

Indirect Inference for Lévy-driven continuous-time GARCH models

Thiago do Rêgo Sousa* Stephan Haug* Claudia Klüppelberg*

November 9, 2018

Abstract

We advocate the use of an Indirect Inference method to estimate the parameter of a COGARCH(1,1) process for equally spaced observations. This requires that the true model can be simulated and a reasonable estimation method for an approximate auxiliary model. We follow previous approaches and use linear projections leading to an auxiliary autoregressive model for the squared COGARCH returns. The asymptotic theory of the Indirect Inference estimator relies on a uniform SLLN and asymptotic normality of the parameter estimates of the auxiliary model, which require continuity and differentiability of the COGARCH process with respect to its parameter and which we prove via Kolmogorov's continuity criterion. This leads to consistent and asymptotically normal Indirect Inference estimates under moment conditions on the driving Lévy process. A simulation study shows that the method yields a substantial finite sample bias reduction compared to previous estimators.

AMS 2010 Subject Classifications: primary: 62F12, 62M10, 91G70, secondary: 37M10, 62P05.

Keywords: Asymptotic normality, Bias reduction, COGARCH, Continuous-time GARCH, Indirect Inference estimation, Projection methods, SLLN

1 Introduction

The COGARCH(1,1) process was introduced in Klüppelberg et al. [23] as a continuous time analog of the discrete time GARCH(1,1) process. It is defined as

$$P_t(\boldsymbol{\theta}) = \int_0^t \sigma_s(\boldsymbol{\theta}) dL_s, \quad t \geq 0, \quad (1.1)$$

with parameter $\boldsymbol{\theta}$ (to be specified in Section 2), L is a Lévy process with Lévy measure $\nu_L \neq 0$ and having càdlàg sample paths. The volatility process $(\sigma_s(\boldsymbol{\theta}))_{s \geq 0}$ is predictable and its stochasticity depends only on L . The COGARCH process satisfies many stylized features of financial time series and is suited for modeling high-frequency data (see Bayracı and Ünal [2], Bibbona and Negri [4], Haug et al. [15], Maller et al. [31], Klüppelberg et al. [25], and Müller [32]).

*Center for Mathematical Sciences, Technical University of Munich, 85748 Garching, Boltzmannstr. 3, Germany, e-mail: thiago.sousa@tum.de, cklu@tum.de, haug@tum.de

In many practical problems, one observes the log-price process $(P_{i\Delta}(\boldsymbol{\theta}_0))_{i=1}^n$ on a fixed grid of size $\Delta > 0$ and the question of interest is how to estimate the true parameter $\boldsymbol{\theta}_0$. The data used for estimation are returns $(G_i(\boldsymbol{\theta}_0))_{i=1}^n$, where

$$G_i(\boldsymbol{\theta}_0) := P_{\Delta i}(\boldsymbol{\theta}_0) - P_{(i-1)\Delta}(\boldsymbol{\theta}_0) = \int_{(i-1)\Delta}^{i\Delta} \sigma_s(\boldsymbol{\theta}_0) dL_s. \quad (1.2)$$

Several methods have been proposed to estimate the parameter of a COGARCH process. A method of moments was proposed in Haug et al. [15], Bibbona and Negri [4] used prediction based estimation as developed in Sørensen [41], and Maller et al. [31] proposed a pseudo maximum likelihood (PML) method which also works for non-equally spaced observations. Both moment and prediction based estimators are consistent and asymptotically normal under certain regularity conditions. The asymptotic properties of the PML estimator were studied in Iannace [18] and in Kim and Lee [22], which require that $\Delta \downarrow 0$ as $n \rightarrow \infty$. For the COGARCH process, Bayracı and Ünal [2] used Indirect Inference with an auxiliary discrete-time GARCH model with Gaussian residuals. No theoretical results were proved, but their simulation study suggests that Indirect Inference estimators achieve a similar performance as the PML estimator of Maller et al. [31] for fixed $\Delta > 0$. Furthermore, Müller [32] proposed a Markov chain Monte Carlo method, when L is a compound Poisson process.

In this paper we advocate an Indirect Inference method, different to the one suggested in Bayracı and Ünal [2], to estimate the COGARCH parameter and derive the asymptotic properties of the estimator. Such methods were introduced in Smith [40] and generalized in Gouriéroux et al. [12], and offer a way to overcome many estimation problems by a clever simulation method. In short, it only requires that the true model can be simulated and a reasonable estimation method for an approximate auxiliary model.

Indirect Inference was originally introduced for complex econometric models to overcome the estimation problem of an intractable likelihood function, as e.g. for continuous time models with stochastic volatility (see Bianchi and Cleur [3], Jiang [20], Laurini and Hotta [28], Raknerud and Skare [37], and Wahlberg et al. [43]). Indirect Inference can also be used as a vehicle to produce estimators which are robust, when there are outliers in the observations (see de Luna and Genton [8] for robust estimation of a discrete time ARMA and Fasen-Hartmann and Kimmig [10] of a continuous time ARMA). Another motivation is given in Gouriéroux et al. [13, 14], where it is shown that Indirect Inference can reduce the finite sample bias considerably. This is our motivation to study the asymptotic properties of Indirect Inference estimators (IIE) in the context of COGARCH estimation.

The Indirect Inference procedure works as follows. Let $\boldsymbol{\pi}$ denote the parameter of an auxiliary model chosen for the COGARCH returns $(G_i(\boldsymbol{\theta}_0))_{i=1}^n$ or some transformed random variables. From this data we estimate $\boldsymbol{\pi}$ and obtain $\hat{\boldsymbol{\pi}}_n$. For many different $\boldsymbol{\theta} \in \Theta$ we simulate $K \geq 1$ independent samples of size n of COGARCH returns $(G_i^{(k)}(\boldsymbol{\theta}))_{i=1}^n$ and compute the estimators $\hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta})$ for $k = 1, \dots, K$. The IIE of $\boldsymbol{\theta}$ is then defined as

$$\hat{\boldsymbol{\theta}}_{n,\text{II}} := \arg \min_{\boldsymbol{\theta} \in \Theta} \left(\hat{\boldsymbol{\pi}}_n - \frac{1}{K} \sum_{k=1}^K \hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta}) \right)^\top \boldsymbol{\Omega} \left(\hat{\boldsymbol{\pi}}_n - \frac{1}{K} \sum_{k=1}^K \hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta}) \right), \quad (1.3)$$

where $\mathbf{\Omega}$ is a symmetric and positive definite weight matrix. Under certain regularity conditions, IIEs are consistent and asymptotically normal. These regularity conditions are mainly related to three aspects: (A) find an auxiliary model whose parameter is connected to the COGARCH parameter through a one-to-one binding function, (B) prove strong consistency and asymptotic normality of $\hat{\boldsymbol{\pi}}_n$, and (C) prove that the estimator $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$, as a function of $\boldsymbol{\theta}$, satisfies conditions for the application of a uniform strong law of large numbers (SLLN) and a delta method for the asymptotic normality.

The starting point (A) is an appropriate auxiliary model that provides a one-to-one binding function. We follow previous approaches and use linear projections leading to an auxiliary autoregressive (AR) model of appropriate order for the squared COGARCH returns $(G_i^2(\boldsymbol{\theta}))_{i \in \mathbb{N}}$. Often the properties of the binding function are assessed via simulation (see Lombardi and Calzolari [30] and Garcia et al. [11]), but for our models the binding function can be proved to be one-to-one.

Part (B), strong consistency and asymptotic normality of the estimator $\hat{\boldsymbol{\pi}}_n$ of the AR model parameter $\boldsymbol{\pi}$, is obtained in a similar way as in classical time series analysis (see e.g. Brockwell and Davis [7]), extending the theory to residuals, which may not be white noise, but an arbitrary stationary and ergodic process with finite variance. The SLLN and asymptotic normality of $\hat{\boldsymbol{\pi}}_n$ will then be a consequence of the fact that $(G_i^2(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ is also strong mixing with appropriate mixing coefficients.

(C) is related to regularity conditions of the map $\boldsymbol{\theta} \mapsto \hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$. To achieve strong consistency of the IIE we need to show that

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}) - \boldsymbol{\pi}_{\boldsymbol{\theta}}\| \xrightarrow{\text{a.s.}} 0, \quad n \rightarrow \infty.$$

To move from point-wise to uniform convergence we use a uniform SLLN in a compact parameter space Θ . For the estimator we study here, the application of a uniform SLLN holds provided that $G_i(\boldsymbol{\theta})$ is a continuous function in $\boldsymbol{\theta}$ and $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} G_i^4(\boldsymbol{\theta}) < \infty$ for all $i \in \mathbb{N}$. The continuity of this map does not follow directly from the continuity of $\sigma_s(\boldsymbol{\theta})$ for fixed s , because the Lévy process in the stochastic integral in (1.2) may have infinite variation. Under conditions on the moments and the characteristic exponent of the driving Lévy process, we find a version of $G_i(\boldsymbol{\theta})$ which is continuous by Kolmogorov's continuity criterion, and as a result we conclude strong consistency of the IIE $\hat{\boldsymbol{\theta}}_{n,\text{II}}$. A Taylor expansion of $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$ around the true parameter $\boldsymbol{\theta}_0$ yields asymptotic normality by the delta method. This will require continuous differentiability of $G_i(\boldsymbol{\theta})$ in $\boldsymbol{\theta}$, which will follow from a result of Hutton and Nelson [16] together with Kolmogorov's continuity criterion.

Our paper is organised as follows. We start in Section 2 with the formal definition of a stationary COGARCH process as returns process, and recall its relevant properties. We also present the autoregressive auxiliary model of the squared returns and define the least squares estimator (LSE) and Yule-Walker estimator (YWE) of the AR parameter, as well as the binding function giving the link to the COGARCH parameter. In Section 2.3 we present the IIE and the conditions, which guarantee a uniform SLLN and asymptotic normality of the IIE. In Section 3 we prove strong consistency and asymptotic normality of the LSE and YWE under the non-standard conditions of stationary ergodicity and a mixing property. Section 4 is dedicated

to strong consistency and asymptotic normality of the IIE of the COGARCH process. Section 5 presents a simulation study and shows that the bias reduction based on the IIE is indeed substantial compared to previous estimators. Technical results like conditions for the existence of a version of the COGARCH returns, which is continuous in its parameter and other auxiliary results are summarized in an Appendix.

Throughout we write $\|\cdot\|$ for the ℓ^1 -norm in \mathbb{R}^d for $d \in \mathbb{N}$ and recall that in \mathbb{R}^d all norms are equivalent. For a matrix $A \in \mathbb{R}^{p \times q}$ we also write $\|A\|$ for the matrix norm generated by the ℓ^1 -norm. For a vector $x \in \mathbb{R}^d$ and a $d \times d$ positive definite matrix Ω we write $\|x\|_\Omega = x^\top \Omega x$. Furthermore, we denote by \mathcal{L}^p the space of p -integrable random variables, and by $\dim(A)$ the dimension of a subset A of \mathbb{R}^d . For a function $f(\boldsymbol{\theta})$ in \mathbb{R} with $\boldsymbol{\theta} \in \mathbb{R}^q$ the gradient with respect to $\boldsymbol{\theta}$ is $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = (\frac{\partial}{\partial \theta_i} f(\boldsymbol{\theta}))_{i=1}^q \in \mathbb{R}^q$ and $\nabla_{\boldsymbol{\theta}}^2 f(\boldsymbol{\theta}) = (\frac{\partial^2}{\partial \theta_k \partial \theta_l} f(\boldsymbol{\theta}))_{k,l=1}^q \in \mathbb{R}^{q \times q}$ denotes the Hessian matrix.

2 COGARCH process, auxiliary AR representation and Indirect Inference Estimation

2.1 Definition of the COGARCH process

For the parameter space of the COGARCH process given as $\{\boldsymbol{\theta} = (\beta, \eta, \varphi)^\top : \beta, \eta, \varphi > 0\}$, we construct a strictly stationary version of the volatility process as in Klüppelberg et al. [23]. First define the process $(Y_s(\boldsymbol{\theta}))_{s \geq 0}$ by

$$Y_s(\boldsymbol{\theta}) := \eta s - \sum_{0 < u \leq s} \log(1 + \varphi(\Delta L_u)^2), \quad s \geq 0, \quad (2.1)$$

with Laplace transform $\mathbb{E}e^{-pY_s(\boldsymbol{\theta})} = e^{s\Psi_{\boldsymbol{\theta}}(p)}$, where

$$\Psi_{\boldsymbol{\theta}}(p) = -p\eta + \int_{\mathbb{R}} ((1 + \varphi x^2)^p - 1) \nu_L(dx), \quad p \geq 0. \quad (2.2)$$

We shall often use the fact that for $p > 0$ by Lemma 4.1(a) in [23],

$$\mathbb{E}|L_1|^{2p} < \infty \quad \text{if and only if} \quad |\Psi_{\boldsymbol{\theta}}(p)| < \infty.$$

Define the volatility process $(\sigma_t^2(\boldsymbol{\theta}))_{t \geq 0}$ by

$$\sigma_t^2(\boldsymbol{\theta}) := \left(\beta \int_0^t e^{Y_s(\boldsymbol{\theta})} ds + \sigma_0^2(\boldsymbol{\theta}) \right) e^{-Y_{t-}(\boldsymbol{\theta})}, \quad t \geq 0, \quad (2.3)$$

where $Y_{t-}(\boldsymbol{\theta})$ denotes the left limit at t and $\sigma_0^2(\boldsymbol{\theta})$ the starting value of the volatility process. If $\mathbb{E}|L_1|^2 < \infty$ and $\Psi_{\boldsymbol{\theta}}(1) < 0$, then by Lemma 4.1(c) of [23], $\sigma_t^2(\boldsymbol{\theta}) \xrightarrow{d} \sigma_\infty^2(\boldsymbol{\theta})$ as $t \rightarrow \infty$, where

$$\sigma_\infty^2(\boldsymbol{\theta}) \stackrel{d}{=} \beta \int_0^\infty e^{-Y_s(\boldsymbol{\theta})} ds.$$

Setting the starting value as

$$\sigma_0^2(\boldsymbol{\theta}) \stackrel{d}{=} \beta \int_0^\infty e^{-Y_s(\boldsymbol{\theta})} ds, \quad \text{independent of } L, \quad (2.4)$$

by Theorem 3.2 of [23] for such $\boldsymbol{\theta}$ the process $(\sigma_t^2(\boldsymbol{\theta}))_{t \geq 0}$ is strictly stationary. Then by Proposition 4.2 of [23] for the stationary process and $k \in \mathbb{N}$,

$$\mathbb{E}\sigma_0^{2k}(\boldsymbol{\theta}) < \infty \quad \text{if and only if} \quad \mathbb{E}L_1^{2k} < \infty \quad \text{and} \quad \Psi_{\boldsymbol{\theta}}(k) < 0. \quad (2.5)$$

Furthermore, for $k = 1, 2$ either of this implies that the squared returns from (1.2) have corresponding finite moments (Proposition 5.1 of [23]). Additionally, by Corollary 3.1 of [23] the process $(P_t(\boldsymbol{\theta}))_{t \geq 0}$ defined in (1.1) with stationary $(\sigma_t(\boldsymbol{\theta}))_{t \geq 0}$ has stationary increments.

2.2 AR representation for the squared returns

We estimate the COGARCH parameter, when the log-price process is observed on a regular grid of fixed size $\Delta > 0$, such that the data are modelled by the returns $(G_i(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ as defined in (1.2).

We state the basic assumptions and recall some properties of the COGARCH process.

Proposition 2.1 (Theorems 3.1 and 3.4 in Haug et al. [15]). *Assume that:*

(A1) *The parameter vector $\boldsymbol{\theta} = (\beta, \eta, \varphi)^\top$ satisfies $\beta, \eta, \varphi > 0$.*

(A2) *$\mathbb{E}L_1 = 0$ and $\text{Var}L_1 = 1$.*

(A3) *The variance c_L of the Brownian component of L is known and satisfies $0 \leq c_L < \text{Var}L_1$.*

(A4) *$\mathbb{E}L_1^4 < \infty$.*

(A5) *$\int_{\mathbb{R}} x^3 \nu_L(dx) = 0$.*

(A6) *$\Psi_{\boldsymbol{\theta}}(2) < 0$.*

Denote the expectation and variance of the squared returns process by

$$\mu_{\boldsymbol{\theta}} = \mathbb{E}G_1^2(\boldsymbol{\theta}) \quad \text{and} \quad \gamma_{\boldsymbol{\theta}}(0) = \text{Var}G_1^2(\boldsymbol{\theta})$$

Then the following holds:

(a) *The autocovariance function of the squared returns process is given by*

$$\gamma_{\boldsymbol{\theta}}(h) = \text{Cov}(G_i^2(\boldsymbol{\theta}), G_{i+h}^2(\boldsymbol{\theta})) = \gamma_{\boldsymbol{\theta}}(0)k_{\boldsymbol{\theta}}e^{-h\rho_{\boldsymbol{\theta}}}, \quad h \in \mathbb{N}. \quad (2.6)$$

(b) *If $\mu_{\boldsymbol{\theta}}, \gamma_{\boldsymbol{\theta}}(0), k_{\boldsymbol{\theta}}, \rho_{\boldsymbol{\theta}} > 0$, then these parameters uniquely determine $\boldsymbol{\theta}$.*

(c) *The process $(G_i(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ is α -mixing with exponentially decaying mixing coefficients.*

Assume that the driving Lévy process satisfies assumptions (A2)-(A5) of Proposition 2.1. We take as parameter space of the COGARCH process a compact set Θ satisfying the relevant conditions of Proposition 2.1; more precisely,

$$\Theta \subset \mathcal{M} := \{\boldsymbol{\theta} = (\beta, \eta, \varphi)^\top : \beta, \eta, \varphi > 0, \Psi_{\boldsymbol{\theta}}(2) < 0 \text{ and } \mu_{\boldsymbol{\theta}}, \gamma_{\boldsymbol{\theta}}(0), k_{\boldsymbol{\theta}}, \rho_{\boldsymbol{\theta}} > 0\}. \quad (2.7)$$

In what follows, we assume that the true model parameter $\boldsymbol{\theta}_0 \in \Theta$. We present the auxiliary AR model using the structure of COGARCH squared returns. Define the centered squared returns for $\boldsymbol{\theta} \in \Theta$ as

$$\tilde{G}_i^2(\boldsymbol{\theta}) := G_i^2(\boldsymbol{\theta}) - \mu_{\boldsymbol{\theta}}, \quad i \in \mathbb{N}. \quad (2.8)$$

Proposition 2.2 (Auxiliary AR(r) model). *Let $\boldsymbol{\theta} \in \Theta$ and $r \geq 2$ be fixed. Define*

$$U_i(\boldsymbol{\theta}) := \tilde{G}_{i+r}^2(\boldsymbol{\theta}) - P_{\mathcal{H}_i} \tilde{G}_{i+r}^2(\boldsymbol{\theta}), \quad i \in \mathbb{N},$$

where $\mathcal{H}_i = \overline{\text{sp}}\{\tilde{G}_{i+r-j}^2(\boldsymbol{\theta}), j = 1, \dots, r\}$ is the closed span in the Hilbert space \mathcal{L}^2 . Then there exist unique real numbers $a_{\boldsymbol{\theta},1}, \dots, a_{\boldsymbol{\theta},r}$ such that

$$U_i(\boldsymbol{\theta}) = \tilde{G}_{i+r}^2(\boldsymbol{\theta}) - \sum_{j=1}^r a_{\boldsymbol{\theta},j} \tilde{G}_{i+r-j}^2(\boldsymbol{\theta}), \quad i \in \mathbb{N}. \quad (2.9)$$

Moreover, the process $(U_i(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ is strictly stationary with $\mathbb{E}U_i(\boldsymbol{\theta}) = 0$ and $\text{Var}U_i(\boldsymbol{\theta}) < \infty$.

Proof. The proof adapts the proof of Proposition 2.2 of Fasen-Hartmann and Kimmig [10] for the COGARCH process. Since $\boldsymbol{\theta} \in \Theta \subset \mathcal{M}$, by Proposition 2.1(a), the autocovariance function of $(\tilde{G}_i^2(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ satisfies $\gamma_{\boldsymbol{\theta}}(0) > 0$ and $\gamma_{\boldsymbol{\theta}}(h) \rightarrow 0$ as $n \rightarrow \infty$. By Proposition 5.1.1 of Brockwell and Davis [7] it follows that the autocovariance matrix of $(\tilde{G}_i^2(\boldsymbol{\theta}))_{i=1}^r$ is non-singular. Hence, the numbers $a_{\boldsymbol{\theta},1}, \dots, a_{\boldsymbol{\theta},r}$ are uniquely given by

$$\begin{pmatrix} a_{\boldsymbol{\theta},1} \\ a_{\boldsymbol{\theta},2} \\ \vdots \\ a_{\boldsymbol{\theta},r} \end{pmatrix} = \begin{pmatrix} \gamma_{\boldsymbol{\theta}}(0) & \gamma_{\boldsymbol{\theta}}(1) & \dots & \gamma_{\boldsymbol{\theta}}(r-1) \\ \gamma_{\boldsymbol{\theta}}(1) & \gamma_{\boldsymbol{\theta}}(0) & \dots & \gamma_{\boldsymbol{\theta}}(r-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{\boldsymbol{\theta}}(r-1) & \gamma_{\boldsymbol{\theta}}(r-2) & \dots & \gamma_{\boldsymbol{\theta}}(0) \end{pmatrix}^{-1} \begin{pmatrix} \gamma_{\boldsymbol{\theta}}(1) \\ \gamma_{\boldsymbol{\theta}}(2) \\ \vdots \\ \gamma_{\boldsymbol{\theta}}(r) \end{pmatrix} \quad (2.10)$$

leading to (2.9). \square

Proposition 2.2 gives an AR(r) representation for $r \geq 2$ for the COGARCH squared returns from (2.8) by rewriting (2.9) as $\tilde{G}_{i+r}^2(\boldsymbol{\theta}) = \sum_{j=1}^r a_{\boldsymbol{\theta},j} \tilde{G}_{i+r-j}^2(\boldsymbol{\theta}) + U_i(\boldsymbol{\theta})$ for $i \in \mathbb{N}$. Let

$$\boldsymbol{\pi}_{\boldsymbol{\theta}} := (\mu_{\boldsymbol{\theta}}, \mathbf{a}_{\boldsymbol{\theta}}, \gamma_{\boldsymbol{\theta}}(0))^{\top} = (\mu_{\boldsymbol{\theta}}, a_{\boldsymbol{\theta},1}, \dots, a_{\boldsymbol{\theta},r}, \gamma_{\boldsymbol{\theta}}(0))^{\top}, \quad (2.11)$$

and let $C \subset \mathbb{R}^r$ be a compact subset of the set containing all possible real coefficients of a strictly stationary AR(r) process. Then we define a compact parameter space of the auxiliary model as

$$\Pi := \left[-\frac{1}{\epsilon}, \frac{1}{\epsilon}\right] \times C \times \left[\epsilon, \frac{1}{\epsilon}\right], \quad (2.12)$$

where ϵ is a small positive constant.

