, where $X_k$ is the $T \times 2$ matrix of a T vector of ones and the T vector $\widehat{\mu}_i$; the posterior model weights are computed via equation (61). Note that the case $\phi = \infty$ is identical to the AIC weighting shown in equation (62). The time period ranges from 1985 to 2011.

| $\phi$ | PE | PEG | MPEG | OJ | CT | GLS | PSS | CDZ |
|---|---|---|---|---|---|---|---|---|
| 0.01 | 12.51 | 12.54 | 12.55 | 12.48 | 12.48 | 12.53 | 12.52 | 12.40 |
| 0.1 | 12.58 | 12.87 | 12.94 | 12.30 | 12.30 | 12.74 | 12.67 | 11.59 |
| 1 | 12.78 | 14.54 | 14.95 | 11.29 | 11.28 | 13.74 | 13.32 | 8.09 |
| 10 | 12.74 | 16.18 | 17.02 | 10.14 | 10.12 | 14.57 | 13.75 | 5.49 |
| 100 | 12.70 | 16.49 | 17.43 | 9.90 | 9.88 | 14.71 | 13.81 | 5.06 |
| $\infty$ | 12.70 | 16.53 | 17.48 | 9.88 | 9.85 | 14.72 | 13.82 | 5.01 |

The more $\phi$ is increased, the more weight is put on the evidence in the data. And for this particular data set, the MPEG method performs the best at predicting subsequent realized returns. In the limiting case in which the researcher discards all prior information ($\phi = \infty$), the posterior model weight of the MPEG method is 17%. Moreover, it is interesting to see that the CDZ method gets almost no support from the data. Furthermore, even in the case in which one discards prior information altogether, a researcher cannot differentiate between the methods with great confidence. Instead, the posterior model weights are rather evenly distributed across the eight methods. This implies that the evidence of any research question has to consider the evidence from all proxies simultaneously as well if one is not willing to put a larger prior weight on one method in particular.

How robust are these posterior model weights though? Table 17 addresses this question. It shows the distribution of the posterior model weights for two different shrinkage parameters ($\phi = 1$ and $\phi = \infty$) and for 10,000 bootstrap runs. In each run, a random sample with replacement and the same size as the original data set (i.e., 324 months) is drawn and the posterior model weights are computed for this bootstrap sample.[103]

Table 17 shows that a researcher faces considerable uncertainty about the performance of the various ICC methods. For instance, in the case of non-informative priors the 1% and 99% percentiles of the

103 Efron and Tibshirani (1993) give a textbook introduction into the bootstrapping method.

Table 17: **Bootstrapped posterior model weights for two shrinkage parameters.** This table shows the distribution of posterior model weights of the ICC methods for two different shrinkage parameters ($\phi = 1$ and $\phi = \infty$) over 10,000 block-bootstrap samples with a block length of 24 months. For each bootstrap sample, the analysis outlined in Table 16 is run.

| Percentile | PE | PEG | MPEG | OJ | CT | GLS | PSS | CDZ |
|---|---|---|---|---|---|---|---|---|
| Posterior model weights for $\phi = 1$ | | | | | | | | |
| 1% | 3.44 | 3.73 | 4.88 | 2.82 | 2.84 | 5.08 | 4.73 | 0.84 |
| 5% | 6.10 | 6.05 | 7.32 | 4.85 | 5.13 | 7.74 | 7.67 | 1.99 |
| 50% | 12.75 | 13.12 | 13.48 | 10.61 | 10.47 | 13.18 | 12.55 | 8.56 |
| 95% | 29.44 | 27.65 | 22.73 | 15.33 | 15.07 | 25.51 | 16.54 | 18.75 |
| 99% | 39.07 | 38.80 | 27.60 | 18.37 | 18.59 | 34.15 | 18.85 | 27.47 |
| Posterior model weights for $\phi = \infty$ | | | | | | | | |
| 1% | 0.44 | 0.47 | 0.80 | 0.25 | 0.29 | 1.05 | 0.88 | 0.03 |
| 5% | 2.04 | 1.97 | 2.84 | 1.15 | 1.25 | 3.41 | 3.01 | 0.20 |
| 50% | 12.35 | 12.95 | 13.68 | 8.26 | 8.05 | 13.03 | 11.65 | 5.47 |
| 95% | 46.04 | 44.44 | 29.27 | 16.79 | 16.95 | 35.98 | 18.62 | 26.61 |
| 99% | 62.13 | 62.24 | 35.56 | 22.72 | 24.16 | 52.50 | 22.45 | 48.38 |

weights for the MPEG method are 0.8% and 36%, respectively. This dispersion is even larger for other methods such as the CDZ, the PE, the PEG, or the GLS method. This large variation is a result of the large noise inherent in realized returns. The true underlying expected return process is clouded by large, unsystematic shocks. Therefore, determining the model weights precisely with predictive regressions requires long samples that we do not have; and if we had them, alternative proxies would be unnecessary in the first place for many research questions. This is an inherent circularity in any expected return proxy that tries to replace realized returns due to high noise in the latter. To determine the quality of any such proxy, one needs the very same realized returns that it intends to replace.

The large noise in realized returns and its consequences are well known in the finance literature. Goyal and Welch (2008) argue that the apparent statistical significance of many predictors is exclusively based on evidence due to years up to and especially on the years of the Oil Shock from 1973 to 1975. Fama and French (2002) find that the high realized returns in the second half of the 20th century are mostly driven by positive unexpected shocks, and not by high expected returns. And Campello, Chen, and Zhang (2008) run the predictive regressions from above with their expected return proxy based on yield spreads and find no relation. They interpret this result as evidence that the shock structure in realized returns in their sample hindered the convergence to their expected return proxy, assumed to be correct, not as evidence that their proxy might be measured with error. From the perspective of the BMA approach advocated here, they have a very informative prior about the correctness of their proxy and therefore, discard any information in the data that casts doubt on this prior.

However, this leaves only one way to evaluate the performance of any proxy, that is, prior information. Of course, since most empirical studies have only one proxy class under consideration, the implicit weight on this proxy is set to one. But this severely overstates, at least in my opinion, the confidence a researcher should have in his proxy. In the example of Campello, Chen, and Zhang (2008), if their proxy is not able to explain subsequent realized returns, why should I choose their proxy instead of proxies based on CDS spreads or the ICC? In other words, a researcher who proposes a new proxy has to compare this proxy with existing ones. The only meaningful method of comparison that I know of are predictive regressions that are highly sensitive, so a researcher might ignore these regressions altogether. But in this case, the researcher has to choose between two options. Either he considers evidence of all proxies simultaneously, which will weaken the power of statistical tests and therefore reduce one of the main advantages of alternative expected return proxies. Or he argues based on prior information why he deems his proxy to be more suitable than other existing proxies and he has to quantify the superiority of his proxy.[104]

This procedure is in sharp contrast to current practice. I am not aware of any study that compares different proxy classes such as expected returns based on the ICC, yield spreads, or CDS data. Most studies ignore the evidence of predictive regressions completely and run ad-hoc robustness checks on their results. That is, they implicitly set the probability that their proxy is correct to one in the main empirical analysis and report their results, conditional on the assumption that their proxy is tracking expected returns perfectly. Afterwards, they repeat their analyses for variations of their proxy under consideration. Thus, the reader has no indication which of these variations

104 Leamer (1978, p. 123) describes issues with the current approach of reporting results based on one proxy and ignoring the evidence based on others as follows: “But much more important is the fact that the output of an interpretive search is an interpretation of the data evidence built on some implicit prior information. This interpretation is relevant to the reader only to the extent that he accepts the implicit prior information as his own, and only then if he understands that it is already built into the result. Publication of the output of an interpretive search is thus equivalent to publication of a posterior distribution without either the sample result or the prior.”

is best supported by the data. It is also difficult for him to combine the evidence from the battery of robustness tests to one coherent picture. And finally, the variations chosen in the robustness section are predominantly selected in an ad-hoc manner with no discernable motivation. This is nicely illustrated in the ICC literature. Most asset pricing studies focus on the PSS approach and its derivatives, while most corporate finance and accounting studies implement abnormal growth in earnings models and residual income models. Furthermore, the studies generally focus on one or a few approaches and thus ignore evidence based on other approaches. This procedure may be motivated by the fact that too many robustness checks will unnecessarily lengthen the empirical part and bore the reader, particularly if the results are quite similar. However, as I show in subsequent examples, omitting such tests can result in quite dramatic misinterpretations.

The model averaging approach introduced in Chapter 4 is a solution to this problem. If a researcher is willing to make the extra effort to motivate this approach shortly, model averaging can take any number of expected return proxies into account without increasing the complexity of the analysis. Also, if one is willing to take predictive regressions into account, despite their sensitivity to large shocks, model averaging will automatically incorporate evidence about the quality of the proxies under consideration. So if one proxy class gets no support in the data, it will not matter in the following empirical analysis. The weighting between the prior information and the data can easily be controlled by the researcher. Furthermore, this approach helps protect a researcher from finding spurious relations between the variable of interest and expected returns by ensuring that a researcher does not select a proxy with a particular measurement error that is related to the variable of interest.

Unfortunately, even this approach is unable to solve the problem of whether any proxy is tracking expected returns well. If no proxy does, the analysis will still be biased, even asymptotically. This is a severe shortcoming of any expected return proxy. Finally, the BMA approach subsumes current approaches, which also apply a model averaging approach implicitly by setting the probability of one proxy to one. So it is also possible to reproduce current studies exactly with this approach, but this requires the researcher to explicitly state his prior.

In the following, I present three empirical examples that show the impact of model uncertainty and apply the model averaging approach to deal with it.

## 7.2 IMPLIED EQUITY RISK PREMIUM

In this section, I replicate the results of Claus and Thomas (2001) for an updated time period and additional ICC methods. Claus and Thomas (2001) were one of the first studies to apply the ICC in empirical research. They use the ICC to compute an implied risk premium, defined as the ICC minus the 10-year government bond yield, and find that the US implied risk premium was only around 3% from 1985 to 1998. Due to a lack of alternative proxies back then, they only apply the CT approach.

I reproduce their analysis for an updated time period from 1985 to 2011 and also incorporate model uncertainty into the analysis by considering seven ICC methods simultaneously. I consider only seven methods here because I set the prior weight on the PE method to zero. Because this method ignores any future earnings growth, I consider this method irrelevant from the outset. Furthermore, I set $\phi = 0$, i.e., I consider each ICC method equally likely to track expected returns correctly. Hence, the posterior model weights are equal to the prior model weights; each of the remaining seven ICC methods gets the same weight. The reason why I do not consider the evidence from predictive regressions as relevant for this research question is because I assume that the level of the ICC, in which we are interested in here, is unrelated to the time series process of the ICC. Only the latter is evaluated with predictive regressions, but since I assume that there is no relation to the former, these regressions do not help me in differentiating between the different methods. As a simple example, take two proxies, one that tracks expected returns perfectly, but is 10 percentage points too high every period, and one that is either 2 percentage points too high or too low, with equal probability. While the former proxy is biased in levels, it will perfectly track the time series of expected returns. The latter, on the other hand, will be unbiased, but not track expected returns reasonable well. In this application, we want to choose the latter, but the predictive regression would choose, at least asymptotically, the former. I therefore ignore the evidence of the predictive regressions.

Hence, our inference for the implied equity risk premium should simultaneously consider the parameter uncertainty within each proxy and the uncertainty across proxies. Figure 7 does exactly that. For each proxy, 10,000 block-bootstrap samples are generated with a block length of 24 months in which the mean of the implied risk premium is calculated. I use block-bootstrapping to preserve the autocorrelation structure of ICCs. The bootstrap samples for each proxy are then combined to one final sample. Based on the 70,000 bootstrapped means, I can compute the mean over all samples, which turns out to be 4.5%, or get the 2.5% and 97.5% percentile (2.4% and 7%, respectively). The plot illustrates that model uncertainty dominates parameter uncertainty considerably in this case. While the range of possible values for the implied risk premium mean, conditional on a specific proxy, is quite narrow – the largest 95% coverage region is 1.6% for the GLS method; it is roughly three times as large when both parameter and model uncertainty are considered. It is therefore of paramount importance to incorporate model uncertainty into the statistical inference.

Three additional points are worth repeating here. First, model uncertainty is not completely eliminated. For instance, all proxies are based on analyst forecasts and these forecasts are probably biased upwards. Second, one could also adjust the model weights based on prior information. For example, the assumption made in the CDZ method that earnings grow with the analysts' long-term growth rate until year 15 is certainly a very aggressive growth assumption. If one deems this assumption to be unreasonable, the prior model weights of the method can be reduced accordingly. Third, this example still proves the usefulness of alternative proxies. The seven ICC methods cover a wide range of earnings growth assumptions and yet, the results imply that the implied risk premium is positive and lies within a reasonable range. Such a statement cannot be made for such a short period based on realized returns. Therefore, the increase in the variance due to model uncertainty is still considerably lower than the decrease due to eliminating the large shocks that affect realized returns.

## 7.3 INTERTEMPORAL RISK-RETURN TRADEOFF

Although finance theory predicts a positive risk-return relation, empirical evidence based on realized returns does not conclusively find a positive sign. In simulations, Lundblad (2007) shows that even if

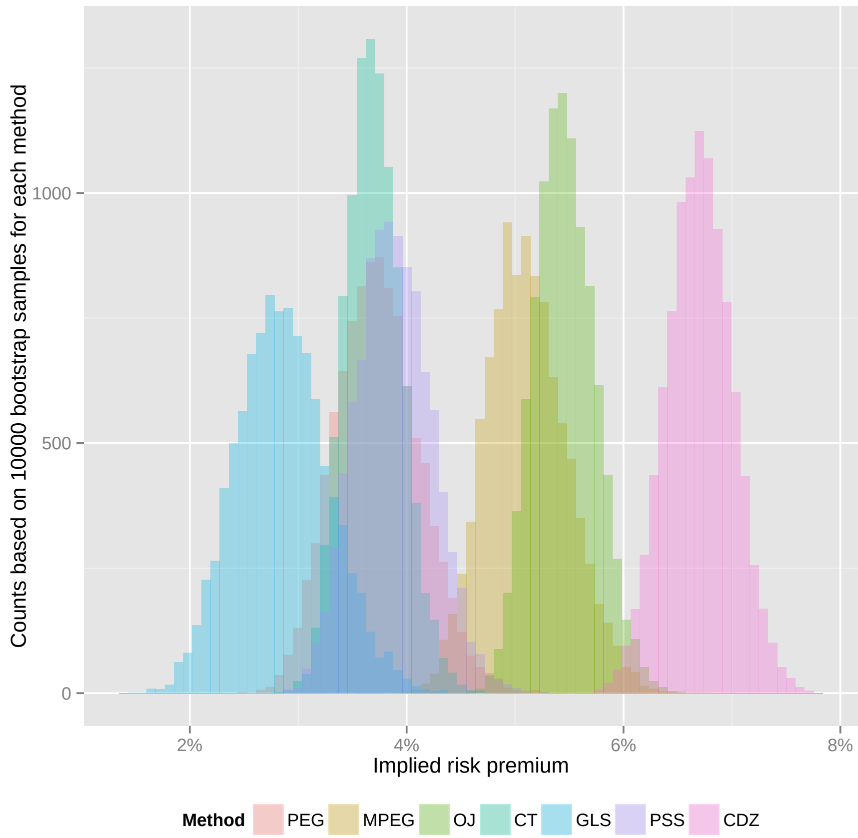


Figure 7: **Histogram of bootstrapped means of implied risk premiums for seven different ICC methods.** This figure overlays the histograms for the means of implied risk premiums computed from seven ICC methods (the PE method is ignored). Each histogram consists of 10,000 means that are computed from block-bootstrapped samples with a block length of 24 months. The monthly sample ranges from 1985 to 2011.

there is a positive relation between the conditional variance and the conditional expected return, it takes very long samples to identify this relation with noisy realized returns.

Consequently, Pástor, Sinha, and Swaminathan (2008) replace realized returns with an ICC measure estimated with the PSS method. They find a positive relation between the conditional mean of market returns, approximated by their ICC, and the variance of market returns for the years 1981 to 2002. Empirically, they run the following regression specifications, which I reproduce and extend with the model averaging approach:[105]

$$\widehat{\mu}_t = a + b\mathrm{Vol}_t + e_t \tag{65}$$

$$\Delta\widehat{\mu}_t = a + b\Delta\mathrm{Vol}_t + e_t, \tag{66}$$

where $\widehat{\mu}_t$ is a proxy for expected excess returns and $\mathrm{Vol}_t$ is either the annualized variance or standard deviation of the daily value-weighted market returns from CRSP for this period. Since the IBES release date is typically a few days after the 15th of each month, I compute the conditional volatility based on returns ranging from the first trading day after the 15th of the previous month to the first trading day after the 15th of the current month.[106] The implied risk premiums are the difference between the ICC minus the 10-year government bond yield. $\Delta\widehat{\mu}_t$ and $\Delta\mathrm{Vol}_t$ are the first difference of the conditional market return mean and volatility proxies, respectively. Because the ICC is highly persistent, I follow Pástor, Sinha, and Swaminathan (2008) and use 12 Newey-West lags in regression (65). Since the first difference of ICCs does not show a persistent autocorrelation structure, I follow Pástor, Sinha, and Swaminathan (2008) and use one lag for specification (66).

The rows labelled "PSS" in Table 18 repeat the analysis of Pástor, Sinha, and Swaminathan (2008) for a different time period (1985 to 2011 instead of 1981 to 2002). Despite the different time periods, the results are very similar. I also find a positive risk-return tradeoff for both the levels and the first difference regressions and for equally and value-weighted implied risk premiums. These results are also robust: the 5th percentile based on Newey-West corrected standard errors is positive in all specifications.

105 They also entertain a third specification in which they model both expected returns and the volatility as AR(1) processes and regress the former's residual on the latter's. To keep the analysis short, I omit this specification here.

106 Using conditional volatilities computed for the calendar month yields very similar results.

Table 18: **Implied premiums regressed on market volatility.** This table replicates and extends the analysis in Pástor, Sinha, and Swaminathan (2008). It reports the slope coefficients as well as the 5% and 95% percentile of the coefficients' distributions from the regressions in equations (65) and (66). The independent variable is return volatility $\sigma_t$, estimated as the realized market volatility from daily returns of the first trading day after the 15th of month $t-1$ to the first trading day after the 15th of month t. I use the CRSP value-weighted index. The column labels $\sigma_t^2$ and $\sigma_t$ denote the regressor. The sampling distributions of the slope coefficients are estimated in three ways. In the PSS case, it is from an OLS with the implied risk premium based on the PSS method and Newey-West corrected standard errors. In the first model averaging approach (MA1), I sample 1,000,000 times from a mixture t-distribution. Each t-distribution's parameters are determined from a regression of the implied risk premium for the eight ICC methods under consideration with Newey-West corrected standard errors. The mixture probabilities are the posterior model weights with $\phi = \infty$. In the second model averaging approach (MA2), the statistics are based on 100,000 block-bootstrap samples with a length of 24 months for equation (65) and 3 months for equation (66). In each bootstrap run, a method-specific implied risk premium is chosen randomly based on the posterior model weights with $\phi = \infty$. The implied risk premiums are the difference between the ICCs and the 10-year government bond yield. In Panel A/B, I use equal/value-weighted premiums. The sample is monthly and ranges from 1985 to 2011.

| | $\sigma_t^2$ | | | $\sigma_t$ | | |
|---|---|---|---|---|---|---|
| | $\widehat{b}$ | 5th perc. | 95th perc. | $\widehat{b}$ | 5th perc. | 95th perc. |
| Panel A: Equal-weighted implied risk premiums | | | | | | |
| Levels: $\widehat{\mu_t} = a + b\sigma_t^{(2)} + e_t$ | | | | | | |
| PSS | 0.105 | 0.053 | 0.157 | 0.087 | 0.062 | 0.113 |
| MA1 | 0.109 | 0.057 | 0.164 | 0.090 | 0.061 | 0.121 |
| MA2 | 0.120 | 0.054 | 0.246 | 0.089 | 0.054 | 0.124 |
| Differences: $\Delta\widehat{\mu_t} = a + b\Delta\sigma_t^{(2)} + e_t$ | | | | | | |
| PSS | 0.028 | 0.014 | 0.042 | 0.027 | 0.017 | 0.036 |
| MA1 | 0.027 | 0.011 | 0.043 | 0.025 | 0.015 | 0.036 |
| MA2 | 0.028 | 0.007 | 0.048 | 0.025 | 0.013 | 0.035 |
| Panel B: Value-weighted implied risk premiums | | | | | | |
| Levels: $\widehat{\mu_t} = a + b\sigma_t^{(2)} + e_t$ | | | | | | |
| PSS | 0.083 | 0.029 | 0.137 | 0.064 | 0.026 | 0.101 |
| MA1 | 0.080 | 0.029 | 0.136 | 0.062 | 0.023 | 0.102 |
| MA2 | 0.079 | 0.004 | 0.154 | 0.057 | -0.000 | 0.101 |
| Differences: $\Delta\widehat{\mu_t} = a + b\Delta\sigma_t^{(2)} + e_t$ | | | | | | |
| PSS | 0.020 | 0.006 | 0.034 | 0.019 | 0.010 | 0.029 |
| MA1 | 0.017 | 0.002 | 0.032 | 0.016 | 0.006 | 0.027 |
| MA2 | 0.016 | -0.003 | 0.033 | 0.016 | 0.005 | 0.026 |

In rows "MA1" and "MA2", I apply the model averaging approach to check whether these robust results could be overestimated due to the ignorance of model uncertainty. Since the density of the slope coefficient $\widehat{b}$, conditional on a specific ICC method, follows a t-distribution, the density across models is a weighted average of these conditional densities. This is simply a mixture t-distribution from which I sample 1,000,000 times. As a robustness check, I also implement a second model averaging approach, MA2, in which I generate 100,000 block-bootstrap samples with a block length of 24 months for equation (65) and 3 months for equation (66). In each sample, an ICC method is chosen randomly based on the posterior model weights. For this example, I decide to use a diffuse prior and set $\phi$ to $\infty$.

Table 18 shows that the consideration of model uncertainty has a negligible effect on the results. In all specifications, the mean of the sampled coefficients is very similar to the regression coefficient from the PSS case. Also, the 90% coverage region widens only marginally. There are now some cases in which 5% of the drawn coefficients are negative, but by and large almost all draws across the eight specifications are positive. This confirms the findings of Pástor, Sinha, and Swaminathan (2008) that there is a positive relation between the conditional market return and the conditional volatility.