We will investigate two well-known estimators of $\boldsymbol{\pi}_{\boldsymbol{\theta}}$ in (2.11), namely the least squares estimator (LSE) and the Yule-Walker estimator (YWE) defined by

$$\hat{\boldsymbol{\pi}}_{n,\text{LS}}(\boldsymbol{\theta}) = \begin{pmatrix} \hat{\mu}_n(\boldsymbol{\theta}) \\ \hat{\mathbf{a}}_{n,\text{LS}}(\boldsymbol{\theta}) \\ \hat{\gamma}_n(0; \boldsymbol{\theta}) \end{pmatrix} \quad \text{and} \quad \hat{\boldsymbol{\pi}}_{n,\text{YW}}(\boldsymbol{\theta}) = \begin{pmatrix} \hat{\mu}_n(\boldsymbol{\theta}) \\ \hat{\mathbf{a}}_{n,\text{YW}}(\boldsymbol{\theta}) \\ \hat{\gamma}_n(0; \boldsymbol{\theta}) \end{pmatrix}, \quad (2.13)$$

respectively, whose components are given as follows.

Definition 2.3 (LSE and YWE). *The estimators of the mean and variance are given by*

$$\hat{\mu}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n G_i^2(\boldsymbol{\theta}) \quad \text{and} \quad \hat{\gamma}_n(0; \boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n (G_i^2(\boldsymbol{\theta}) - \hat{\mu}_n(\boldsymbol{\theta}))^2.$$

(a) The LSE of $(a_{\boldsymbol{\theta},1}, \dots, a_{\boldsymbol{\theta},r})^\top$ is given by

$$\hat{\mathbf{a}}_{n,\text{LS}}(\boldsymbol{\theta}) = \arg \min_{\mathbf{c} \in C} S_n(\mathbf{c}; \boldsymbol{\theta}),$$

for C as in (2.12), and

$$S_n(\mathbf{c}; \boldsymbol{\theta}) := \frac{1}{n-r} \sum_{i=1}^{n-r} \left((G_{i+r}^2(\boldsymbol{\theta}) - \hat{\mu}_n(\boldsymbol{\theta})) - c_1(G_{i+r-1}^2(\boldsymbol{\theta}) - \hat{\mu}_n(\boldsymbol{\theta})) - \dots - c_r(G_i^2(\boldsymbol{\theta}) - \hat{\mu}_n(\boldsymbol{\theta})) \right)^2.$$

(b) The YWE of $(a_{\boldsymbol{\theta},1}, \dots, a_{\boldsymbol{\theta},r})^\top$ is given by

$$\hat{\mathbf{a}}_{n,\text{YW}}(\boldsymbol{\theta}) = \hat{\mathbf{\Gamma}}_n^{-1}(\boldsymbol{\theta}) \hat{\boldsymbol{\gamma}}_n(\boldsymbol{\theta}), \quad n \in \mathbb{N}, \quad (2.14)$$

where $\hat{\mathbf{\Gamma}}_n^{-1}(\boldsymbol{\theta}) = (\hat{\gamma}_n(i-j; \boldsymbol{\theta}))_{i,j=1}^r$ and $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\theta}) = (\hat{\gamma}_n(1; \boldsymbol{\theta}), \dots, \hat{\gamma}_n(r; \boldsymbol{\theta}))^\top$ are defined in terms of the empirical autocovariance function

$$\hat{\gamma}_n(h; \boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n-h} (G_i^2(\boldsymbol{\theta}) - \hat{\mu}_n(\boldsymbol{\theta}))(G_{i+h}^2(\boldsymbol{\theta}) - \hat{\mu}_n(\boldsymbol{\theta})), \quad h, n \in \mathbb{N}, n > h.$$

We now define a function that will connect the COGARCH process to its auxiliary AR model from Proposition 2.2.

Proposition 2.4 (Binding function). *Define the binding function $\pi : \Theta \rightarrow \Pi$ by $\pi(\boldsymbol{\theta}) = \boldsymbol{\pi}_\theta$ as in (2.11). Then π is injective and continuously differentiable for $r \geq 2$.*

Proof. As in the proof of Lemma 2.5 in Fasen-Hartmann and Kimmig [10], we decompose $\pi : \Theta \rightarrow \Pi$ into three maps $\pi = \pi_1 \circ \pi_2 \circ \pi_3$. Define $\pi_1 : \Theta \rightarrow \mathbb{R}^4$ by

$$\pi_1(\boldsymbol{\theta}) = (\mu_\theta, k_\theta, \rho_\theta, \gamma_\theta(0))^\top,$$

which is by Proposition 2.1(b) injective. Next define $\pi_2 : \pi_1(\Theta) \rightarrow \mathbb{R}^{r+2}$ by

$$\pi_2(\mu_\theta, k_\theta, \rho_\theta, \gamma_\theta(0)) = (\mu_\theta, \gamma_\theta(1), \dots, \gamma_\theta(r), \gamma_\theta(0))^\top.$$

By (2.6), $\gamma_\theta(h) = \gamma_\theta(0)k_\theta e^{-h\rho_\theta}$ for every $h \in \mathbb{N}$, and simple algebra shows that k_θ and ρ_θ are uniquely determined by

$$k_\theta = \frac{\gamma_\theta^2(1)}{\gamma_\theta(0)\gamma_\theta(2)} \quad \text{and} \quad \rho_\theta = \log \left(\frac{\gamma_\theta(1)}{\gamma_\theta(2)} \right), \quad (2.15)$$

and, therefore, π_2 is injective. Finally, define the map $\pi_3 : \pi_2(\pi_1(\Theta)) \rightarrow \Pi$, by

$$\pi_3(\mu_\theta, \gamma_\theta(1), \dots, \gamma_\theta(r), \gamma_\theta(0)) = (\mu_\theta, a_{\boldsymbol{\theta},1}, \dots, a_{\boldsymbol{\theta},r}, \gamma_\theta(0))^\top.$$

The map π_3 is injective, since $\gamma_\theta(1), \dots, \gamma_\theta(r)$ are uniquely determined by $a_{\boldsymbol{\theta},1}, \dots, a_{\boldsymbol{\theta},r}$ and $\gamma_\theta(0)$. We need $r \geq 2$ in order to recover $(\gamma_\theta(1), \gamma_\theta(2))$ from $(\gamma_\theta(0), a_{\boldsymbol{\theta},1}, a_{\boldsymbol{\theta},2})$ using the system of Yule-Walker equations (2.10), so that (2.15) remains valid. This implies the injectivity of the composition π .

Now we prove that π is continuously differentiable. The map π_1 is given in terms of equations (3.6)-(3.9) of Theorem 3.1 in Haug et al. [15], which are continuously differentiable maps of $\Psi_{\boldsymbol{\theta}}(1)$ and $\Psi_{\boldsymbol{\theta}}(2)$ as defined in (2.2). By assumption (A4) of Proposition 2.1 the Lévy process L has finite fourth moment and, therefore, both $\Psi_{\boldsymbol{\theta}}(1)$ and $\Psi_{\boldsymbol{\theta}}(2)$ exist and are continuously differentiable in $\boldsymbol{\theta}$. By (2.6), π_2 is continuously differentiable. Finally, π_3 is also continuously differentiable since it is defined recursively by means of the Yule-Walker equations (2.10). This proves that the composition π is continuously differentiable. \square

2.3 Indirect Inference Estimation

Let $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$ denote an estimator of the auxiliary AR(r) model for $r \geq 2$ based on the returns $(G_i(\boldsymbol{\theta}))_{i=1}^n$, where $\boldsymbol{\theta}$ lies in a compact subset Θ of \mathcal{M} as in (2.7). We define now the IIE for the COGARCH process.

Definition 2.5 (IIE). *Let $\mathbf{G}_n := (G_i(\boldsymbol{\theta}_0))_{i=1}^n$ be the returns as defined in (1.2). Let $\hat{\boldsymbol{\pi}}_n$ be one of the estimators given in (2.13) of $\boldsymbol{\pi}_{\boldsymbol{\theta}_0}$ as defined in (2.11). For arbitrary $\boldsymbol{\theta} \in \Theta$ and $k = 1, \dots, K$ let $\hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta})$ be estimators of $\boldsymbol{\pi}_{\boldsymbol{\theta}}$ based on independent simulated paths $\mathbf{G}_{n,k}(\boldsymbol{\theta}) := (G_i^{(k)}(\boldsymbol{\theta}))_{i=1}^n$. Let $\boldsymbol{\Omega}$ be a symmetric and positive definite weight matrix. Define the function*

$$\hat{L}_{\text{II}} : \Theta \rightarrow [0, \infty) \quad \text{based on } \mathbf{G}_n \text{ by } \hat{L}_{\text{II}}(\boldsymbol{\theta}, \mathbf{G}_n) := \left\| \hat{\boldsymbol{\pi}}_n - \frac{1}{K} \sum_{k=1}^K \hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta}) \right\|_{\boldsymbol{\Omega}}.$$

Then the IIE of $\boldsymbol{\theta}_0$ is defined as

$$\hat{\boldsymbol{\theta}}_{n,\text{II}} := \arg \min_{\boldsymbol{\theta} \in \Theta} \hat{L}_{\text{II}}(\boldsymbol{\theta}, \mathbf{G}_n). \quad (2.16)$$

Concerning the asymptotic behavior of the IIE one would hope that strong consistency and asymptotic normality of the auxiliary model estimators also implies strong consistency and asymptotic normality of the IIE. However, as the Indirect Inference method is based on the simulation of the COGARCH process for many different parameters, we need a stronger (uniform) consistency result and also additional regularity conditions to ensure this. The following is a modification of Propositions 1 and 3 of Gourieroux et al. [12], and it is the analog of Theorem 3.2 of Fasen-Hartmann and Kimmig [10] in the context of our model.

Proposition 2.6. *Assume the setting of Definition 2.5 and $r \geq 2$.*

(a) *If the uniform SLLN*

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}) - \boldsymbol{\pi}_{\boldsymbol{\theta}}\| \xrightarrow{\text{a.s.}} 0, \quad n \rightarrow \infty, \quad (2.17)$$

holds, then the IIE (2.16) is strongly consistent:

$$\hat{\boldsymbol{\theta}}_{n,\text{II}} \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0, \quad n \rightarrow \infty.$$

(b) *Assume additionally to (2.17) that the following hold:*

(b.1) *for every $n \in \mathbb{N}$ the map $\boldsymbol{\theta} \mapsto \hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$ is continuously differentiable,*

(b.2) *for every $\boldsymbol{\theta} \in \Theta$ we have $\sqrt{n}(\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}) - \boldsymbol{\pi}_{\boldsymbol{\theta}}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ as $n \rightarrow \infty$, and*

(b.3) for every sequence $(\boldsymbol{\theta}_n)_{n \in \mathbb{N}}$ with $\boldsymbol{\theta}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$ it also holds that

$$\nabla_{\boldsymbol{\theta}} \hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}_n) \xrightarrow{P} \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}_0}, \quad n \rightarrow \infty,$$

and $\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}(\boldsymbol{\theta}_0)$ has full column rank 3.

(b.4) The true parameter $\boldsymbol{\theta}_0$ lies in the interior of Θ .

Then the IIE (2.17) is asymptotically normal:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{n,\text{II}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \Xi_{\boldsymbol{\theta}_0}), \quad n \rightarrow \infty,$$

where the asymptotic variance is given by

$$\Xi_{\boldsymbol{\theta}_0} = (\mathcal{J}_{\boldsymbol{\theta}_0})^{-1} \mathcal{I}_{\boldsymbol{\theta}_0} (\mathcal{J}_{\boldsymbol{\theta}_0})^{-1} \quad (2.18)$$

with

$$\begin{aligned} \mathcal{J}_{\boldsymbol{\theta}_0} &= (\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}_0})^\top \boldsymbol{\Omega} (\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}_0}) \text{ and} \\ \mathcal{I}_{\boldsymbol{\theta}_0} &= (\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}_0})^\top \boldsymbol{\Omega} \left(1 + \frac{1}{K}\right) \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \boldsymbol{\Omega} (\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}_0}). \end{aligned} \quad (2.19)$$

Proof. Part (a) follows as a particular case of the proof of Theorem 3.2 of [10]. For part (b), we need to check assumptions (C.3)-(C.5) of Theorem 3.2 in [10]. By construction of the estimator (2.13) the asymptotic covariance matrices in (C.3) and (C.4) are identical, so that (b.2) implies (C.3) and (C.4). Instead of verifying (C.5) we modify the argument in [10] (under (b.1) and (b.4)) when manipulating the first order condition

$$0 = \nabla_{\boldsymbol{\theta}} \hat{\mathcal{L}}_{\text{II}}(\hat{\boldsymbol{\theta}}_{n,\text{II}}, \mathbf{G}_n) = 2(\nabla_{\boldsymbol{\theta}} \hat{\boldsymbol{\pi}}_n(\hat{\boldsymbol{\theta}}_{n,\text{II}}))^T \boldsymbol{\Omega} (\hat{\boldsymbol{\pi}}_n(\hat{\boldsymbol{\theta}}_{n,\text{II}}) - \hat{\boldsymbol{\pi}}_n). \quad (2.20)$$

We perform, as in Theorem 3.2 in Newey and McFadden [33], a Taylor expansion of order 1 around the true value $\boldsymbol{\theta}_0$ of the function $\hat{\boldsymbol{\pi}}_n(\hat{\boldsymbol{\theta}}_{n,\text{II}})$ in (2.20). After rearranging the terms this leads to

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{n,\text{II}} - \boldsymbol{\theta}_0) = - \left((\nabla_{\boldsymbol{\theta}} \hat{\boldsymbol{\pi}}_n(\hat{\boldsymbol{\theta}}_{n,\text{II}}))^T \boldsymbol{\Omega} (\nabla_{\boldsymbol{\theta}} \hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}_n)) \right)^{-1} (\nabla_{\boldsymbol{\theta}} \hat{\boldsymbol{\pi}}_n(\hat{\boldsymbol{\theta}}_{n,\text{II}})) \boldsymbol{\Omega} \sqrt{n}(\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}_0) - \hat{\boldsymbol{\pi}}_n),$$

where $\boldsymbol{\theta}_n$ is such that $\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\| \leq \|\hat{\boldsymbol{\theta}}_{n,\text{II}} - \boldsymbol{\theta}_0\|$. The asymptotic normality is now a consequence of taking the limit for $n \rightarrow \infty$ and using (a), (b.2) and (b.3). \square

3 Auxiliary AR model - strong consistency and asymptotic normality

Our objective is to investigate the asymptotic behavior of the IIE for the COGARCH parameter $\boldsymbol{\theta}$ using an AR(r) model for fixed $r \geq 2$ as auxiliary model. This amounts to verifying all assumptions of Proposition 2.6. As a first step we investigate consistency and joint asymptotic normality of the parameter estimator of the auxiliary AR(r) model, which results from the projection of Proposition 2.2, and may have a non-zero mean. It is worth noticing that the noise

$(U_i(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ from (2.9) is defined as the projection error over the finite past. Thus, $U_i(\boldsymbol{\theta})$ is orthogonal to $\mathcal{H}_i = \overline{\text{sp}}\{\tilde{G}_{i+r-j}^2(\boldsymbol{\theta}), j = 1, \dots, r\}$, but we cannot guarantee that it is also orthogonal to $\overline{\text{sp}}\{\tilde{G}_{i+r-j}^2(\boldsymbol{\theta}), j \in \mathbb{N}\}$, so it may not be a white noise process. Therefore, the classical asymptotic theory for the estimation of autoregressive processes (when data come from an AR model with white noise residuals) does not apply directly. Since the residuals are stationary and ergodic with zero mean and finite variance, and since $U_i(\boldsymbol{\theta})$ is orthogonal to $\mathcal{H}_i = \overline{\text{sp}}\{\tilde{G}_{i+r-j}^2(\boldsymbol{\theta}), j = 1, \dots, r\}$, we can obtain results by modifying the classical arguments.

We shall do this for the two estimators from Definition 2.3 and recall that in classical time series theory they are asymptotically equivalent (cf. the proof of Theorem 8.1.2 in Brockwell and Davis [7]). To the best of our knowledge, this has not yet been covered in the literature.

Remark 3.1. *This section provides asymptotic results for the estimators of the auxiliary AR model for some arbitrary, but fixed COGARCh parameter $\boldsymbol{\theta}$, where the dependence on $\boldsymbol{\theta}$ is irrelevant, and we omit it for ease of notation. We define $(W_i)_{i \in \mathbb{N}} := (G_i^2)_{i \in \mathbb{N}}$ and rewrite the auxiliary AR(r) model of Proposition 2.2 with parameter $\boldsymbol{\pi} = (\mu, \mathbf{a}, \gamma(0))$ as*

$$\tilde{W}_{i+r} = \sum_{j=1}^r a_j \tilde{W}_{i+r-j} + U_i, \quad i \in \mathbb{N},$$

where $\tilde{W}_i = \tilde{G}_i^2 = W_i - \mu$, $\mu = \mathbb{E}W_1$ and $\gamma(0) = \text{Var}W_1$.