Figure 8 gives the answer to the question why model uncertainty does not affect the results. It shows the histogram for 100,000 draws from the mixture t-distribution of the MA1 approach. In particular, the histogram plots draws from the case in which the first difference of the value-weighted implied risk premium is regressed on the first difference of the variance (lower left block in Table 18). It becomes clear from this figure that all methods lead to very similar results, with only minor variation in the mean and the variance of the slope coefficient's distribution. Not surprisingly, an inference based on a weighted average of similar distributions is similar to an inference based on any of these distributions.

In summary, this example shows that model averaging is an easy to use and flexible approach to incorporate model uncertainty. The results can be presented in a more concise way than based on separate evidence for each of the methods. It is also straightforward to extend this approach to more specifications of a specific proxy class or even across proxy classes. It also emphasizes that alternative expected return proxies have their merits over realized returns. In cases in which reasonable alterations of an expected return proxy lead to

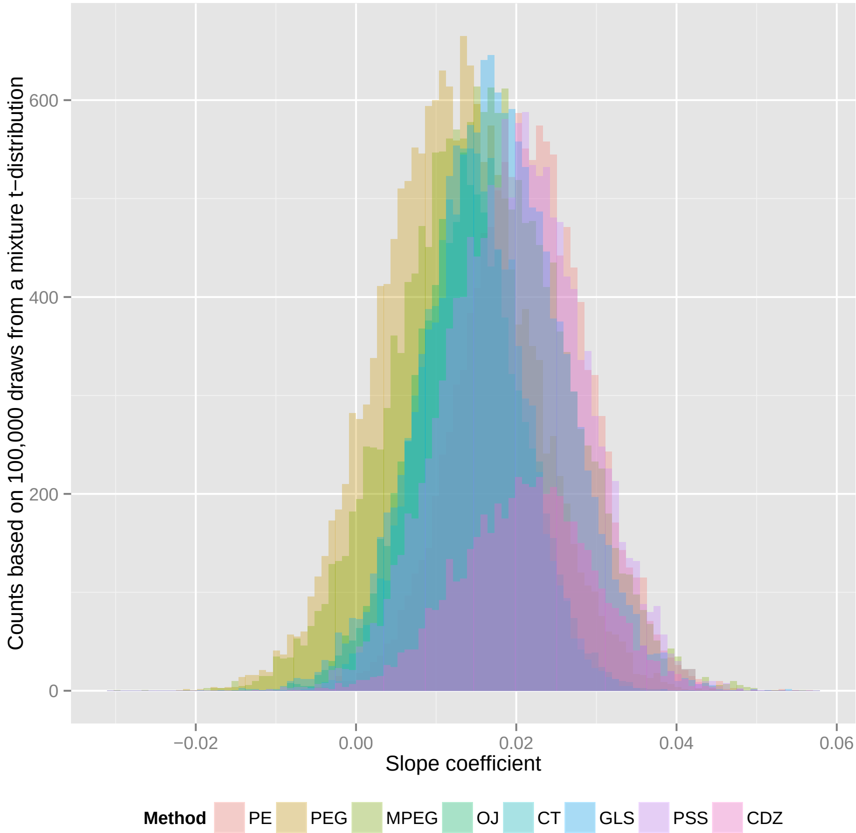


Figure 8: **Mixture t-distribution of slope coefficients from regressing implied risk premiums on market volatility.** This plot shows 100,000 draws from a mixture t-distribution. Each t-distribution represents the sampling distribution of the slope coefficient from regressing the first differences of the value-weighted implied risk premium of the specific method on the first differences of the market volatility, which is measured here as the variance of daily stock returns. For more information, see the description in Table 18 of the MA1 model averaging approach. The monthly sample period begins in January 1985 and ends in December 2011.

similar conclusions, model uncertainty has a negligible effect on the results. However, it is vital to check this, as the next example shows.

## 7.4 IMPORTANCE OF CASH FLOW AND DISCOUNT RATE NEWS

In a recent study, Chen, Da, and Zhao (2013) entertain the ICC to determine whether stock prices move because of revisions in expected cash flows or discount rates. Other studies predominantly entertain a VAR approach to estimate the time series of expected returns and back out cash flow news as the residual. Instead, Chen, Da, and Zhao (2013) use direct expected cash flow measures, namely analyst forecasts. They show that capital gain returns $Retx$ between $t + j$ and $t$ can be separated into two parts. First, a cash flow part $CF_{j,k}$, which is the part that explains changes in stock prices due to changes in analyst forecasts between $t + j$ and $t$, holding the discount rate constant. Second, a discount rate part $DR_{j,k}$, which is the part that explains changes in stock prices due to changes in discount rates, holding the cash flows constant. As the subscript $k$ indicates, both parts are dependent on the specific ICC method. In their paper, they estimate the discount rates with the CDZ method.

Recall from equation (2) and its derivatives that the stock price can be expressed as a function of the vector of future expected earnings $veps_k^t$, some transformation of this vector, and an ICC proxy, $R_t^k$. Both are estimated at time $t$. $Retx$ over horizon $j$ can then be expressed as

$$\begin{aligned} Retx_j &= \frac{P_{t+j} - P_t}{P_t} \\ &= \frac{f(veps_k^{t+j}, R_{t+j}^k) - f(veps_k^t, R_t^k)}{P_t} \\ &= CF_{j,k} + DR_{j,k}, \end{aligned} \tag{67}$$

where

$$\begin{aligned} CF_{j,k} = \Bigg( &\frac{f(veps_k^{t+j}, R_{t+j}^k) - f(veps_k^t, R_{t+j}^k)}{P_t} + \\ &\frac{f(veps_k^{t+j}, R_t^k) - f(veps_k^t, R_t^k)}{P_t} \Bigg) / 2 \end{aligned} \tag{68}$$

and

$$DR_{j,k} = \left( \frac{f(veps_k^t, R_{t+j}^k) - f(veps_k^t, R_t^k)}{P_t} + \right.$$
$$\left. \frac{f(veps_k^{t+j}, R_{t+j}^k) - f(veps_k^{t+j}, R_t^k)}{P_t} \right) /2. \tag{69}$$

Finally, we can study the variance of capital gain returns through CF and DR news:

$$\begin{aligned} Var(Retx_t) &= Cov(CF_{t,k}, Retx_t) + Cov(DR_{t,k}, Retx_t) \\ 1 &= \frac{Cov(CF_{t,k}, Retx_t)}{Var(Retx_t)} + \frac{Cov(DR_{t,k}, Retx_t)}{Var(Retx_t)}, \end{aligned} \tag{70}$$

where capital gain returns, CF news, and DR news are computed for a specific horizon j (omitted here). As can be seen from equation (70), the slope coefficients obtained from regressing $CF_{t,k}$ and $DR_{t,k}$, respectively, on $Retx_j$ represent the portion of capital gain returns driven by CF news and DR news for a specific ICC method k.

### 7.4.1 *Introductory example*

It is instructive to make a short example on how this return decomposition by Chen, Da, and Zhao (2013) works. For simplicity, suppose that investors do not incorporate any growth assumptions into the firm valuation. Furthermore, suppose that expected returns for all future periods are constant, which implies that expected returns are equal to the ICC.[107] Thus the stock price is given as the forecast for next period's expected dividends divided by the ICC: $P_0 = \frac{E[DPS_1]}{R_0^e}$. Suppose further that a researcher knows that investors make these assumptions to value their stocks. Because he can observe the current stock price, the only uncertainty lies in the estimation of the dividends for the next period.

Table 19 provides all relevant information for a small numerical example. Panel A contains the observable data for a stock. For the two periods, the stock had large capital gains with a total return of 200%. The question we are interested in is how much of these capital gains are due to changes in expected cash flows and how much are due

107 I discuss this point in detail in Section 2.1.

Table 19: **Numerical example to illustrate the return decomposition approach by Chen, Da, and Zhao (2013).** In this numerical example, it is assumed that the stock price is derived by a simple perpetuity, i.e., $P_t = \frac{E_t[DPS_{t+1}]}{R_t^e}$. Furthermore, it is assumed that an econometrician knows this, that is, there is only uncertainty about the correct estimation of the expected dividends next period. The values in Panel A are observable to an econometrician, while Panel B shows the true, but latent expectations of investors. Panel C and D show the results for two different assumptions made by an econometrician. The rows "CF" and "DR" are computed with the formulas given in equation (68) and (69). Retx, CF, and DR refer to a one period horizon.

| | Time | | |
|---|---|---|---|
| Variable | 0 | 1 | 2 |
| Panel A: Observable data | | | |
| Price | 100 | 160 | 300 |
| Retx | | 60% | 87.5% |
| Panel B: Investors' expectations | | | |
| E[DPS] | 10 | 8 | 12 |
| $R^e$ | 10% | 5% | 4% |
| CF | | -30% | 56.25% |
| DR | | 90% | 31.25% |
| Panel C: Results for method 1 | | | |
| E[DPS] | 10 | 16 | 30 |
| $R^e$ | 10% | 10% | 10% |
| CF | | 60% | 87.5% |
| DR | | 0% | 0% |
| Panel D: Results for method 2 | | | |
| E[DPS] | 10 | 10 | 10 |
| $R^e$ | 10% | 6.25% | 3.33% |
| CF | | 0% | 0% |
| DR | | 60% | 87.5% |

to changes in expected returns.[108] Panel B gives the answer to this question. From time 0 to time 1, investors actually *lowered* their expectations for all future dividends. Because they also expected much lower returns, this led to an increase in the stock price. For period 1, the CF part was therefore actually negative with -30% and the DR part was even larger than the returns with 90%.[109] The sum of those two parts is equal to the capital gain returns. This has to hold because of a tautological relation similar to the CS return decomposition.[110] In period 2, investors simultaneously raise their cash flow expectations from 8 to 12 and lower their expected returns, which leads to a positive capital gain return of 87.5%. Roughly two third of this return is explained by changes in cash flows and the rest by changes in discount rates.

Unfortunately, the information in Panel B is unobservable for an econometrician. As a consequence, he has to estimate the expected dividends and Panel C and D show the estimates for two different methods. Method 1 in Panel C assumes that the dividend expectations increase substantially over time. This assumption implies that

108 Note that Chen, Da, and Zhao (2013) only focus on capital gain returns and ignore returns from dividend payments.

109 To compute these values, one has to compute hypothetical stock prices first, one with the expected dividends at time 1 and the ICC at time 0 and one with the expected dividends at time 0 and the ICC at time 1. This yields

$$f(E[DPS_1], R_0^e) = \frac{8}{0.1} = 80$$

and

$$f(E[DPS_0], R_1^e) = \frac{10}{0.05} = 200.$$

CF and DR are then easily computed via equation (68) and (69) as

$$CF = (160 - 200 + 80 - 100)/(2 \times 100) = \frac{-60}{200} = -30\%$$
$$DR = (200 - 100 + 160 - 80)/(2 \times 100) = \frac{180}{200} = 90\%.$$

110 The decomposition by Chen, Da, and Zhao (2013) is, however, in three respects different from the more standard CS return decomposition that I introduce in detail in Appendix A.1. First, Chen, Da, and Zhao (2013) focus only on capital gain returns, while the CS decomposition includes dividends. Second, they do not approximate the present value formula by a loglinearization. Third, CF and DR news sum up to the *unexpected* return (realized return minus expected return), while they add up to the *realized* price change in the framework of Chen, Da, and Zhao (2013). However, the variation in expected returns is typically small relative to the variation in realized returns for stocks. In their appendix, Chen, Da, and Zhao (2013) show that the differences between the two approaches are minor as long as similar forecasts for future cash flows are used. Of course, the use of different forecasts is the main contribution of their paper and explains why their results differ from prior studies.

the capital gain returns are completely explained by the CF news part. A researcher that applies this method would conclude, incorrectly, that stock prices move due to changes in cash flows only. At the other extreme, method 2 in Panel D attributes the capital gain returns exclusively to changes in the DR part because the expected dividends stay constant over time.

This simple example illustrates that the choice of a specific ICC method determines the results. Methods that leave the cash flow expectations rather constant over time, such as the GLS method, will attribute most of the movement in stock prices to the DR news part. To the contrary, methods that update cash flow expectations a lot through time will attribute this movement to the CF news part instead. Because we do not know which of these assumptions is actually correct, we cannot make statements about the importance of the two parts based on only one method. Only if all methods would yield similar results, this would be preliminary evidence.

Of course, this is the same argument I made previously with respect to the usefulness of the Easton and Monahan (2005) approach. With any ICC method, we can always incorrectly define CF and DR news in such a way that they are consistent with realized returns. My critique towards Easton and Monahan (2005) is that they do not consider this tautological relation consistently. My critique towards Chen, Da, and Zhao (2013) will be that they ignore the large uncertainty they have about the correct specification of the ICC method and pretend that there is only one reasonable way to compute consistent news proxies.

Before I do so in detail, I shortly want to show how one can compute $Retx$, $CF$, and $DR$ for longer horizons. Above, I set $j = 1$. In this simple example with two periods, we can also compute these values for $j = 2$, for example for the true investor expectations. Plugging the values of Table 19 into equation (67), (68) and (69), we obtain

$$\begin{aligned} Retx_2 &= \frac{300 - 100}{100} = 200\% \\ CF_2 &= \frac{300 - 250 + 120 - 100}{2 \times 100} = 35\% \\ DR_2 &= \frac{250 - 100 + 300 - 120}{2 \times 100} = 165\%. \end{aligned}$$

Again, the sum of CF and DR news adds up to the capital gain returns.

### 7.4.2 *Replication of main results on the aggregate portfolio*

Table 20 is a replication of Table 2 in Chen, Da, and Zhao (2013) for the aggregate market. For each quarter throughout my sample, I compute $Retx$, $CF_k$, and $DR_k$ for each firm and different horizons via equation (67), (68) and (69), where in this case $k$ denotes the CDZ method.[111] In a next step, I aggregate the $Retx$, $CF$ and $DR$ for each quarter and horizon (value-weighting). For the aggregated time series, I show the summary statistics in Panel A of Table 20 and the results of regression (70) in Panel B, together with confidence intervals that are based on Newey-West standard errors with the lag set to the number of overlapping quarters. Note that the number of observations across horizons differ. For larger horizons, there are more overlapping periods and hence less observations. However, in untabulated results I show that the requirement that all the time series for all horizons have the same number of observations has a small effect on the results. For this requirement, the sample ends implicitly in 2004, even for shorter horizons, because one needs additional seven years to compute the variables for a horizon of 28 quarters.

Although I use a different sample – for instance, I require that for every observation all ICCs are available –, the results are very similar. In both samples, the variation in capital gain returns is mostly explained by the DR news part for shorter horizons and by the CF news part for longer horizons. At a quarterly horizon, only 17%/16% of the return variation of the market portfolio is explained by CF news in my/their sample. This fraction increases to 72%/59% at a seven-year horizon. Also, the results are robust: at twelve quarters and beyond, the fraction of CF news is above 50%, even for the 5% percentile. In summary, these results imply that cash flow news is important in driving the stock price movements, based on evidence of the CDZ method.

The results from Table 20 can be easily obtained from the BMA approach. In fact, my argument is that Chen, Da, and Zhao (2013) use such an averaging approach, but implicitly set the prior model weight of their CDZ method to one and the weights for all other specifications to zero. That is, they assume that their method is cor-

111 I winsorize $Retx$, $DR$, and $CF$ for each horizon at the 1% and 99% breakpoints, in accordance with Chen, Da, and Zhao (2013). This is the reason why the tautological relation in equation (67) is broken and the slope coefficients in Panel B do not add up to 1, but the deviations are marginal. In line with Chen, Da, and Zhao (2013), I also use quarterly data, i.e., I only consider observations from March, June, September, and December of each year.

Table 20: **Return decomposition using CDZ method.** This table replicates Table 2 in Chen, Da, and Zhao (2013). Panel A reports for the value-weighted market portfolio the mean as well as the variance of capital gain returns (Retx), cash flow (CF) news, and discount rate (DR) news, from one quarter up to 28 quarters. Panel B reports the portion of capital gain returns that can be explained by CF and DR news, respectively. These are determined by regressing CF news and DR news on aggregate Retx. The rows 5% and 95% report the confidence intervals around the coefficients and are based on Newey-West standard errors with the lag set to the number of overlapping quarters. The sample is quarterly and ranges from 1985 to 2011. All numbers are in percent.

| | Horizon (Quarter) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 4 | 8 | 12 | 16 | 20 | 24 | 28 |
| Panel A: Summary statistics | | | | | | | | | |
| Mean(CF) | 2.06 | 4.25 | 7.90 | 15.06 | 21.61 | 31.11 | 43.11 | 55.59 | 68.48 |
| Mean(DR) | 0.24 | 0.49 | 1.86 | 4.95 | 7.96 | 10.88 | 13.69 | 17.19 | 22.42 |
| Mean(Retx) | 2.30 | 4.70 | 9.68 | 19.92 | 29.73 | 42.48 | 57.85 | 74.35 | 92.79 |
| Var(CF) | 0.37 | 0.84 | 2.07 | 5.08 | 8.67 | 13.42 | 19.25 | 25.94 | 32.75 |
| Var(DR) | 0.76 | 1.27 | 2.45 | 3.42 | 3.70 | 4.34 | 4.67 | 4.35 | 6.58 |
| Var(Retx) | 0.59 | 1.27 | 2.53 | 6.26 | 11.72 | 20.69 | 31.39 | 42.69 | 56.58 |
| Panel B: Decomposition | | | | | | | | | |
| 5% | 0.54 | 14.88 | 21.08 | 42.36 | 56.23 | 58.85 | 61.45 | 65.08 | 64.61 |
| CF | 16.53 | 32.32 | 42.45 | 63.19 | 71.02 | 71.37 | 72.19 | 74.33 | 72.37 |
| 95% | 32.52 | 49.77 | 63.82 | 84.02 | 85.80 | 83.89 | 82.93 | 83.59 | 80.13 |
| 5% | 66.50 | 48.27 | 34.33 | 13.35 | 11.19 | 12.78 | 12.18 | 10.15 | 13.87 |
| DR | 82.38 | 66.02 | 56.04 | 34.64 | 26.19 | 25.39 | 23.16 | 20.48 | 22.91 |
| 95% | 98.27 | 83.77 | 77.76 | 55.93 | 41.19 | 37.99 | 34.13 | 30.80 | 31.96 |

Table 21: **Return decomposition using the model averaging approach.** This table updates Table 2 in Chen, Da, and Zhao (2013) by applying the model averaging approach proposed. The posterior model weights are based on $\phi = 1$. In Panel A, the slope coefficients are sampled from a mixture t-distribution where each t-distribution is scaled by the Newey-West standard errors with the lag set to the number of overlapping horizons and the slope coefficient added. The weighting across the t-distribution is based on the posterior model weights. 1,000,000 draws are taken. Panel B is based on 100,000 block-bootstrap samples with a block length of 20 quarters, drawn with replacement. In each run, the CF news based on an ICC method is chosen randomly, subject to the posterior model weights. Then, the slope coefficient for the specific bootstrap sample and an ICC-specific CF news part is returned. Both panels show the 5% percentile, the mean, and the 95% percentile of the generated samples. The sample is quarterly and ranges from 1985 to 2011. All numbers are in percent.

| | Horizon (Quarter) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 4 | 8 | 12 | 16 | 20 | 24 | 28 |
| Panel A: Mixture t-distribution with Newey-West standard errors | | | | | | | | | |
| 5% | -2.23 | 10.44 | 7.45 | 17.46 | 24.70 | 27.64 | 28.79 | 26.72 | 21.93 |
| CF | 15.54 | 26.29 | 27.48 | 36.24 | 39.51 | 39.91 | 40.13 | 40.61 | 38.58 |
| 95% | 35.00 | 43.85 | 48.72 | 61.44 | 68.36 | 69.08 | 70.17 | 72.66 | 70.95 |
| Panel B: Bootstrapped samples | | | | | | | | | |
| 5% | 1.77 | 14.76 | 2.50 | 15.47 | 22.30 | 23.86 | 22.94 | 18.28 | 11.76 |
| CF | 15.53 | 25.89 | 26.45 | 36.00 | 39.76 | 39.86 | 39.50 | 39.28 | 37.22 |
| 95% | 30.77 | 38.03 | 47.75 | 62.12 | 69.92 | 68.27 | 69.34 | 71.63 | 68.84 |

rect on prior grounds, although they do not give any arguments at all as to why their method is definitely correct and other methods such as the PSS or CT method are definitely misspecified. Next, I use more reasonable priors that acknowledge the fact that there are no convincing economic reasons that would allow the dismissal of any ICC method on prior grounds. That is, I set my prior belief that all ICC methods should be equally likely. Also, I set $\phi$ to 1, which gives roughly equal weight to the evidence in the data and my prior beliefs. The sampling from the parameter densities are done as in the previous example. That is, in the first case I sample from a mixture t-distribution, where each t-distribution's parameters are estimated from an OLS with Newey-West corrected standard errors. In the second case, I apply a bootstrap approach in which I choose an ICC method randomly in each run, based on the posterior model weights, and obtain the regression coefficient for the specific bootstrap sample.