3.1 Strong consistency of LSE and YWE

Lemma 3.2. *Let the assumptions of Proposition 2.1 hold. Then as $n \rightarrow \infty$, $\hat{\mu}_n \xrightarrow{\text{a.s.}} \mu$ and $\hat{\gamma}_n(h) \xrightarrow{\text{a.s.}} \gamma(h)$ for all $h \in \mathbb{N}_0$.*

Proof. From Proposition 2.1 we know that $\mathbb{E}|W_1| < \infty$ and $(W_i)_{i \in \mathbb{N}}$ is ergodic, so that Birkhoff's ergodic theorem (see e.g. Theorem 4.4 in Krengel [27]) gives immediately $\hat{\mu}_n \xrightarrow{\text{a.s.}} \mu$ as $n \rightarrow \infty$. To prove almost sure convergence of the empirical autocovariance function, we first investigate it, when the mean μ is known:

$$\gamma_n^*(h) := \frac{1}{n} \sum_{i=1}^{n-h} (W_i - \mu)(W_{i+h} - \mu), \quad h \in \mathbb{N}_0. \quad (3.1)$$

Since $W_i W_{i+h}$ is for every $i \in \mathbb{N}$ a measurable map of finitely many values of $(W_i)_{i \in \mathbb{N}}$, the sequence $(W_i W_{i+h})_{i \in \mathbb{N}}$ is ergodic. From Proposition 2.1(a), $\mathbb{E}|W_1 W_{1+h}| < \infty$, so that Birkhoff's ergodic theorem gives $\gamma_n^*(h) \xrightarrow{\text{a.s.}} \gamma(h)$ as $n \rightarrow \infty$. Simple algebra shows that

$$\hat{\gamma}_n(h) - \gamma_n^*(h) = \frac{1}{n} \sum_{i=1}^{n-h} (W_i + W_{i+h} - \hat{\mu}_n - \mu)(\mu - \hat{\mu}_n). \quad (3.2)$$

Since $\hat{\mu}_n \xrightarrow{\text{a.s.}} \mu$, the difference $\gamma_n^*(h) - \hat{\gamma}_n(h) \xrightarrow{\text{a.s.}} 0$ as $n \rightarrow \infty$; hence, $\hat{\gamma}_n(h) \xrightarrow{\text{a.s.}} \gamma(h)$ as $n \rightarrow \infty$. \square

Theorem 3.3 (Consistency of LSE and YWE). *Let the assumptions of Proposition 2.1 hold. Then as $n \rightarrow \infty$, $\hat{\mathbf{a}}_{n,\text{LS}} \xrightarrow{\text{a.s.}} \mathbf{a}$ and $\hat{\mathbf{a}}_{n,\text{YW}} \xrightarrow{\text{a.s.}} \mathbf{a}$.*

Proof. We start by proving strong consistency of the LSE, when the mean μ is known:

$$\mathbf{a}_{n,\text{LS}}^* = \arg \min_{\mathbf{c} \in C} S_n^*(\mathbf{c}), \quad (3.3)$$

for C as in (2.12), and

$$S_n^*(\mathbf{c}) = \frac{1}{n-r} \sum_{i=1}^{n-r} ((W_{i+r} - \mu) - c_r(W_{i+r-1} - \mu) - \cdots - c_1(W_i - \mu))^2.$$

As in Section 8.10* of [7] we write the auxiliary AR(r) model in matrix form as

$$\tilde{\mathbf{Y}}_n = \tilde{\mathbf{W}}_n \mathbf{a} + \mathbf{U}_n, \quad n \in \mathbb{N},$$

where $\tilde{\mathbf{Y}}_n = (\tilde{W}_{r+1}, \dots, \tilde{W}_n)^\top$, $\mathbf{U}_n = (U_1, \dots, U_{n-r})^\top$ and $\tilde{\mathbf{W}}_n$ is the $n \times r$ design matrix,

$$\tilde{\mathbf{W}}_n = \begin{pmatrix} \tilde{W}_r & \tilde{W}_{r-1} & \cdots & \tilde{W}_1 \\ \tilde{W}_{r+1} & \tilde{W}_r & \cdots & \tilde{W}_2 \\ \vdots & \vdots & & \vdots \\ \tilde{W}_{n-1} & \tilde{W}_{n-2} & \cdots & \tilde{W}_{n-r} \end{pmatrix}. \quad (3.4)$$

Then notice that

$$(n-r)S_n^*(\mathbf{c}) = (\tilde{\mathbf{Y}}_n - \tilde{\mathbf{W}}_n \mathbf{c})^\top (\tilde{\mathbf{Y}}_n - \tilde{\mathbf{W}}_n \mathbf{c}),$$

revealing the LSE as a linear regression-type estimator given by

$$\mathbf{a}_{n,\text{LS}}^* = (\tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n)^{-1} \tilde{\mathbf{W}}_n^\top \tilde{\mathbf{Y}}_n, \quad (3.5)$$

provided that the $r \times r$ matrix $\tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n$ is invertible. We prove that $n^{-1} \tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n$ converges a.s. to an invertible matrix. For each fixed $u, v \in \{1, \dots, r\}$ the (u, v) -th entry of this matrix is

$$\frac{1}{n} \sum_{i=0}^{n-r-1} \tilde{W}_{r+1-u+i} \tilde{W}_{r+1-v+i}.$$

Since $\tilde{W}_{r+1-u+i} \tilde{W}_{r+1-v+i}$ is for every $i \in \mathbb{N}_0$ a measurable map of finitely many values of $(W_i)_{i \in \mathbb{N}}$, the sequence $(\tilde{W}_{r+1-u+i} \tilde{W}_{r+1-v+i})_{i \in \mathbb{N}_0}$ is ergodic. Since $\mathbb{E}W_1^2 < \infty$ Birkhoff's ergodic theorem gives

$$\frac{1}{n} \sum_{i=0}^{n-r-1} \tilde{W}_{r+1-u+i} \tilde{W}_{r+1-v+i} \xrightarrow{\text{a.s.}} \mathbb{E} \tilde{W}_1 \tilde{W}_{1+|u-v|}, \quad n \rightarrow \infty, \quad (3.6)$$

and thus $n^{-1} \tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n \xrightarrow{\text{a.s.}} \mathbf{\Gamma}$ as $n \rightarrow \infty$, where $\mathbf{\Gamma}$ is the autocovariance matrix of the squared COGARCH returns, which is non-singular (cf. the proof of Proposition 2.2). Thus, $\mathbf{\Gamma}$ is invertible and therefore the estimator given in (3.5) is well defined for n large enough. With (3.5) we calculate

$$\begin{aligned} \mathbf{a}_{n,\text{LS}}^* - \mathbf{a} &= (\tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n)^{-1} \tilde{\mathbf{W}}_n^\top (\tilde{\mathbf{W}}_n \mathbf{a} + \mathbf{U}_n) - \mathbf{a} \\ &= n (\tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n)^{-1} \frac{1}{n} \tilde{\mathbf{W}}_n^\top \mathbf{U}_n \\ &= n (\tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n)^{-1} \frac{1}{n} \sum_{i=1}^n \left(\tilde{W}_{i+r} - \sum_{j=1}^r a_j \tilde{W}_{i+r-j} \right) \begin{pmatrix} \tilde{W}_{i+r-1} \\ \vdots \\ \tilde{W}_i \end{pmatrix} \\ &=: (n^{-1} \tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n)^{-1} \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i. \end{aligned} \quad (3.7)$$

Since \mathbf{Z}_i is for every $i \in \mathbb{N}$ a measurable map of finitely many values of $(W_i)_{i \in \mathbb{N}}$, the sequence $(\mathbf{Z}_i)_{i \in \mathbb{N}}$ is ergodic. According to Proposition 2.2, $\tilde{W}_{i+r} - \sum_{j=1}^r a_j \tilde{W}_{i+r-j}$ is uncorrelated with $\tilde{W}_i, \dots, \tilde{W}_{i+r-1}$ for all $i \in \mathbb{N}$. Since $\mathbb{E}|\mathbf{Z}_1| < \infty$ Birkhoff's ergodic theorem gives

$$\frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \xrightarrow{\text{a.s.}} \mathbb{E}\mathbf{Z}_1 = \mathbf{0}.$$

This together with the fact that the first term of (3.7) converges a.s. to $\mathbf{\Gamma}^{-1}$ shows that

$$\mathbf{a}_{n,\text{LS}}^* \xrightarrow{\text{a.s.}} \mathbf{a}, \quad n \rightarrow \infty.$$

It remains to prove that $(\hat{\mathbf{a}}_{n,\text{LS}} - \mathbf{a}_{n,\text{LS}}^*) \xrightarrow{\text{a.s.}} \mathbf{0}$ as $n \rightarrow \infty$. Write the LSE in the matrix form

$$\hat{\mathbf{a}}_{n,\text{LS}} = (\bar{\mathbf{W}}_n^\top \bar{\mathbf{W}}_n)^{-1} \bar{\mathbf{W}}_n^\top \bar{\mathbf{Y}}_n,$$

where $\bar{\mathbf{W}}_n$ and $\bar{\mathbf{Y}}_n$ denote the matrix and vector defined in Eq. (3.4), with entries of the form $\bar{W}_i = W_i - \hat{\mu}_n$. Using the matrix identity $A^{-1}x - C^{-1}y = A^{-1}(x - y) + A^{-1}(C - A)C^{-1}y$ gives

$$\begin{aligned} (\hat{\mathbf{a}}_{n,\text{LS}} - \mathbf{a}_{n,\text{LS}}^*) &= \left(\frac{\bar{\mathbf{W}}_n^\top \bar{\mathbf{W}}_n}{n} \right)^{-1} \left(\frac{\bar{\mathbf{W}}_n^\top \bar{\mathbf{Y}}_n}{n} - \frac{\tilde{\mathbf{W}}_n^\top \tilde{\mathbf{Y}}_n}{n} \right) \\ &\quad + \left(\frac{\bar{\mathbf{W}}_n^\top \bar{\mathbf{W}}_n}{n} \right)^{-1} \left(\frac{\tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n}{n} - \frac{\bar{\mathbf{W}}_n^\top \bar{\mathbf{W}}_n}{n} \right) \left(\frac{\tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n}{n} \right)^{-1} \left(\frac{\tilde{\mathbf{W}}_n^\top \tilde{\mathbf{Y}}_n}{n} \right). \end{aligned} \quad (3.8)$$

By Birkhoff's ergodic theorem $n^{-1} \bar{\mathbf{W}}_n^\top \bar{\mathbf{W}}_n, n^{-1} \tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n$ and $n^{-1} \tilde{\mathbf{W}}_n^\top \tilde{\mathbf{Y}}_n$ converge a.s. to two matrices and a vector, respectively. Additionally, by (3.2) we can apply Birkhoff's ergodic theorem to obtain as $n \rightarrow \infty$,

$$\left(\frac{\bar{\mathbf{W}}_n^\top \bar{\mathbf{Y}}_n}{n} - \frac{\tilde{\mathbf{W}}_n^\top \tilde{\mathbf{Y}}_n}{n} \right) \xrightarrow{\text{a.s.}} \mathbf{0} \quad \text{and} \quad \left(\frac{\tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n}{n} - \frac{\bar{\mathbf{W}}_n^\top \bar{\mathbf{W}}_n}{n} \right) \xrightarrow{\text{a.s.}} \mathbf{0},$$

showing that the LSE is consistent. For the YWE the proof is a direct consequence of Lemma 3.2 and the continuous mapping theorem. \square

3.2 Asymptotic normality of the LSE and YWE

One of the requirements for asymptotic normality of the IIE of the COGARCH parameter $\boldsymbol{\theta}$ is condition (b.2) of Proposition 2.6. This means we have to prove asymptotic normality of $\hat{\boldsymbol{\pi}}_{n,\text{LS}}$ and $\hat{\boldsymbol{\pi}}_{n,\text{YW}}$. We start with an auxiliary result.

Lemma 3.4. *Let the assumptions of Proposition 2.1 hold. Let $\mathbf{a}_{n,\text{LS}}^*$ be the LSE defined in (3.3) and $\mathbf{a}_{n,\text{YW}}^*$ be the modification of the YWE defined in (2.14), when the true mean μ is known, i.e., with $\hat{\gamma}_n(\cdot)$ replaced by $\gamma^*(\cdot)$ from (3.1). Then as $n \rightarrow \infty$,*

- (a) $\sqrt{n}(\hat{\mu}_n^2 - \mu^2) \xrightarrow{P} 0$,
- (b) $\sqrt{n}(\mathbf{a}_{n,\text{YW}}^* - \hat{\mathbf{a}}_{n,\text{YW}}) \xrightarrow{P} 0$,
- (c) $\sqrt{n}(\mathbf{a}_{n,\text{YW}}^* - \mathbf{a}_{n,\text{LS}}^*) \xrightarrow{P} 0$, and
- (d) $\sqrt{n}(\mathbf{a}_{n,\text{LS}}^* - \hat{\mathbf{a}}_{n,\text{LS}}) \xrightarrow{P} 0$.

Proof. (a) Write $\sqrt{n}(\hat{\mu}_n^2 - \mu^2) = \sqrt{n}(\hat{\mu}_n + \mu)(\hat{\mu}_n - \mu)$ and notice that by Lemma 3.2 we only need to show that $\sqrt{n}(\hat{\mu}_n + \mu)$ is bounded in probability. It follows from (2.6) that $\gamma(h)$ decays exponentially with h and thus $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$. Let $\epsilon > 0$ be fixed and apply Chebyshev's inequality to get

$$\mathbb{P}(\sqrt{n}|\hat{\mu}_n + \mu| > \epsilon) \leq \epsilon^{-2} n \mathbb{V}\text{ar}(\hat{\mu}_n) \rightarrow \epsilon^{-2} \sum_{h=-\infty}^{\infty} \gamma(h) < \infty, \quad n \rightarrow \infty,$$

where the convergence follows from Theorem 7.1.1 in Brockwell and Davis [7].

(b) Write $\mathbf{a}_{n,\text{YW}}^* = (\mathbf{\Gamma}_n^*)^{-1} \boldsymbol{\gamma}_n^*$ with autocovariance function $\gamma^*(\cdot)$ defined in (3.1). Using properties of the inverse matrix we get

$$\begin{aligned} & \sqrt{n}(\mathbf{a}_{n,\text{YW}}^* - \hat{\mathbf{a}}_{n,\text{YW}}) \\ &= \sqrt{n}(\hat{\mathbf{\Gamma}}_n^{-1} \hat{\boldsymbol{\gamma}}_n - (\mathbf{\Gamma}_n^*)^{-1} \boldsymbol{\gamma}_n^*) \\ &= \hat{\mathbf{\Gamma}}_n^{-1} \sqrt{n}(\mathbf{\Gamma}_n^* - \hat{\mathbf{\Gamma}}_n)(\mathbf{\Gamma}_n^*)^{-1} \hat{\boldsymbol{\gamma}}_n + (\mathbf{\Gamma}_n^*)^{-1} \sqrt{n}(\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_n^*). \end{aligned}$$

The estimators $\hat{\mathbf{\Gamma}}_n, \mathbf{\Gamma}_n^*$ and $\hat{\boldsymbol{\gamma}}_n$ are all bounded in probability. For fixed $h \in \mathbb{N}_0$ it follows from (3.2) and Lemma 3.4(a) that $\sqrt{n}(\hat{\boldsymbol{\gamma}}_n(h) - \boldsymbol{\gamma}_n^*(h)) \xrightarrow{P} 0$ as $n \rightarrow \infty$. Therefore, $\sqrt{n}(\mathbf{\Gamma}_n^* - \hat{\mathbf{\Gamma}}_n)$ and $\sqrt{n}(\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_n^*)$ also converge to zero in probability, which entails (b).

(c) This follows similarly as in the proof of Theorem 8.1.1 in Brockwell and Davis [7].

(d) By (3.8) and observing that $n^{-1} \bar{\mathbf{W}}_n^\top \bar{\mathbf{W}}_n, n^{-1} \tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n$ and $n^{-1} \tilde{\mathbf{W}}_n^\top \tilde{\mathbf{Y}}_n$ are bounded in probability, we only need to show that $n^{-\frac{1}{2}} \{ \bar{\mathbf{W}}_n^\top \bar{\mathbf{Y}}_n - \tilde{\mathbf{W}}_n^\top \tilde{\mathbf{Y}}_n \}$ and $n^{-\frac{1}{2}} \{ \tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n - \bar{\mathbf{W}}_n^\top \bar{\mathbf{W}}_n \}$ converge to zero in probability as $n \rightarrow \infty$. These terms only depend on the autocovariance function of the process $(W_i)_{i \in \mathbb{N}}$ and therefore convergence in probability to zero follows from $\sqrt{n}(\hat{\boldsymbol{\gamma}}_n(h) - \boldsymbol{\gamma}_n^*(h)) \xrightarrow{P} 0$ as $n \rightarrow \infty$ as can be seen from (3.2), and the fact that $\sqrt{n}\hat{\mu}_n$ is bounded in probability. \square

The following is the main result of this Section and proves Proposition 2.6(b.2).

Theorem 3.5 (Asymptotic normality of the LSE and YWE). *Let the assumptions of Proposition 2.1 hold. Assume additionally that $\mathbb{E}|L_1|^{8+\epsilon} < \infty$ and $\Psi_{\boldsymbol{\theta}}(4 + \frac{\epsilon}{2}) < 0$ for some $\epsilon > 0$ and that the matrix $\boldsymbol{\Sigma}$ defined in (3.9) is positive definite. Then, both LSE and YWE for the AR(r) model for $r \geq 2$ are asymptotically normal with covariance matrix*

$$\boldsymbol{\Sigma} = \left(\begin{array}{c|ccc|c} 1 & 0 & \dots & 0 & 0 \\ \hline 0 & & & & 0 \\ \hline \vdots & & \mathbf{\Gamma}^{-1} & & \vdots \\ \hline 0 & & & & 0 \\ \hline 0 & 0 & \dots & 0 & 1 \end{array} \right) \boldsymbol{\Sigma}^*, \quad (3.9)$$

where $\mathbf{\Gamma}$ is the autocovariance matrix of $(W_i)_{i=1}^r$,

$$\boldsymbol{\Sigma}^* = \mathbb{E} \mathbf{C}_1 \mathbf{C}_1^\top + 2 \sum_{i=1}^{\infty} \mathbb{E} \mathbf{C}_1 \mathbf{C}_{1+i}^\top, \quad (3.10)$$

with $C_i \in \mathbb{R}^{r+2}$ given by

$$C_i = \begin{pmatrix} \tilde{W}_i \\ (\tilde{W}_{i+r} - \sum_{j=1}^r a_j \tilde{W}_{i+r-j}) \tilde{W}_{i+r-1} \\ \vdots \\ (\tilde{W}_{i+r} - \sum_{j=1}^r a_j \tilde{W}_{i+r-j}) \tilde{W}_i \\ W_i^2 - \mu^2 \end{pmatrix} \quad (3.11)$$

Proof. Write

$$\sqrt{n}(\hat{\boldsymbol{\pi}}_{n,\text{LS}} - \boldsymbol{\pi}) = \sqrt{n} \begin{pmatrix} \hat{\mu}_n - \mu \\ \hat{\mathbf{a}}_{n,\text{LS}} - \mathbf{a} \\ \hat{\gamma}_n(0) - \gamma(0) \end{pmatrix} = \sqrt{n} \begin{pmatrix} 0 \\ \hat{\mathbf{a}}_{n,\text{LS}} - \mathbf{a}_{n,\text{LS}}^* \\ \mu^2 - \hat{\mu}_n^2 \end{pmatrix} + \sqrt{n} \begin{pmatrix} \hat{\mu}_n - \mu \\ \mathbf{a}_{n,\text{LS}}^* - \mathbf{a} \\ \hat{\gamma}_n(0) + \hat{\mu}_n^2 - \mathbb{E}W_1^2 \end{pmatrix} \quad (3.12)$$

and

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\pi}}_{n,\text{YW}} - \boldsymbol{\pi}) &= \sqrt{n} \begin{pmatrix} \hat{\mu}_n - \mu \\ \hat{\mathbf{a}}_{n,\text{YW}} - \mathbf{a} \\ \hat{\gamma}_n(0) - \gamma(0) \end{pmatrix} \\ &= \sqrt{n} \begin{pmatrix} 0 \\ \hat{\mathbf{a}}_{n,\text{YW}} - \mathbf{a}_{n,\text{YW}}^* \\ \mu^2 - \hat{\mu}_n^2 \end{pmatrix} + \sqrt{n} \begin{pmatrix} 0 \\ \mathbf{a}_{n,\text{YW}}^* - \mathbf{a}_{n,\text{LS}}^* \\ 0 \end{pmatrix} + \sqrt{n} \begin{pmatrix} \hat{\mu}_n - \mu \\ \mathbf{a}_{n,\text{LS}}^* - \mathbf{a} \\ \hat{\gamma}_n(0) + \hat{\mu}_n^2 - \mathbb{E}W_1^2 \end{pmatrix}. \end{aligned} \quad (3.13)$$

We apply Lemma 3.4 to the right-hand side of (3.12) and (3.13) and find that it suffices to prove that

$$\sqrt{n} \begin{pmatrix} \hat{\mu}_n - \mu \\ \mathbf{a}_{n,\text{LS}}^* - \mathbf{a} \\ \hat{\gamma}_n(0) + \hat{\mu}_n^2 - \mathbb{E}W_1^2 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}), \quad n \rightarrow \infty.$$

Using (3.7) we write

$$\begin{aligned} \sqrt{n} \begin{pmatrix} \hat{\mu}_n - \mu \\ \mathbf{a}_{n,\text{LS}}^* - \mathbf{a} \\ \hat{\gamma}_n(0) + \hat{\mu}_n^2 - \mathbb{E}W_1^2 \end{pmatrix} &= \sqrt{n} \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n (W_i - \mu) \\ n(\tilde{\mathbf{W}}_n^\top \tilde{\mathbf{W}}_n)^{-1} \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \\ \frac{1}{n} \sum_{i=1}^n (W_i - \hat{\mu}_n)^2 + \hat{\mu}_n^2 - \mathbb{E}W_1^2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & & & & 0 \\ \vdots & & & & \vdots \\ 0 & & & & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} W_i - \mu \\ \mathbf{Z}_i \\ W_i^2 - \mathbb{E}W_1^2 \end{pmatrix} \\ &=: \mathbf{B}_n \frac{1}{\sqrt{n}} \sum_{i=1}^n C_i. \end{aligned} \quad (3.14)$$

For the asymptotic normality of (3.14) we use the Cramér-Wold device and show that

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \boldsymbol{\lambda}^\top C_i \right) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^* \boldsymbol{\lambda}), \quad n \rightarrow \infty,$$

for all vectors $\boldsymbol{\lambda} \in \mathbb{R}^{r+2}$ such that $\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^* \boldsymbol{\lambda} > 0$. It follows from Proposition 2.1(c) that the squared returns process $(W_i)_{i \in \mathbb{N}}$ is α -mixing with exponentially decaying mixing coefficients. Since each \mathbf{C}_i is a measurable function of W_i, \dots, W_{i-r} it follows from Remark 1.8 of Bradley [6] that $(\boldsymbol{\lambda}^\top \mathbf{C}_i)_{i \in \mathbb{N}}$ is also α -mixing with mixing coefficients satisfying $\alpha_{\mathbf{C}}(n) \leq \alpha_{\mathbf{W}}(n - (r + 1))$ for all $n \geq r + 2$. Therefore $\sum_{n=0}^{\infty} (\alpha_{\mathbf{C}}(n))^{\frac{\epsilon}{2+\epsilon}} < \infty$ for all $\epsilon > 0$. Since $\mathbb{E}|L_1|^{8+\epsilon} < \infty$ and $\Psi_{\boldsymbol{\theta}}(4 + \frac{\epsilon}{2}) < 0$ it follows from (2.5) that $\mathbb{E}|W_1|^{4+\epsilon/2} < \infty$ and, as a consequence, $\mathbb{E}|\boldsymbol{\lambda}^\top \mathbf{C}_1|^{2+\epsilon/4} < \infty$. Thus, the CLT for α -mixing sequences applies (see Theorem 18.5.3 of Ibragimov and Linnik [19]) so that

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \boldsymbol{\lambda}^\top \mathbf{C}_i \right) \xrightarrow{d} \mathcal{N}(0, \zeta), \quad n \rightarrow \infty,$$

where

$$\zeta = \mathbb{E} \boldsymbol{\lambda}^\top \mathbf{C}_1 \mathbf{C}_1^\top \boldsymbol{\lambda} + 2 \sum_{i=1}^{\infty} \mathbb{E} \boldsymbol{\lambda}^\top \mathbf{C}_1 \mathbf{C}_{1+i}^\top \boldsymbol{\lambda}.$$

After rearranging this equation we find (3.10). Let $\mathbf{B}_n = (b_{u,v}^n)_{u,v=1}^{r+2}$ denote the matrix as defined in (3.14). Using (3.6) we get, as $n \rightarrow \infty$,

$$b_{u,v}^n \xrightarrow{\text{a.s.}} \mathbb{E}(W_1 - \mu)(W_{1+|u-v|} - \mu), \quad 2 \leq u, v \leq r + 1.$$

Then the inner block of the matrix \mathbf{B}_n converges a.s. to $\boldsymbol{\Gamma}^{-1}$. This gives (3.9) which finishes the proof. \square

4 IIE of the COGARCh process - strong consistency and asymptotic normality

The objective of this section is to prove strong consistency and asymptotic normality of the IIE of the COGARCh parameter $\boldsymbol{\theta}$. Let $(G_i(\boldsymbol{\theta}_0))_{i=1}^n$ be the returns originating from a COGARCh log-price process (1.1). As auxiliary model we will use an AR(r) model for fixed $r \geq 2$ as in Proposition 2.2, whose parameters are estimated by one of the estimators $\hat{\boldsymbol{\pi}}_n$ from Definition 2.3, which we consider as functions of the COGARCh parameter $\boldsymbol{\theta}$.

4.1 Preliminary results

We begin with an auxiliary result, which is a consequence of Theorem 3.2 of [23].

Lemma 4.1. *Assume that $\mathbb{E}|L_1|^2 < \infty$ and $\Psi_{\boldsymbol{\theta}_i}(1) < 0$ for $i = 1, \dots, d$. Then for every $t > 0$,*

$$(\sigma_t^2(\boldsymbol{\theta}_1), \dots, \sigma_t^2(\boldsymbol{\theta}_d)) \stackrel{d}{=} (\sigma_0^2(\boldsymbol{\theta}_1), \dots, \sigma_0^2(\boldsymbol{\theta}_d))$$

In what follows we shall need for fixed $\varphi > 0$ the stochastic process

$$K_s(\varphi) = \sum_{0 < u \leq s} \frac{(\Delta L_u)^2}{1 + \varphi(\Delta L_u)^2}, \quad s \geq 0. \quad (4.1)$$

Lemma 4.2. *The process $(K_s(\varphi))_{s \geq 0}$ is a Lévy process and $\mathbb{E}|K_s(\varphi)|^p < \infty$ for all $p \in \mathbb{N}$.*

Proof. That $(K_s(\varphi))_{s \geq 0}$ is a Lévy process is clear. Since

$$\sup_{s \geq 0} |\Delta K_s(\varphi)| = \sup_{s \geq 0} \frac{(\Delta L_s)^2}{1 + \varphi(\Delta L_s)^2} \leq \frac{1}{\varphi} < \infty \quad (4.2)$$

it follows that $(K_s(\varphi))_{s \geq 0}$ has bounded jumps and therefore it has moments of all orders (see e.g. Theorem 2.4.7 of [1]). \square

For $p \geq 1$ consider the sets

$$\Theta^{(p)} \subset \mathcal{M}^{(p)} := \{\boldsymbol{\theta} \in \mathcal{M} : \Psi_{\boldsymbol{\theta}}(p) < 0\}, \quad (4.3)$$

where $\Theta^{(p)}$ is compact, and recall from (2.5) that the condition $\Psi_{\boldsymbol{\theta}}(p) < 0$ implicitly requires $\mathbb{E}|L_1|^{2p} < \infty$.

Lemma 4.3. *Let $p \geq 1$ and $t \geq 0$ be fixed and consider the sets $\Theta^{(p)}$ and $\mathcal{M}^{(p)}$ as in (4.3). Then*

(a) *there exist numbers $\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_N^* \in \mathcal{M}^{(p)}$ such that*

$$\sup_{\boldsymbol{\theta} \in \Theta^{(p)}} e^{-Y_t(\boldsymbol{\theta})} \leq \sum_{j=1}^N e^{-Y_t(\boldsymbol{\theta}_j^*)}.$$

(b) (Proposition 2 in Klüppelberg et al. [24]) *there exists some $\sigma^* > 0$ such that $\sigma_0(\boldsymbol{\theta}) \geq \sigma^*$ a.s. for all $\boldsymbol{\theta} \in \Theta$.*

Proof. (a) We use a Heine-Borel argument to control the exponential term. Since $\Theta^{(p)}$ is compact we can find a finite collection of open sets $(\Theta_j^{(p)})_{j=1}^N$ such that $\Theta^{(p)} \subseteq \cup_{j=1}^N \Theta_j^{(p)} \subset \mathcal{M}^{(p)}$. For each fixed j the closure $\overline{\Theta_j^{(p)}}$ is a subset of $\mathcal{M}^{(p)}$ and therefore there exists a point $\boldsymbol{\theta}_j^* = (\beta_j^*, \eta_j^*, \varphi_j^*)^\top \in \mathcal{M}^{(p)}$ such that $\eta \geq \eta_j^*, \varphi \leq \varphi_j^*$ for all $\boldsymbol{\theta} \in \Theta_j^{(p)}$. This implies that for all $\boldsymbol{\theta} \in \Theta_j^{(p)}$:

$$Y_t(\boldsymbol{\theta}) = \eta t - \sum_{0 < u \leq t} \log(1 + \varphi(\Delta L_u)^2) \geq \eta_j^* t - \sum_{0 < u \leq t} \log(1 + \varphi_j^*(\Delta L_u)^2) = Y_t(\boldsymbol{\theta}_j^*), \quad t \geq 0.$$

Therefore,

$$\sup_{\boldsymbol{\theta} \in \Theta^{(p)}} e^{-Y_t(\boldsymbol{\theta})} \leq \sum_{j=1}^N e^{-Y_t(\boldsymbol{\theta}_j^*)},$$

proving the statement. \square

Remark 4.4. *Both $\hat{\pi}_{n, \text{YW}}$ (by (2.14)) and $\hat{\pi}_{n, \text{LS}}$ (as in the proof of Proposition 5.6 in Fasen-Hartmann and Kimmig [10]) can be written as a map $g : \mathbb{R}^{r+2} \rightarrow \mathbb{R}^{r+2}$ for $r \geq 2$ with $g(\mathbf{x}) = (g_1(\mathbf{x}), \dots, g_{r+2}(\mathbf{x}))$ for $\mathbf{x} = (x_1, \dots, x_{r+2})$ applied to the vector*

$$f_n(\boldsymbol{\theta}) = \left(\frac{1}{n} \sum_{i=1}^n G_i^2(\boldsymbol{\theta}), \frac{1}{n} \sum_{i=1}^{n-h} G_i^2(\boldsymbol{\theta}) G_{i+h}^2(\boldsymbol{\theta}), h = 0, \dots, r \right), \quad n \in \mathbb{N}. \quad (4.4)$$

Since g involves only matrix multiplications and matrix inversion of non-singular matrices, it inherits the smoothness properties of $G_i(\boldsymbol{\theta})$ for $i \in \mathbb{N}$. Since $(G_i(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ is stationary and ergodic, Birkhoff's ergodic theorem applies and $f_n(\boldsymbol{\theta})$ converges as $n \rightarrow \infty$ pointwise to

$$f(\boldsymbol{\theta}) = (\mathbb{E}G_1^2(\boldsymbol{\theta}), \mathbb{E}G_1^2(\boldsymbol{\theta})G_{1+h}^2(\boldsymbol{\theta}), h = 0, \dots, r). \quad (4.5)$$

Remark 4.5. *The results that follow are related to continuity and differentiability of the random elements $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)$ for $i \in \mathbb{N}$ with respect to $\boldsymbol{\theta}$. According to (2.3) and (2.4) we find*

$$G_i(\boldsymbol{\theta}) = \int_{(i-1)\Delta}^{i\Delta} \sigma_s((\beta, \eta, \varphi)) dL_s = \sqrt{\beta} \int_{(i-1)\Delta}^{i\Delta} \sigma_s((1, \eta, \varphi)) dL_s = \sqrt{\beta} G_i((1, \eta, \varphi)),$$

which is linear in $\sqrt{\beta}$, hence, $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)$ is obviously continuous in β and has a partial derivative with respect to $\beta > 0$.

4.2 Strong consistency of the IIE

To ensure strong consistency of $\hat{\boldsymbol{\theta}}_{n,\text{II}}$, we need to verify that $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$ satisfies the uniform SLLN of Proposition 2.6(a). The results of Lemma 3.2 and Theorem 3.3 guarantee point-wise strong consistency. Uniform strong consistency will hold by continuity of g (cf. Lemma A.2), if we can apply a uniform SLLN to the sequence in (4.4).

Since the sequence of random elements $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)_{i \in \mathbb{N}}$ is stationary and ergodic, we need to show (cf. Theorem 7 in Straumann and Mikosch [42]) that $G_i(\boldsymbol{\theta})$ is for every $i \in \mathbb{N}$ a continuous function of $\boldsymbol{\theta}$ on Θ or on some compact subset $\Theta^{(p)}$ of Θ and that

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(p)}} G_i^4(\boldsymbol{\theta}) < \infty.$$

Proving that $G_i(\boldsymbol{\theta})$ is ω -wise continuous in its parameter $\boldsymbol{\theta}$ is not straightforward, since

$$G_i(\boldsymbol{\theta}) = \int_{(i-1)\Delta}^{i\Delta} \sigma_s(\boldsymbol{\theta}) dL_s$$

is a stochastic integral, driven by an arbitrary Lévy process, which also drives the stochastic volatility process. If L has finite variation, we can use dominated convergence to show continuity, but this is not possible when L has infinite variation sample paths; cf. Remark 4.9 below. However, as we shall show in the next result, applying Kolmogorov's continuity criterion, we can always find a version $(G_i^{(c)}(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ of the sequence $(G_i(\boldsymbol{\theta}))_{i \in \mathbb{N}}$, which is continuous on a possibly smaller compact parameter space $\Theta^{(p)} \subseteq \Theta$ for Θ .

Theorem 4.6 (Hölder continuity). *Assume that $\mathbb{E}|L_1|^{2p(1+\epsilon)} < \infty$ for some $p > 2$ and $\epsilon > 0$. Then there exists a version $(G_i^{(c)}(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ of the random elements $(G_i(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ which is Hölder continuous of every order $\gamma \in [0, (p-2)/(2p))$ on $\Theta^{(p(1+\epsilon))}$ as defined in (4.3). Additionally, for every $q \in [0, 2p)$, $i \in \mathbb{N}$, and for*

$$U_i = \sup_{\substack{0 < \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| < 1 \\ \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(p(1+\epsilon))}}} \frac{|G_i^{(c)}(\boldsymbol{\theta}_1) - G_i^{(c)}(\boldsymbol{\theta}_2)|}{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma} \quad (4.6)$$

we have

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} |G_i^{(c)}(\boldsymbol{\theta})|^q < \infty \quad \text{and} \quad \mathbb{E} U_i^q < \infty. \quad (4.7)$$

Proof. Without loss of generality we prove this for $i = 1$. We find a continuous version of the random element $G_1(\boldsymbol{\theta})$ on $\Theta^{(p(1+\epsilon))}$. We first prove continuity with respect to (η, φ) and assume

that $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(p(1+\epsilon))}$ with $\beta_1, \beta_2 = 1$. Using the simple inequality $|a - b|^{2p} \leq |a^2 - b^2|^p$, the stationarity of $\sigma_0(\boldsymbol{\theta})$ in Lemma 4.1, its differentiability in (A.9) and the mean value theorem gives

$$\begin{aligned} \int_0^\Delta \mathbb{E}|\sigma_s(\boldsymbol{\theta}_1) - \sigma_s(\boldsymbol{\theta}_2)|^{2p} ds &\leq \int_0^\Delta \mathbb{E}|\sigma_s^2(\boldsymbol{\theta}_1) - \sigma_s^2(\boldsymbol{\theta}_2)|^p \\ &= \Delta \mathbb{E}|\sigma_0^2(\boldsymbol{\theta}_1) - \sigma_0^2(\boldsymbol{\theta}_2)|^p \\ &\leq \Delta \left(\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} |\nabla_{\eta, \varphi} \sigma_0^2(\boldsymbol{\theta})|^p \right) \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^p < \infty \end{aligned} \quad (4.8)$$

by Lemma A.5 with $k = 1$. By (A2) of Proposition 2.1, $(L_t)_{t \geq 0}$ is a martingale. Since $\mathbb{E}|L_1|^{2p} < \infty$ and $\int_0^\Delta \mathbb{E}|\sigma_s(\boldsymbol{\theta}_1) - \sigma_s(\boldsymbol{\theta}_2)|^{2p} ds < \infty$ we can apply Theorem 66 of Ch. 5 in Protter [35] to the stochastic integral in (1.1) and obtain

$$\mathbb{E}|G_1(\boldsymbol{\theta}_1) - G_1(\boldsymbol{\theta}_2)|^{2p} = \mathbb{E} \left| \int_0^\Delta (\sigma_s(\boldsymbol{\theta}_1) - \sigma_s(\boldsymbol{\theta}_2)) dL_s \right|^{2p} \leq c^* \int_0^\Delta \mathbb{E}|\sigma_s(\boldsymbol{\theta}_1) - \sigma_s(\boldsymbol{\theta}_2)|^{2p} ds,$$

where c^* is a positive constant. This combined with (4.8) gives

$$\mathbb{E}|G_1(\boldsymbol{\theta}_1) - G_1(\boldsymbol{\theta}_2)|^{2p} \leq c\Delta \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^p, \quad (4.9)$$

where $c = c^* \Delta \mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} |\nabla_{\eta, \varphi} \sigma_0^2(\boldsymbol{\theta})|^p$. Since $\beta_1, \beta_2 = 1$ we show continuity with respect to (η, φ) ; i.e. the parameter space has dimension $d = 2$. Since $p > 2 = d$ we can apply Kolmogorov's continuity criterion (Theorem 10.1 in Schilling and Partzsch [38], see also Theorem 2.5.1 of Ch. 5 in Khoshnevisan [21]). Then there exists a version $(G_1^{(c)}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))})$ of $(G_1(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))})$ which is Hölder continuous of every order $\gamma \in [0, (p-2)/(2p)]$; hence, also continuous. Since Θ is compact, Lemma A.1 together with (4.9) gives $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} |G_1^{(c)}(\boldsymbol{\theta})|^q < \infty$ for every $q \in [0, 2p)$. Finally, the second expectation in (4.7) is finite by Theorem 10.1 in [38].

Now because of Remark 4.5, $G_1^{(c)}(\boldsymbol{\theta})$ is linear in $\sqrt{\beta}$ and therefore the results can be generalized for the map $\boldsymbol{\theta} \mapsto G_1^{(c)}(\boldsymbol{\theta})$ on $\Theta^{(p(1+\epsilon))}$. Indeed, let β^* be as in (A.15) and

$$\beta_* = \inf\{\beta > 0 : (\beta, \eta, \varphi) \in \Theta\} > 0.$$

Now, for arbitrary $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(p(1+\epsilon))}$ we can use Remark 4.5, the mean value theorem for $\beta \mapsto \sqrt{\beta}$ and the Hölder continuity of order γ of $(\eta, \varphi) \mapsto G_1^{(c)}((1, \eta, \varphi))$, the definition of the ℓ^1 -norm and the fact that $\gamma \in (0, 1)$ to get

$$\begin{aligned} &|G_1^{(c)}(\boldsymbol{\theta}_1) - G_1^{(c)}(\boldsymbol{\theta}_2)| \\ &\leq |G_1^{(c)}((1, \eta_1, \varphi_1)) - G_1^{(c)}((1, \eta_2, \varphi_2))| \sqrt{\beta^*} + \frac{1}{2\sqrt{\beta_*}} |\beta_1 - \beta_2| \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} |G_1^{(c)}(\boldsymbol{\theta})| \\ &\leq K \|(\eta_1, \varphi_1) - (\eta_2, \varphi_2)\|^\gamma \sqrt{\beta^*} + \frac{1}{2\sqrt{\beta_*}} |\beta_1 - \beta_2|^\gamma |2\beta^*|^{1-\gamma} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} |G_1^{(c)}(\boldsymbol{\theta})| \\ &\leq \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma \left(K \sqrt{\beta^*} + \frac{1}{2\sqrt{\beta_*}} |2\beta^*|^{1-\gamma} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} |G_1^{(c)}(\boldsymbol{\theta})| \right), \end{aligned} \quad (4.10)$$

showing the Hölder continuity of $\boldsymbol{\theta} \mapsto G_1^{(c)}(\boldsymbol{\theta})$ on $\Theta^{(p(1+\epsilon))}$. Now, the first expectation in (4.7) is finite since $|G_1^{(c)}(\boldsymbol{\theta})| \leq \beta^* |G_1^{(c)}((1, \eta, \varphi))|$. Now let $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(p(1+\epsilon))}$ be such that $0 < \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| <$

1. Using the inequality at the first line of (4.10) and the definition of the ℓ^1 -norm gives

$$\begin{aligned}
& \sup_{\substack{0 < \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| < 1 \\ \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(p(1+\epsilon))}}} \frac{|G_1^{(c)}(\boldsymbol{\theta}_1) - G_1^{(c)}(\boldsymbol{\theta}_2)|}{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma} \\
& \leq \left(\sup_{\substack{0 < \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| < 1 \\ \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(p(1+\epsilon))}}} \frac{|G_1^{(c)}((1, \eta_1, \varphi_1) - G_1^{(c)}((1, \eta_2, \varphi_2))|}{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma} \right) \sqrt{\beta^*} \\
& \quad + \sup_{\substack{0 < \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| < 1 \\ \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(p(1+\epsilon))}}} \frac{|\beta_1 - \beta_2|}{2\sqrt{\beta^*} \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma} |G_1^{(c)}(\boldsymbol{\theta})| \\
& \leq \left(\sup_{\substack{0 < \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| < 1 \\ \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(p(1+\epsilon))}}} \frac{|G_1^{(c)}((1, \eta_1, \varphi_1) - G_1^{(c)}((1, \eta_2, \varphi_2))|}{\|(\eta_1, \varphi_1) - (\eta_2, \varphi_2)\|^\gamma} \right) \sqrt{\beta^*} + \frac{1}{2\sqrt{\beta^*}} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} |G_1^{(c)}(\boldsymbol{\theta})|
\end{aligned} \tag{4.11}$$

Applying the supremum and raising both sides of (4.11) to the power q gives the result. \square

Remark 4.7. *In view of Theorem 4.6 we will from now on work with a continuous version of the returns $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))})_{i \in \mathbb{N}}$.*

Theorem 4.8 (Strong consistency of the IIE). *Assume that $\mathbb{E}|L_1|^{2p(1+\epsilon)} < \infty$ for some $p > 2$ and $\epsilon > 0$ and let $(G_i(\boldsymbol{\theta}_0))_{i=1}^n$ be the returns (1.2) with parameter $\boldsymbol{\theta}_0 \in \Theta^{(p(1+\epsilon))}$ from (4.1). Suppose that the auxiliary AR(r) model for $r \geq 2$ is estimated by the LSE or the YWE of Definition 2.3. Then*

$$\hat{\boldsymbol{\theta}}_{n, \text{II}} \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0, \quad n \rightarrow \infty.$$

Proof. According to Proposition 2.6(a), strong consistency of the IIE will follow if as $n \rightarrow \infty$,

$$\sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} \|\hat{\boldsymbol{\pi}}_{n, \text{LS}}(\boldsymbol{\theta}) - \boldsymbol{\pi}_\boldsymbol{\theta}\| \xrightarrow{\text{a.s.}} 0 \quad \text{and} \quad \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} \|\hat{\boldsymbol{\pi}}_{n, \text{YW}}(\boldsymbol{\theta}) - \boldsymbol{\pi}_\boldsymbol{\theta}\| \xrightarrow{\text{a.s.}} 0.$$

By Remark 4.4 and Lemma A.2 it suffices to prove that

$$\sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} \|f_n(\boldsymbol{\theta}) - f(\boldsymbol{\theta})\| \xrightarrow{\text{a.s.}} 0, \quad n \rightarrow \infty, \tag{4.12}$$

for f_n and f as defined in (4.4) and (4.5), respectively. The Cauchy-Schwarz inequality gives for every $h \in \mathbb{N}_0$,

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} G_1^2(\boldsymbol{\theta}) G_{1+h}^2(\boldsymbol{\theta}) \leq \left(\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} G_1^4(\boldsymbol{\theta}) \right)^{\frac{1}{2}} \left(\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} G_{1+h}^4(\boldsymbol{\theta}) \right)^{\frac{1}{2}} < \infty. \tag{4.13}$$

The right-hand side of (4.13) is finite by Theorem 4.6. It also follows from the same theorem that $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} G_1^2(\boldsymbol{\theta}) < \infty$ and, hence, by Theorem 7 in Straumann and Mikosch [42] the uniform SLLN holds and we obtain for all $h \in \mathbb{N}_0$ as $n \rightarrow \infty$,

$$\begin{aligned}
& \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} \left| \frac{1}{n} \sum_{i=1}^n G_i^2(\boldsymbol{\theta}) - \mathbb{E} G_1^2(\boldsymbol{\theta}) \right| \xrightarrow{\text{a.s.}} 0 \quad \text{and} \\
& \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} \left| \frac{1}{n} \sum_{i=1}^n G_i^2(\boldsymbol{\theta}) G_{i+h}^2(\boldsymbol{\theta}) - \mathbb{E} G_1^2(\boldsymbol{\theta}) G_{1+h}^2(\boldsymbol{\theta}) \right| \xrightarrow{\text{a.s.}} 0.
\end{aligned} \tag{4.14}$$

Hence (4.12) follows from (4.14) finishing the proof. \square

Remark 4.9. *If the Lévy process $(L_t)_{t \geq 0}$ has finite variation sample paths, then the stochastic integral in (1.2) can be treated pathwise as a Riemann-Stieltjes integral, such that continuity of $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)_{i \in \mathbb{N}}$ follows from Lemma 4.3(c) and dominated convergence. Therefore, Theorem 4.6 is valid for $\boldsymbol{\theta}_0 \in \Theta \supseteq \Theta^{(p(1+\epsilon))}$ for $p > 2$ and some $\epsilon > 0$. Additionally, since the total variation process is also a Lévy process we can use Theorem 66 of Ch. 5 in Protter [35] to show that $\mathbb{E}L_1^4 < \infty$ implies $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} G_i^4(\boldsymbol{\theta}) < \infty$ for all $i \in \mathbb{N}$. Therefore, also Theorem 4.8 is valid for $\boldsymbol{\theta}_0 \in \Theta \supseteq \Theta^{(p(1+\epsilon))}$.*

4.3 Asymptotic normality of the IIE

In order to prove asymptotic normality of the IIE, we need to verify the conditions (b.1), (b.2) and (b.3) of Proposition 2.6. We recall that (b.2) has been proved in Theorem 3.5, and it remains to prove (b.1) and (b.3), which are related to the smoothness of $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$.