Table 21 presents the results. Incorporating uncertainty about the correct ICC specification widens the coverage regions considerably. Only for shorter horizons one can be reasonably sure that returns are

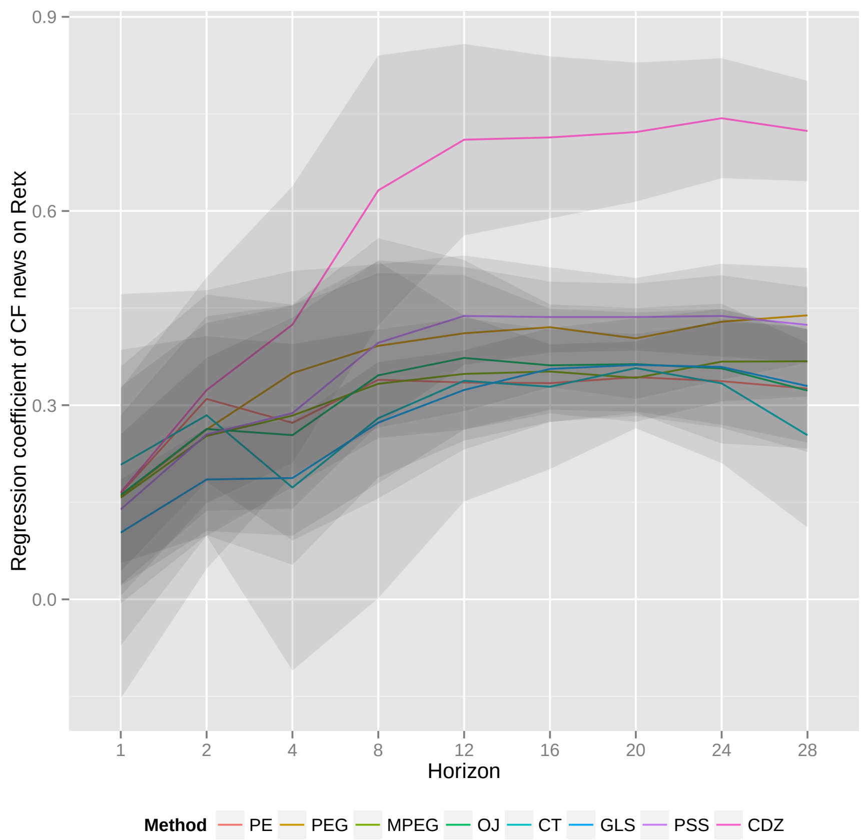


Figure 9: **Fraction of Retx driven by CF news for different ICC methods and horizons.** This figure shows the fraction of variation in Retx attributable to CF news for different ICC methods and horizons. The fraction is defined as the regression coefficient of CF news on Retx. The shaded area around each line represents the 90% confidence bands around the coefficients. The bands are computed via Newey-West standard errors with the lag set to the number of overlapping quarters. The sample is quarterly and ranges from 1985 to 2011.

mostly driven by DR news. For longer horizons, the intervals become too large to draw any reasonable conclusions. The results also show that the two approaches of model averaging yield similar results here. Furthermore, CF news is less important in explaining capital gain returns for longer horizons according to the averaged results, and in stark contrast to the evidence reported in Table 20.

Figure 9 shows why the results conditional on only the CDZ method on the one hand and conditional on the eight ICC methods on the other hand differ so dramatically. It plots the fraction of the variation in capital gain returns that is explained by CF news over various horizons and for different ICC methods. As becomes apparent, the view that most of the variation in capital gain returns is driven by CF news is only supported by the CDZ method. All other methods come to the conclusion that DR news, even for longer horizons, is more important. However, there is also large variation across the remaining methods, which means that the return decomposition approach based on ICCs is sensitive to the specific model.

The rationale behind this finding is that every ICC method equates the current stock price with a transformation of discounted expected dividends. Differences arise on how the second part is transformed. In this particular research question, the specific assumptions have a large impact. Chen, Da, and Zhao (2013) assume that the earnings growth rate converges to the industry long-term growth rate provided by analysts over the next 15 years, although these growth rates are commonly interpreted to represent the next five years (see Claus and Thomas 2001) and are probably affected by analyst bias. Obviously, these growth assumptions are sensitive to the current market environment. For example, during the Dot-com bubble in 2001 the mean across the industry growth rates within my sample was as high as 24%. Assuming that investors expected earnings growth rates to converge to these growth rates for the next 15 years will obviously explain almost all of the capital gains that accrued over this period. Such an extreme assumption is not made by the other methods. For example, the PSS method assumes that the earnings growth rate in period 3 is the earnings growth rate provided by analysts and extrapolates this growth rate over 15 years to the historical average of the nominal GDP growth rate. This much more conservative assumption about expected earnings leaves a much larger part of capital gains unexplained and the ICC has to step in to fill the gap. Similar arguments can be made for the other methods as well.

This, of course, is an outcome of model uncertainty. We do not know if investors updated their long-term earnings growth assumptions or if they updated their expected returns. It is the question we want to answer. Each method emphasizes the two parts differently and hence, results conditional on only one method ignore the uncertainty we have about these assumptions.

Furthermore, a comparison with the Easton and Monahan (2005) evaluation approach that I introduce in Chapter 3 is instructive because it illustrates the circularity we are facing here. Easton and Monahan (2005) define a CF news proxy that is independent of the specific ICC method. Next, they evaluate how much each of the ICC methods fails to explain subsequent realized returns. In their original analysis, all proxies fail massively and Easton and Monahan (2005) take this finding as evidence that all proxies are grossly misspecified. I offer an alternative explanation: It could very well be that their CF news proxy is completely misspecified. In contrast, Chen, Da, and Zhao (2013) follow the approach of defining consistent CF and DR news proxies. They derive them in such a way that the two, irrespective of the specific ICC method, always add up to the capital gain returns. Then, they act as if those proxies are defined correctly and try to answer the question which part is more important. Unfortunately, they only do this for one method and fail to acknowledge the uncertainty they have in the correctness of this method.

### 7.4.3 *Replication of robustness checks*

This example highlights that it is important to implement a broad set of reasonable alternative specifications to check the robustness of results based on the ICC. However, Chen, Da, and Zhao (2013) do not use alternative ICC methods, but only change input parameters of their CDZ method. As I show next, these changes do have a minor impact on the results, which is why the issue of model uncertainty has gone unnoticed in the original study.

There are two broad robustness tests run by Chen, Da, and Zhao (2013). In one test, they allow the steady-state earnings growth rates and plowback rates to be functions of firm characteristics and estimate those using historical data, instead of using the rolling average of the historical nominal GDP growth rate for all firms. Within each of the 12 Fama-French industries, they classify stocks into eight portfolios by an independent triple-sort according to their sizes, book-to-market ratios, and ages. For each of these portfolios, they compute the average earnings growth rates and plowback rates 15 years later. Finally, they use a rolling window of 18 years to calculate portfolio averages of those long-run growth rates and plowback rates and rerun their analysis with these portfolio-level earnings growth rates and plowback rates. The results are very similar to the base case on an aggregate level, on which I focus here. Over a horizon of 4/24 quarters, CF news explains 0.41/0.55 of the variation in capital gain returns in their sample. The values in the base case are 0.36/0.63. Instead of replicating this robustness test exactly, I choose to implement an easier, but in my opinion instructive, robustness check. I simply use arbitrary long-run earnings growth rates.[112] I do so because I hypothesize that the steady-state assumptions do not have such an important impact in the first place and thus any robustness check dealing with them, reasonable or not, should have a minor impact on the results. I use two different steady-state growth rates. First, I assign every industry an integer number $i$ ranging from 1 to 48.[113] For each firm, I then use a steady-state earnings growth rate of $2 \times i/10\%$. For instance, if a firm falls into industry assigned the integer 30, I assume a growth rate of 6% for this firm. This results in a large cross-sectional variation of the steady-state earnings growth rates from 0.1% to 9.6%. Second, for each firm within one industry and quarter, I assign a random number that is drawn from a uniform distribution with a minimum of 0.1% and a maximum of 9.6%. This is an extremely unreasonable assumption because it allows large random shocks in the steady-state growth rates. For example, it is possible that firms within one industry are assumed to have a steady-state growth rate of 1% in one quarter and 9% in the next. Also, this test allows for a large variation both over time and firms.

In another set of robustness tests, they control for analyst forecast bias. In one test, they simply replace the consensus mean forecast with the highest or lowest analyst forecast. Additionally, they adjust the forecasts based on known relations between the forecast bias and other variables. To be more precise, they cite studies that document that analysts are optimistic in particular for firms for which there is large investment banking demand. To control for this effect, they construct earnings forecasts as weighted averages between the most pessimistic and most optimistic forecast, where the weights are dependent on the amount of cash raised through external financing, which is a proxy for investment banking demand. Furthermore, they compute the weights based on recent forecast errors because these forecast errors have been found to be persistent. In their Table 3, they find CF news even more important if they use the highest analyst forecast. For example, over a horizon of 24 quarters the slope coefficient of CF news on capital gain returns is 0.85 instead of 0.63. With the lowest forecasts, it decreases slightly to 0.51, but CF news still explains more than half of the subsequent capital gain returns. Because the adjusted

112 Note that the long-run earnings growth rates implicitly determine the steady-state plowback rates. Refer to Section 2.2.3 for a description of the relation between the steady-state earnings growth rate and plowback rate.

113 Because I use the 48 Fama-French industry classification, I have 48 industries to start with.

forecasts are a weighted average of the highest and lowest forecasts, it is not surprising that the results based on these adjusted forecasts are closer to the base case than the results based on the highest or lowest forecasts, respectively. Therefore, I only replicate the robustness tests with the highest and lowest forecasts.

In the following, I use an alternative data set that only requires that the CDZ methods and its derivatives described in this section return a value for the ICC. This allows me to better compare my results with Chen, Da, and Zhao (2013). Table 22 shows the summary statistics for the CDZ derivatives as well as the number of firms, the median market capitalization, and the median steady-state plowback rate for each year, which I can now compare with Panel A in Table 1 of Chen, Da, and Zhao (2013). The two samples are very similar, both in the levels as well as the dynamics over time. For example, the median market capitalization at the beginning of my sample is 273 million dollars and increases to 1203 million dollars by 2010. They report values of 238 and 1343 million dollars, respectively. Furthermore, the mean of the median base ICC time series between my and their sample is almost identical (13.4% vs. 13.5%) and the correlation is almost perfect (99.6%). Finally, the cross-sectional variation is very similar as well, which can be seen by a comparison between the Q1 and Q3 columns in my and their table. Alternatively, one can look at the statistics for the cross-sectional standard deviation in each year for both samples. The correlation between the time series of these cross-sectional standard deviations is 97.1% and the mean of the two time series differs only by 0.07 percentage points.

The ICCs computed with unreasonable steady-state growth rates do not differ much from the base method. This is a first indication that changes in the steady-state growth rate do not have a large impact on the results. In contrast, using high or low forecasts instead of the consensus mean has a large impact on the level of the ICC, a result that was already shown in Figure 4 and Figure 5 for a different sample composition. However, these time series have a downward trend in the ICC over time as well.

Table 23 reports the decomposition results for the different derivatives of the CDZ method. Panel A is almost identical with the results reported in Table 20, which is based on a sample for which all ICC methods have to have a numerical value. This suggests the conclusion that sample selection issues discussed in Chapter 5 are not relevant. The results of Panel B and Panel C for different steady-state earnings

Table 22: **Summary statistics for the CDZ method by year.** This table replicates Panel A of Table 1 in Chen, Da, and Zhao (2013). For each year, this table shows the average number of firms per quarter, the median market capitalization (in millions of dollars), and the median steady-state plowback rate (PB) for the sample. The sample consists of all firms for which an ICC is available for each derivative of the CDZ method. The derivatives are: the base CDZ method (Base method); an ICC based on steady-state earnings growth rates that vary by industry (SS 1); an ICC based on steady-state earnings growth rates that vary by industry and quarter (SS 2); an ICC based on the highest earnings forecasts (Highest EPS); an ICC based on the lowest earnings forecasts (Lowest EPS). For these ICC derivatives, this table shows the first quartile (Q1), the median, and the third quartile (Q3) as well as the standard deviation (SD) of the cross-sectional distribution of the ICCs. The steady-state growth rate as well as the columns about the ICC are in percent. The sample is quarterly and ranges from 1985 to 2011.

| | Number | Market | Steady | Base method | | | | SS 1 | | | | SS 2 | | | | Highest EPS | | | | Lowest EPS | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | of firms | Cap | -state PB | Q1 | Median | Q3 | SD | Q1 | Median | Q3 | SD | Q1 | Median | Q3 | SD | Q1 | Median | Q3 | SD | Q1 | Median | Q3 | SD |
| 1985 | 1615 | 272.9 | 51.2 | 13.6 | 15.5 | 17.9 | 4.2 | 14.1 | 16.1 | 18.6 | 4.2 | 14.3 | 16.3 | 18.7 | 4.3 | 16.2 | 18.7 | 22.0 | 5.5 | 10.7 | 12.7 | 14.9 | 4.1 |
| 1986 | 1620 | 298.2 | 57.3 | 11.9 | 13.8 | 16.0 | 3.9 | 12.5 | 14.3 | 16.6 | 3.9 | 12.4 | 14.4 | 16.7 | 3.9 | 14.6 | 17.0 | 20.2 | 5.2 | 8.9 | 10.9 | 13.0 | 3.9 |
| 1987 | 1657 | 316.4 | 56.7 | 11.7 | 13.8 | 16.4 | 4.3 | 12.3 | 14.4 | 16.8 | 4.2 | 12.4 | 14.4 | 17.0 | 4.3 | 14.4 | 17.0 | 20.4 | 5.3 | 8.8 | 11.0 | 13.3 | 4.3 |
| 1988 | 1660 | 295.0 | 53.7 | 12.6 | 14.6 | 17.0 | 4.2 | 13.2 | 15.1 | 17.5 | 4.2 | 13.1 | 15.3 | 17.7 | 4.2 | 15.0 | 17.6 | 20.6 | 5.1 | 9.8 | 11.9 | 14.3 | 4.3 |
| 1989 | 1692 | 310.7 | 57.1 | 11.7 | 13.7 | 16.0 | 3.9 | 12.3 | 14.3 | 16.6 | 4.0 | 12.3 | 14.4 | 16.7 | 4.0 | 14.1 | 16.4 | 19.4 | 4.8 | 9.1 | 11.2 | 13.4 | 3.9 |
| 1990 | 1710 | 281.6 | 53.0 | 12.4 | 14.7 | 17.4 | 4.4 | 13.0 | 15.2 | 18.0 | 4.4 | 13.0 | 15.3 | 18.0 | 4.4 | 15.0 | 17.7 | 20.9 | 5.2 | 9.6 | 12.0 | 14.5 | 4.3 |
| 1991 | 1748 | 344.2 | 57.3 | 11.5 | 13.4 | 15.7 | 3.7 | 12.0 | 14.0 | 16.3 | 3.8 | 12.0 | 14.0 | 16.3 | 3.8 | 14.0 | 16.4 | 19.3 | 4.7 | 8.8 | 10.7 | 12.9 | 3.8 |
| 1992 | 1904 | 378.3 | 57.2 | 11.5 | 13.4 | 15.7 | 3.8 | 11.9 | 13.8 | 16.2 | 3.9 | 12.2 | 14.1 | 16.4 | 3.9 | 13.7 | 16.1 | 18.9 | 4.5 | 9.0 | 10.9 | 13.2 | 3.8 |
| 1993 | 2159 | 387.6 | 57.1 | 11.4 | 13.3 | 15.8 | 3.9 | 11.8 | 13.7 | 16.3 | 4.0 | 11.9 | 13.9 | 16.3 | 3.9 | 13.6 | 15.9 | 18.8 | 4.5 | 8.9 | 11.0 | 13.3 | 3.8 |
| 1994 | 2552 | 326.0 | 54.2 | 12.0 | 14.0 | 16.4 | 4.1 | 12.3 | 14.4 | 16.9 | 4.2 | 12.4 | 14.6 | 17.1 | 4.3 | 14.0 | 16.4 | 19.2 | 4.7 | 9.8 | 11.8 | 14.2 | 4.1 |
| 1995 | 2680 | 360.4 | 53.7 | 12.0 | 14.0 | 16.3 | 3.9 | 12.3 | 14.5 | 16.8 | 4.0 | 12.4 | 14.5 | 16.9 | 4.0 | 13.9 | 16.3 | 19.0 | 4.6 | 9.8 | 11.8 | 14.0 | 3.9 |
| 1996 | 2940 | 383.3 | 53.7 | 11.8 | 13.9 | 16.6 | 4.3 | 12.1 | 14.4 | 17.0 | 4.3 | 12.5 | 14.6 | 17.3 | 4.3 | 13.8 | 16.4 | 19.6 | 5.1 | 9.6 | 11.6 | 14.1 | 4.1 |
| 1997 | 3240 | 416.8 | 52.5 | 11.6 | 14.2 | 17.4 | 4.9 | 11.9 | 14.7 | 17.9 | 4.9 | 12.0 | 14.8 | 18.0 | 5.0 | 13.7 | 16.8 | 20.7 | 5.8 | 9.4 | 11.8 | 14.7 | 4.6 |
| 1998 | 3216 | 416.1 | 50.0 | 11.7 | 14.8 | 18.3 | 5.7 | 12.1 | 15.3 | 18.7 | 5.8 | 12.3 | 15.3 | 18.8 | 5.7 | 14.2 | 17.9 | 22.1 | 7.0 | 9.2 | 12.0 | 15.3 | 5.3 |
| 1999 | 2974 | 434.1 | 50.1 | 11.9 | 14.8 | 18.4 | 5.4 | 12.2 | 15.2 | 18.8 | 5.5 | 12.5 | 15.3 | 18.9 | 5.4 | 14.4 | 17.9 | 22.1 | 6.7 | 9.3 | 11.9 | 15.2 | 5.2 |
| 2000 | 2637 | 614.7 | 47.8 | 12.6 | 15.4 | 19.0 | 6.2 | 13.0 | 15.8 | 19.4 | 6.2 | 13.1 | 15.9 | 19.5 | 6.2 | 15.2 | 18.7 | 23.2 | 7.7 | 9.8 | 12.6 | 15.8 | 5.9 |
| 2001 | 2281 | 686.6 | 53.0 | 11.4 | 13.8 | 17.6 | 6.9 | 11.6 | 14.2 | 18.0 | 6.9 | 12.0 | 14.4 | 18.1 | 6.9 | 13.7 | 17.2 | 22.5 | 9.0 | 8.7 | 10.9 | 13.8 | 6.2 |
| 2002 | 2324 | 654.7 | 55.4 | 11.1 | 13.1 | 15.7 | 4.8 | 11.3 | 13.5 | 16.2 | 4.9 | 11.5 | 13.6 | 16.3 | 4.9 | 13.5 | 16.7 | 20.8 | 6.9 | 8.2 | 10.2 | 12.3 | 4.4 |
| 2003 | 2364 | 720.4 | 61.7 | 9.9 | 11.6 | 13.7 | 3.7 | 10.0 | 12.0 | 14.2 | 3.8 | 10.5 | 12.2 | 14.3 | 3.8 | 11.9 | 14.5 | 17.5 | 5.1 | 7.3 | 9.2 | 11.1 | 3.5 |
| 2004 | 2423 | 909.6 | 63.4 | 9.6 | 11.3 | 13.1 | 3.6 | 9.8 | 11.6 | 13.5 | 3.7 | 9.9 | 11.7 | 13.6 | 3.6 | 11.4 | 13.9 | 16.6 | 4.9 | 7.3 | 9.0 | 10.6 | 3.3 |
| 2005 | 2497 | 1031.6 | 63.3 | 9.7 | 11.3 | 13.1 | 3.5 | 9.9 | 11.6 | 13.5 | 3.7 | 10.0 | 11.8 | 13.5 | 3.6 | 11.5 | 13.8 | 16.5 | 4.8 | 7.4 | 9.0 | 10.6 | 3.2 |
| 2006 | 2451 | 1109.2 | 62.6 | 9.8 | 11.4 | 13.3 | 3.5 | 9.9 | 11.7 | 13.7 | 3.6 | 10.3 | 11.9 | 13.8 | 3.5 | 11.5 | 13.8 | 16.6 | 4.8 | 7.6 | 9.1 | 10.7 | 3.1 |
| 2007 | 2369 | 1203.4 | 61.5 | 9.9 | 11.6 | 13.4 | 3.3 | 10.1 | 11.9 | 13.8 | 3.4 | 10.2 | 12.0 | 13.8 | 3.4 | 11.6 | 13.8 | 16.4 | 4.4 | 7.7 | 9.4 | 11.0 | 3.1 |
| 2008 | 2210 | 894.6 | 54.8 | 10.8 | 12.9 | 15.2 | 4.8 | 11.0 | 13.2 | 15.7 | 4.9 | 11.1 | 13.3 | 15.7 | 4.8 | 12.8 | 15.4 | 18.5 | 6.1 | 8.4 | 10.4 | 12.5 | 4.4 |
| 2009 | 1777 | 867.3 | 61.0 | 9.4 | 11.3 | 13.5 | 4.5 | 9.6 | 11.6 | 13.9 | 4.6 | 9.8 | 11.7 | 14.0 | 4.6 | 11.5 | 13.9 | 16.6 | 5.4 | 6.8 | 8.7 | 10.9 | 4.3 |
| 2010 | 1861 | 1202.9 | 60.4 | 9.6 | 11.3 | 13.3 | 3.9 | 9.7 | 11.6 | 13.7 | 4.0 | 9.9 | 11.7 | 13.7 | 4.0 | 11.5 | 13.7 | 16.3 | 5.0 | 7.1 | 8.9 | 10.8 | 3.7 |
| 2011 | 1912 | 1340.7 | 56.3 | 10.2 | 12.1 | 14.2 | 4.3 | 10.3 | 12.3 | 14.6 | 4.5 | 10.5 | 12.5 | 14.7 | 4.4 | 12.2 | 14.6 | 17.4 | 5.6 | 7.6 | 9.5 | 11.6 | 4.0 |

Table 23: **Return decomposition using CDZ method (robustness).** This table performs robustness checks similar to Table 3 in Chen, Da, and Zhao (2013). In Panel A, I use the base CDZ method to compute the ICC. In Panel B, an alternative steady-state earnings growth rate is used that varies over industries. In Panel C, the steady-state earnings growth rate is varied over both industries and time. Panel D and E are based on ICCs computed with the highest and lowest earnings forecasts, respectively. In all panels I report the portion of capital gain returns that can be explained by CF and DR news, respectively. These are determined by regressing CF news and DR news on aggregate Retx. The rows 5% and 95% report the confidence intervals around the coefficients and are based on Newey-West standard errors with the lag set to the number of overlapping quarters. The sample is quarterly and ranges from 1985 to 2011. All numbers are in percent.

| | Horizon (Quarter) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 4 | 8 | 12 | 16 | 20 | 24 | 28 |
| Panel A: Base CDZ method | | | | | | | | | |
| 5% | -1.64 | 16.57 | 16.17 | 36.44 | 52.70 | 58.79 | 64.57 | 67.14 | 63.64 |
| CF | 12.62 | 31.46 | 36.59 | 60.63 | 71.75 | 72.78 | 73.49 | 74.71 | 71.38 |
| 95% | 26.87 | 46.34 | 57.00 | 84.82 | 90.80 | 86.77 | 82.40 | 82.29 | 79.12 |
| 5% | 72.05 | 51.75 | 40.88 | 11.62 | 5.16 | 9.90 | 12.38 | 10.63 | 14.69 |
| DR | 86.26 | 66.88 | 61.89 | 36.71 | 24.58 | 23.61 | 21.34 | 19.50 | 23.71 |
| 95% | 100.47 | 82.01 | 82.90 | 61.80 | 43.99 | 37.31 | 30.29 | 28.38 | 32.74 |
| Panel B: Arbitrary steady-state growth rates by industry | | | | | | | | | |
| 5% | -1.44 | 16.62 | 16.03 | 35.87 | 51.79 | 57.91 | 63.61 | 65.93 | 62.36 |
| CF | 12.67 | 31.31 | 36.26 | 59.78 | 70.67 | 71.70 | 72.31 | 73.37 | 69.97 |
| 95% | 26.78 | 46.00 | 56.49 | 83.70 | 89.56 | 85.49 | 81.01 | 80.80 | 77.57 |
| 5% | 72.14 | 52.07 | 41.37 | 12.73 | 6.39 | 11.17 | 13.71 | 11.99 | 16.14 |
| DR | 86.20 | 67.02 | 62.20 | 37.55 | 25.65 | 24.67 | 22.47 | 20.80 | 25.11 |
| 95% | 100.27 | 81.97 | 83.04 | 62.38 | 44.90 | 38.17 | 31.24 | 29.62 | 34.08 |
| Panel C: Arbitrary steady-state growth rates over time and industries | | | | | | | | | |
| 5% | -0.57 | 18.95 | 15.76 | 36.73 | 53.25 | 58.77 | 64.48 | 67.44 | 63.86 |
| CF | 15.10 | 32.93 | 35.98 | 60.46 | 71.25 | 72.40 | 73.28 | 74.62 | 71.21 |
| 95% | 30.76 | 46.90 | 56.21 | 84.18 | 89.26 | 86.04 | 82.07 | 81.81 | 78.57 |
| 5% | 68.12 | 51.24 | 41.74 | 12.73 | 6.99 | 10.79 | 12.68 | 11.33 | 15.24 |
| DR | 83.81 | 65.46 | 62.50 | 37.08 | 25.15 | 24.04 | 21.51 | 19.68 | 23.83 |
| 95% | 99.49 | 79.68 | 83.26 | 61.44 | 43.31 | 37.29 | 30.34 | 28.03 | 32.42 |
| Panel D: CDZ method with highest forecasts | | | | | | | | | |
| 5% | -16.50 | 11.22 | 20.96 | 52.28 | 71.85 | 77.25 | 80.38 | 82.25 | 75.16 |
| CF | 3.30 | 28.63 | 36.67 | 71.34 | 86.89 | 89.02 | 89.35 | 91.36 | 85.74 |
| 95% | 23.09 | 46.04 | 52.37 | 90.39 | 101.92 | 100.79 | 98.31 | 100.47 | 96.31 |
| 5% | 75.34 | 51.85 | 45.23 | 5.61 | -5.97 | -4.80 | -3.69 | -5.74 | -1.75 |
| DR | 95.34 | 69.10 | 61.30 | 25.80 | 9.98 | 7.33 | 5.85 | 4.08 | 9.54 |
| 95% | 115.35 | 86.34 | 77.36 | 45.99 | 25.92 | 19.47 | 15.38 | 13.91 | 20.83 |
| Panel E: CDZ method with lowest forecasts | | | | | | | | | |
| 5% | -0.19 | 10.14 | 5.09 | 19.25 | 35.30 | 44.12 | 55.42 | 61.26 | 59.62 |
| CF | 13.32 | 28.13 | 34.48 | 54.88 | 63.19 | 62.97 | 65.69 | 68.22 | 65.88 |
| 95% | 26.83 | 46.11 | 63.86 | 90.51 | 91.07 | 81.82 | 75.96 | 75.17 | 72.15 |
| 5% | 71.23 | 51.51 | 31.92 | 5.90 | 5.23 | 14.62 | 19.54 | 18.41 | 21.76 |
| DR | 85.02 | 69.89 | 62.78 | 42.38 | 33.23 | 33.08 | 29.19 | 26.33 | 29.55 |
| 95% | 98.81 | 88.27 | 93.64 | 78.87 | 61.23 | 51.55 | 38.85 | 34.25 | 37.34 |

growth assumptions are virtually identical to Panel A. This is a confirmation of my hypothesis that changes in the steady-state earnings growth assumption have a negligible effect on the return decomposition approach. It is therefore not surprising that the robustness test performed by Chen, Da, and Zhao (2013) does not change their conclusion that CF news is the more important part of stock price movements in the long run. The same is true for robustness checks with respect to analyst forecast bias. Panel D and E report the results using the highest and lowest analyst forecasts. In line with Chen, Da, and Zhao (2013), I find CF news to be even more important based on the highest forecasts. For example, they report that CF news explains 39% of return variance at the annual frequency for the aggregate portfolio and 85% at six years. I find values of 37% and 91%, respectively. Using the lowest forecasts, the portion of CF news decreases in both samples. They now report that CF news explains 33% of the return variance at the annual frequency and 51% at the six-year horizon, while I obtain values of 34% and 68%.

In summary, I find that the return decomposition results are robust to variations *within* the CDZ method, in line with the robustness tests of Chen, Da, and Zhao (2013). Nevertheless, the results are not robust with respect to variations *between* different ICC methods. It is therefore of paramount importance to consider the model uncertainty in the choice of an ICC method in the statistical inference because otherwise the results can imply too much confidence about the relative importance of the drivers of stock price movements. Furthermore, this example has shown that the BMA approach is an easy to use and easy to interpret way to incorporate model uncertainty.

# 8 CONCLUSION

## 8.1 SUMMARY OF MAIN RESULTS

The main contribution of this thesis is the introduction of a Bayesian model averaging (BMA) approach into the variable selection problem that a researcher faces who can choose between a multitude of expected return proxies. While this approach is applicable to any expected return proxy, I focus on the implied cost of capital (ICC) in my discussions, which is defined as the internal rate of return that equates the current stock price with expected future dividends. To this day, it is the most prominent measure of return expectations that is not based, directly or indirectly, on realized returns.

In Chapter 2, I first summarize earlier work by Pástor, Sinha, and Swaminathan (2008) and Hughes, Liu, and Liu (2009) who examine the underlying theoretical relation between expected returns on the one hand and the ICC on the other hand. It turns out that there are noteworthy differences between the two concepts even if the vector of future expected dividends is measured without error. Nevertheless, the true, but unobservable ICC is a useful proxy for expected returns and under certain simplifying assumptions even perfectly correlated with it. Yet, the vector of expected dividends is as unobservable as expected returns and therefore additional assumptions have to be made to approximate the ICC. The last decade has seen an explosion of research that proposes adjustments to these simplifying assumptions, both minor and major ones. The second part of Chapter 2 explains and categorizes them. Finally, I give a non-exhaustive overview of empirical studies that apply the ICC with a focus on how these studies select their specific ICC implementation from the numerous options available. I come to the conclusion that the current approach is ad-hoc and does not follow any clear guidelines.

It is well known from the model selection and model averaging literature that such an ad-hoc selection of a proxy raises concerns of data fishing. Maybe a researcher simply selects the proxy that is most in line with his research question, instead of the proxy that is the best in tracking true, but latent expected returns. To alleviate such concerns, Easton and Monahan (2005) and Lee, So, and Wang (2011), among others, recommend evaluating the different proxies to identify the best one in a preceding step. While there are minor differences between the approaches, which I present in Chapter 3, the fundamental idea is the same for all of them: any proxy that tracks expected returns next period has to explain subsequent realized returns eventually. This directly follows from the Campbell and Shiller (1988) (CS) return decomposition that shows that log realized returns over a period are the sum of expected returns at the beginning of the period and cash flow and discount rate news that arrive in the period. Because the news parts are uncorrelated with expected returns and zero on average, it directly follows that a proxy is better, the better it is able to explain subsequent realized returns. Most studies run univariate predictive regressions of subsequent realized returns on their expected return proxy to test the quality of this proxy.

By contrast, Easton and Monahan (2005) recommend extending such regressions by additional proxies for cash flow news and discount rate news. I object to this recommendation. Easton and Monahan (2005) propose cash flow and discount rate news proxies based on theoretical arguments and simplifying assumptions. Therefore, the quality of these proxies is dependent on these assumptions and I consider it very unlikely that it is possible to approximate the news part reasonably. We already have a very hard time estimating expected returns next period, but the news parts contain changes in expectations about all future periods. This seems to be an impossible task and I confirm this presumption later in my empirical analysis. As a consequence, their evaluation method introduces additional issues of measurement error and also ignores the major positive feature of the news parts, namely their asymptotic unbiasedness and uncorrelatedness with expected returns next period. Furthermore, I show that it is always possible to define the proxies for cash flow news and discount rate news in such a way that the evaluation method of Easton and Monahan (2005) indicates *incorrectly* a perfect relation between *any* proxy and true expected returns. In summary, the main takeaway from this chapter is that there is great uncertainty about which proxy measures expected returns best.

Ignoring such uncertainty in the statistical inference can produce severe biases in all statistical measures for the classical linear model, which has prompted Breiman (1992) to label such ignorance the *quiet scandal* in the statistical community. BMA allows the incorporation of such uncertainty and has found increased popularity in recent years. Nevertheless, to the best of my knowledge it has not been applied to the variable selection problem in the expected return proxy literature. Chapter 4 fills this gap. It shows how reasonable assumptions about the prior and the likelihood lead to posterior model weights with an analytical solution. In other words, the posterior model weights denote the degree of support for each proxy, given the prior beliefs of the researcher and its association with subsequent realized returns. Because the model weights are dependent on the ability of each proxy to explain subsequent realized returns, this provides a nice link to the discussion in Chapter 3: any alternative measure of expected returns has to rely on subsequent realized returns for its empirical evaluation. The only way to circumvent this reliance is to argue on *prior grounds* that the measure tracks expected returns perfectly, thereby setting the model weight of this proxy implicitly to one.

The theoretical explanations of this chapter show that in large samples the BMA approach yields results similar to a model selection approach that tries to pick the best proxy. This holds because it is not a problem to identify the best proxy in large samples. The chapter also shows that the issue of model uncertainty is only of second-order importance if all proxies come to similar conclusions. Intuitively, if all proxies tell the same story, an average across all proxies will also do so. However, I emphasize the importance of model averaging in cases in which different proxies lead to different results. And finally, the BMA framework allows me to evaluate the impact on the results if all proxies are biased. Not surprisingly, the average across all of these proxies will then also be biased. This is, at least in my opinion, a severe shortcoming of any alternative proxy in comparison to realized returns and important for the applied researcher to keep in mind.

Chapter 5 introduces the data set that is used in the following chapters to empirically test the statements about the Easton and Monahan (2005) evaluation approach and to apply the BMA approach to three different research questions. This data set is based on the complete universe of listed US companies for which the relevant variables from IBES and Compustat are available. With respect to the valuation model, I focus on two derivatives of the residual income model (CT, GLS), two derivatives of the dividend discount model (PSS, CDZ), and three derivatives of the abnormal earnings growth model (PEG, MPEG, OJ). I also implement the forward PE ratio as a naïve benchmark. Because the latter ignores any growth after the next period, it yields the lowest estimates for expected returns by far.

In general, the results confirm the findings of previous studies. First, the standard deviation of each ICC method is an order of magnitude smaller than the standard deviation of realized returns (cf., e.g., Lee, Ng, and Swaminathan 2009). Second, the abnormal earnings growth models typically lead to somewhat larger estimates than the estimates based on the residual income models (cf., e.g., Daske, Van Halteren, and Maug 2010 and Hail and Leuz 2009). Third, the correlation in the aggregated time series is very high, while the cross-sectional correlation is substantially lower. This confirms claims by Li, Ng, and Swaminathan (2013) that the aggregate ICC is likely to be less noisy because the averaging of the firm-level ICCs reduces the impact of idiosyncratic errors. Finally, I perform sensitivity analyses with respect to the number of observations per method, time misalignment issues, and analyst forecast bias. While these sensitivity checks are rather simplistic, they provide a first indication that the dynamics of the ICC are not substantially affected by those issues. Of course, the analyst forecast bias has a large impact on the levels of the ICC estimates.

In Chapter 6, I confirm the findings of Chapter 3. I show that it is possible for each of the eight ICC estimates to establish an apparent and incorrect perfect relation between the ICC estimate and subsequent realized returns by defining the news proxies accordingly. Also, I show that the news proxies defined by Easton and Monahan (2005) are severely misspecified.

Finally, in Chapter 7 I apply the BMA approach to three research questions that have previously been examined with the help of the aggregate ICC, but only conditional on one method. To do so, I compute the posterior model weights first, which already reveal two interesting findings. First, there is no clear winner between the different methods. For example, the MPEG method gets the highest support from the data. But even in the limiting case in which the researcher discards all prior information ($\phi = \infty$), the posterior model weight of the MPEG method is only 17%. Also, the CDZ method gets the least support from the data, which is an indication that the extreme earnings growth assumptions made by Chen, Da, and Zhao (2013) do not coincide with investors' expectations. Second, there is large uncertainty about the posterior model weights, as a simple bootstrap exercise shows. For instance, in the case of non-informative priors the

1% and 99% percentiles of the weights for the MPEG method are 0.8% and 36%, respectively. This finding is again driven by the large shocks that affect realized returns and emphasizes the general point that it is not possible to precisely identify a single best ICC proxy from a set of proxies in small samples.

Equipped with the posterior model weights, I replicate three studies from the ICC literature. I start with the study by Claus and Thomas (2001) who use the CT method to approximate the implied risk premium. In this example, model uncertainty dominates parameter uncertainty, which illustrates the importance of incorporating the former by a model averaging approach. The confidence bands are roughly three times larger for the implied risk premium mean over all ICC methods than the confidence bands conditional on only a specific ICC method. The reason why model uncertainty is so important here is the fact that the levels of different ICC estimates differ widely, as has been shown in the summary statistics in Chapter 5. Taken for face value, the mean implied risk premium for the US from 1985 to 2011 was 4.5%. In the second example, I revisit the study of Pástor, Sinha, and Swaminathan (2008) who report a positive risk-return tradeoff. That is, they regress the ICC based on the PSS method on a volatility measure – either the daily standard deviation or variance over a month – and find a positive slope coefficient. Incorporating model uncertainty into this research question does not alter the main conclusion. Because all methods indicate such a positive risk-return tradeoff, an average across all methods does so as well. This example illustrates nicely that BMA per se does not destroy the advantages of alternative expected return proxies. If all proxies come to similar conclusions, the results are robust to specification errors. Only if the research question is sensitive to the specific proxy used, BMA does correctly widen the confidence bands to indicate the uncertainty a researcher should have on the results conditional on a specific proxy. This is exactly what happens in the third research question I look at. I repeat the analysis of Chen, Da, and Zhao (2013) who want to answer the question whether stock price movements are driven to a larger extent by changes in cash flow expectations or by changes in discount rate expectations. To do so, they rely on the ICC and define changes in cash flow expectations as hypothetical price changes due to updated earnings forecasts, while leaving the ICC constant; and they define changes in discount rate expectations as hypothetical price changes due to an updated ICC, while leaving the earnings forecasts constant. As a proxy for the true, but unobservable ICC, they use the CDZ method and report that most of the changes in stock prices are driven by CF news. On the contrary, I show that this result is not robust to the specific ICC method. Conditional on a specific method, a researcher could claim that the majority of stock price movements is driven by CF news or, at the other extreme, that they are almost exclusively driven by DR news. Therefore, it is of paramount importance to average across the evidence of all methods and in this case the coverage regions are too wide to make any meaningful statement.

At the very least, this dissertation has shown that researchers have to identify and control for the major parameters of an expected return proxy. In the case of the ICC, those are the differences in long-term or steady-state growth assumptions and different versions of the dividend discount model. Only controlling for minor parameters, such as the analyst bias in the first two or three forecasts, has typically a far smaller effect. As a consequence, even robustness sections that look extensive can be misleading. Additionally, I propose a method that allows a researcher to directly incorporate model uncertainty into his statistical inference. The results are easy to understand and interpret.

## 8.2 Limitations

A first limitation of this thesis is my focus on eight ICC specifications in the empirical part of this thesis. While I implement the most common approaches from the literature, one could criticize my choice of the eight different methods as ad-hoc and arbitrary. Thus it would have been interesting to implement more ICC specifications such as ICCs computed with regression-based earnings forecasts, with adjusted earnings forecasts or simultaneously estimated long-term earnings growth rates. Of course, incorporating more proxies is time-consuming, both for the researcher and the reader in the case of current approaches which present separate results for several specifications that are chosen ad-hoc. The former has to implement the methods, obtain additional variables, etc. The latter has to work through a larger robustness section and summarize the results that are spread out in many tables. A researcher therefore faces a tradeoff between the consideration of model uncertainty on the one hand and his and his readers time constraints on the other hand. It is obvious that a researcher cannot implement all specifications proposed, but my research also showed that he should at least take the evidence of four to five major different specifications into account. It is therefore an interesting open question about how a researcher should deal with this tradeoff. Because the BMA approach is easily extended to as many proxies a researcher considers reasonable without any lengthening of the empirical analysis, this tradeoff is now solely determined by the time constraints of the researcher.

Another caveat of my dissertation is a theoretical one. To derive the formulas necessary to compute the posterior model weights, I rely on certain simplifying restrictions such as normality, linearity, and the choice of priors. This is in line with current practice in the literature, but a better understanding about the impact of these restrictions on the results is certainly desirable. In particular, I make the assumption that the exogenous variables, which are the expected return proxies in my case, are either fixed in repeated samples or independent of the error term. In the case of predictive regressions, which I have to deal with here, this assumption is clearly violated (cf., e.g., Stambaugh 1999). My approach to this problem is similar to Wright (2008), who acknowledges the issue, but simply wants to assess the results obtained by BMA. My simulation results provide additional evidence that the BMA approach does indeed yield better results than ignoring model uncertainty altogether, which is in my opinion quite intuitive in examples such as the replication of Chen, Da, and Zhao (2013). Because I need this assumption to *relatively* compare the predictive power of each expected return proxy, and not to evaluate the *absolute* quality of one predictor, I do not see how the weights could be biased towards one proxy in particular. Moreover, one can always increase the informativeness of the priors so that the information in the data is discarded. In this case, the posterior model weights are equally weighted and problems from the predictive regressions are non-existent. Given that the posterior model weights are rather evenly distributed in my sample anyways, for many research questions it should be a reasonable approach to ignore the evidence of predictive regressions and simply use an average across different proxies. This simplifies the analysis substantially.

Finally, I tackled the variable selection problem encountered in the expected return proxy literature from a model uncertainty perspective and settled on the BMA approach as a solution to this problem. It might be more natural to tackle the variable selection problem from a measurement error perspective.

## 8.3 OUTLOOK

In this dissertation, I focused on only three exemplary research questions to show the implementation of the BMA approach. Naturally, it would be interesting to revisit many more studies that do not account for model uncertainty sufficiently and check the robustness of the original results. In particular, one could focus on studies that deal with the cross-sectional variation in ICC estimates. As I have shown in Chapter 5, the cross-sectional summary statistics differ much more between different ICC methods than in the case of the aggregate ICC. Thus I think that model uncertainty is an even more important issue in cross-sectional studies. Because these studies often focus on regression models with many control variables, one would have to sample the coefficients for each variable separately.

Furthermore, the BMA approach allows a comparison across different proxy classes, which is something that has not been done to the best of my knowledge. Maybe there is only great uncertainty between different ICC methods, but the ICC approach is clearly dominated by expected returns from CDS or bond yield data, or vice versa. The BMA can help answering such a question.

Also, predictive regressions are the most common approach to measuring aggregate return expectations (cf. Kelly and Pruitt 2013) and great advances have been made in the last decade to improve this approach. In particular, Pástor and Stambaugh (2009) deal with the problem of imperfect predictors, that is, predictors that are correlated with true expected returns but cannot deliver it perfectly. Of course, any alternative expected return proxy can be interpreted as such an imperfect predictor. Therefore, one can use expected return proxies to try to predict subsequent realized returns, as recently done by Li, Ng, and Swaminathan (2013). Furthermore, one can then use the results from these regressions to back out the expected return process again. Alternatively, one can directly take the estimates of the alternative proxies, but use predictive regressions to evaluate the quality of such proxies. This shows that predictive regressions on the one hand and alternative expected return proxies on the other hand are both competing approaches to measure expected returns, but the former is also necessary to evaluate the latter. In brief, there is an inherent link between the two approaches and I consider my thesis only as a first step in working this relation out.

Finally, the BMA framework nicely illustrates how to evaluate any expected return proxy class. All these proxies are derived from an underlying theoretical model that links unobservable expected returns to observable input parameters. It is then possible to back out the unobservable expected returns, subject to additional simplifying assumptions. The more leeway a method has in setting these assumptions, the larger is the uncertainty a researcher faces about the correctness of one specific implementation of this method. Take the ICC as an example. It relies on the correct estimation of the vector of all future expected dividends. There is large uncertainty about this vector and hence large differences between different ICC specifications. If we could estimate this vector more precisely, we could substantially lower the uncertainty in the ICC estimates. Recently, Binsbergen, Brandt, and Koijen (2012) and Binsbergen, Hueskes, et al. (2013) made great progress in that respect. Using novel data sets from derivative markets, they are able to extract expectations about future dividends from market data. In this case, one does not have to rely on potentially biased analyst or regression-based forecasts, which reduces the uncertainty considerably. Unfortunately, this data is only available for short time periods and only for a few firms, but it shows that model uncertainty can be reduced. In summary, we can draw conclusions with more confidence if we are able to reduce the degrees of freedom in the estimation of a specific expected return proxy. I am confident that future research will make important contributions in this respect.