4.3.1 Differentiability properties of $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))})$

Condition (b.1), refers to the differentiability of the map $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$. By Remark 4.4 and the chain rule we only need to prove differentiability of $G_i(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ for every $i \in \mathbb{N}$. Since $G_i(\boldsymbol{\theta})$ is defined in terms of a stochastic integral we can not simply interchange the order of the Riemann differentiation and the stochastic integration, however, under appropriate regularity conditions formulated in Hutton and Nelson [16] this is possible.

We start by investigating the candidate for the differential of $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))})$ with $\Theta^{(p(1+\epsilon))}$ as in (4.3), namely the map

$$\boldsymbol{\theta} \mapsto \int_{(i-1)\Delta}^{i\Delta} \nabla_{\boldsymbol{\theta}} \sigma_s(\boldsymbol{\theta}) dL_s := \nabla_{\boldsymbol{\theta}} G_i(\boldsymbol{\theta}). \quad (4.15)$$

We show in Lemma 4.10 that we can find a version of the integral on the rhs, which is continuous on a subset $\Theta^{(2p(1+\epsilon))}$ of $\Theta^{(p(1+\epsilon))}$. Then, Theorem 4.11 asserts that $G_i(\boldsymbol{\theta})$ is differentiable on $\Theta^{(2p(1+\epsilon))}$ and that its differential is indeed given by (4.15).

Lemma 4.10 (Hölder continuity of derivatives). *Assume that $\mathbb{E}|L_1|^{4p(1+\epsilon)} < \infty$ for some $p > 2$ and $\epsilon > 0$. Then there exists a version $(\nabla_{\boldsymbol{\theta}} G_i^{(c)}(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ of the random elements $(\nabla_{\boldsymbol{\theta}} G_i(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ which is Hölder continuous of every order $\gamma \in [0, (p-2)/p)$ on $\Theta^{(2p(1+\epsilon))}$ as defined in (4.3). Additionally, for every $q \in [0, p)$, $i \in \mathbb{N}$, $l \in \{1, 2, 3\}$, and for*

$$V_i = \sup_{\substack{0 < \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| < 1 \\ \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(2p(1+\epsilon))}}} \frac{|\frac{\partial}{\partial \boldsymbol{\theta}_l} G_i(\boldsymbol{\theta}_1) - \frac{\partial}{\partial \boldsymbol{\theta}_l} G_i(\boldsymbol{\theta}_2)|}{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma} \quad (4.16)$$

we have

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}} \left| \frac{\partial}{\partial \boldsymbol{\theta}_l} G_i(\boldsymbol{\theta}) \right|^q < \infty \quad \text{and} \quad \mathbb{E} V_i^q < \infty.$$

Proof. Without loss of generality we consider $i = 1$. Note that in view of Remark 4.5 we can write (4.15) as

$$\left(\frac{1}{2\sqrt{\beta}} \int_0^\Delta \sigma_s((1, \eta, \varphi)) dL_s, \sqrt{\beta} \int_0^\Delta \frac{\partial}{\partial \eta} \sigma_s((1, \eta, \varphi)) dL_s, \sqrt{\beta} \int_0^\Delta \frac{\partial}{\partial \varphi} \sigma_s((1, \eta, \varphi)) dL_s \right)^\top. \quad (4.17)$$

From Remark 4.7 the first component of (4.17) is continuous in β even on Θ . For the remaining two components, we show continuity with respect to (η, ϕ) . Thus, assume that $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(2p(1+\epsilon))}$ with $\beta_1, \beta_2 = 1$.

Using the distributional property of $\sigma_s(\cdot)$ in Lemma 4.1 and the differentiability of $\boldsymbol{\theta} \mapsto \sigma_s(\boldsymbol{\theta})$ in Lemma A.4 gives for every Borel set $B \in \mathcal{B}(\mathbb{R})$ that

$$\begin{aligned} & \mathbb{P}\left(\frac{\partial}{\partial\eta}(\sigma_s(\boldsymbol{\theta}_1) - \sigma_s(\boldsymbol{\theta}_2)) \in B\right) \\ &= \mathbb{P}\left(\lim_{h \rightarrow 0, h \in \mathbb{Q}} \frac{[\sigma_s(\boldsymbol{\theta}_1 + (h, 0)) - \sigma_s(\boldsymbol{\theta}_1)] - [\sigma_s(\boldsymbol{\theta}_2 + (h, 0)) - \sigma_s(\boldsymbol{\theta}_2)]}{h} \in B\right) \\ &= \lim_{h \rightarrow 0, h \in \mathbb{Q}} \mathbb{P}\left(\frac{[\sigma_s(\boldsymbol{\theta}_1 + (h, 0)) - \sigma_s(\boldsymbol{\theta}_1)] - [\sigma_s(\boldsymbol{\theta}_2 + (h, 0)) - \sigma_s(\boldsymbol{\theta}_2)]}{h} \in B\right) \\ &= \lim_{h \rightarrow 0, h \in \mathbb{Q}} \mathbb{P}\left(\frac{[\sigma_0(\boldsymbol{\theta}_1 + (h, 0)) - \sigma_0(\boldsymbol{\theta}_1)] - [\sigma_0(\boldsymbol{\theta}_2 + (h, 0)) - \sigma_0(\boldsymbol{\theta}_2)]}{h} \in B\right) \\ &= \mathbb{P}\left(\frac{\partial}{\partial\eta}(\sigma_0(\boldsymbol{\theta}_1) - \sigma_0(\boldsymbol{\theta}_2)) \in B\right), \end{aligned}$$

so that

$$\frac{\partial}{\partial\eta}(\sigma_s(\boldsymbol{\theta}_1) - \sigma_s(\boldsymbol{\theta}_2)) \stackrel{d}{=} \frac{\partial}{\partial\eta}(\sigma_0(\boldsymbol{\theta}_1) - \sigma_0(\boldsymbol{\theta}_2)). \quad (4.18)$$

Similar calculations show that (4.18) is also valid for $\frac{\partial}{\partial\eta}$ replaced by $\frac{\partial}{\partial\varphi}$. Thus, $(\nabla_{\eta, \varphi} \sigma_s(\boldsymbol{\theta}))_{s \geq 0}$ is stationary and it follows from its differentiability in (A.10) and the mean value theorem that

$$\begin{aligned} \int_0^\Delta \mathbb{E} \|\nabla_{\eta, \varphi} \sigma_s(\boldsymbol{\theta}_1) - \nabla_{\eta, \varphi} \sigma_s(\boldsymbol{\theta}_2)\|^p ds &= \Delta \mathbb{E} \|\nabla_{\eta, \varphi} \sigma_0(\boldsymbol{\theta}_1) - \nabla_{\eta, \varphi} \sigma_0(\boldsymbol{\theta}_2)\|^p \\ &\leq \Delta \left(\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}} \|\nabla_{\eta, \varphi}^2 \sigma_0(\boldsymbol{\theta})\|^p \right) \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^p < \infty, \end{aligned}$$

by Lemma A.5 with $k = 2$. The rest of the proof follows along the same lines as in Theorem 4.6. \square

Theorem 4.11 (Differentiable version of $(G_i(\boldsymbol{\theta}))_{i \in \mathbb{N}}$). *Assume the conditions of Lemma 4.10. Then there is a version $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))})_{i \in \mathbb{N}}$ for $\Theta^{(2p(1+\epsilon))}$ as in (4.3), which is continuously differentiable and its derivative is given a.s. by $(\nabla_{\boldsymbol{\theta}} G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))})_{i \in \mathbb{N}}$.*

Proof. Without loss of generality we consider $i = 1$. From Remark 4.5 it follows that $G_1(\boldsymbol{\theta}) = \sqrt{\beta} G_1((1, \eta, \varphi))$ so that obviously

$$\frac{\partial}{\partial\beta} G_1(\boldsymbol{\theta}) = \frac{1}{2\sqrt{\beta}} G_1((1, \eta, \varphi)) = \int_0^\Delta \frac{\partial}{\partial\beta} \sigma_s(\boldsymbol{\theta}) dL_s.$$

Interchanging the partial differentiation with respect to (η, ϕ) and the stochastic integral requires the four regularity conditions of Theorem 2.2 in Hutton and Nelson [16]. Let $\mathcal{F}_t := \sigma(\{L_s, 0 \leq s \leq t\})$, such that $(\mathcal{F}_t)_{t \geq 0}$ is the filtration generated by the Lévy process L . Condition (i) of that paper is satisfied, since $(\sigma_t(\boldsymbol{\theta}))_{t \geq 0}$ is predictable, we consider the parameter space \mathcal{M} with the Borel σ -algebra, and the parameter $\boldsymbol{\theta}$ is independent of t . Since $\sigma_s(\boldsymbol{\theta}) = \sqrt{\beta} \sigma_s((1, \eta, \varphi))$ these regularity conditions need only to be checked for the map $(\eta, \varphi) \mapsto \sigma_s((1, \eta, \varphi))$. Condition (ii) requires that $\int_0^\Delta \sigma_s^2(\boldsymbol{\theta}) d\langle L \rangle_s < \infty$ a.s. for every $\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}$, where $\langle L \rangle = (\langle L \rangle_s)_{s \geq 0}$ is the

characteristic of the martingale L . Since L is a square integrable Lévy process, $\langle L \rangle_s = s\mathbb{E}L_1^2$ and thus this condition holds since $s \mapsto \sigma_s(\boldsymbol{\theta})$ has bounded sample paths on the compact $[0, \Delta]$. The first part of condition (iii) requires that for every fixed s , the map $\boldsymbol{\theta} \mapsto \sigma_s(\boldsymbol{\theta})$ is absolutely continuous. From the definition of $\sigma_s^2(\boldsymbol{\theta})$ in (2.3) we have for $\beta_1 = \beta_2 = 1$,

$$\sigma_s^2(\boldsymbol{\theta}) = e^{-Y_s - (\boldsymbol{\theta})} \left(\int_0^s e^{Y_v(\boldsymbol{\theta})} dv + \int_0^\infty e^{-Y_v(\boldsymbol{\theta})} dv \right) =: h(\boldsymbol{\theta})(f(\boldsymbol{\theta}) + g(\boldsymbol{\theta})). \quad (4.19)$$

Then for fixed $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(2p(1+\epsilon))}$ we use Lemma A.4 in combination with the mean value theorem and Lemma 4.3(a) to get

$$\begin{aligned} & |\sigma_s^2(\boldsymbol{\theta}_1) - \sigma_s^2(\boldsymbol{\theta}_2)| \quad (4.20) \\ & \leq \left| (f(\boldsymbol{\theta}_1) + g(\boldsymbol{\theta}_1)) \left(h(\boldsymbol{\theta}_1) - h(\boldsymbol{\theta}_2) + h(\boldsymbol{\theta}_2)(f(\boldsymbol{\theta}_1) - f(\boldsymbol{\theta}_2) + g(\boldsymbol{\theta}_1) - g(\boldsymbol{\theta}_2)) \right) \right| \\ & \leq (h(\boldsymbol{\theta}_2) + f(\boldsymbol{\theta}_1) + g(\boldsymbol{\theta}_1)) \left(|h(\boldsymbol{\theta}_1) - h(\boldsymbol{\theta}_2)| + |f(\boldsymbol{\theta}_1) - f(\boldsymbol{\theta}_2)| + |g(\boldsymbol{\theta}_1) - g(\boldsymbol{\theta}_2)| \right) \\ & \leq \sup_{\boldsymbol{\theta} \in \Theta} \{h(\boldsymbol{\theta}) + f(\boldsymbol{\theta}) + g(\boldsymbol{\theta})\} \left(|h(\boldsymbol{\theta}_1) - h(\boldsymbol{\theta}_2)| + |f(\boldsymbol{\theta}_1) - f(\boldsymbol{\theta}_2)| + |g(\boldsymbol{\theta}_1) - g(\boldsymbol{\theta}_2)| \right) \\ & \leq \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \sum_{j=1}^N \left\{ \sup_{\boldsymbol{\theta} \in \Theta} (h(\boldsymbol{\theta}) + f(\boldsymbol{\theta}) + g(\boldsymbol{\theta})) \right\} \left\{ e^{-Y_s(\boldsymbol{\theta}_j^*)} (s + K_s(\varphi_*)) \right. \\ & \quad \left. + \int_0^s e^{Y_v(\boldsymbol{\theta}_j^*)} (v + K_v(\varphi_*)) dv + \int_0^\infty e^{-Y_v(\boldsymbol{\theta}_j^*)} (v + K_v(\varphi_*)) dv \right\}, \end{aligned}$$

where $(\boldsymbol{\theta}_j^*)_{j=1}^N \in \mathcal{M}^{(2p(1+\epsilon))}$. Since Θ is compact and for each fixed $s \geq 0$, $\boldsymbol{\theta} \mapsto \sigma_s(\boldsymbol{\theta})$ is continuous, $\sup_{\boldsymbol{\theta} \in \Theta} \{h(\boldsymbol{\theta}) + f(\boldsymbol{\theta}) + g(\boldsymbol{\theta})\}$ is finite. Furthermore, Lemma A.3 implies that the other three random variables at the right-hand side of (4.20) have finite first moment, and are therefore also a.s. finite. Thus (4.20) implies that the map $\boldsymbol{\theta} \mapsto \sigma_s^2(\boldsymbol{\theta})$ is a.s. Lipschitz continuous on $\Theta^{(2p(1+\epsilon))}$ and, as a consequence, absolutely continuous on $\Theta^{(2p(1+\epsilon))}$. For the second part of condition (iii) we recall first that we have assumed that $\beta = 1$, such that we focus on the partial differentiation of the parameter $(\eta, \varphi)^\top$. A non-decreasing predictable process $(\lambda_t)_{t \geq 0}$ is needed such that for every t and $\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}$

$$\int_0^t \|\nabla_{\eta, \varphi} \sigma_s(\boldsymbol{\theta})\|^2 d\langle L \rangle_s < \lambda_t, \quad \text{a.s.}$$

From (4.19), the product rule and Proposition 2 in [24] we find

$$\|\nabla_{\eta, \varphi} \sigma_s(\boldsymbol{\theta})\| \leq \frac{1}{2\sigma^*} \{ \|\nabla_{\eta, \varphi} h(\boldsymbol{\theta})\| (f(\boldsymbol{\theta}) + g(\boldsymbol{\theta})) + h(\boldsymbol{\theta}) \|\nabla_{\eta, \varphi} f(\boldsymbol{\theta}) + \nabla_{\eta, \varphi} g(\boldsymbol{\theta})\| \}. \quad (4.21)$$

We use Lemma A.4 and the definition of the process $(Y_t(\boldsymbol{\theta}))_{t \geq 0}$ in (2.1). First note that

$$\eta \leq \sup\{\eta > 0 : (\beta, \eta, \varphi) \in \Theta\} =: \eta^* < \infty, \quad \varphi \geq \inf\{\varphi > 0 : (\beta, \eta, \varphi) \in \Theta\} =: \varphi_* > 0 \quad (4.22)$$

and we get the bound

$$f(\boldsymbol{\theta}) = \int_0^s e^{Y_v(\boldsymbol{\theta})} dv = \int_0^s \exp \left\{ \eta v - \sum_{0 < u \leq s} \log(1 + \varphi(\Delta L_u)^2) \right\} dv \leq s e^{\eta^* s}. \quad (4.23)$$

Hence it follows from (4.19) that

$$\|\nabla_{\eta,\varphi}f(\boldsymbol{\theta})\| = \int_0^s ve^{Y_v(\boldsymbol{\theta})}dv + \int_0^s e^{Y_v(\boldsymbol{\theta})}K_v(\varphi)dv \leq (s + K_s(\varphi_*))se^{\eta^*s}, \quad (4.24)$$

$$\|\nabla_{\eta,\varphi}h(\boldsymbol{\theta})\| = se^{-Y_{s-}(\boldsymbol{\theta})} + e^{-Y_{s-}(\boldsymbol{\theta})}K_s(\varphi) \leq e^{-Y_s(\boldsymbol{\theta})}(s + K_s(\varphi_*)), \quad (4.25)$$

$$\|\nabla_{\eta,\varphi}g(\boldsymbol{\theta})\| = \int_0^\infty ve^{-Y_v(\boldsymbol{\theta})}dv + \int_0^\infty e^{-Y_v(\boldsymbol{\theta})}K_v(\varphi)dv \leq \int_0^\infty (v + K_v(\varphi_*))e^{-Y_v(\boldsymbol{\theta})}dv \quad (4.26)$$

From (4.21) and the bounds given in (4.23), (4.24), (4.25), and (4.26) we obtain

$$\begin{aligned} \|\nabla_{\eta,\varphi}\sigma_s(\boldsymbol{\theta})\| &\leq \frac{1}{2\sigma^*}e^{-Y_s(\boldsymbol{\theta})}(s + K_s(\varphi_*))\left(se^{\eta^*s} + \int_0^\infty e^{-Y_v(\boldsymbol{\theta})}dv\right) \\ &+ \frac{1}{2\sigma^*}e^{-Y_{s-}(\boldsymbol{\theta})}\left\{(s + K_s(\varphi_*))se^{\eta^*s} + \int_0^\infty e^{-Y_v(\boldsymbol{\theta})}(v + K_v(\varphi_*))dv\right\} =: l_s(\boldsymbol{\theta}). \end{aligned} \quad (4.27)$$

Using the compactness of $\Theta^{(2p(1+\epsilon))}$, (4.27) and Lemma 4.3(a) gives

$$\sup_{(\eta,\varphi) \in \Theta^{(2p(1+\epsilon))}} \|\nabla_{\eta,\varphi}\sigma_s(\boldsymbol{\theta})\| \leq \sup_{(\eta,\varphi) \in \Theta^{(2p(1+\epsilon))}} l_s(\boldsymbol{\theta}) \leq \sum_{j=1}^N l_s(\boldsymbol{\theta}_j^*),$$

where $(\boldsymbol{\theta}_j^*)_{j=1}^N$ in $\mathcal{M}^{(2p(1+\epsilon))}$. Thus,

$$\int_0^t \|\nabla_{\eta,\varphi}\sigma_s(\boldsymbol{\theta})\|^2 d\langle L \rangle_s < 1 + \mathbb{E}L_1^2 \int_0^t \left| \sum_{j=1}^N l_s(\boldsymbol{\theta}_j^*) \right|^2 ds := \lambda_t, \quad 0 \leq t \leq \Delta,$$

which is a well defined process. Since $(\lambda_t)_{t \geq 0}$ is adapted to the filtration $(\mathcal{F}_t)_{t \geq 0}$ and continuous, it is predictable. The fourth regularity condition we need to check is that the maps

$$\boldsymbol{\theta} \mapsto \int_0^\Delta \sigma_s(\boldsymbol{\theta})dL_s \quad \text{and} \quad \boldsymbol{\theta} \mapsto \int_0^\Delta \nabla_{\eta,\varphi}\sigma_s(\boldsymbol{\theta})dL_s$$

are continuous, which has been proved in Theorem 4.6 and Lemma 4.10. This concludes the proof. \square

Remark 4.12. In view of Theorem 4.11 we will from now on work with returns $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))})_{i \in \mathbb{N}}$ with $\Theta^{(2p(1+\epsilon))}$ as in (4.3), which are continuously differentiable with

$$\nabla_{\boldsymbol{\theta}}G_i(\boldsymbol{\theta}) = \int_{(i-1)\Delta}^{i\Delta} \nabla_{\boldsymbol{\theta}}\sigma_s(\boldsymbol{\theta})dL_s, \quad i \in \mathbb{N}.$$

As a consequence, also the map $\boldsymbol{\theta} \mapsto \hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta})$ is continuously differentiable on $\Theta^{(2p(1+\epsilon))}$, hence, condition (b.1) of Proposition 2.6 holds.