# A

# APPENDIX

## A.1 The Campbell and Shiller (1988) loglinearization

The log return $r_{t+1}$ on an investment such as a single stock or a stock market index from t to $t+1$ is defined as[114]

$$r_{t+1} = \log(1 + R_{t+1}) = \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right), \tag{71}$$

where $P_t$ is the market price at t and $D_{t+1}$ is the dividend paid at $t+1$. This equation can also be written as follows:[115]

$$\begin{aligned} r_{t+1} &= \log(P_{t+1} + D_{t+1}) - \log(P_t) \\ &= \log\left(P_{t+1}\left(1 + \frac{D_{t+1}}{P_{t+1}}\right)\right) - \log(P_t) \\ &= p_{t+1} - p_t + \log\left(1 + e^{\log\left(\frac{D_{t+1}}{P_{t+1}}\right)}\right) \\ &= p_{t+1} - p_t + \log\left(1 + e^{d_{t+1} - p_{t+1}}\right). \end{aligned} \tag{72}$$

The last term in this equation is a nonlinear function of the log dividend-price ratio, $f(d_{t+1} - p_{t+1})$, and can be further simplified by taking a first-order Taylor approximation around the mean of

114 For a more detailed textbook treatment of these derivations, see Campbell, Lo, and MacKinlay (1997, Chapter 7).

115 Throughout this chapter I use lowercase letters to denote log variables, e.g., $\log(D_t) \equiv d_t$.

the log dividend-price ratio, $(\overline{d-p})$. Also, I define $\rho = \frac{1}{1+e^{\overline{d-p}}}$ and $\kappa = -\log(\rho) - (1-\rho)\log\left(\frac{1}{\rho} - 1\right)$, which results in

$$\begin{aligned}
r_{t+1} &\approx p_{t+1} - p_t + \log\left(1 + e^{\overline{d-p}}\right) \\
&\quad + \frac{e^{\overline{d-p}}}{1+e^{\overline{d-p}}}(d_{t+1} - p_{t+1} - \overline{d-p}) \\
&\approx p_{t+1} - p_t + \log\left(\rho^{-1}\right) \\
&\quad + \left(1 - \frac{1}{1+e^{\overline{d-p}}}\right)(d_{t+1} - p_{t+1} - \overline{d-p}) \\
&\approx p_{t+1} - p_t - \log(\rho) + (1-\rho)(d_{t+1} - p_{t+1} - \overline{d-p}) \\
&\approx -\log(\rho) - (1-\rho)(\overline{d-p}) + \rho\, p_{t+1} + (1-\rho)d_{t+1} - p_t \\
&\approx \kappa + \rho\, p_{t+1} + (1-\rho)d_{t+1} - p_t.
\end{aligned} \tag{73}$$

The Taylor approximation replaces the sum of the log price and the log dividend with a weighted average of the two, whereby the former is weighted by $\rho$ and the latter by $(1-\rho)$. This approximation is mostly applied to stock market securities, such as a single stock or a stock market index, for which the dividend is normally much smaller than the price. This results in a $\rho$ close to 1.

Solving equation (73) for $p_t$ and iterating forward yields

$$\begin{aligned}
p_t &\approx \kappa + \rho\, p_{t+1} + (1-\rho)d_{t+1} - r_{t+1} \\
&\approx \kappa + \rho\left(\kappa + \rho\, p_{t+2} + (1-\rho)d_{t+2} - r_{t+2}\right) \\
&\quad + (1-\rho)d_{t+1} - r_{t+1} \\
&\approx \kappa + \rho\kappa + \rho^2 p_{t+2} + \rho(1-\rho)d_{t+2} + (1-\rho)d_{t+1} \\
&\quad - (\rho r_{t+2} + r_{t+1}) \\
&\approx \lim_{T\to\infty} \sum_{j=0}^{T} \rho^j \kappa + \sum_{j=0}^{T} \rho^j \left[(1-\rho)d_{t+1+j} - r_{t+j+1}\right] + \rho^T p_{t+T}.
\end{aligned}$$

Because $\rho$ is smaller than one, the first term can be rewritten as $\kappa/(1-\rho)$. Furthermore, rational bubbles are ruled out, i.e., it is assumed that the term $\lim_{T\to\infty} \rho^T p_{t+T} \to 0$. Also, taking expectation on both sides results in

$$p_t \approx \frac{\kappa}{1-\rho} + E_t\left[\sum_{j=0}^{\infty} \rho^j \left[(1-\rho)d_{t+j+1} - r_{t+j+1}\right]\right]. \tag{74}$$

Equation (74) shows that the current price is high when expected cash flows are high and low when expected returns are low. This equation is identical to equation (37) in the main text.

Finally, we can substitute equation (74) into equation (73):

$$
\begin{aligned}
r_{t+1} &\approx \kappa + \rho\, p_{t+1} + (1-\rho) d_{t+1} - p_t \\
&\approx \kappa - \frac{\kappa}{1-\rho} + (1-\rho) d_{t+1} \\
&\quad + \rho \left( \frac{\kappa}{1-\rho} + E_{t+1}\left[ \sum_{j=0}^{\infty} \rho^j \left[ (1-\rho) d_{t+j+2} - r_{t+j+2} \right] \right] \right) \\
&\quad - E_t \left[ \sum_{j=0}^{\infty} \rho^j \left[ (1-\rho) d_{t+j+1} - r_{t+j+1} \right] \right] \\
&\approx \kappa - \frac{\kappa}{1-\rho} + \frac{\kappa\rho}{1-\rho} + \rho^0 (1-\rho) d_{t+1+0} \\
&\quad + E_{t+1}\left[ \sum_{j=1}^{\infty} \rho^j \left[ (1-\rho) d_{t+j+1} - r_{t+j+1} \right] \right] \\
&\quad - E_t \left[ \sum_{j=0}^{\infty} \rho^j \left[ (1-\rho) d_{t+j+1} - r_{t+j+1} \right] \right] \\
&\approx E_t[r_{t+1}] \\
&\quad + E_{t+1}\left[ \sum_{j=0}^{\infty} \rho^j (1-\rho) d_{t+j+1} \right] - E_{t+1}\left[ \sum_{j=1}^{\infty} \rho^j r_{t+j+1} \right] \\
&\quad - E_t\left[ \sum_{j=0}^{\infty} \rho^j (1-\rho) d_{t+j+1} \right] + E_t\left[ \sum_{j=1}^{\infty} \rho^j r_{t+j+1} \right] \\
&\approx E_t[r_{t+1}] + d_t - d_t \\
&\quad + E_{t+1}\left[ \sum_{j=0}^{\infty} \rho^j d_{t+j+1} - \rho^j d_{t+j} \right] - E_t\left[ \sum_{j=0}^{\infty} \rho^j d_{t+j+1} - \rho^j d_{t+j} \right] \\
&\quad - E_{t+1}\left[ \sum_{j=1}^{\infty} \rho^j r_{t+j+1} \right] + E_t\left[ \sum_{j=1}^{\infty} \rho^j r_{t+j+1} \right] \\
&\approx E_t[r_{t+1}] \\
&\quad + \left( E_{t+1}\left[ \sum_{j=0}^{\infty} \rho^j \Delta d_{t+j+1} \right] - E_t\left[ \sum_{j=0}^{\infty} \rho^j \Delta d_{t+j+1} \right] \right) \\
&\quad - \left( E_{t+1}\left[ \sum_{j=1}^{\infty} \rho^j r_{t+j+1} \right] - E_t\left[ \sum_{j=1}^{\infty} \rho^j r_{t+j+1} \right] \right),
\end{aligned} \tag{75}
$$

where $\Delta d_{t+j+1} \equiv d_{t+j+1} - d_{t+j}$. This equation shows that unexpected stock returns, i.e., $r_{t+1} - E_t[r_{t+1}]$, can only come from two sources: changes in expectations of future dividends (cash flow news) or future returns (discount rate news). A decrease in expected future dividends is associated with a capital loss today, while a decrease in expected future returns is associated with a capital gain today.

## A.2 THEORETICAL RELATION BETWEEN EXPECTED RETURNS AND THE ICC IN THE MODEL OF PÁSTOR, SINHA, AND SWAMINATHAN (2008)

In this appendix, I map the simple framework of Pástor, Sinha, and Swaminathan (2008) into the predictive regression framework and discuss the impact of the difference between expected returns on the one hand and the ICC on the other hand on the slope coefficient in such predictive regressions.

The CS approximation, introduced in Appendix A.1, of the present value formula expresses the log price $p_t \equiv \log(P_t)$ as[116]

$$p_t = \frac{\kappa}{1-\rho} + (1-\rho)\sum_{j=0}^{\infty} \rho^j E_t(d_{t+j+1}) - \sum_{j=0}^{\infty} \rho^j E_t(r_{t+1+j}), \tag{76}$$

where $\kappa = -\log(\rho) - (1-\rho)\log(1/\rho - 1)$. As Pástor, Sinha, and Swaminathan (2008) point out, in this framework the log ICC is defined as the value of $r_t^e$ that solves

$$p_t = \frac{\kappa}{1-\rho} + (1-\rho)\sum_{j=0}^{\infty} \rho^j E_t(d_{t+j+1}) - r_t^e \sum_{j=0}^{\infty} \rho^j. \tag{77}$$

It directly follows from equation (76) and (77) that the ICC is just a scaled version of the vector of all future discounted expected returns:[117]

$$r_t^e = (1-\rho)\sum_{j=0}^{\infty} \rho^j E_t(r_{t+1+j}). \tag{78}$$

This also implies that the log ICC is not identical to the expected return next period, $\mu_t \equiv E_t[r_{t+1}]$, in cases in which expected returns

116 As already explained in footnote 37, I ignore the approximation error in the following and use an equal sign.

117 This holds exactly in the CS framework, but Chen, Da, and Zhao (2013) provide evidence that the approximation error is close to zero even if the ICC is computed from the classical present value formula in equation (2).

are time-varying. To provide more insight into the relation between the ICC and expected returns next period, I assume in accordance with Pástor, Sinha, and Swaminathan (2008) that the conditional expected return, $\mu_t$, follows a stationary AR(1) process:

$$\mu_{t+1} = \lambda_\mu + \tau_\mu \mu_t + \nu_{t+1}, \quad 0 < \tau_\mu < 1, \quad \nu_{t+1} \sim N(0, \sigma_\nu^2). \quad (79)$$

Under this assumption, they show that

$$E_t\left[\sum_{j=0}^{\infty} \rho^j r_{t+j+1}\right] = \frac{\lambda_\mu}{(1-\tau_\mu)(1-\rho)} + \left(\mu_t - \frac{\lambda_\mu}{1-\tau_\mu}\right)\frac{1}{1-\rho\tau_\mu}. \quad (80)$$

Replacing the LHS of equation (80) with the relevant part of the RHS in equation (78) gives a relation between the ICC and the conditional expected return today, $\mu_t$:

$$r_t^e = \frac{\lambda_\mu}{1-\tau_\mu} + \left(\mu_t - \frac{\lambda_\mu}{1-\tau_\mu}\right)\frac{1-\rho}{1-\rho\tau_\mu}. \quad (81)$$

Equation (81) shows that the ICC would be perfectly correlated with the conditional expected return today given that the AR(1) assumption is appropriate and that the vector of expected dividends is measured without error. This let Pástor, Sinha, and Swaminathan (2008) to conclude that the ICC, at least theoretically, is a good proxy for conditional expected returns.

However, it is obvious from equation (81) that a regression of realized returns on the true log ICC will not result in a slope of one anymore, which is the criterion for a perfect proxy of the expected return period next period (see Chapter 3). Instead, the regression coefficient will be

$$\begin{aligned}\beta_{r_t^e} &= \frac{Cov\left(r_t^e, r_{t+1}\right)}{Var\left(r_t^e\right)} \\ &= \frac{Cov\left(\frac{\lambda_\mu}{1-\tau_\mu} + \left(\mu_t - \frac{\lambda_\mu}{1-\tau_\mu}\right)\frac{1-\rho}{1-\rho\tau_\mu}, \mu_t + u_{t+1}\right)}{Var\left(\frac{\lambda_\mu}{1-\tau_\mu} + \left(\mu_t - \frac{\lambda_\mu}{1-\tau_\mu}\right)\frac{1-\rho}{1-\rho\tau_\mu}\right)} \\ &= \frac{1-\rho\tau_\mu}{1-\rho}. \end{aligned} \quad (82)$$

Equation (82) shows that $\beta_{r_t^e}$ will be larger than one as long as $\tau_\mu$ is below one. This finding is consistent with the study of Li, Ng, and Swaminathan (2013) who report betas between 1.5 and 2.5 for predictive regressions of realized returns on aggregated ICCs for a US sample ranging from 1977 to 2011.

To understand the difference between the log ICC and the log expected return next period in this framework better, it is instructive to solve equation (79) forward. Further ahead expected returns for period $t+1+j$ can then be expressed as

$$E_t[r_{t+1+j}] = \lambda_\mu \frac{1-\tau_\mu^j}{1-\tau_\mu} + \tau_\mu^j \mu_t, \tag{83}$$

where $E[\mu] = \lambda_\mu/(1-\tau_\mu)$ is the unconditional mean of expected returns. Thus, further ahead expected returns are a weighted average of the expected return for next period, $\mu_t$, and the unconditional or long-run mean, whereby the weight for the latter converges to 1 for further ahead periods. In this simple AR(1) framework, the current state gets less and less important, i.e., it is transitory, and for periods that lie far ahead it does not have any relevance anymore. The ICC is a geometric average over all future expected returns and is therefore less volatile than the expected return next period because it always takes the expectation of further ahead returns close to the long-run mean into account as well. As a consequence, a regression of subsequent realized returns on the ICC instead of $\mu_t$ leads to a slope coefficient larger than 1.

As a small numerical example of this effect, consider the following parameter values:

- $\rho = 0.95$.
- $\tau_\mu = 0.8$.
- $\lambda_\mu = 2\%$ per year, which implies an unconditional mean return of 10% per year.
- Two scenarios:
  1. $\mu_{t,low} = 5\%$ per year.
  2. $\mu_{t,high} = 15\%$ per year.

Figure 10 shows the expected return over time for both scenarios as well as the corresponding ICC. It illustrates that investors expect returns in the far ahead future to converge to the long-run mean, irrespective of the current level of expected returns. Because the ICC contains information about expected returns for *all* periods, it is less

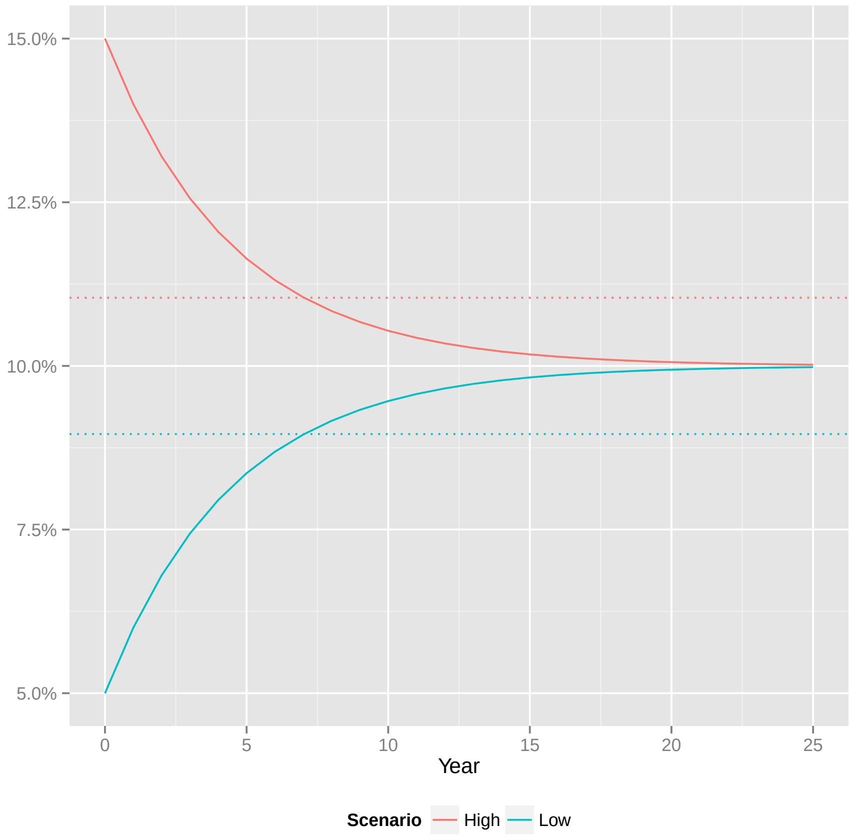


Figure 10: **Expected returns and the ICC for different periods and scenarios in the Pástor, Sinha, and Swaminathan (2008) framework.** This plot shows the expected returns for the first 25 years as well as the ICC for two scenarios. In scenario "High", the current expected return $\mu_t$ is 15% per year, in scenario "Low" 5% per year. Expected returns are computed from equation (83). The ICCs are computed from equation (81) and represented as horizontal, dotted lines. The other parameters are set to $\rho = 0.95$, $\tau_\mu = 0.8$, and $\lambda_\mu = 2\%$ per year.

volatile than expected returns next period. With this specific calibration, the regression coefficient of subsequent realized returns on the true, but unobservable ICC would be 4.8. In empirical applications this coefficient is likely to be lower due to additional measurement error in estimated ICCs.

## A.3 UPPER BOUND OF $R^2$ IN UNIVARIATE ERRORS-IN-VARIABLES PROBLEM

Suppose a variable $y_{t+1}$ is driven only by one single factor, $x_t$, plus additional random noise $\epsilon_{t+1}$. This error term is independent with any variable from time $t$. That is, we know that $Cov(x_t, \epsilon_{t+1}) = 0$:

$$y_{t+1} = x_t + \epsilon_{t+1}. \tag{84}$$

Because $x_t$ is a latent variable, we only observe proxies of it that are both scaled by a factor $s$ and have additional measurement error $w_t$. $w_t$ can be correlated with $x_t$, but is uncorrelated with $\epsilon_{t+1}$. A proxy $\widehat{x}_t$ is then given as

$$\widehat{x}_t = s x_t + w_t. \tag{85}$$

Here, I show that the $R^2$ in the case in which $Var(w_t) = 0$ is the upper bound for any proxy.

The $R^2$ in the univariate regression framework with intercept is just the squared correlation between $\widehat{x}_t$ and $y_{t+1}$:

$$\begin{aligned} R^2 &= \left( \frac{Cov(y_{t+1}, \widehat{x}_t)}{\sigma_y \sigma_{\widehat{x}}} \right)^2 \\ &= \frac{Cov(x_t + \epsilon_{t+1}, s x_t + w_t)^2}{Var(y_{t+1}) Var(\widehat{x}_t)} \\ &= \frac{(s Var(x_t) + Cov(x_t, w_t))^2}{Var(y_{t+1}) Var(s x_t + w_t)} \end{aligned} \tag{86}$$

Equation (86) simplifies to $R^2 = Var(x_t)/Var(y_{t+1})$ if $Var(w_t)$ equals 0. Next, I want to show that this is an upper bound.

To do so, I split up the measurement error in two parts, one which is correlated with $x_t$ and one which is uncorrelated. This can be done by regressing $w_t$ on $x_t$:

$$w_t = \tau_0 + \tau_1 x_t + v_t. \tag{87}$$

By construction, $Cov(x_t, v_t) = 0$. Plugging equation (87) into (86), we get:

$$
\begin{aligned}
R^2 &= \frac{(sVar(x_t) + Cov(x_t, w_t))^2}{Var(y_{t+1})Var(sx_t + w_t)} \\
&= \frac{(sVar(x_t) + Cov(x_t, \tau_0 + \tau_1 x_t + v_t))^2}{Var(y_{t+1})Var(sx_t + \tau_0 + \tau_1 x_t + v_t)} \\
&= \frac{(s + \tau_1)^2 Var(x_t)^2}{Var(y_{t+1})((s + \tau_1)^2 Var(x_t) + Var(v_t))} \qquad (88)
\end{aligned}
$$

Equation (88) equals the $R^2$ of the true, but latent explanatory variable $x_t$ if the variance of $v_t$ is zero. In all other cases, however, the denominator is larger due to the additional term $Var(v_t)$. This, in turn, decreases $R^2$, which shows that any proxy measured with additional error that is at least partly orthogonal to expected returns results in lower $R^2$.

## A.4 A SHORT INTRODUCTION TO BAYESIAN STATISTICS

This section gives a short introduction into Bayesian statistics. Using the example of a linear regression, it shows how the statistical inference works in a Bayesian setting. Furthermore, some equations used in the main part are derived.

For a more detailed treatment of Bayesian statistics in general and regression models in particular, the interested reader is referred to textbooks such as Poirier (1995), Koop (2003), Lunn et al. (2012), and Gelman et al. (2013), from which this section heavily draws on.

### A.4.1 *Theory*

All of Bayesian statistics is founded on one simple principle, Bayes' theorem. This theorem is usually expressed in terms of probabilities for observable events $A$ and $B$. It states that

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}. \qquad (89)$$

In words: the *conditional* probability of $A$ given $B$, also referred to as the *posterior* probability of $A$ after taking into account the value of $B$, is given as the *conditional* probability of $B$ given $A$ multiplied by the *marginal* probability of $A$ and divided by the *marginal* probability of $B$. $p(A)$ is also often referred to as the *prior* probability of $A$, where "prior" indicates "before taking account of the information in B" (cf. Lunn et al. 2012). Bayes' theorem is applicable to any probability distribution $p(\cdot)$.

In frequentist statistics it is a commonly made assumption that parameters of a model are unknown, but fixed quantities. Only the data are assumed to be random draws from probability distributions. In contrast, both data and parameters have a probability distribution in Bayesian statistics and so Bayes' theorem can also be applied to estimate a vector of parameters $\theta$.[118]

Let $p(\cdot)$ now denote a probability density rather than a simple probability of an event. Further, let $y$ denote the data. Then, equation (89) can be used to make inferences about $\theta$:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{90}$$

where $p(\theta)$ is the prior distribution of $\theta$, $p(y|\theta)$ is the distribution of the data $y$, conditional on the parameters of the model $\theta$, and $p(\theta|y)$ is the posterior distribution of $\theta$, conditional on the data $y$. $p(y|\theta)$ is taken here as a function of $\theta$, not of $y$. This function is commonly called the *likelihood function*.

$p(y)$ is obtained by integrating out $\theta$ of the likelihood function:

$$p(y) = \int p(y|\theta)p(\theta)d\theta. \tag{91}$$

$p(y)$ is also called the *marginal* distribution of $y$ because $\theta$ is swept out. Thus it is independent of $\theta$ and can be considered as a constant in equation (90), which normalizes the posterior distribution $p(\theta|y)$ so that it integrates to 1 and is therefore a valid probability distribution. Most relevant properties of the posterior can be obtained without this normalization and since the computation of $p(y)$ can be complicated,

118 Leamer (1978, Chapter 2) discusses the underlying differences between frequentist and Bayesian inference in detail. In the introduction of this chapter on page 21, he summarizes the main points as follows: "An inference is a logical conclusion drawn from a set of facts. Statistical inference is concerned with drawing conclusions about unobservables $\boldsymbol{\theta}$ from a set of facts, including observed data $\mathbf{x}$ and a conditional probability distribution $f(\mathbf{x}|\boldsymbol{\theta})$, that indicates the probability of various values of $\mathbf{x}$ given various values of $\boldsymbol{\theta}$. Bayesian inference is distinguished from classical inference by its inclusion of a "prior" probability function $f(\boldsymbol{\theta})$ in the set of facts. To a Bayesian there is no sound logical reason why the distribution $f(\mathbf{x}|\boldsymbol{\theta})$ should be regarded to be more of a "fact" than the distribution $f(\boldsymbol{\theta})$. A classicist, however, argues that the distribution $f(\mathbf{x}|\boldsymbol{\theta})$ is an objectively verifiable feature of the world, whereas any distribution $f(\boldsymbol{\theta})$ is purely a figment of someone's imagination."

it is often ignored. In this case, the posterior is just proportional to the product of the likelihood and the prior:

$$p(\theta|y) \propto p(y|\theta)p(\theta). \tag{92}$$

In short, in Bayesian inference one updates his prior beliefs into posterior beliefs conditional on the observed data (cf. Koop, Poirier, and Tobias 2007). To compute the posterior $p(\theta|y)$, a researcher has to set up a full probability model for all relevant parameters and data in a problem. Because the posterior can only be derived analytically for a few simple cases, most of Bayesian statistics relies on computational tools to sample from the posterior distribution (cf., e.g., Koop 2003 or Lunn et al. 2012). However, the linear regression model with additional assumptions about the prior and the likelihood is such a simple case. Because the BMA approach introduced in Chapter 4 is based on such a linear regression model, I discuss it here in detail.

This model posits a linear relationship between the response or outcome variable $y_i$ and a $1 \times k$ vector of explanatory variables, $x_i$, where $i = 1, \ldots, N$ indexes the relevant observation unit. In matrix notation, $y$ denotes the vector of outcomes for the N subjects and X denotes the $N \times k$ matrix of predictors:

$$y = X\beta + \epsilon, \tag{93}$$

where $\beta = \beta_1, \ldots, \beta_k$ is a vector of length $k$ that holds the parameters of interest. For many applications, the variable $x_{i,1}$ is set to 1.

I assume here that the classical assumptions of a linear model hold.[119] Since one of these assumptions is that the explanatory variable $x_t$ is not random, the likelihood function is defined by the probability distribution of the error terms. A commonly made assumption is that $\epsilon_i$ is normally distributed with mean zero and constant variance across units $i$:

$$\epsilon \sim N(0, \sigma^2 I_N), \tag{94}$$

where $I_N$ is the $N \times N$ unit or identity matrix. Often, one works with the error precision parameter instead of the error variance, that is, $h \equiv \sigma^{-2}$. This model is also called the *normal linear model* because $y$, given X, is normally distributed.

119 For a summary of these assumptions, refer to, e.g., Kennedy (2008, Chapter 3).

The linear regression model is now completely specified. The parameter vector is $\theta = (\beta, h)$, and the likelihood function, subject to $\theta$ and equation (94), is given as

$$\begin{aligned} L(\theta) &= p(y|\theta) \\ &= \prod_{i=1}^{N} p(\epsilon_i) \\ &= \prod_{i=1}^{N} \frac{h^{1/2}}{\sqrt{2\pi}} \exp\left(-\frac{h}{2}\left(y_i - \sum_{j=1}^{k} \beta_j x_{i,k}\right)^2\right) \\ &= \frac{h^{N/2}}{(2\pi)^{N/2}} \exp\left(-\frac{h}{2}(y - X\beta)'(y - X\beta)\right). \end{aligned} \tag{95}$$

Finally, we have to specify the prior distributions of $\theta$, which is one of the main objections of frequentists against Bayesian analysis because it introduces subjective beliefs into the analysis (cf. Kass and Wasserman 1996).

Often, conjugate priors for the likelihood function are used, which ensure that the posterior distribution is in the same family as the prior distribution. A natural conjugate prior has the additional property that it follows the same functional form as the likelihood function (cf. Koop 2003). In addition to computational convenience, Koop, Poirier, and Tobias (2007, p. 19) describe another advantage of such natural conjugate priors:

> Natural conjugate priors have the desirable feature that prior information can be viewed as "fictitious sample information" in that it is combined with the sample in exactly the same way that additional sample information would be combined. The only difference is that the prior information is "observed" in the mind of the researcher, not in the real world.

In the regression model used here, the natural conjugate prior is the normal-gamma distribution, which implies by the definition of conjugacy that the posterior distribution follows a normal-gamma distribution as well.[120] Formally, it is assumed that the precision parameter $h$ follows a gamma distribution ($h \sim G(\underline{h}, \underline{\nu})$) and that the conditional distribution of $\beta$, given $h$, is a multivariate normal distribution ($\beta \sim N(\underline{\beta}, h^{-1}\underline{V})$). That is, the prior for $\beta$ and $h$ is $NG(\underline{\beta}, \underline{V}, \underline{h}, \underline{\nu})$.[121]

120 Koop, Poirier, and Tobias (2007) provide a derivation in Exercise 10.1.
121 I denote prior parameters with underlines and posterior parameters with overlines.

It can be shown that in this case (cf., e.g., Poirier 1995) the marginal prior distribution of $\underline{\beta}$, i.e., the distribution for which $\sigma^2$ is integrated out, follows a multivariate t-distribution $t(\underline{\beta}, \underline{s}^2\underline{V}, \underline{\nu})$ with mean and variance[122]

$$E(\underline{\beta}) = \underline{\beta}, \quad \text{if } \underline{\nu} > 1, \tag{96}$$

and

$$Var(\underline{\beta}) = \frac{\underline{\nu}}{\underline{\nu} - 2}\underline{s}^2\underline{V}, \quad \text{if } \underline{\nu} > 2. \tag{97}$$

The parameters of the prior, which are often referred to as hyper-parameters (cf. Gelman et al. 2013), have to be specified by the researcher *a priori*.

Since this specification is natural conjugate, the posterior distribution of $\beta$ and $h$ is the normal-gamma distribution $NG(\overline{\beta}, \overline{V}, \overline{h}, \overline{\nu})$, where

$$\overline{\beta} = \overline{V}(\underline{V}^{-1}\underline{\beta} + X'X\beta_{OLS}), \tag{98}$$

$$\overline{V} = (\underline{V}^{-1} + X'X)^{-1}, \tag{99}$$

$$\overline{\nu} = \underline{\nu} + N, \tag{100}$$

$$\overline{\nu s}^2 = \underline{\nu s}^2 + SSE + (\beta_{OLS} - \underline{\beta})'[X'X\overline{V}\underline{V}^{-1}](\beta_{OLS} - \underline{\beta}). \tag{101}$$

$\beta_{OLS}$ is the OLS estimator of $\beta$ and $SSE$ is the sum of squared error from an OLS regression.

Equation (98) to (101) illustrate nicely the view of the prior as a "fictitious" sample.[123] In this case, $\underline{\nu}$ can be interpreted as the sample size of this "fictitious" sample and the posterior mean of $\beta$, $\overline{\beta}$, is just a weighted average between the prior mean, $\underline{\beta}$, and the evidence from the data, $\beta_{OLS}$. The weights are proportional to $X'X$ and $\underline{V}^{-1}$, respectively, where the latter reflects the confidence in the prior.[124] The more confident a researcher is in his prior, the more $\overline{\beta}$ is biased towards it. Furthermore, equation (101) shows that the posterior sum of squares, $\overline{\nu s}^2$, combines the prior sum of squares, $\underline{\nu s}^2$, the sample

122 Throughout this section, I use the same parameterization for distributions as Koop (2003), which can be found in his appendix.

123 Leamer (1978) provides a derivation that shows that the posteriors are nothing else than the results obtained from pooling two samples, one which is the data and one which comes from the fictitious sample of the prior assumptions.

124 The variance-covariance matrix of $\beta_{OLS}$ is given as $var(\beta_{OLS}) = \sigma_\epsilon^2(X'X)^{-1}$. $\underline{V}$ has the same interpretation as $(X'X)^{-1}$ for the fictitious prior sample. That is, the smaller $\underline{V}$, the smaller is the uncertainty in the prior estimate and the higher is the precision. A higher precision translates into a higher weight.

sum of squares, $SSE$, and a term that measures the conflict between prior and data information. Due to the assumptions made about the prior and the likelihood, the marginal likelihood and the marginal posterior density of $\beta$ follow multivariate t-distributions as well (cf. Koop 2003). Therefore, sampling from these distributions is straightforward.

### A.4.2 *Numerical example*

The above discussion is best understood with a small numerical example. Suppose that the outcomes $y$ are generated by the data generating process with the following underlying parameters:[125]

- $\beta = 2$, i.e., we have a univariate linear model with no intercept.
- $h = 1$.

A researcher is interested in the slope coefficient $\beta$, but does only observe N draws from the data generating process. I simulate 50 data points by drawing both the explanatory variable $x_i$ and the errors $\epsilon_i$ (since $h = 1$) from a standard normal distribution and computing $y$ from equation (93).

The first prior specification is as follows:

- $\underline{\beta} = 1.5$.
- $\underline{V} = 0.25$
- $\underline{\nu} = 10$.
- $\underline{h} = 1$.

This information allows us to sample from the marginal distributions of the prior, the likelihood, and the posterior, which are shown in Figure 11. The prior for $\beta$ is centered around 1.5, but has a large variance. A researcher who would have specified such a prior would not be very certain about his prior information. The likelihood is centered around 1.94, which is the OLS estimate for this specific random sample. In contrast to the fictitious prior sample, the evidence from the data allows a much more precise inference about $\beta$. This can be seen from the much smaller dispersion of the likelihood in Figure 11. The posterior combines the evidence of the prior and the likelihood.

125 This example is taken from Koop (2003, Chapter 2).

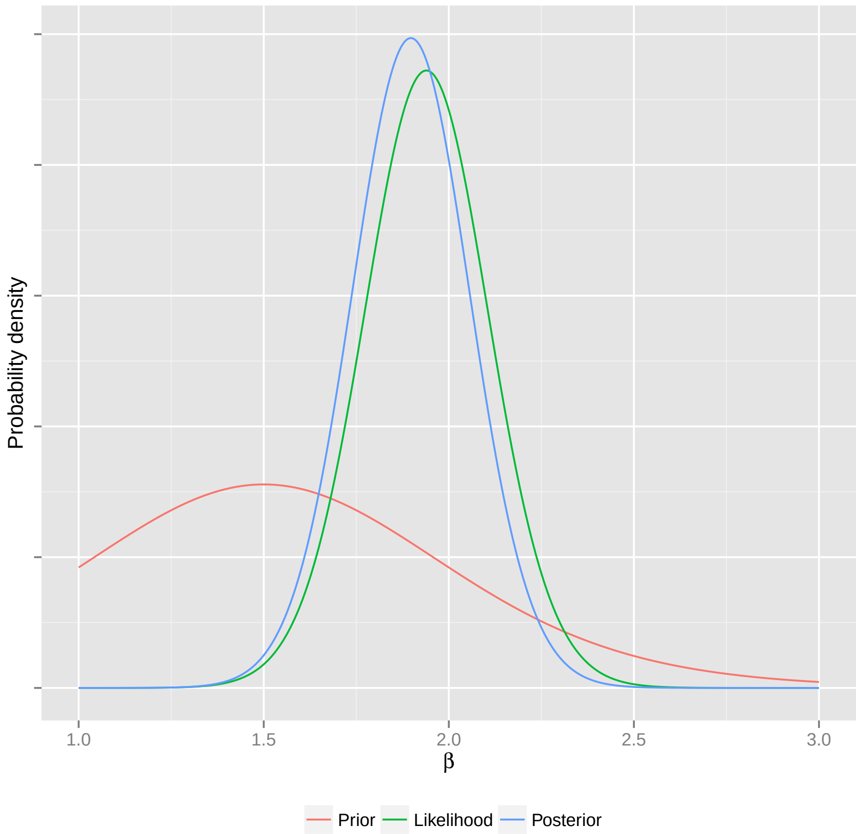


Figure 11: **Marginal prior, likelihood, and posterior for $\beta$ (case 1).** The prior specification for this plot is given by $\beta, h \sim NG(\underline{\beta}, \underline{V}, \underline{h}, \underline{\nu})$ with $\underline{\beta} = 1.5$, $\underline{V} = 0.25$, $\underline{\nu} = 10$ and $\underline{h} = 1$.

Because the evidence in the data is given more weight here, the posterior is similar to the likelihood. Koop (2003) shows that if we would further decrease the importance of the prior information, the likelihood and the posterior would eventually converge. In this case, because the prior does not play any role anymore, it is referred to as a *noninformative prior*.

It is easy to go the other way and make the prior more informative. Maybe a researcher has done a similar analysis many times before and is fairly certain that his prior of a coefficient of 1.5 is a good estimate. In this case, he can increase the fictitious sample size and decrease the uncertainty in his prior. For example, he could increase $\underline{\nu}$ by a factor of 10 and decrease $\underline{V}$ accordingly by the same factor:

- $\underline{\beta} = 1.5$.
- $\underline{V} = 0.025$
- $\underline{\nu} = 100$.
- $\underline{h} = 1$.

Figure 12 shows the densities for the same data, but updated prior beliefs. Because the information in the prior and in the data is now roughly equally important to the researcher, the posterior is in the middle of the two.

### A.4.3 *Discussion of priors*

A main issue of the prior specification in the linear model with normal-gamma distributed priors is the specification of $\underline{V}$. One easy to implement solution to this problem is the use of the *g-prior*, as introduced in Zellner (1986). In this case, $\underline{V}$ is set to $\phi(X'X)^{-1}$, where $\phi$ is a shrinkage parameter that controls the informativeness of the prior. Note that the researcher only has to set $\phi$, a single parameter.

With the g-prior specification, $\overline{V}$ simplifies to

$$\begin{aligned}
\overline{V} &= (\underline{V}^{-1} + X'X)^{-1} \\
&= ((\phi(X'X)^{-1})^{-1} + X'X)^{-1} \\
&= ((\phi)^{-1}X'X + X'X)^{-1} \\
&= (\phi)(X'X + \phi X'X)^{-1} \\
&= \frac{\phi}{1+\phi}(X'X)^{-1}.
\end{aligned} \tag{102}$$

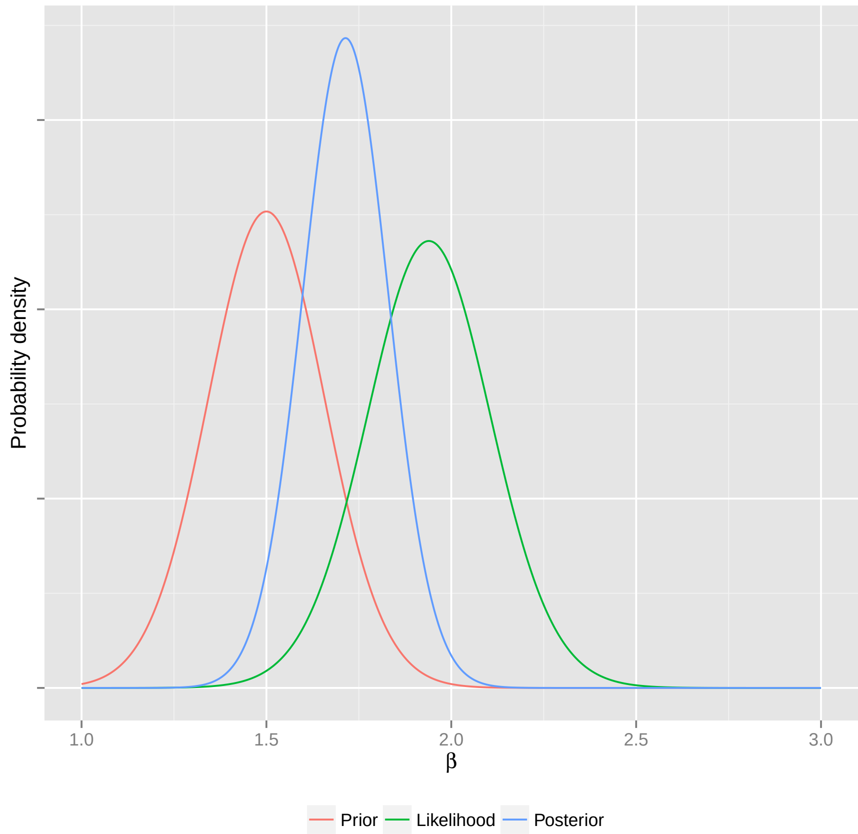


Figure 12: **Marginal prior, likelihood, and posterior for $\beta$ (case 2).** The prior specification for this plot is given by $\beta, h \sim NG(\underline{\beta}, \underline{V}, \underline{h}, \underline{\nu})$ with $\underline{\beta} = 1.5$, $\underline{V} = 0.025$, $\underline{\nu} = 100$ and $\underline{h} = 1$.

Another commonly made simplification is the assumption that $\underline{\beta}$ is set to zero. This assumption mimics a researcher that believes in no relation between the explanatory variables and the outcome a priori. Furthermore, since the subjectivism in the priors is often an objection against Bayesian statistics, it is tried to reduce its impact as much as possible. In the case of the g-prior specification, it is desirable to control the informativeness of the prior only via $\phi$, so $\underline{\nu}$ is often set to 0.[126] $\overline{\nu s}^2$ is then given as

$$\begin{aligned}
\overline{\nu s}^2 &= \underline{\nu s}^2 + \mathrm{SSE} + (\beta_{OLS} - \underline{\beta})'[(X'X)\overline{V}\underline{V}^{-1}](\beta_{OLS} - \underline{\beta}) \\
&= \underline{\nu s}^2 + \mathrm{SSE} \\
&\quad + (\beta_{OLS} - 0)'[(X'X)\frac{\phi}{1+\phi}(X'X)^{-1}\frac{1}{\phi}(X'X)](\beta_{OLS} - 0) \\
&= \underline{\nu s}^2 + \mathrm{SSE} + \frac{1}{1+\phi}\beta_{OLS}'X'X\beta_{OLS} \\
&= 0 + \mathrm{SSE} + \frac{1}{1+\phi}\beta_{OLS}'X'X\beta_{OLS} \\
&= (y - X\beta_{OLS})'(y - X\beta_{OLS}) + \frac{1}{1+\phi}\beta_{OLS}'X'X\beta_{OLS} \\
&= y'y - 2y'X\beta_{OLS} + \beta_{OLS}X'X\beta_{OLS} + \frac{1}{1+\phi}\beta_{OLS}'X'X\beta_{OLS} \\
&= y'y - 2(X\beta_{OLS} + \epsilon)'X\beta_{OLS} + \beta_{OLS}'X'X\beta_{OLS} \\
&\quad + \frac{1}{1+\phi}\beta_{OLS}'X'X\beta_{OLS} \\
&= y'y - \beta_{OLS}'X'X\beta_{OLS} + \frac{1}{1+\phi}\beta_{OLS}'X'X\beta_{OLS} \\
&= y'y + \left(\frac{1}{1+\phi} - 1\right)\beta_{OLS}'X'X\beta_{OLS} \\
&= y'y - \left(\frac{\phi}{1+\phi}\right)\beta_{OLS}'X'X\beta_{OLS} \\
&= y'y - \left(\frac{\phi}{1+\phi}\right)\left((X'X)^{-1}X'y\right)'X'X(X'X)^{-1}X'y \\
&= y'y - \left(\frac{\phi}{1+\phi}\right)(X'y)'\left((X'X)^{-1}\right)'X'y \\
&= y'y - \left(\frac{\phi}{1+\phi}\right)y'X(X'X)^{-1}X'y. \qquad (103)
\end{aligned}$$

Finally, plugging equation (102) into equation (98), it is easy to show that

$$\overline{\beta} = \frac{1}{1+\phi}\underline{\beta} + \frac{\phi}{1+\phi}\beta_{OLS}. \qquad (104)$$

126 As Wright (2008) highlights, this assumption also implies that the prior on $\sigma^2$ simplifies to a uniform distribution. Such a distribution is called *improper* because it does not integrate to 1. Fortunately, this is not an issue here because the posterior distribution is still proper in this case (for a discussion see Gelman et al. 2013). Improper priors give equal weights to all possible values and are therefore a nice way to be as non-informative as possible in the priors. This is often of great importance to researchers because it provides a nice link to classical/frequentist approaches. See, for instance, Leamer (1978, p. 110): "The critical defect of a Bayesian analysis of data is that prior distributions are both personally difficult to specify and also subject to variation among interested people. As a consequence, a Bayesian analysis based on any particular prior is of little interest."

Because $\underline{\beta}$ is typically set to zero, this further simplifies to $\overline{\beta} = \phi/(1+\phi)\beta_{OLS}$. Equation (104) shows that the posterior mean converges to $\beta_{OLS}$ if $\phi$ becomes large. In this case, the information in the prior is ignored and the classical OLS results are reproduced.

### A.4.4 *Derivation of posterior model weights*

With the above derivations, it is easy to derive equation (61) from the main text. To do so, I assume now that the independent variables in equation (93) are model specific:

$$M_k : y = X_k \beta_k + \epsilon_k. \tag{105}$$

Otherwise, the assumptions remain the same. The posterior model weights are proportional to the product of the model likelihood and the model prior:

$$p(M_k|D) \propto p(D|M_k)p(M_k), \tag{106}$$

where D is the data. In general, the marginal likelihood, conditional on model k, is given as[127]

$$p(D|M_k) = c_k \left[\frac{|\overline{V}_k|}{|\underline{V}_k|}\right]^{1/2} [\overline{\nu}_k \overline{s}_k^2]^{-\overline{\nu}_k/2}, \tag{107}$$

where

$$c_k = \frac{\Gamma(\overline{\nu}_k/2)[\underline{\nu}_k \underline{s}_k^2]^{\underline{\nu}_k/2}}{\Gamma(\underline{\nu}_k/2)\pi^{N/2}}. \tag{108}$$

$\Gamma(\cdot)$ denotes the gamma function. All that is left to do is to compute $p(D|M_k)$ for the special case with a g-prior specification. Set-

127 See, e.g., Poirier (1995) or Koop (2003).

ting $\underline{V} = \phi(X'X)^{-1}$ and $\overline{V} = \phi/(1+\phi)(X'X)^{-1}$ (see equation 102) in equation (107), we obtain[128]

$$\begin{aligned}
\left[\frac{|\overline{V}_k|}{|\underline{V}_k|}\right]^{1/2} [\overline{\nu}_k \overline{s}_k^2]^{-\overline{\nu}_k/2} &= \left[\frac{|\frac{\phi}{1+\phi}(X'X)^{-1}|}{|\phi(X'X)^{-1}|}\right]^{1/2} [\overline{\nu}_k \overline{s}_k^2]^{-\overline{\nu}_k/2} \\
&= \left[\frac{\left(\frac{\phi}{1+\phi}\right)^K |(X'X)^{-1}|}{\phi^K |(X'X)^{-1}|}\right]^{1/2} [\overline{\nu}_k \overline{s}_k^2]^{-\overline{\nu}_k/2} \\
&= \left[\frac{1}{1+\phi}\right]^{K/2} [\overline{\nu}_k \overline{s}_k^2]^{-\overline{\nu}_k/2} \qquad (109)
\end{aligned}$$

Plugging equation (109) into (107) and ignoring terms that are constant across models, $p(D|M_k)$ simplifies to

$$p(D|M_k) \propto c_k (\overline{\nu s}_k^2)^{-\overline{\nu}_k/2}. \qquad (110)$$

Finally, if we set $\underline{\nu}$ to zero, as done before, $c_k$ is equal across models as well and can be ignored. Additionally, we can replace $\overline{\nu s}$ with the term given in equation (103) to obtain

$$p(D|M_k) \propto \left(y'y - \left(\frac{\phi}{1+\phi}\right) y'X_k(X_k'X_k)^{-1}X_k'y\right)^{-N/2}. \qquad (111)$$

Equation (111) is identical to equation (61) of the main text, only the notation differs. Note that the prior model weights, $p(M_k)$, are set to $1/k$ in the main text, which implies that $p(M_k|D)$ is proportional to $p(D|M_k)$. Because the prior model weights are equal across models, they cancel out in the computation and the posterior model weights are only driven by the evidence in the data for each model.