4.3.2 Convergence of the derivatives

Finally, we prove condition (b.3) of Proposition 2.6

Proposition 4.13 (Consistency of the derivatives). *Assume that $\mathbb{E}|L_1|^{4p(1+\epsilon)} < \infty$ for some $p > 2/5$ and $\epsilon > 0$. Let $\hat{\boldsymbol{\pi}}_n$ be one of the estimators $\hat{\boldsymbol{\pi}}_{n,\text{LS}}$ and $\hat{\boldsymbol{\pi}}_{n,\text{YW}}$ from Definition 2.3. Then for every sequence $(\boldsymbol{\theta}_n)_{n \in \mathbb{N}} \subset \Theta^{(2p(1+\epsilon))}$ and $\boldsymbol{\theta}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$ we have $\nabla_{\boldsymbol{\theta}}\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}_n) \xrightarrow{P} \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}_{\boldsymbol{\theta}_0}$ as $n \rightarrow \infty$.*

Proof. Recall from Remark 4.4 that we can write the two estimators $\hat{\boldsymbol{\pi}}_{n,\text{LS}}$ and $\hat{\boldsymbol{\pi}}_{n,\text{YW}}$ as a continuously differentiable map $g : \mathbb{R}^{r+2} \rightarrow \mathbb{R}^{r+2}$, whose Jacobi matrix exists and all partial derivatives of g are continuous. Hence, $\hat{\boldsymbol{\pi}}_n(\boldsymbol{\theta}) = (g_1(f_n(\boldsymbol{\theta})), \dots, g_{r+2}(f_n(\boldsymbol{\theta})))^\top$ for $\boldsymbol{\theta} = (\beta, \eta, \varphi) =: (\theta_1, \theta_2, \theta_3)$, and we obtain for the partial derivatives by the chain rule

$$\frac{\partial}{\partial \theta_l} g_k(f_n(\boldsymbol{\theta})) = \left(\frac{\partial g_k(f_n(\boldsymbol{\theta}))}{\partial x_1}, \dots, \frac{\partial g_k(f_n(\boldsymbol{\theta}))}{\partial x_{r+2}} \right) \left(\frac{\partial}{\partial \theta_l} f_n(\boldsymbol{\theta}) \right). \quad (4.28)$$

for every $l = 1, 2, 3$ and $k = 1, \dots, r+2$. By the continuous mapping theorem and (4.28) it suffices to prove that as $n \rightarrow \infty$,

$$f_n(\boldsymbol{\theta}_n) \xrightarrow{P} f(\boldsymbol{\theta}_0) \quad \text{and} \quad \frac{\partial}{\partial \theta_j} f_n(\boldsymbol{\theta}_n) \xrightarrow{P} \frac{\partial}{\partial \theta_l} f(\boldsymbol{\theta}_0), \quad l = 1, 2, 3.$$

Let $l \in \{1, 2, 3\}$ be fixed. It follows from (4.14) and from Lemma B.2 that as $n \rightarrow \infty$,

$$\sup_{\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}} \|f_n(\boldsymbol{\theta}) - f(\boldsymbol{\theta})\| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}} \left\| \frac{\partial}{\partial \theta_l} f_n(\boldsymbol{\theta}) - \frac{\partial}{\partial \theta_l} f(\boldsymbol{\theta}) \right\| \xrightarrow{P} 0. \quad (4.29)$$

Since

$$\begin{aligned} \|f_n(\boldsymbol{\theta}_n) - f(\boldsymbol{\theta}_0)\| &\leq \|f_n(\boldsymbol{\theta}_n) - f(\boldsymbol{\theta}_n)\| + \|f(\boldsymbol{\theta}_n) - f(\boldsymbol{\theta}_0)\| \\ &\leq \sup_{\Theta^{(2p(1+\epsilon))}} \|f_n(\boldsymbol{\theta}) - f(\boldsymbol{\theta})\| + \|f(\boldsymbol{\theta}_n) - f(\boldsymbol{\theta}_0)\|, \end{aligned} \quad (4.30)$$

from the continuity of f on $\Theta^{(2p(1+\epsilon))}$, the fact that $\boldsymbol{\theta}_n \xrightarrow{P} \boldsymbol{\theta}_0$, and (4.29) it follows that $f_n(\boldsymbol{\theta}_n) \xrightarrow{P} f(\boldsymbol{\theta}_0)$. Similar calculations as in (4.30) show that

$$\frac{\partial}{\partial \theta_l} f_n(\boldsymbol{\theta}_n) \xrightarrow{P} \frac{\partial}{\partial \theta_l} f(\boldsymbol{\theta}_0)$$

concluding the proof. □

We are now ready to state asymptotic normality of the IIE.

Theorem 4.14 (Asymptotic normality of the IIE). *Assume that $\mathbb{E}|L_1|^{4p(1+\epsilon)} < \infty$ for some $p > 2/5$ and $\epsilon > 0$. Let $(G_i(\boldsymbol{\theta}_0))_{i=1}^n$ be the returns (1.2) with true parameter $\boldsymbol{\theta}_0$ is an element of the interior of $\Theta^{(2p(1+\epsilon))}$ as defined in (4.3). Suppose that the auxiliary AR(r) model for $r \geq 2$ is estimated by the LSE or the YWE of Definition 2.3. If the matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}$ defined in Theorem 3.5 is positive definite and $\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}(\boldsymbol{\theta}_0)$ has full column rank 3, then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{n,\text{II}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \Xi_{\boldsymbol{\theta}_0}), \quad n \rightarrow \infty,$$

where $\Xi_{\boldsymbol{\theta}_0}$ is defined in (2.18).

Proof. The asymptotic normality follows from Proposition 2.6. Since $\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))} \subseteq \Theta^{(p(1+\epsilon))} \subseteq \Theta$, Theorem 4.8 implies condition (a). Conditions (b.1) and (b.3) are valid by Proposition 4.13 and the fact that $\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}(\boldsymbol{\theta}_0)$ has full column rank 3. Furthermore, (b.2) holds by Theorem 3.5, since $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}$ is positive definite. □

Remark 4.15 (Estimation of the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_{n,\Pi}$). *The asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_{n,\Pi}$ given in Theorem 4.14 depends on $K, \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}_{\boldsymbol{\theta}_0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}$ and Ω . Using the map $\boldsymbol{\theta} \mapsto \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}_{\boldsymbol{\theta}}$ from (2.11) we compute $\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}_{\hat{\boldsymbol{\theta}}_{n,\Pi}}$. An application of the continuous mapping Theorem in combination with the continuity of $\boldsymbol{\theta} \mapsto \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}_{\boldsymbol{\theta}}$ and Theorem 4.8 gives $\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}_{\hat{\boldsymbol{\theta}}_{n,\Pi}} \xrightarrow{\text{a.s.}} \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}_{\boldsymbol{\theta}_0}$. Recall $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}$ as in (3.9) which depends on the inverse of the autocovariance function $\Gamma_{\boldsymbol{\theta}_0}$ and on $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}^*$ as defined in (3.10). A strongly consistent estimator of $\Gamma_{\boldsymbol{\theta}_0}$ is given by $\Gamma_{\hat{\boldsymbol{\theta}}_{n,\Pi}}$. Let \hat{C}_k be as in (3.11) with W_k replaced by $G_k^2(\boldsymbol{\theta}_0)$ and $\boldsymbol{\pi} = (\mu, \mathbf{a}, \gamma(0))$ replaced by $\hat{\boldsymbol{\pi}}_n$. Then we estimate $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}^*$ by*

$$\hat{\mu}_{n,C_1,C_1^T} + 2 \sum_{i=1}^{n-r-1} \hat{\mu}_{n,C_1,C_{1+i}^T},$$

where

$$\hat{\mu}_{n,C_1,C_{1+i}^T} = \frac{1}{n-i-r} \sum_{k=1}^{n-i-r} C_k C_{k+i}^T, \quad i = 0, \dots, n-r-1.$$

Remark 4.16. *If the Lévy process $(L_t)_{t \geq 0}$ has finite variation sample paths, then the stochastic integral in (1.2) can be treated pathwise as a Riemann-Stieltjes integral, such that continuous differentiability of $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)_{i \in \mathbb{N}}$ follows by dominated convergence with dominating function as in (4.27). Therefore, Theorem 4.11 is valid for $\boldsymbol{\theta}_0 \in \Theta \supseteq \Theta^{(2p(1+\epsilon))}$ for some $p > 2$ and $\epsilon > 0$. Additionally, since the total variation process is also a Lévy process we can use Theorem 66 of Ch. 5 in Protter [35] to show that, if $\mathbb{E}L_1^{8+\delta} < \infty$ for some $\delta > 0$, then $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} \|\nabla_{\boldsymbol{\theta}} G_i(\boldsymbol{\theta})\|^{4+\delta/2} < \infty$ for all $i \in \mathbb{N}$. This combined with Remark 4.9 and a dominated convergence argument can be applied to show that Lemma 4.13 is valid for $\boldsymbol{\theta}_0 \in \Theta^{(2p(1+\epsilon))}$, and, as a consequence, also Theorem 4.14.*

5 Simulation study

The data used for estimation is a sample of COGARCH squared returns $\mathbf{G}_n^2 = (G_i^2(\boldsymbol{\theta}_0))_{i=1}^n$ as defined in (1.2) with true parameter value $\boldsymbol{\theta}_0 \in \Theta$ as in (2.7) observed on a fixed grid of size $\Delta = 1$. We choose a pure jump Variance Gamma (VG) process as the Lévy process, which has infinite activity and has been used successfully for modeling stock prices (see Haug et al. [15] and reference therein). The Lévy measure of the VG process with parameter $C > 0$ has Lebesgue density

$$\nu_L(dx) = \frac{C}{|x|} \exp\{-(2C)^{1/2}|x|\} dx, \quad x \neq 0. \quad (5.1)$$

The Indirect Inference method of Gourieroux et al. [12] based on simulations was originally proposed to estimate models where the binding function is difficult or impossible to compute. However, the binding function $\boldsymbol{\theta} \mapsto \boldsymbol{\pi}_{\boldsymbol{\theta}}$ from Proposition 2.4 can be computed explicitly from the formulas given in Theorem 3.1 of Haug et al. [15] and the Yule-Walker equations in (2.10), leading to the IIE

$$\hat{\boldsymbol{\theta}}_{n,\Pi^*} := \arg \min_{\boldsymbol{\theta} \in \Theta} \|\hat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_{\boldsymbol{\theta}}\|_{\Omega}. \quad (5.2)$$

We perform a simulation study to evaluate the finite sample performance of the IIE $\hat{\boldsymbol{\theta}}_{n,\Pi^*}$ in (5.2) and also to compare it with the method of moments (MM) estimator $\hat{\boldsymbol{\theta}}_{n,\text{MM}}$ (Algorithm 1 in Haug

et al. [15]) and the optimal prediction based (OPB) estimator $\hat{\theta}_{n,\text{OPB}}$ (equation (7) in Bibbona and Negri [4]). As in the simulation studies in [4, 15], we take the VG process with true parameter value $\theta_0 = (0.04, 0.053, 0.038)$ and $C = 1$ in (5.1), which implies $\Psi_{\theta_0}(4) = -0.0261 < 0$. Under these conditions, all three estimators $\hat{\theta}_{n,\text{MM}}$ (Theorem 3.8 in [15]), $\hat{\theta}_{n,\text{OPB}}$ (Theorem 3.1 in [4]) and IIE $\hat{\theta}_{n,\text{II}^*}$ (Theorem 4.8) are consistent.

The MM is based on r empirical autocovariances, OPB based on r predictors, and IIE based on an $\text{AR}(r)$ auxiliary model. Inspection of several empirical autocovariance functions of the squared returns \mathbf{G}_n^2 with $n = 10\,000$ revealed $r = 70$ as a suitable number of lags in most of the cases. Since we have to fix r in a simulation study, we choose $r = 70$ for all three estimators.

We compare the three estimators in a simulation study in Section 5.1. Then we show in Section 5.2 how the IIE based on simulations defined in (2.16) can reduce the finite sample bias of $\hat{\theta}_{n,\text{II}^*}$ considerably. Finally, to understand how the condition $\Psi_{\theta_0}(4) < 0$ affects the estimation, we investigate the finite sample bias of both IIEs with two different true parameter values $\theta_0^{(1)}$ and $\theta_0^{(2)}$ satisfying $\Psi_{\theta_0^{(2)}}(4) < \Psi_{\theta_0}(4) < \Psi_{\theta_0^{(1)}}(4) < 0$, where $\Psi_{\theta_0^{(1)}}(4)$ is near zero.

5.1 Simulation results

The computations were performed using the R software (R Core Team [36]). Simulation of the COGARCH process and computation of $\hat{\theta}_{n,\text{MM}}$ and $\hat{\theta}_{n,\text{OPB}}$ are performed with the COGARCH R package from Bibbona et al. [5] (see also the YUIMA R package in Iacus et al. [17] for the simulation and estimation of higher order COGARCH models). We first compute $\hat{\theta}_{n,\text{MM}}$ based on the sample \mathbf{G}_n^2 . The estimators $\hat{\theta}_{n,\text{OPB}}$ and $\hat{\theta}_{n,\text{II}^*}$ are computed via the optimization routine *optim* in R, which requires an initial parameter value and we take $\hat{\theta}_{n,\text{MM}}$. To compute $\hat{\pi}_n$ in (5.2) we use the YWE from Definition 2.3 and take the identity matrix for Ω to compute $\hat{\theta}_{n,\text{II}^*}$. In principle, there is an optimal choice of Ω (see Remark 3 of de Luna and Genton [8] and Prop. 4 of Gouriéroux et al. [12]). It depends on the covariance matrix Σ of the auxiliary model in (3.9) (see also Remark 4.24(b) in [10]). This matrix depends on an infinite series and on covariances between COGARCH returns to the powers 2,4,6 and 8, and has no explicit expression. According to Remark 3 of [8] and empirical evidence reported on p. S97f of Gouriéroux et al. [12] the gain of efficiency when using the optimal weight matrix is negligible, so that we only consider estimators based on the identity matrix for Ω .

We focus on the YWE for the auxiliary model, a comparison including the LSE will be given in Do Rêgo Sousa [9]. The estimator $\hat{\theta}_{n,\text{OPB}}$ only returns a result when $\Psi_{\hat{\theta}_{n,\text{OPB}}}(4) < 0$. The estimators $\hat{\theta}_{n,\text{MM}}$ and $\hat{\theta}_{n,\text{II}^*}$ always return a value. The results are given in Table 5.1, where we excluded those paths for which the condition $\Psi_{\hat{\theta}_n}(4) < 0$ is not satisfied for at least one of the estimators compared here. The results are based on 1 000 independent samples of COGARCH squared returns.

The results in Table 5.1 for the estimators $\hat{\theta}_{n,\text{MM}}$ and $\hat{\theta}_{n,\text{OPB}}$ are similar to those of Table 2 in Bibbona and Negri [4]. The OPB estimator has the smallest RMSE. The MM has the smallest relative bias for the parameter φ , and the OPB the smallest for β and ϕ . The estimator $\hat{\theta}_{n,\text{II}^*}$ performed similarly to $\hat{\theta}_{n,\text{MM}}$, but it has a large bias for the parameters β and φ . This is probably due to the fact that $\hat{\theta}_{n,\text{II}^*}$ depends on $\hat{\pi}_n$, which is a biased estimator of π even for AR models

		Mean	Std	RMSE	RB
	$\hat{\beta}$	0.04698	0.02032	0.02148	0.17457
$\hat{\theta}_{n,\text{II}^*}$	$\hat{\eta}$	0.05038	0.01482	0.01504	-0.04939
	$\hat{\varphi}$	0.03243	0.00994	0.01139	-0.14663
	$\hat{\beta}$	0.05226	0.01805	0.02182	0.30658
$\hat{\theta}_{n,\text{MM}}$	$\hat{\eta}$	0.05662	0.01576	0.01616	0.06827
	$\hat{\varphi}$	0.03667	0.01023	0.01031	-0.03513
	$\hat{\beta}$	0.04439	0.01609	0.01667	0.10965
$\hat{\theta}_{n,\text{OPB}}$	$\hat{\eta}$	0.05274	0.01317	0.01317	-0.00489
	$\hat{\varphi}$	0.03583	0.00815	0.00843	-0.05712
	$\hat{\beta}$	0.04204	0.02032	0.02041	0.05105
$\hat{\theta}_{n,\text{II}}$	$\hat{\eta}$	0.05318	0.01623	0.01622	0.00336
	$\hat{\varphi}$	0.03661	0.00955	0.00965	-0.03661

Table 5.1: Performance assessment based on 1000 independent samples of COGARARCH squared returns \mathbf{G}_n^2 for $n = 10\,000$, sampled with parameter values $\beta_0 = 0.04$, $\eta_0 = 0.053$ and $\varphi_0 = 0.038$: mean, standard deviation (Std), root mean squared error (RMSE) and relative bias (RB). Both IIEs $\hat{\theta}_{n,\text{II}^*}$ in (5.2) and $\hat{\theta}_{n,\text{II}}$ in (5.3) used the identity matrix for Ω . The IIE $\hat{\theta}_{n,\text{II}}$ is based on $K = 100$ simulated paths.

with i.i.d. noise as shown in Shaman and Stine [39]. The auxiliary AR model from Proposition 2.2 has stationary and ergodic residuals, and certainly $\hat{\pi}_n$ has a bias, which propagates to the IIE. As a remedy, we use the IIE based on simulations and show that it can reduce the bias of $\hat{\theta}_{n,\text{II}^*}$, it also outperforms $\hat{\theta}_{n,\text{MM}}$ and $\hat{\theta}_{n,\text{OPB}}$.

5.2 Finite sample bias

In Gourieroux et al. [13, 14] it is shown that Indirect Inference based on simulations can reduce the finite sample bias considerably, in particular, when the bias originates from the estimator of the auxiliary model. The idea of the bias reduction is that the IIE

$$\hat{\theta}_{n,\text{II}} = \arg \min_{\theta \in \Theta} \left\| \hat{\pi}_n - \frac{1}{K} \sum_{k=1}^K \hat{\pi}_{n,k}(\theta) \right\|_{\Omega}, \quad K \in \mathbb{N}, \quad (5.3)$$

from Definition 2.5 finds a $\theta \in \Theta$ which minimizes the distance between two biased estimators, $\hat{\pi}_n$ and $\frac{1}{K} \sum_{k=1}^K \hat{\pi}_{n,k}(\theta)$. As they have a similar bias, they have a chance to cancel. We proceed to investigate the finite sample performance of the estimator $\hat{\theta}_{n,\text{II}}$ in (5.3).

According to [14], the number of simulated paths K in (5.3) has to be large enough to ensure that $\mathbb{E} \hat{\pi}_n(\theta)$ is well approximated by $\frac{1}{K} \sum_{k=1}^K \hat{\pi}_{n,k}(\theta)$ for all θ appearing in the optimization algorithm. Furthermore, the asymptotic variance of the IIE decreases with K (see Eq. (2.19)). To compute $\hat{\theta}_{n,\text{II}}$ we need to evaluate the function

$$\theta \mapsto \frac{1}{K} \sum_{k=1}^K \hat{\pi}_{n,k}(\theta) \quad (5.4)$$

for all θ giving a representation of the parameter space. To compute (5.4) for a fixed θ , we simulate K independent samples $\mathbf{G}_{n,k}(\theta) := (G_i^{(k)}(\theta))_{i=1}^n$ for $k = 1, \dots, K$. For different θ we

$n = 5\,000$					
		Mean	Std	RMSE	RB
	$\hat{\beta}$	0.04710	0.02196	0.02307	0.17738
$\hat{\boldsymbol{\theta}}_{n,\text{II}^*}$	$\hat{\eta}$	0.04977	0.02036	0.02061	-0.06094
	$\hat{\varphi}$	0.03168	0.01317	0.01461	-0.16637
	$\hat{\beta}$	0.04999	0.03228	0.03377	0.24968
$\hat{\boldsymbol{\theta}}_{n,\text{II}}$	$\hat{\eta}$	0.05935	0.02458	0.02538	0.11974
	$\hat{\varphi}$	0.03990	0.01379	0.01391	0.04989
$n = 7\,500$					
		Mean	Std	RMSE	RB
	$\hat{\beta}$	0.05093	0.02015	0.02291	0.27323
$\hat{\boldsymbol{\theta}}_{n,\text{II}^*}$	$\hat{\eta}$	0.05401	0.01786	0.01788	0.01896
	$\hat{\varphi}$	0.03439	0.01158	0.01212	-0.09502
	$\hat{\beta}$	0.04181	0.02375	0.02381	0.04537
$\hat{\boldsymbol{\theta}}_{n,\text{II}}$	$\hat{\eta}$	0.05322	0.01897	0.01896	0.00408
	$\hat{\varphi}$	0.03668	0.01093	0.01101	-0.03487

Table 5.2: Performance assessment based on 1 000 independent samples of COGARCH squared returns \mathbf{G}_n^2 for $n = 5\,000$ and $n = 7\,500$, sampled with parameter values $\beta_0 = 0.04$, $\eta_0 = 0.053$ and $\varphi_0 = 0.038$: mean, standard deviation (Std), root mean squared error (RMSE) and relative bias (RB). Both IIEs $\hat{\boldsymbol{\theta}}_{n,\text{II}^*}$ in (5.2) and $\hat{\boldsymbol{\theta}}_{n,\text{II}}$ in (5.3) used the identity matrix for $\boldsymbol{\Omega}$. The IIE $\hat{\boldsymbol{\theta}}_{n,\text{II}}$ is based on $K = 100$ simulated paths.

use the same pseudo-random numbers to generate the K independent samples, which turns (5.4) into a deterministic function of $\boldsymbol{\theta}$ and thus suitable for optimization.