## A.5 SIMULATION RESULTS OF BMA PERFORMANCE

The aim of this simulation is to evaluate the performance of the model averaging approach in comparison to other applied or recommended approaches of choosing among a set of alternative expected return proxies in the outlined research question of equation (47). Since we do not know the exact properties of latent expected returns, I decide to simulate four different specifications that cover a wide range of reasonable assumptions.

128 For this derivation, the following calculation rule for a matrix $A$ with rank $j$ is needed: $|cA| = c^j|A|$.

We are interested in the following regression:

$$\mu_t = \gamma_0 + \gamma_1 x_t + \epsilon_t. \tag{112}$$

More precisely, we want to determine if $\gamma_1$ deviates from zero, but we face the problem that $\mu_t$ is unobservable. Instead, we can only rely on $k = 8$ different proxies available. I consider the following approaches to estimate $\gamma_1$:

- **Best:** This approach runs the regression specified in equation (112) for each available proxy and takes the results of the proxy that yields the most significant results. In this simulation, it means that I take the slope coefficient with the highest positive value, assuming that a researcher wants to establish a positive relation between true expected returns and $x_t$.

- **Ave:** This approach averages across all proxies *before the regression is run* and reports the statistics for the averaged proxy. This approach was proposed by Hail and Leuz (2009).[129]

- **Highest $R^2$:** In a first step, this approach selects the proxy that explains subsequent realized returns best. Afterwards, regression (112) is only run for this proxy. This is a classical model selection approach in which the model selection and the statistical inference step are separated. This approach is based on the work of Easton and Monahan (2005) and Lee, So, and Wang (2011).

- **BMA:** First, weights for each proxy are computed based on how well this proxy is able to explain subsequent realized returns. The weights are computed with the equations provided in Chapter 4. I set $\phi = \infty$ and give each model equal weight a priori. Second, the posterior distribution of the quantity of interest is computed via equation (51). Since every estimated slope coefficient follows a t-distribution, the marginal density across all models is a mixture t-distribution.

- **Real:** This approach uses realized returns as a proxy for expected returns.

I want to evaluate the performance of these approaches for different scenarios. In particular, I want to find out how each of them performs

129 They also report their results for all of their four methods separately.

in cases in which the true expected return process is or is not within the set of proxies under consideration and in cases in which the measurement error is or is not correlated with the variable of interest $x_t$. Another requirement I want to fulfill is that $\mu_t$ follows an AR(1) process since this is a commonly made assumption in the literature.[130]

Therefore, I simulate $x_t$ as a monthly AR(1) process:[131]

$$x_{t+1} = \tau_x x_t + \varepsilon_{x,t+1}, \quad \varepsilon_{x,t+1} \sim N(0, \sigma^2_{\varepsilon_x}). \tag{113}$$

For the case in which $x_t$ and $\mu_t$ are related, I compute $\mu_t$ – based on equation (112) – as

$$\mu_t = 1 \times x_t + \varepsilon_{\mu,t}, \quad \varepsilon_{\mu,t} \sim N(0, \sigma^2_{\varepsilon_\mu}). \tag{114}$$

In this case, an econometrician wants to obtain $\gamma_1 = 1$.

For the case in which there is no relation between $x_t$ and $\mu_t$, i.e., $\gamma_1 = 0$, I simulate expected returns as

$$\mu_{t+1} = \tau_\mu \mu_t + \varepsilon_{\mu,t+1}, \quad \varepsilon_{\mu,t} \sim N(0, \sigma^2_{\varepsilon_\mu}). \tag{115}$$

Subsequent realized returns are then simulated from the following equation:

$$r_{t+1} = \mu_t + u_{t+1}, \tag{116}$$

where the variance of $u_{t+1}$ is set to $(4.8\%)^2$, which results in a variance of roughly $(5\%)^2$ per month, consistent with empirical evidence.

Finally, $k$ proxies are constructed from the following equation:

$$\widehat{\mu}_{t,k} = \mu_t + w_{t,k}, \tag{117}$$

where $w_{t,k}$ is generated as $x_t + m_t$ and $m_t$ is white noise. The ratio $\sigma^2_x/(\sigma^2_x + \sigma^2_m)$ controls how systematic the measurement error varies with $x_t$. If it is one, all variation in the measurement error is driven by $x_t$; if it is zero, $w_{t,k}$ is just white noise. The ratio $\sigma^2_\mu/(\sigma^2_\mu + \sigma^2_w)$ determines how much of the variation in each proxy is due to $\mu_t$ and how much variation is due to $w_{t,k}$. Each proxy $\widehat{\mu}_{t,k}$ is then rescaled so that its variance is equal to the variance of $\mu_t$. This ensures that for changing ratios $\sigma^2_\mu/(\sigma^2_\mu + \sigma^2_w)$ the variance of the proxies remains the

130 See for example Pástor, Sinha, and Swaminathan (2008), Pástor and Stambaugh (2009), and Binsbergen and Koijen (2010).

131 All variables are demeaned and thus the intercept is irrelevant.

same. If true expected returns are part of the set of proxies, the ratio $\sigma_\mu^2/(\sigma_\mu^2 + \sigma_w^2)$ is set to 1 for one out of the $k$ proxies.

Table 24 shows the values of the input parameters for the four different scenarios that I run. Let's first discuss the parameters that remain constant over the scenarios. First, the unexpected shocks to realized returns are calibrated in such a way that the standard deviation of realized returns is around 5%, roughly matching the evidence in the data. Second, $x_t$, the independent variable, is in all cases a rather persistent AR(1) process. Third, I simulate in every scenario eight proxies an econometrician can choose from. This is the same number of proxies that I use in the empirical part of my thesis.

Next, let's focus on the differences between the scenarios. The first and second scenario both model a relation between $x_t$ and $\mu_t$. Additionally, the explanatory power of predictive regressions is rather high in the first scenario, since the variance of expected returns is roughly 8% of the variation in realized returns. Therefore, the predictive regression should be able, even in small samples, to identify the correct proxy. The measurement errors of all proxies (except for the true expected return process) are equally driven by $x_t$ and white noise, but the importance of measurement error in relation to $\mu_t$ is distributed uniformly across the proxies. Therefore, there should be quite some variation between the proxies. In the second scenario, the variance of the shocks to expected returns is reduced by a factor of ten. This means that almost all the variation of $\mu_t$ is driven by $x_t$ and that the expected-to-unexpected ratio in realized returns is lower. Also, in this case the true proxy is not within the set of proxies under consideration. Both the signal-to-noise in the measurement error (how much of the variation in measurement error is driven by $x_t$) and in the proxies (how much of the variation in each proxy is driven by $\mu_t$) are uniformly distributed from 0 to 1, adding even more variation between each proxy in each run.

In the third and fourth scenario, there is no relation between $x_t$ and $\mu_t$. $\mu_t$ is in this case an equally persistent, but uncorrelated, AR(1) process.[132] The other parameters are identical to the first two scenarios.

I run each scenario 10,000 times with 300 months sampled in each run. Due to data requirements in empirical research, this sample size is on the upper bound of sample sizes found in empirical time se-

132 Note that this case implies autocorrelated residuals, which is why I use adjusted standard errors with Newey-West lag corrections. The number of lags is computed via the method proposed in Newey and West (1994).

Table 24: **Values in simulation for different scenarios.** This table shows the values for the Monte Carlo simulation for the four different scenarios. $\mathtt{unif}(0,1)$ means that the parameter is drawn from a uniform distribution with minimum 0 and maximum 1.

| | Value in scenario | | | |
|---|---|---|---|---|
| Parameter | 1 | 2 | 3 | 4 |
| $\sigma_u^2$ | $(4.8\%)^2$ | $(4.8\%)^2$ | $(4.8\%)^2$ | $(4.8\%)^2$ |
| $\tau_x$ | 0.8 | 0.8 | 0.8 | 0.8 |
| $\sigma_{\varepsilon_{x,t+1}}^2$ | $(0.6\%)^2$ | $(0.6\%)^2$ | $(0.6\%)^2$ | $(0.6\%)^2$ |
| $\gamma_1$ | 1 | 1 | 0 | 0 |
| $\tau_\mu$ | - | - | 0.8 | 0.8 |
| $\sigma_{\varepsilon_{\mu,t+1}}^2$ | $(1\%)^2$ | $(0.1\%)^2$ | $(1\%)^2$ | $(0.1\%)^2$ |
| k | 8 | 8 | 8 | 8 |
| $\sigma_x^2/(\sigma_x^2+\sigma_m^2)$ | $0.5\ \forall\ k$ | $\mathtt{unif}(0,1)\ \forall\ k$ | $0.5\ \forall\ k$ | $\mathtt{unif}(0,1)\ \forall\ k$ |
| $\sigma_\mu^2/(\sigma_\mu^2+\sigma_{w_k}^2)$ | $\mathtt{unif}(0,1)$ for $k-1$, 1 for kth proxy | $\mathtt{unif}(0,1)\ \forall\ k$ | $\mathtt{unif}(0,1)$ for $k-1$, 1 for kth proxy | $\mathtt{unif}(0,1)\ \forall\ k$ |

ries studies that apply expected return proxies and therefore, larger sample sizes are of little interest. Also, as Sala-I-Martin, Doppelhofer, and Miller (2004) highlight, model selection and model averaging is strictly a small-sample problem. Asymptotically, a researcher is able to identify the best proxy almost surely, so the model selection and the model averaging approach converge. Note, however, that this does not imply that a researcher gets rid of the potential bias in his proxies. Even asymptotically, we can only make *relative* statements about the performance of proxies, not *absolute* ones. In other words, we might identify the best proxy, but we can still not be sure that this proxy measures expected returns without any error. This is another shortcoming of any other proxy in comparison to realized returns. However, in very large samples we can often just use realized returns because it is an unbiased estimate of expected returns and its noise is not an issue anymore.

For each scenario, I compute the bias as well as range of the estimator and the coverage. The bias is the difference between the estimator, i.e., the weighted average slope coefficient $\widehat{\gamma}_a$ that approach $a$ obtained over all Monte Carlo runs,[133] minus the true $\gamma$. The range is the average difference between the 97.5th percentile and the 2.5th percentile of the regression coefficient. The coverage is the number of runs in which the confidence or credible intervals of the specific approach cover $\gamma$ divided by the number of all runs. Note that I use Newey-West corrected standard errors throughout the simulation.

Table 25 presents the results. The column labelled "True" shows the results if one could observe expected returns. Not surprisingly, in this case the bias is close to zero and the coverage is around 95%. However, an econometrician has to settle with proxies and their performance is shown in the remaining columns. The advantages and disadvantages of realized returns are obvious from Table 25. Its application leads to unbiased results (the bias is close to zero and the coverage is close to 95%), but with a lot of parameter uncertainty. The confidence interval that we can impose with realized returns can be more than 40 times as large as the confidence interval we would obtain if we could observe expected returns. This is the main motivation for any alternative expected return proxy and as we can see from the remaining columns, all other approaches lead to much lower intervals. In other words, these approaches are much more certain about the estimated

133 In the case in which only one one regression is run (True, Real, Best, Ave, R2), it is just the regression coefficient obtained from an OLS.

Table 25: **Summary statistics of Monte Carlo simulation.** This table presents summary statistics for four different scenarios (see Table 24). The bias in Panel A is defined as the mean of the OLS estimator (for approaches True, Real, Best, Ave, and R2) or of the weighted average across proxies (for approach BMA) minus the true coefficient $\gamma$. Range in Panel B is the difference between the 97.5th percentile and the 2.5th percentile. The coverage in Panel C shows the ratio (in percent) of how often the 95% percentile of the specific approach covered the true coefficient $\gamma$. The statistics are based on 10,000 Monte Carlo runs. In each run, a time series of 300 months is simulated.

| Scenario | True | Real | Best | Ave | R2 | BMA |
|---|---|---|---|---|---|---|
| Panel A: Bias | | | | | | |
| 1 | 0.0003 | -0.0020 | 0.2860 | 0.2069 | 0.0897 | 0.1197 |
| 2 | -0.0000 | 0.0010 | 0.0252 | -0.0393 | -0.0223 | -0.0169 |
| 3 | -0.0001 | 0.0006 | 1.3617 | 0.8499 | 0.0991 | 0.1534 |
| 4 | -0.0001 | -0.0027 | 0.1438 | 0.0895 | 0.0859 | 0.0882 |
| Panel B: Range | | | | | | |
| 1 | 0.2241 | 1.0970 | 0.1465 | 0.1405 | 0.2024 | 0.3697 |
| 2 | 0.0225 | 1.0779 | 0.0241 | 0.0299 | 0.0621 | 0.1676 |
| 3 | 0.7686 | 1.2505 | 0.2882 | 0.5390 | 0.7570 | 1.0364 |
| 4 | 0.0768 | 1.0779 | 0.0302 | 0.0505 | 0.0499 | 0.1345 |
| Panel C: Coverage | | | | | | |
| 1 | 93.0400 | 93.3100 | 0.3000 | 14.0523 | 57.1900 | 86.7800 |
| 2 | 93.3500 | 93.3300 | 10.4100 | 2.9885 | 22.3700 | 54.6200 |
| 3 | 92.2200 | 91.5200 | 0.0100 | 53.9041 | 80.7700 | 90.9800 |
| 4 | 92.1900 | 93.3000 | 0.0000 | 5.0548 | 13.3900 | 28.7700 |

parameter. However, since BMA incorporates model uncertainty, it has the largest range of the alternative approaches.

Are the methods correct in being certain though? The answer is no. All methods can lead to substantial biases and overconfident results. The worst performer is clearly the "Best" approach. In this case, one simply picks the proxy that yields the results that favor one's hypothesis the most. This is certainly an extreme case in which a researcher goes fishing for the most significant results. The coverage here is close to zero. This is because one always finds too strong of a relation, even in cases in which there is a relation between true expected returns and the variable of interest.

Both averaging across proxies *before* the regressions are run, as proposed by Hail and Leuz (2009), and selecting the proxy based on predictive regressions prevent a researcher from fishing for the "best" proxy and are clearly superior to this approach. They result in better coverages and less bias, but have still far narrower confidence/credible regions than results based on realized returns. Yet, they do not incorporate the uncertainty an econometrician has about which of the proxies is the correct one. Hence, the regions are too narrow and coverage is far too low. That is, we get precise results, but often those results are precisely wrong. A researcher that entertains any of those three approaches severely overestimates the power of the results and makes biased inferences. In cases in which the true proxy is actually within the set of expected return proxies (scenario 1 and 3), BMA is a solution to this problem. The coverage intervals are close to 95%, so model uncertainty is correctly incorporated. Still, the ranges are lower than in the case of realized returns and therefore, applying alternative proxies pays off. However, in cases in which the true proxy is not within the set of proxies under consideration (scenario 2 and 4), the results are still biased and the coverage is far too low, even in the case of BMA.[134] If all proxies are biased, averaging across them will also lead to biased results. Since I am unaware of any reasonable test that establishes the unbiasedness of the proxies under consideration, particularly in short samples, this is a severe shortcoming for any such proxy.

134 Note that the absolute bias is lower in scenario 2 and 4 for the BMA approach. This is due to the fact that the shocks that affect true expected returns are lower which increases the precision in detecting a relation between $x_t$ and $\mu_t$. This can be seen from a comparison of the ranges in Panel B of column "True". However, because the confidence intervals are even smaller in absolute terms, they often do not cover the correct coefficient.

Table 26: **Posterior model weights for different shrinkage parameters (simple realized returns).** This table shows the posterior model weights of the ICC methods for different shrinkage parameters $\phi$. The weights are based on predictive regressions of subsequent realized returns for the next month on the ICCs. In comparison to Table 16, the realized returns are not logged here. The following priors are specified: Equal prior model probabilities $p(M_k)$ across ICC methods, an improper prior on $\sigma^2$, and the natural conjugate g-prior specification for $\beta$: $N(0, \phi\sigma^2(X_k'X_k)^{-1})$, where $X_k$ is the $T \times 2$ matrix of a T vector of ones and the T vector $\widehat{\mu}_i$; the posterior model weights are computed via equation (61). Note that the case $\phi = \infty$ is identical to the AIC weighting shown in equation (62). The time period ranges from 1985 to 2011.

| $\phi$ | PE | PEG | MPEG | OJ | CT | GLS | PSS | CDZ |
|---|---|---|---|---|---|---|---|---|
| 0.01 | 12.51 | 12.55 | 12.55 | 12.48 | 12.48 | 12.53 | 12.52 | 12.40 |
| 0.1 | 12.54 | 12.95 | 12.97 | 12.29 | 12.27 | 12.75 | 12.65 | 11.57 |
| 1 | 12.55 | 15.03 | 15.15 | 11.21 | 11.11 | 13.77 | 13.19 | 7.99 |
| 10 | 12.28 | 17.16 | 17.42 | 9.96 | 9.80 | 14.58 | 13.47 | 5.33 |
| 100 | 12.19 | 17.59 | 17.88 | 9.70 | 9.53 | 14.72 | 13.49 | 4.90 |
| $\infty$ | 12.18 | 17.64 | 17.94 | 9.67 | 9.50 | 14.73 | 13.50 | 4.85 |

In summary, those results show that using alternative expected return proxies without the incorporation of model uncertainty can severely overestimate the confidence a researcher should have in the results. It can also lead to biased results. BMA helps in cases in which the true expected return process is within the set of proxies of the researcher. It must fail, like all other methods, if this is not the case.

## A.6 SENSITIVITY OF THE POSTERIOR MODEL WEIGHTS IN THE PREDICTIVE REGRESSION FRAMEWORK

Table 26 is identical to Table 16 with the only exception that I do not continuously compound the realized returns here. A comparison between the two tables reveals that the posterior model weights are hardly unaffected by this decision.

Table 27 replicates Table 16, but regresses monthly *excess* log realized returns on implied risk premiums instead of log realized returns on implied costs of capital. In general, the weights are more evenly distributed, which indicates that the model uncertainty is even larger in this case. Nevertheless, the results are similar. The GLS method is by far the best performing method, while the CDZ method is worst in explaining subsequent realized returns.

Table 27: **Posterior model weights for different shrinkage parameters (excess continuously compounded realized returns and implied risk premiums).** This table shows the posterior model weights of the ICC methods for different shrinkage parameters $\phi$. The weights are based on predictive regressions of subsequent *excess* realized returns for the next month on the implied risk premiums. The following priors are specified: Equal prior model probabilities $p(M_k)$ across ICC methods, an improper prior on $\sigma^2$, and the natural conjugate g-prior specification for $\beta$: $N(0, \phi\sigma^2(X_k'X_k)^{-1})$, where $X_k$ is the $T \times 2$ matrix of a $T$ vector of ones and the $T$ vector $\widehat{\mu}_i$; the posterior model weights are computed via equation (61). Note that the case $\phi = \infty$ is identical to the AIC weighting shown in equation (62). The time period ranges from 1985 to 2011.

| $\phi$ | PE | PEG | MPEG | OJ | CT | GLS | PSS | CDZ |
|---|---|---|---|---|---|---|---|---|
| 0.01 | 12.50 | 12.49 | 12.52 | 12.50 | 12.51 | 12.50 | 12.50 | 12.47 |
| 0.1 | 12.52 | 12.43 | 12.68 | 12.52 | 12.59 | 12.53 | 12.53 | 12.20 |
| 1 | 12.60 | 12.08 | 13.49 | 12.61 | 13.02 | 12.63 | 12.64 | 10.92 |
| 10 | 12.65 | 11.72 | 14.34 | 12.67 | 13.43 | 12.71 | 12.73 | 9.75 |
| 100 | 12.66 | 11.64 | 14.51 | 12.68 | 13.51 | 12.72 | 12.75 | 9.53 |
| $\infty$ | 12.66 | 11.64 | 14.53 | 12.68 | 13.52 | 12.73 | 12.75 | 9.50 |

# BIBLIOGRAPHY


Abarbanell, Jeffery, and Reuven Lehavy. 2003. "Biased forecasts or biased earnings? The role of reported earnings in explaining apparent bias and over/underreaction in analysts' earnings forecasts." *Journal of Accounting and Economics* 36 (1): 105–146.

Ang, Andrew, and Jun Liu. 2007. "Risk, return, and dividends." *Journal of Financial Economics* 85 (1): 1–38.

Ashton, David, and Pengguo Wang. 2013. "Terminal valuations, growth rates and the implied cost of capital." *Review of Accounting Studies* 18 (1): 261–290.

Attig, Najah, Omrane Guedhami, and Dev Mishra. 2008. "Multiple large shareholders, control contests, and implied cost of equity." *Journal of Corporate Finance* 14 (5): 721–737.

Avramov, Doron. 2002. "Stock return predictability and model uncertainty." *Journal of Financial Economics* 64 (3): 423–458.

Bakshi, Gurdip, and Chen Zhiwu. 2008. "Cash flow risk, discounting risk, and the equity risk premium puzzle." In *Handbook of the Equity Risk Premium,* edited by Rajnish Mehra. Elsevier, Amsterdam, The Netherlands.

Bansal, Ravi, and Amir Yaron. 2004. "Risks for the long run: A potential resolution of asset pricing puzzles." *Journal of Finance* 59 (4): 1481–1509.

Berg, Tobias, and Christoph Kaserer. 2013. "Extracting the Equity Premium from CDS Spreads." *Journal of Derivatives* 21 (1): 8–26.

Beyer, Anne, Daniel Cohen, Thomas Lys, and Beverly Walther. 2010. "The financial reporting environment: Review of the recent literature." *Journal of Accounting and Economics* 50 (2-3): 296–343.

Binsbergen, Jules van, Michael Brandt, and Ralph Koijen. 2012. "On the timing and pricing of dividends." *American Economic Review* 102 (4): 1596–1618.