In order to save computation time when computing (5.4) we use for every simulated path the fact that $\mathbf{G}_{n,k}(\boldsymbol{\theta}) = \sqrt{\beta}\mathbf{G}_{n,k}((1, \eta, \phi))$ (see Remark 4.5) and thus it follows from Definition 2.3 that

$$\hat{\boldsymbol{\pi}}_{n,k}(\boldsymbol{\theta}) = \begin{pmatrix} \hat{\mu}_n(\boldsymbol{\theta}) \\ \hat{\boldsymbol{a}}_{n,k}(\boldsymbol{\theta}) \\ \hat{\gamma}_{n,k}(0; \boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \beta\hat{\mu}_{n,k}((1, \eta, \phi)) \\ \hat{\boldsymbol{a}}_{n,k}((1, \eta, \phi)) \\ \beta^2\hat{\gamma}_{n,k}(0; (1, \eta, \phi)) \end{pmatrix}. \quad (5.5)$$

As it is computationally impossible to perform the optimization (5.3) for all $\boldsymbol{\theta} \in \Theta$, we have to restrict Θ in a reasonable way, and we restrict Θ to values in the set

$$\Theta_{\text{rest}} := \{\boldsymbol{\theta} \in \Theta : \Psi_{\boldsymbol{\theta}}(4) < 0, \boldsymbol{\theta} \in (0, \hat{\beta}_{\text{max}}) \times (0, \hat{\eta}_{\text{max}}) \times (0, \hat{\varphi}_{\text{max}})\}$$

where $\hat{\beta}_{\text{max}}$, $\hat{\eta}_{\text{max}}$ and $\hat{\varphi}_{\text{max}}$ are upper bounds for the estimated parameters from Table 5.1 for all 1 000 independent samples \mathbf{G}_n^2 and all estimators.

For $K = 100$ and $n = 10\,000$, every evaluation of (5.4) takes approximately 13 minutes on a personal computer. The next goal would be to evaluate (5.3) using a gradient based routine. This is out of reach with respect to computation time. As a remedy we adopt the strategy of precomputing (5.4) on a fine grid $\Theta_{\text{grid}} \subset \Theta_{\text{rest}}$. The set Θ_{grid} was created by generating an equally spaced grid on Θ_{rest} with componentwise distance for the parameters η and φ equal to

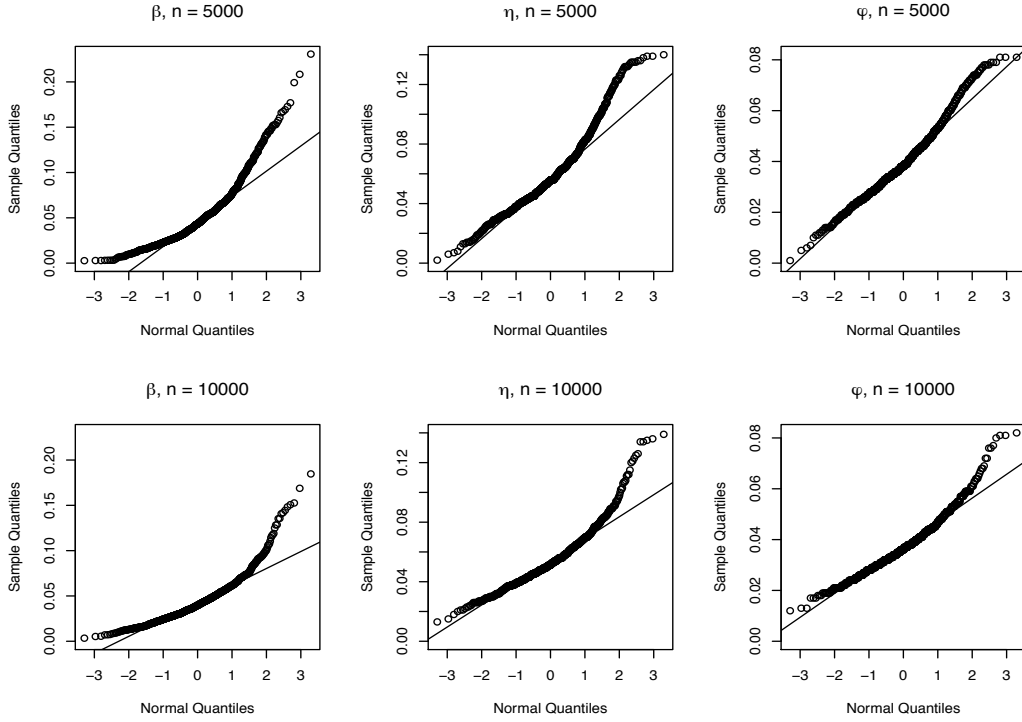


Figure 1: QQ plots of the estimators $\hat{\theta}_{n,\text{II}}$ of θ_0 as in Table 5.1 for $n = 5000$ (top line) and $n = 10000$ (bottom line).

0.001 (resulting in about 6.000 different points). The grid for the component β was then created with spacing 0.001, but without the need to simulate the COGARCH path again by using the relation in (5.5). Afterwards, with COGARCH returns \mathbf{G}_n^2 generated independently from the samples $\mathbf{G}_{n,k}(\theta)$, $k = 1, \dots, K$, applied to compute (5.4), we compute $\hat{\pi}_n$, and the estimator $\hat{\theta}_{n,\text{II}}$ is then simply given by

$$\arg \min_{\theta \in \Theta_{\text{grid}}} \left\| \hat{\pi}_n - \frac{1}{K} \sum_{k=1}^K \hat{\pi}_{n,k}(\theta) \right\|_{\Omega},$$

where we choose the identity matrix for Ω . The results are presented in the bottom line of Table 5.1. We notice a significant bias reduction for the simulation based estimator $\hat{\theta}_{n,\text{II}}$ compared to $\hat{\theta}_{n,\text{II}}^*$. The standard deviation of the estimator for η is slightly larger for $\hat{\theta}_{n,\text{II}}$, but this is expected since the simulations increase the asymptotic variance by a factor of $(1 + \frac{1}{K})$ as can be seen from (2.19). The relative bias of $\hat{\theta}_{n,\text{II}}$ is also smaller than those of the estimators $\hat{\theta}_{n,\text{MM}}$ and $\hat{\theta}_{n,\text{OPB}}$. Since the standard deviations of the components of $\hat{\theta}_{n,\text{II}}$ are larger than for those of $\hat{\theta}_{n,\text{OPB}}$ and the bias reduction is comparable for the parameters η and φ , the RMSE does not seem to improve, even though the bias of $\hat{\theta}_{n,\text{II}}$ is smaller.

We also compare the performance of the IIE with and without simulation for different sample sizes n with θ_0 as in Table 5.1. The results are given in Table 5.2. For $n = 5000$ we only observe a bias reduction of $\hat{\theta}_{n,\text{II}}$ for $\hat{\eta}$, whereas the bias reduction of $\hat{\theta}_{n,\text{II}}$ is noticeable for all three components already for $n = 7500$ and of course for $n = 10000$; cf. Table 5.1.

We also can see in Figure 1 that for $n = 5000$ and $n = 10000$ the asymptotic normality of

$\hat{\theta}_{n,\text{II}}$ has not yet been reached, although some improvement for growing sample sizes is visible in the QQ plots of $\hat{\beta}$ and $\hat{\eta}$, however not for $\hat{\varphi}$.

To clarify if the bias reduction of $\hat{\theta}_{n,\text{II}}$ depends on the choice of the true parameter values we perform a simulation study with two different values: $\theta_0^{(1)} = (0.04, 0.051, 0.040)$ and $\theta_0^{(2)} = (0.04, 0.055, 0.036)$. Both values are in the stationarity region with

$$\Psi_{\theta_0^{(1)}}(4) = -0.0060, \quad \Psi_{\theta_0}(4) = -0.0261, \quad \Psi_{\theta_0^{(2)}}(4) = -0.0460.$$

The results are presented in Table 5.3. As for θ_0 in Table 5.1, they also show significant bias reduction for both values for the estimator $\hat{\theta}_{n,\text{II}}$ based on simulations, when compared to $\hat{\theta}_{n,\text{II}^*}$. However, the bias for $\hat{\beta}^{(1)}$ is much higher than for $\hat{\beta}$ and $\hat{\beta}^{(2)}$ reflecting the fact that $\Psi_{\theta_0^{(1)}}(4)$ is very close to zero. The estimators $\hat{\eta}^{(1)}$ and $\hat{\varphi}^{(1)}$ seem to be robust with respect to this fact. Additionally, the relative biases for $\hat{\beta}^{(2)}$ and $\hat{\varphi}^{(2)}$ are even smaller than those for $\hat{\beta}$ and $\hat{\varphi}$ and $\hat{\beta}^{(1)}$ and $\hat{\varphi}^{(1)}$.

$\theta_0^{(1)} = (0.04, 0.051, 0.040)$					
		Mean	Std	RMSE	RB
$\hat{\theta}_{n,\text{II}^*}$	$\hat{\beta}$	0.05452	0.02341	0.02754	0.36298
	$\hat{\eta}$	0.05027	0.01294	0.01296	-0.01433
	$\hat{\varphi}$	0.03478	0.00857	0.01003	-0.13046
$\hat{\theta}_{n,\text{II}}$	$\hat{\beta}$	0.04586	0.02133	0.02211	0.14658
	$\hat{\eta}$	0.05142	0.01421	0.01421	0.00827
	$\hat{\varphi}$	0.03788	0.00872	0.00897	-0.05300
$\theta_0^{(2)} = (0.04, 0.055, 0.036)$					
		Mean	Std	RMSE	RB
$\hat{\theta}_{n,\text{II}^*}$	$\hat{\beta}$	0.04315	0.01858	0.01883	0.07886
	$\hat{\eta}$	0.05177	0.01603	0.01635	-0.05867
	$\hat{\varphi}$	0.03109	0.01057	0.01165	-0.13643
$\hat{\theta}_{n,\text{II}}$	$\hat{\beta}$	0.04084	0.01829	0.01830	0.02090
	$\hat{\eta}$	0.05571	0.01666	0.01667	0.01295
	$\hat{\varphi}$	0.03570	0.00948	0.00948	-0.00828

Table 5.3: Performance assessment based on 1 000 independent samples of COGARCH squared returns \mathbf{G}_n^2 for $n = 10\,000$, sampled with parameter values $\theta_0^{(1)} = (0.04, 0.051, 0.040)$ and $\theta_0^{(2)} = (0.04, 0.055, 0.036)$: mean, standard deviation (Std), root mean squared error (RMSE) and relative bias (RB). Both IIEs $\hat{\theta}_{n,\text{II}^*}$ in (5.2) and $\hat{\theta}_{n,\text{II}}$ in (5.3) used the identity matrix for Ω . The IIE $\hat{\theta}_{n,\text{II}}$ is based on $K = 100$ simulated paths.

Acknowledgement

Thiago do Rêgo Sousa gratefully acknowledges support from CNPq, the National Council for Scientific and Technological Development, Brazil and of the TUM Graduate School. All authors take pleasure to thank Carsten Chong, Jean Jacod, Viet Son Pham, and Robert Stelzer for

helpful discussions. We thank Vicky Fasen for careful reading and some critical remarks, which lead to an improvement of our paper.

A Appendix to Section 4.1

The first Lemma states important properties about moments of a continuous version of a stochastic process found via Kolmogorov's continuity criterion. The property stated in (A.1) is used to apply a uniform SLLN in Theorem 4.8. Lemma A.4 is used to compute $\nabla_{\boldsymbol{\theta}}\sigma_0^2(\boldsymbol{\theta})$ and $\nabla_{\boldsymbol{\theta}}^2\sigma_0^2(\boldsymbol{\theta})$, needed to find a continuous version of the map $\boldsymbol{\theta} \mapsto \int_{(i-1)\Delta}^{i\Delta} \sigma_s(\boldsymbol{\theta})dL_s$ in Theorem 4.6, and of $\boldsymbol{\theta} \mapsto \int_{(i-1)\Delta}^{i\Delta} \nabla_{\boldsymbol{\theta}}\sigma_s(\boldsymbol{\theta})dL_s$ in Lemma 4.10.

Lemma A.1. *Let $(X(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)$ be a stochastic process with $\Theta \subset \mathbb{R}_+^d$ compact for $d \in \mathbb{N}$. Assume that there exist positive constants p, c, ϵ such that for all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$:*

$$\mathbb{E}|X^{(c)}(\boldsymbol{\theta}_1) - X^{(c)}(\boldsymbol{\theta}_2)|^p \leq c\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^{d+\epsilon}.$$

Then there exists a continuous version $(X^{(c)}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)$ of $(X(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)$ such that

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |X^{(c)}(\boldsymbol{\theta})|^q < \infty. \tag{A.1}$$

Proof. Since Θ is compact we can use the Heine-Borel theorem to find a finite collection of open sets $(\Theta_j)_{j=1}^N$ such that $\Theta \subset \cup_{j=1}^N \Theta_j$ and $\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \leq \delta^*$ for every $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta_j$. Choosing an arbitrary $\boldsymbol{\theta}_j \in \Theta_j \cap \Theta$ for $j = 1, \dots, N$ and using $|a - b|^q \leq 2^{q-1}|a^q - b^q|$ gives for $q < p$,

$$\begin{aligned} \mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |X^{(c)}(\boldsymbol{\theta})|^q &\leq \sum_{j=1}^N \mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta_j} |X^{(c)}(\boldsymbol{\theta})|^q \\ &\leq \sum_{j=1}^N 2^{q-1} \mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta_j} \{|X^{(c)}(\boldsymbol{\theta}) - X^{(c)}(\boldsymbol{\theta}_j)|^q + |X^{(c)}(\boldsymbol{\theta}_j)|^q\} \\ &\leq 2^{q-1} \sum_{j=1}^N (1 + \mathbb{E}|X^{(c)}(\boldsymbol{\theta}_j)|^q) < \infty, \end{aligned}$$

since $\mathbb{E}|X^{(c)}(\boldsymbol{\theta})|^p < \infty$ for all $\boldsymbol{\theta} \in \Theta$. □

The next Lemma gives necessary conditions for the existence of a continuous version of a stochastic process, and of fractional moments of order $p \geq 1$ of a random variable that appears when computing inequalities involving moments of $\sigma_0^2(\boldsymbol{\theta})$.

The following Lemma is well-known from Analysis, and can be found for instance as Exercise 6 in Ch. 15.7 of [26].

Lemma A.2. *Suppose that $g : \mathbb{R}^p \rightarrow \mathbb{R}^q$ is continuous and that*

$$\sup_{\boldsymbol{\theta} \in \Theta} \|f_n(\boldsymbol{\theta}) - f(\boldsymbol{\theta})\| \xrightarrow{\text{a.s.}} 0, \quad n \rightarrow \infty,$$

where $(f_n(\boldsymbol{\theta}))_{n \in \mathbb{N}}$ is a sequence of random vectors in \mathbb{R}^p , $f : \Theta \in \mathbb{R}^d \mapsto \mathbb{R}^p$ is a deterministic function and Θ is compact. Then as $n \rightarrow \infty$,

$$\sup_{\boldsymbol{\theta} \in \Theta} \|g(f_n(\boldsymbol{\theta})) - g(f(\boldsymbol{\theta}))\| \xrightarrow{\text{a.s.}} 0.$$

Lemma A.3. Let $p, b \geq 1$, $a, k \geq 0$, $\boldsymbol{\theta} \in \mathcal{M}$ be fixed and $(K_s(\tilde{\varphi}))_{s \geq 0}$ as defined in (4.1) for fixed $\tilde{\varphi} > 0$. If $\mathbb{E}|L_1|^{2p(1+\epsilon)} < \infty$ and $\Psi_{\boldsymbol{\theta}}(p(1+\epsilon)) < 0$ for some $\epsilon > 0$, then

$$\mathbb{E} \left(\int_0^\infty (s^a + s^k K_s^b(\tilde{\varphi})) e^{-Y_s(\boldsymbol{\theta})} ds \right)^p < \infty.$$

Proof. The proof is similar to the proof of Proposition 4.1 in Lindner and Maller [29]. For every $j \in \mathbb{N}_0$ define $Q_j(\boldsymbol{\theta}) := \int_j^{j+1} (s^a + s^k K_s^b(\tilde{\varphi})) e^{-Y_s(\boldsymbol{\theta})} ds$. Then

$$\begin{aligned} \mathbb{E} Q_j^p(\boldsymbol{\theta}) &= \mathbb{E} \left(\int_j^{j+1} (s^a + s^k K_s^b(\tilde{\varphi})) e^{-Y_s(\boldsymbol{\theta})} ds \right)^p \\ &\leq \mathbb{E} \left(\sup_{j \leq s \leq j+1} (s^a + s^k K_s^b(\tilde{\varphi})) e^{-Y_s(\boldsymbol{\theta})} \right)^p \\ &\leq \mathbb{E} \left(((j+1)^a + (j+1)^k K_{j+1}^b(\tilde{\varphi}))^p \sup_{j \leq s \leq j+1} e^{-pY_s(\boldsymbol{\theta})} \right) \quad (\text{A.2}) \\ &\leq \left(\mathbb{E} ((j+1)^a + (j+1)^k K_{j+1}^b(\tilde{\varphi}))^{p(1+\epsilon)/\epsilon} \right)^{\epsilon/(1+\epsilon)} \left(\mathbb{E} \sup_{j \leq s \leq j+1} e^{-p(1+\epsilon)Y_s(\boldsymbol{\theta})} \right)^{1/(1+\epsilon)} \end{aligned}$$

by the Hölder inequality. Since by Lemma 4.3 $(K_s(\tilde{\varphi}))_{s \geq 0}$ is a Lévy process with moments of all orders, repeated differentiation of the characteristic function of $K_{j+1}(\tilde{\varphi})$ gives a constant $c > 0$ such that

$$\mathbb{E} ((j+1)^a + (j+1)^k K_{j+1}^b(\tilde{\varphi}))^{p(1+\epsilon)/\epsilon} \leq c(j+1)^{mp(1+\epsilon)/\epsilon}, \quad (\text{A.3})$$

where $m = a + k + b$. Since the process $(e^{Y_s(\boldsymbol{\theta}) - s\Psi_{\boldsymbol{\theta}}(1)})_{s \geq 0}$ is a martingale, we can use Doob's martingale inequality, the Laplace transform in (2.2) and the fact that $\Psi_{\boldsymbol{\theta}}(1) < 0$ to get

$$\begin{aligned} \mathbb{E} \sup_{j \leq s \leq j+1} e^{-p(1+\epsilon)Y_s(\boldsymbol{\theta})} &\leq e^{-(j+1)p(1+\epsilon)\Psi_{\boldsymbol{\theta}}(1)} \mathbb{E} \sup_{j \leq s \leq j+1} e^{-p(1+\epsilon)Y_s(\boldsymbol{\theta}) + sp(1+\epsilon)\Psi_{\boldsymbol{\theta}}(1)} \\ &\leq e^{-(j+1)p(1+\epsilon)\Psi_{\boldsymbol{\theta}}(1)} \mathbb{E} e^{-p(1+\epsilon)Y_{j+1}(\boldsymbol{\theta}) + p(1+\epsilon)(j+1)\Psi_{\boldsymbol{\theta}}(1)} \\ &= \mathbb{E} e^{-p(1+\epsilon)Y_{j+1}(\boldsymbol{\theta})} \\ &= e^{(j+1)\Psi_{\boldsymbol{\theta}}(p(1+\epsilon))}. \end{aligned} \quad (\text{A.4})$$

Equation (A.2) together with (A.3) and (A.4) gives

$$\mathbb{E} Q_j^p(\boldsymbol{\theta}) \leq c^*(j+1)^{mp} e^{(j+1)\Psi_{\boldsymbol{\theta}}(p(1+\epsilon))/(1+\epsilon)} < \infty, \quad (\text{A.5})$$

where $c^* = c^{\epsilon/(1+\epsilon)}$. Let $\alpha := \lfloor p \rfloor$ be the integer part of p and suppose that $p > \alpha$. Then

$$\begin{aligned} \left(\int_0^n (s^a + s^k K_s^b(\tilde{\varphi})) e^{-Y_s(\boldsymbol{\theta})} ds \right)^p &= \left(\sum_{j=0}^{n-1} Q_j(\boldsymbol{\theta}) \right)^p \\ &= \sum_{j_1=0}^{n-1} \cdots \sum_{j_\alpha=0}^{n-1} Q_{j_1}(\boldsymbol{\theta}) \cdots Q_{j_\alpha}(\boldsymbol{\theta}) \left(\sum_{j_{\alpha+1}=0}^{n-1} Q_{j_{\alpha+1}}(\boldsymbol{\theta}) \right)^{p-\alpha} \quad (\text{A.6}) \\ &\leq \sum_{j_1=0}^{n-1} \cdots \sum_{j_\alpha=0}^{n-1} \sum_{j_{\alpha+1}=0}^{n-1} Q_{j_1}(\boldsymbol{\theta}) \cdots Q_{j_\alpha}(\boldsymbol{\theta}) Q_{j_{\alpha+1}}^{p-\alpha}(\boldsymbol{\theta}). \end{aligned}$$

If p is an integer the last sum in (A.6) disappears. By (A.5), for each $j = 1, \dots, \alpha + 1$, $Q_j \in \mathcal{L}^p$ so we can apply the Hölder inequality with $\frac{1}{p} + \cdots + \frac{1}{p} + \frac{p-\alpha}{p} = 1$ to the right-hand side of (A.6).