Binsbergen, Jules van, Wouter Hueskes, Ralph Koijen, and Evert Vrugt. 2013. "Equity yields." *Journal of Financial Economics* 110 (3): 503–519.

Binsbergen, Jules van, and Ralph Koijen. 2010. "Predictive Regressions: A Present-Value Approach." *Journal of Finance* 65 (4): 1439–1471.

Botosan, Christine, and Marlene Plumlee. 2005. "Assessing alternative proxies for the expected risk premium." *Accounting Review* 80 (1): 21–53.

Botosan, Christine, Marlene Plumlee, and He Wen. 2011. "The Relation between Expected Returns, Realized Returns, and Firm Risk Characteristics." *Contemporary Accounting Research* 28 (4): 1085–1122.

Bradshaw, Mark. 2011. "Analysts' Forecasts: What Do We Know After Decades of Work?" *Working Paper.*

Bradshaw, Mark, Michael Drake, James Myers, and Linda Myers. 2012. "A re-examination of analysts' superiority over time-series forecasts of annual earnings." *Review of Accounting Studies* 17 (4): 944–968.

Brav, Alon, Reuven Lehavy, and Roni Michaely. 2005. "Using expectations to test asset pricing models." *Financial Management* 34 (3): 31–64.

Breiman, Leo. 1992. "The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error." *Journal of the American Statistical Association* 87 (419): 738–754.

Brown, Lawrence. 1993. "Earnings forecasting research: Its implications for capital markets research." *International Journal of Forecasting* 9 (3): 295–320.

Brown, Lawrence, Robert Hagerman, Paul Griffin, and Mark Zmijewski. 1987. "Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings." *Journal of Accounting and Economics* 9 (1): 61–87.

Brown, Lawrence, and Michael Rozeff. 1978. "The superiority of analyst forecasts as measures of expectations: Evidence from earnings." *Journal of Finance* 33 (1): 1–16.

Burnham, Kenneth, and David Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach.* 2nd ed. Springer, New York, USA.

Campbell, John. 1991. "A variance decomposition for stock returns." *Economic Journal* 101 (405): 157–179.

Campbell, John, and John Ammer. 1993. "What Moves the Stock and Bond Markets? A Variance Decomposition for Long-Term Asset Returns." *Journal of Finance* 48 (1): 3–37.

Campbell, John, and John Cochrane. 1999. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy* 107 (2): 205–251.

Campbell, John, Andrew Wen-Chuan Lo, and Archie Craig MacKinlay. 1997. *The econometrics of financial markets.* Princeton University Press, Princeton, USA.

Campbell, John, Christopher Polk, and Tuomo Vuolteenaho. 2010. "Growth or glamour? Fundamentals and systematic risk in stock returns." *Review of Financial Studies* 23 (1): 305–344.

Campbell, John, and Robert Shiller. 1988. "The dividend-price ratio and expectations of future dividends and discount factors." *Review of Financial Studies* 1 (3): 195–228.

Campbell, John, and Motohiro Yogo. 2006. "Efficient tests of stock return predictability." *Journal of Financial Economics* 81 (1): 27–60.

Campello, Murillo, Long Chen, and Lu Zhang. 2008. "Expected returns, yield spreads, and asset pricing tests." *Review of Financial Studies* 21 (3): 1297–1338.

Chava, Sudheer, and Amiyatosh Purnanandam. 2010. "Is default risk negatively related to stock returns?" *Review of Financial Studies* 23 (6): 2523–2559.

Chen, Kevin, Zhihong Chen, and K.C. John Wei. 2011. "Agency costs of free cash flow and the effect of shareholder rights on the implied cost of equity capital." *Journal of Financial and Quantitative Analysis* 46 (1): 171–207.

Chen, Long, Zhi Da, and Xinlei Zhao. 2013. "What Drives Stock Price Movements?" *Review of Financial Studies* 26 (4): 841–876.

Chen, Long, and Xinlei Zhao. 2009. "Return decomposition." *Review of Financial Studies* 22 (12): 5213–5249.

Claeskens, Gerda, and Nils Lid Hjort. 2008. *Model selection and model averaging.* Cambridge University Press, New York, USA.

Claus, James, and Jacob Thomas. 2001. "Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets." *Journal of Finance* 56 (5): 1629–1666.

Cochrane, John. 2005. *Asset Pricing.* Princeton University Press, Princeton, USA.

———. 2008. "The dog that did not bark: A defense of return predictability." *Review of Financial Studies* 21 (4): 1533–1575.

———. 2011. "Presidential address: Discount rates." *Journal of Finance* 66 (4): 1047–1108.

Cremers, Martijn. 2002. "Stock return predictability: A Bayesian model selection perspective." *Review of Financial Studies* 15 (4): 1223–1249.

Daske, Holger. 2006. "Economic Benefits of Adopting IFRS or US-GAAP – Have the Expected Cost of Equity Capital Really Decreased?" *Journal of Business Finance & Accounting* 33 (3-4): 329–373.

Daske, Holger, Günther Gebhardt, and Stefan Klein. 2006. "Estimating the expected cost of equity capital using analysts' consensus forecasts." *Schmalenbach Business Review* 58:2–36.

Daske, Holger, Jörn Van Halteren, and Ernst Maug. 2010. "Evaluating methods to estimate the implied cost of equity capital: a simulation study." *Working Paper.*

Das, Somnath, Carolyn Levine, and Ki Sivaramakrishnan. 1998. "Earnings predictability and bias in analysts' earnings forecasts." *Accounting Review* 73 (2): 277–294.

Dhaliwal, Dan, Shane Heitzman, and Oliver Zhen Li. 2006. "Taxes, leverage, and the cost of equity capital." *Journal of Accounting Research* 44 (4): 691–723.

Dijk, Mathijs van. 2011. "Is size dead? A review of the size effect in equity returns." *Journal of Banking & Finance* 35 (12): 3263–3274.

Dugar, Amitabh, and Siva Nathan. 1995. "The Effect of Investment Banking Relationships on Financial Analysts' Earnings Forecasts and Investment Recommendations." *Contemporary Accounting Research* 12 (1): 131–160.

Easton, Peter. 2004. "PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital." *Accounting Review* 79 (1): 73–95.

———. 2006. "Use of forecasts of earnings to estimate and compare cost of capital across regimes." *Journal of Business Finance & Accounting* 33 (3-4): 374–394.

———. 2007. "Estimating the Cost of Capital Implied by Market Prices and Accounting Data." *Foundations and Trends® in Accounting* 2 (4): 241–364.

Easton, Peter, and Steven Monahan. 2005. "An evaluation of accounting-based measures of expected returns." *Accounting Review* 80 (2): 501–538.

———. 2010. "Evaluating accounting-based measures of expected returns: Easton and Monahan and Botosan and Plumlee Redux." *Working Paper.*

Easton, Peter, and Gregory Sommers. 2007. "Effect of analysts' optimism on estimates of the expected rate of return implied by earnings forecasts." *Journal of Accounting Research* 45 (5): 983–1015.

Easton, Peter, Gary Taylor, Pervin Shroff, and Theodore Sougiannis. 2002. "Using forecasts of earnings to simultaneously estimate growth and the rate of return on equity investment." *Journal of Accounting Research* 40 (3): 657–676.

Ecker, Frank, Jennifer Francis, Per Olsson, and Katherine Schipper. 2013. "Associations between realized returns and risk proxies using non-random samples." *Working Paper.*

Efron, Bradley, and Robert Tibshirani. 1993. *An introduction to the bootstrap.* CRC Press, Boca Raton, USA.

Elton, Edwin. 1999. "Presidential address: expected return, realized return, and asset pricing tests." *Journal of Finance* 54 (4): 1199–1220.

Engsted, Tom, Thomas Pedersen, and Carsten Tanggaard. 2012. "Pitfalls in VAR based return decompositions: A clarification." *Journal of Banking & Finance* 36 (5): 1255–1265.

Fama, Eugene. 1970. "Efficient Capital Markets: A Review Of Theory And Empirical Work." *Journal of Finance* 25 (2): 383–417.

———. 1991. "Efficient capital markets: II." *Journal of Finance* 46 (5): 1575–1617.

Fama, Eugene, and Kenneth French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* 33 (1): 3–56.

———. 1997. "Industry costs of equity." *Journal of Financial Economics* 43 (2): 153–193.

———. 2001. "Disappearing dividends: changing firm characteristics or lower propensity to pay?" *Journal of Financial Economics* 60 (1): 3–43.

———. 2002. "The equity premium." *Journal of Finance* 57 (2): 637–659.

Fama, Eugene, and James MacBeth. 1973. "Risk, return, and equilibrium: Empirical tests." *Journal of Political Economy* 81 (3): 607–636.

Fernandez, Carmen, Eduardo Ley, and Mark Steel. 2001. "Model uncertainty in cross-country growth regressions." *Journal of Applied Econometrics* 16 (5): 563–576.

Francis, Jennifer, Ryan LaFond, Per Olsson, and Katherine Schipper. 2004. "Costs of equity and earnings attributes." *Accounting Review* 79 (4): 967–1010.

Frankel, Richard, and Charles Lee. 1999. "Accounting diversity and international valuation." *Working Paper.*

Freedman, David. 1983. "A note on screening regression equations." *American Statistician* 37 (2): 152–155.

Friewald, Nils, Christian Wagner, and Josef Zechner. 2013. "The Cross-Section of Credit Risk Premia and Equity Returns." *Journal of Finance* Forthcoming.

Gebhardt, William, Charles Lee, and Bhaskaran Swaminathan. 2001. "Toward an implied cost of capital." *Journal of Accounting Research* 39 (1): 135–176.

Gelman, Andrew, John Carlin, David Stern Hal Dunson, Aki Vehtari, and Donald Rubin. 2013. *Bayesian data analysis.* 3rd ed. CRC Press, Boca Raton, USA.

George, Edward, and Dean Foster. 2000. "Calibration and empirical Bayes variable selection." *Biometrika* 87 (4): 731–747.

Gode, Dan, and Partha Mohanram. 2003. "Inferring the cost of capital using the Ohlson–Juettner model." *Review of Accounting Studies* 8 (4): 399–431.

Goyal, Amit, and Ivo Welch. 2008. "A comprehensive look at the empirical performance of equity premium prediction." *Review of Financial Studies* 21 (4): 1455–1508.

Guay, Wayne, SP Kothari, and Susan Shu. 2011. "Properties of implied cost of capital using analysts' forecasts." *Australian Journal of Management* 36 (2): 125–149.

Hail, Luzi, and Christian Leuz. 2006. "International differences in the cost of equity capital: Do legal institutions and securities regulation matter?" *Journal of Accounting Research* 44 (3): 485–531.

———. 2009. "Cost of capital effects and changes in growth expectations around U.S. cross-listings." *Journal of Financial Economics* 93 (3): 428–454.

Hanauer, Matthias, Christoph Jäckel, and Christoph Kaserer. 2013. "A New Look at the Fama-French-Model: Evidence Based on Expected Returns." *Working Paper.*

Hann, Rebecca, Maria Ogneva, and Oguzhan Ozbas. 2013. "Corporate diversification and the cost of capital." *Journal of Finance* 68 (5): 1961–1999.

Harvey, Campbell, Yan Liu, and Heqing Zhu. 2013. "...and the cross-section of expected returns." *Working Paper.*

Hoeting, Jennifer, David Madigan, Adrian Raftery, and Chris Volinsky. 1999. "Bayesian model averaging: a tutorial." *Statistical Science* 14 (4): 382–401.

Hou, Kewei, Mathijs van Dijk, and Yinglei Zhang. 2012. "The implied cost of capital: A new approach." *Journal of Accounting and Economics* 53 (3): 504–526.

Hovakimian, Armen, and Ekkachai Saenyasiri. 2012. "US Analyst Regulation and the Earnings Forecast Bias Around the World." *European Financial Management.*

Hribar, Paul, and Nicole Jenkins. 2004. "The effect of accounting restatements on earnings revisions and the estimated cost of capital." *Review of Accounting Studies* 9 (2-3): 337–356.

Hughes, John, Jing Liu, and Jun Liu. 2009. "On the relation between expected returns and implied cost of capital." *Review of Accounting Studies* 14 (2-3): 246–259.

Hughes, John, Jing Liu, and Wei Su. 2008. "On the relation between predictable market returns and predictable analyst forecast errors." *Review of Accounting Studies* 13 (2-3): 266–291.

Isidro, Helena, John O'Hanlon, and Steven Young. 2004. "Dirty surplus accounting flows: international evidence." *Accounting and Business Research* 34 (4): 383–410.

———. 2006. "Dirty surplus accounting flows and valuation errors." *Abacus* 42 (3-4): 302–344.

Jäckel, Christoph. 2013. "A critique of the implied cost of capital evaluation method by Easton and Monahan (2005)." *Unpublished Working Paper.*

———. 2014. "Model uncertainty and expected return proxies." *Working Paper.*

Jäckel, Christoph, Christoph Kaserer, and Katja Mühlhäuser. 2013. "Analystenschätzungen und zeitvariable Marktrisikoprämien – Eine Betrachtung der europäischen Kapitalmärkte." *Wirtschaftsprüfung* 8:365–383.

Jäckel, Christoph, and Katja Mühlhäuser. 2011. "The equity risk premium across European markets: An analysis using the implied cost of capital." *Working Paper.*

Kass, Robert, and Adrian Raftery. 1995. "Bayes factors." *Journal of the American Statistical Association* 90 (430): 773–795.

Kass, Robert, and Larry Wasserman. 1996. "The selection of prior distributions by formal rules." *Journal of the American Statistical Association* 91 (435): 1343–1370.

Kelly, Bryan, and Seth Pruitt. 2013. "Market Expectations in the Cross-Section of Present Values." *Journal of Finance* 68 (5): 1721–1756.

Kennedy, Peter. 2008. *A Guide to Econometrics.* 6th ed. Blackwell Publishing, Oxford, UK.

Koijen, Ralph, and Stijn Van Nieuwerburgh. 2011. "Predictability of returns and cash flows." *Annual Review of Financial Economics* 3:467–491.

Koop, Gary. 2003. *Bayesian Econometrics.* Wiley, Chichester, UK.

Koop, Gary, Dale Poirier, and Justin Tobias. 2007. *Bayesian econometric methods.* Cambridge University Press, Cambridge, USA.

Lambert, Richard. 2009. "Discussion of "on the relation between expected returns and implied cost of capital"." *Review of Accounting Studies* 14 (2-3): 260–268.

Larocque, Stephannie. 2013. "Analysts' earnings forecast errors and cost of equity capital estimates." *Review of Accounting Studies* 18:135–166.

Leamer, Edward. 1978. *Specification searches: ad hoc inference with nonexperimental data.* Wiley, New York, USA.

Lee, Charles, David Ng, and Bhaskaran Swaminathan. 2009. "Testing international asset pricing models using implied costs of capital." *Journal of Financial and Quantitative Analysis* 44 (2): 307–335.

Lee, Charles, Eric So, and Charles Wang. 2011. "Evaluating implied cost of capital estimates." *Working Paper.*

Lewellen, Jonathan. 2004. "Predicting returns with financial ratios." *Journal of Financial Economics* 74 (2): 209–235.

Ley, Eduardo, and Mark Steel. 2009. "On the effect of prior assumptions in Bayesian model averaging with applications to growth regression." *Journal of Applied Econometrics* 24 (4): 651–674.

Lim, Terence. 2001. "Rationality and analysts' forecast bias." *Journal of Finance* 56 (1): 369–385.

Lintner, John. 1965. "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets." *Review of Economics and Statistics* 47 (1): 13–37.

Li, Yan, David Ng, and Bhaskaran Swaminathan. 2013. "Predicting Market Returns Using Aggregate Implied Cost of Capital." *Journal of Financial Economics* 110 (2): 419–436.

Ljungqvist, Alexander, Christopher Malloy, and Felicia Marston. 2009. "Rewriting history." *Journal of Finance* 64 (4): 1935–1960.

Lukacs, Paul, Kenneth Burnham, and David Anderson. 2010. "Model selection bias and Freedman's paradox." *Annals of the Institute of Statistical Mathematics* 62 (1): 117–125.

Lundblad, Christian. 2007. "The risk return tradeoff in the long run: 1836–2003." *Journal of Financial Economics* 85 (1): 123–150.

Lunn, David, Chris Jackson, David Spiegelhalter, Nicky Best, and Andrew Thomas. 2012. *The BUGS book: A practical introduction to Bayesian analysis.* CRC Press, Boca Raton, USA.

Malmendier, Ulrike, and Devin Shanthikumar. 2013. "Do security analysts speak in two tongues?" *Review of Financial Studies* Forthcoming.

Manski, Charles. 2007. *Identification for prediction and decision.* Harvard University Press, Cambridge, USA.

Mehra, Rajnish. 2008. *Handbook of the equity risk premium.* Elsevier, Amsterdam, The Netherlands.

Mehra, Rajnish, and Edward Prescott. 1985. "The equity premium: A puzzle." *Journal of Monetary Economics* 15 (2): 145–161.

Merton, Robert. 1973. "An intertemporal capital asset pricing model." *Econometrica* 41 (5): 867–887.

Miller, Merton, and Franco Modigliani. 1961. "Dividend policy, growth, and the valuation of shares." *Journal of Business* 34 (4): 411–433.

Mohanram, Partha, and Dan Gode. 2013. "Removing predictable analyst forecast errors to improve implied cost of equity estimates." *Review of Accounting Studies* 18:1–36.

Moral-Benito, Enrique. 2012. "Model averaging in economics." *Working Paper.*

Mossin, Jan. 1966. "Equilibrium in a capital asset market." *Econometrica* 34 (4): 768–783.

Mühlhäuser, Katja. 2013. "Measuring expected stock returns - The implied cost of capital and its applications." *Dissertation.*

Nekrasov, Alexander, and Maria Ogneva. 2011. "Using earnings forecasts to simultaneously estimate firm-specific cost of equity and long-term growth." *Review of Accounting Studies* 16 (3): 414–457.

Newey, Whitney, and Kenneth West. 1994. "Automatic lag selection in covariance matrix estimation." *Review of Economic Studies* 61 (4): 631–653.

Ohlson, James. 2005. "On Accounting-Based Valuation Formulae." *Review of Accounting Studies* 10 (2-3): 323–347.

Ohlson, James, and Beate Juettner-Nauroth. 2005. "Expected EPS and EPS growth as determinants of value." *Review of Accounting Studies* 10 (2-3): 349–365.

Ortiz-Molina, Hernán, and Gordon Phillips. 2013. "Real asset liquidity and the cost of capital." *Journal of Financial and Quantitative Analysis* Forthcoming.

Pástor, Ľuboš. 2000. "Portfolio selection and asset pricing models." *Journal of Finance* 55 (1): 179–223.

Pástor, Ľuboš, Meenakshi Sinha, and Bhaskaran Swaminathan. 2008. "Estimating the intertemporal risk–return tradeoff using the implied cost of capital." *Journal of Finance* 63 (6): 2859–2897.

Pástor, Ľuboš, and Robert Stambaugh. 1999. "Costs of equity capital and model mispricing." *Journal of Finance* 54 (1): 67–121.

———. 2000. "Comparing assset pricing models: an investment perspective." *Journal of Financial Economics* 56 (3): 335–381.

———. 2009. "Predictive systems: Living with imperfect predictors." *Journal of Finance* 64 (4): 1583–1628.

———. 2012. "Are stocks really less volatile in the long run?" *Journal of Finance* 67 (2): 431–478.

Pástor, Ľuboš, and Pietro Veronesi. 2009. "Learning in Financial Markets." *Annual Review of Financial Economics* 1:361–381.

Penman, Stephen. 2005. "Discussion of "On Accounting-Based Valuation Formulae" and "Expected EPS and EPS Growth as Determinants of Value"." *Review of Accounting Studies* 10 (2-3): 367–378.

———. 2007. *Financial statement analysis and security valuation.* 3rd ed. McGraw-Hill/Irwin, New York, USA.

Poirier, Dale. 1995. *Intermediate Statistics and Econometrics: A Comparative Approach.* The MIT press, Cambridge, USA.

Raftery, Adrian, David Madigan, and Jennifer Hoeting. 1997. "Bayesian model averaging for linear regression models." *Journal of the American Statistical Association* 92 (437): 179–191.

Raftery, Adrian, and Yingye Zheng. 2003. "Discussion of "Performance of Bayesian model averaging"." *Journal of the American Statistical Association* 98 (464): 931–938.

Ramnath, Sundaresh, Steve Rock, and Philip Shane. 2005. "Value line and I/B/E/S earnings forecasts." *International Journal of Forecasting* 21 (1): 185–198.

———. 2008. "The financial analyst forecasting literature: A taxonomy with suggestions for further research." *International Journal of Forecasting* 24 (1): 34–75.

Ritter, Jay, and Ivo Welch. 2002. "A review of IPO activity, pricing, and allocations." *Journal of Finance* 57 (4): 1795–1828.

Sadka, Gil, and Ronnie Sadka. 2009. "Predictability and the earnings–returns relation." *Journal of Financial Economics* 94 (1): 87–106.

Sala-I-Martin, Xavier, Gernot Doppelhofer, and Ronald Miller. 2004. "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." *American Economic Review* 94 (4): 813–835.

Sharpe, William. 1964. "CAPITAL ASSET PRICES: A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK." *Journal of Finance* 19 (3): 425–442.

Shiller, Robert. 1981. "Do stock Prices move too much to be justified by subsequent changes in dividends?" *American Economic Review* 71 (3): 421–436.

So, Eric. 2013. "A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts?" *Journal of Financial Economics* 108 (3): 615–640.

Stambaugh, Robert. 1999. "Predictive regressions." *Journal of Financial Economics* 54 (3): 375–421.

Tang, Yue, Jin Wu, and Lu Zhang. 2013. "Do Anomalies Exist Ex Ante?" *Review of Finance* Forthcoming.

Vuolteenaho, Tuomo. 2002. "What Drives Firm-Level Stock Returns?" *Journal of Finance* 57 (1): 233–264.

Wansbeek, Tom, and Erik Meijer. 2000. *Measurement error and latent variables in econometrics.* Advanced Textbooks in Economics. Elsevier, Amsterdam, The Netherlands.

Wright, Jonathan. 2008. "Bayesian model averaging and exchange rate forecasts." *Journal of Econometrics* 146 (2): 329–341.

Zellner, Arnold. 1986. "On assessing prior distributions and Bayesian regression analysis with g-prior distributions." In *Bayesian inference and decision techniques: Essays in Honor of Bruno De Finetti,* edited by P. K. Goel and Arnold Zellner. North Holland, Amsterdam, The Netherlands.