This together with (A.5) gives

$$\begin{aligned}
& \mathbb{E} \left(\int_0^m (s^a + s^k K_s^b(\tilde{\varphi})) e^{-Y_s(\boldsymbol{\theta})} ds \right)^p \\
& \leq \sum_{j_1=0}^{n-1} \cdots \sum_{j_\alpha=0}^{n-1} \sum_{j_{\alpha+1}=0}^{n-1} (\mathbb{E}(Q_{j_1}^p(\boldsymbol{\theta}))^{\frac{1}{p}} \cdots (\mathbb{E}(Q_{j_\alpha}^p(\boldsymbol{\theta}))^{\frac{1}{p}} (\mathbb{E}(Q_{j_{\alpha+1}}^p(\boldsymbol{\theta}))^{\frac{p-\alpha}{p}} \\
& \leq c^* \left(\sum_{j=0}^{n-1} (j+1)^m e^{(j+1)/(p(1+\epsilon))\Psi_{\boldsymbol{\theta}}(p(1+\epsilon))} \right)^\alpha \left(\sum_{j=0}^{n-1} (j+1)^{m(p-\alpha)} e^{(j+1)(p-\alpha)/(p(1+\epsilon))\Psi_{\boldsymbol{\theta}}(p(1+\epsilon))} \right).
\end{aligned} \tag{A.7}$$

Since $\Psi_{\boldsymbol{\theta}}(p(1+\epsilon)) < 0$ both series in (A.7) converge. The monotone convergence theorem applied to the expectation in the first line of (A.7) gives the result. \square

Lemma A.4. *Let $\boldsymbol{\theta} = (\beta, \eta, \varphi)$ with $\beta, \eta, \varphi > 0$ and consider the process $(Y_s(\boldsymbol{\theta}))_{s \geq 0}$ as in (2.1). Let $K_s(\varphi)$ be as defined in (4.1). Then:*

(a) For every fixed $s > 0$,

$$\nabla_{\eta, \varphi}(e^{-Y_s(\boldsymbol{\theta})}) = e^{-Y_s(\boldsymbol{\theta})} \begin{pmatrix} -s \\ K_s(\varphi) \end{pmatrix}. \tag{A.8}$$

(b) If $\mathbb{E}|L_1|^{2(1+\epsilon)} < \infty$ and $\Psi_{\boldsymbol{\theta}}(1+\epsilon) < 0$ for some $\epsilon > 0$, then

$$\nabla_{\eta, \varphi} \left(\int_0^\infty e^{-Y_s(\boldsymbol{\theta})} ds \right) = \int_0^\infty e^{-Y_s(\boldsymbol{\theta})} \begin{pmatrix} -s \\ K_s(\varphi) \end{pmatrix} ds \tag{A.9}$$

and

$$\nabla_{\eta, \varphi}^2 \left(\int_0^\infty e^{-Y_s(\boldsymbol{\theta})} ds \right) = \int_0^\infty e^{-Y_s(\boldsymbol{\theta})} \begin{pmatrix} s^2 & -sK_s(\varphi) \\ -sK_s(\varphi) & (d_s^2(\varphi) + d_s'(\varphi)) \end{pmatrix} ds. \tag{A.10}$$

Proof. (a) The partial derivatives of $Y_s(\boldsymbol{\theta}) = \eta s - \sum_{0 < u \leq s} \log(1 + \varphi(\Delta L_u)^2)$ are given by

$$\frac{\partial Y_s(\boldsymbol{\theta})}{\partial \eta} = s \quad \text{and} \quad \frac{\partial Y_s(\boldsymbol{\theta})}{\partial \varphi} = -K_s(\varphi),$$

where the derivative with respect to φ follows by dominated convergence since we have the following bound independent of φ :

$$K_s(\varphi) \leq \sum_{0 < u \leq s} (\Delta L_u)^2 < \infty. \tag{A.11}$$

A simple application of the chain rule gives (A.8).

(b) It follows from Lemma 4.3(a) that we can find a collection of points $(\boldsymbol{\theta}_j^*)_{j=1}^N$ in \mathcal{M} such that

$$\sup_{\boldsymbol{\theta} \in \Theta} e^{-Y_s(\boldsymbol{\theta})} \leq \sum_{j=1}^N e^{-Y_s(\boldsymbol{\theta}_j^*)}, \quad s \geq 0. \tag{A.12}$$

The first derivative of $K_s(\varphi)$ follows from dominated convergence with the upper bound in (A.11) and is given by

$$K_s'(\varphi) = - \sum_{0 < u \leq s} \frac{(\Delta L_u)^4}{(1 + \varphi(\Delta L_u)^2)^2}, \quad s \geq 0.$$

Now, similar calculations as in (4.2) show that $|K'_s(\varphi)| \leq K_s(\varphi_*)/\varphi_*$ for φ_* as defined in (4.22). This combined with (A.12) allows us to obtain an upper bound for the sum of the bounds of the absolute values of the integrals at the r.h.s. of (A.9) and (A.10) given by

$$\sum_{j=1}^N \int_0^\infty e^{-Y_s(\boldsymbol{\theta}_j^*)} (s + s^2 + K_s(\varphi_*)(1 + 2s + 1/\phi_*) + d_s^2(\varphi_*)) ds. \quad (\text{A.13})$$

Since $\mathbb{E}|L_1|^{2(1+\epsilon)} < \infty$ and $\Psi_{\boldsymbol{\theta}_j^*}(1 + \epsilon) < 0$ for all $j = 1, \dots, N$ we can apply Lemma A.3 with $p = 1$ to prove that the integral in (A.13) has finite first moment and is therefore well defined. This allows us to use dominated convergence to differentiate under the integral sign and then use the chain and product rule combined with (A.8) to obtain (A.9) and (A.10). \square

Lemma A.5. *Let $p \geq 1$ and $k \in \{1, 2\}$. If $\mathbb{E}|L_1|^{2kp(1+\epsilon)} < \infty$ for some $\epsilon > 0$ then*

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} \|\nabla_{\eta, \varphi}^k \sigma_0^2(\boldsymbol{\theta})\|^p < \infty \quad \text{and} \quad \mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(kp(1+\epsilon))}} \|\nabla_{\eta, \varphi}^k \sigma_0(\boldsymbol{\theta})\|^p < \infty.$$

Proof. For $k \in \{1, 2\}$ let R_{kp} denote the integral defined in (A.13) with $(\boldsymbol{\theta}_j^*)_{j=1}^N \in \mathcal{M}^{(kp(1+\epsilon))}$ as in (4.3). By the same argument preceding (A.13) and from (2.4) we get

$$\sup_{\boldsymbol{\theta} \in \Theta^{(kp(1+\epsilon))}} \left(\|\nabla_{\eta, \varphi} \sigma_0^2(\boldsymbol{\theta})\| + \|\nabla_{\eta, \varphi}^2 \sigma_0^2(\boldsymbol{\theta})\| \right) \leq c\beta^* R_{kp}, \quad k = 1, 2, \quad (\text{A.14})$$

where $c > 0$ and

$$\beta^* = \sup\{\beta > 0 : (\beta, \eta, \varphi) \in \Theta\} < \infty. \quad (\text{A.15})$$

Since from Lemma 4.3(b) we know that $\sigma_0(\boldsymbol{\theta}) \geq \sigma^* > 0$, the chain rule implies that

$$\|\nabla_{\eta, \varphi} \sigma_0(\boldsymbol{\theta})\| \leq \frac{1}{\sigma^*} \|\nabla_{\eta, \varphi} \sigma_0^2(\boldsymbol{\theta})\|. \quad (\text{A.16})$$

Using (A.16) combined with the chain rule for the second order derivative gives

$$\|\nabla_{\eta, \varphi}^2 \sigma_0(\boldsymbol{\theta})\| \leq \frac{1}{4\sigma^*} \|\nabla_{\eta, \varphi}^2 \sigma_0^2(\boldsymbol{\theta})\| + \frac{1}{8(\sigma^*)^3} \|\nabla_{\eta, \varphi} \sigma_0^2(\boldsymbol{\theta})\|^2. \quad (\text{A.17})$$

Using (A.14) combined with (A.16) gives

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(p(1+\epsilon))}} \|\nabla_{\eta, \varphi} \sigma_0(\boldsymbol{\theta})\|^p \leq \mathbb{E} \left(\frac{1}{\sigma^*} c\beta^* R_p \right)^p < \infty,$$

by an application of Lemma A.3. Now, (A.14) combined with (A.17) gives

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}} \|\nabla_{\eta, \varphi}^2 \sigma_0(\boldsymbol{\theta})\|^p \leq \mathbb{E} \left(\frac{1}{4\sigma^*} c\beta^* R_{2p} + \frac{1}{8(\sigma^*)^3} (c\beta^* R_{2p})^2 \right)^p < \infty,$$

by an application of the Cauchy-Schwartz inequality and Lemma A.3 with p replaced by $2p$. \square

B Appendix to Section 4.2

Lemmas B.2 and B.1 are used in the proof of Proposition 4.13 to control the convergence of arithmetic means defined in terms of the sequences $(G_i(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ and $(\nabla_{\boldsymbol{\theta}} G_i(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ with $\nabla_{\boldsymbol{\theta}} G_i(\boldsymbol{\theta})$ defined in the sense of Remark 4.12.

Lemma B.1. *Let $\boldsymbol{\theta} = (\beta, \eta, \phi) =: (\theta_1, \theta_2, \theta_3)$ with $\beta, \eta, \phi > 0$ and $\Delta > 0$. Suppose that $\mathbb{E}|L_1|^2 < \infty$ and $\Psi_{\boldsymbol{\theta}}(1) < 0$. Let $(\sigma_t(\boldsymbol{\theta}))_{t \geq 0}$ be the stationary volatility process starting with $\sigma_0(\boldsymbol{\theta})$ as in (2.4) independent of L . Then for all three components of $\boldsymbol{\theta}$ the sequences*

$$\left(\int_{(i-1)\Delta}^{i\Delta} \sigma_s(\boldsymbol{\theta}) dL_s, \int_{(i-1)\Delta}^{i\Delta} \frac{\partial}{\partial \theta_j} \sigma_s(\boldsymbol{\theta}) dL_s \right)_{i \in \mathbb{N}}$$

are stationary and ergodic.

Proof. Consider without loss of generality $j = 1$. Define the i.i.d. sequence $(S_k)_{k \in \mathbb{Z}}$ with

$$S_k = (\Delta L_u, (k-1)\Delta < u \leq k\Delta).$$

We consider

$$((\sigma_s(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta), (i-1)\Delta < s \leq i\Delta) =: g(\boldsymbol{\theta}, \boldsymbol{\theta} \in \Theta, (S_k)_{k=-\infty}^i)$$

as a measurable function of all relevant jumps ΔL_u . Additionally, since limits of differentiable functions are measurable, there exists a measurable map h such that

$$\left(\frac{\partial}{\partial \theta_1} \sigma_s(\boldsymbol{\theta}), (i-1)\Delta < s \leq i\Delta \right) = h((S_k)_{k=-\infty}^i, (\boldsymbol{\theta} + (c, 0, 0))_{c \in \mathbb{Q}}).$$

By observing that a stochastic integral is defined as a measurable map depending on the integrand and integrator processes, we can write

$$\int_{(i-1)\Delta}^{i\Delta} \sigma_s(\boldsymbol{\theta}) dL_s = g((S_k)_{k=-\infty}^i, \boldsymbol{\theta}) \quad \text{and} \quad \int_{(i-1)\Delta}^{i\Delta} \frac{\partial}{\partial \theta_1} \sigma_s(\boldsymbol{\theta}) dL_s = h((S_k)_{k=-\infty}^i, (\boldsymbol{\theta} + (c, 0, 0))_{c \in \mathbb{Q}}).$$

Using Proposition 5 in Straumann and Mikosch [42] (see also Theorem 2.1 in Krengel [27]) we can conclude the stationarity and ergodicity of the process $(G_i(\boldsymbol{\theta}), \nabla G_i(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ based on the stationarity and ergodicity of the sequence $(S_i)_{i \in \mathbb{Z}}$ and the measurability of g and h . \square

Lemma B.2. *If $\mathbb{E}|L_1|^{4p(1+\epsilon)} < \infty$ for some $p > 5/2$ and $\epsilon > 0$ then for every $l \in \{1, 2, 3\}$ and $h \in \mathbb{N}_0$ we have*

$$\sup_{\boldsymbol{\theta} \in \Theta(2p(1+\epsilon))} \left| \frac{\partial}{\partial \theta_l} \left(\frac{1}{n} \sum_{i=1}^{n-h} G_i^2(\boldsymbol{\theta}) G_{i+h}^2(\boldsymbol{\theta}) \right) - \frac{\partial}{\partial \theta_l} (\mathbb{E} G_1^2(\boldsymbol{\theta}) G_{1+h}^2(\boldsymbol{\theta})) \right| \xrightarrow{P} 0, \quad n \rightarrow \infty. \quad (\text{B.1})$$

Proof. The proof follows closely the strategy in the proof of Proposition 5.5 in Fasen-Hartmann and Kimmig [10], which divides the proof into three steps: Pointwise convergence, local Hölder continuity, and stochastic equicontinuity. Let $l \in \{1, 2, 3\}$ and $h \in \mathbb{N}_0$ be fixed. Write $\hat{\mu}_n(h; \boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n-h} G_i^2(\boldsymbol{\theta}) G_{i+h}^2(\boldsymbol{\theta})$. Then, a simple application of the chain and product rule gives

$$\frac{\partial}{\partial \theta_l} \hat{\mu}_n(h; \boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n-h} \left[2G_i(\boldsymbol{\theta}) \left(\frac{\partial}{\partial \theta_l} G_i(\boldsymbol{\theta}) \right) G_{i+h}^2(\boldsymbol{\theta}) + 2G_i^2(\boldsymbol{\theta}) G_{i+h}(\boldsymbol{\theta}) \left(\frac{\partial}{\partial \theta_l} G_{i+h}(\boldsymbol{\theta}) \right) \right]. \quad (\text{B.2})$$

Step 1. Pointwise convergence. Let $\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}$ be fixed. It follows from Lemma B.1 that the sequence $(G_i(\boldsymbol{\theta}), \frac{\partial}{\partial \boldsymbol{\theta}_l} G_i(\boldsymbol{\theta}))_{i \in \mathbb{N}}$ is stationary and ergodic. Additionally, it follows from the lemma's assumptions combined with Theorem 4.6, Lemma 4.10 and the Hölder inequality with $\frac{1}{5} + \frac{2}{5} + \frac{2}{5} = 1$ that

$$\begin{aligned} & \mathbb{E} G_1(\boldsymbol{\theta}) \left(\frac{\partial}{\partial \boldsymbol{\theta}_l} G_1(\boldsymbol{\theta}) \right) G_{1+h}^2(\boldsymbol{\theta}) \\ & \leq (\mathbb{E} G_1^5(\boldsymbol{\theta}))^{1/5} \left(\mathbb{E} \left(\frac{\partial}{\partial \boldsymbol{\theta}_l} G_1(\boldsymbol{\theta}) \right)^{5/2} \right)^{2/5} (\mathbb{E} G_{1+h}^5(\boldsymbol{\theta}))^{2/5} < \infty. \end{aligned} \quad (\text{B.3})$$

The same calculations in (B.3) can be applied to show that the expectation of the second term in the summation (B.2) is also finite. This allows us to apply Birkhoff convergence theorem to conclude that

$$\frac{\partial}{\partial \boldsymbol{\theta}_l} \hat{\mu}_n(h; \boldsymbol{\theta}) \xrightarrow{P} \mathbb{E} G_1^2(\boldsymbol{\theta}) G_{1+h}^2(\boldsymbol{\theta}), \quad n \rightarrow \infty.$$

Step 2. $\frac{\partial}{\partial \boldsymbol{\theta}_l} \hat{\mu}_n(h; \boldsymbol{\theta})$ is locally Hölder-continuous on $\Theta^{(2p(1+\epsilon))}$. For $i \in \mathbb{N}$ let U_i and V_i be as defined in (4.6) and (4.16), respectively. By stationarity of $(G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)_{i \in \mathbb{N}}$ and $(\frac{\partial}{\partial \boldsymbol{\theta}_l} G_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)_{i \in \mathbb{N}}$, $U_i \stackrel{d}{=} U_1$, $V_i \stackrel{d}{=} V_1$ and for every $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(2p(1+\epsilon))}$ with $\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| < 1$ it follows from Theorem 4.6 and Lemma 4.10 that there exists $\gamma \in (0, 1)$ such that for all $i \in \mathbb{N}$:

$$|G_i(\boldsymbol{\theta}_1) - G_i(\boldsymbol{\theta}_2)| \leq U_i \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma$$

and

$$\left| \frac{\partial}{\partial \boldsymbol{\theta}_l} G_i(\boldsymbol{\theta}_1) - \frac{\partial}{\partial \boldsymbol{\theta}_l} G_i(\boldsymbol{\theta}_2) \right| \leq V_i \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma.$$

Using the inequality

$$|a_1 b_1 c_1^2 - a_2 b_2 c_2^2| \leq |a_1| |b_1| |c_1 + c_2| |c_1 - c_2| + |a_1| |c_2^2| |b_1 - b_2| + |b_2 c_2^2| |a_1 - a_2|,$$

valid for every $a_1, a_2, b_1, b_2, c_1, c_2 \in \mathbb{R}$ gives for all $i \in \mathbb{N}$,

$$\begin{aligned} & \left| G_i(\boldsymbol{\theta}_1) \left(\frac{\partial}{\partial \boldsymbol{\theta}_l} G_i(\boldsymbol{\theta}_1) \right) G_{i+h}^2(\boldsymbol{\theta}_1) - G_i(\boldsymbol{\theta}_2) \left(\frac{\partial}{\partial \boldsymbol{\theta}_l} G_i(\boldsymbol{\theta}_2) \right) G_{i+h}^2(\boldsymbol{\theta}_2) \right| \\ & \leq 2 \left(\sup_{\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}} |G_i(\boldsymbol{\theta})| \left| \frac{\partial}{\partial \boldsymbol{\theta}_l} G_i(\boldsymbol{\theta}) \right| |G_{i+h}(\boldsymbol{\theta})| \right) U_{i+h} \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma \\ & \quad + \left(\sup_{\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}} |G_i(\boldsymbol{\theta})| |G_{i+h}^2(\boldsymbol{\theta})| \right) V_i \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma \\ & \quad + \left(\sup_{\boldsymbol{\theta} \in \Theta^{(2p(1+\epsilon))}} \left| \frac{\partial}{\partial \boldsymbol{\theta}_l} G_i(\boldsymbol{\theta}) \right| |G_{i+h}^2(\boldsymbol{\theta})| \right) U_i \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma \\ & =: I_{i,h} \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma. \end{aligned} \quad (\text{B.4})$$

Another application of the Hölder inequality combined with (4.7) and an analogous result for $\nabla_{\boldsymbol{\theta}} G_i(\boldsymbol{\theta})$ gives $\mathbb{E} I_{1,h} < \infty$. Similar calculations as in (B.4) can be used to show that for all $i \in \mathbb{N}$

$$\left| G_i^2(\boldsymbol{\theta}_1) G_{i+h}(\boldsymbol{\theta}_1) \left(\frac{\partial}{\partial \boldsymbol{\theta}_l} G_{i+h}(\boldsymbol{\theta}_1) \right) - G_i^2(\boldsymbol{\theta}_2) G_{i+h}(\boldsymbol{\theta}_2) \left(\frac{\partial}{\partial \boldsymbol{\theta}_l} G_{i+h}(\boldsymbol{\theta}_2) \right) \right| \leq I_{i,h}^* \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^\gamma, \quad (\text{B.5})$$

with $\mathbb{E}I_{1,h}^* < \infty$.

Step 3. Stochastic equicontinuity. Let $\xi, \nu > 0$ and $0 < \delta < \min\{1, \eta\xi/\mathbb{E}(I_{1,h} + I_{1,h}^*)\}$. Then, it follows from (B.2), (B.4), (B.5) and Markov's inequality that

$$\mathbb{P}\left(\sup_{\substack{0 < \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| < \delta \\ \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^{(2p(1+\epsilon))}}} \left| \frac{\partial}{\partial \boldsymbol{\theta}_l} \hat{\mu}_n(h; \boldsymbol{\theta}_1) - \frac{\partial}{\partial \boldsymbol{\theta}_l} \hat{\mu}_n(h; \boldsymbol{\theta}_2) \right| > \eta\right) \leq \mathbb{E}(I_{1,h} + I_{1,h}^*) \frac{\delta^\gamma}{\eta} < \xi.$$

This together with the pointwise convergence in Step 1 allow us to conclude the uniform convergence in (B.1) by means of Theorem 10.2 in Pollard [34]. \square

References

- [1] D. Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge University Press, 2009.
- [2] S. Bayracı and G. Ünal. Stochastic interest rate volatility modeling with a continuous-time GARCH(1,1) model. *Journal of Computational and Applied Mathematics*, 259:464–473, 2014.
- [3] C. Bianchi and E. M. Cleur. Indirect estimation of stochastic differential equation models: some computational experiments. *Computational Economics*, 9(3):257–274, 1996.
- [4] E. Bibbona and I. Negri. Higher moments and prediction-based estimation for the COGARCH(1,1) model. *Scandinavian Journal of Statistics*, 42(4):891–910, 2015.
- [5] E. Bibbona, I. Negri, A. Arcagni, and L. Mercuri. *COGARCH: Simulation and inference for COGARCH(1,1) processes*, 2014. URL <http://R-Forge.R-project.org/projects/cogarch/>. R package version 1.3-1/r32.
- [6] R. Bradley. *Introduction to Strong Mixing Conditions*, volume 1. Kendrick Press, Heber City, 2007.
- [7] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, New York, 2013.
- [8] X. de Luna and M. G. Genton. Robust simulation-based estimation of ARMA models. *Journal of Computational and Graphical Statistics*, 10(2):370–387, 2001.
- [9] T. Do Rêgo Sousa. *Simulation-based Estimation of Stochastic Volatility Processes*. PhD thesis, Department of Mathematics, Technical University of Munich. In preparation.
- [10] V. Fasen-Hartmann and S. Kimmig. Robust estimation of continuous-time arma models via indirect inference. *arXiv:1804.00849*, 2018.
- [11] R. Garcia, E. Renault, and D. Veredas. Estimation of stable distributions by Indirect Inference. *Journal of Econometrics*, 161(2):325–337, 2011.
- [12] C. Gouriéroux, A. Monfort, and E. Renault. Indirect inference. *Journal of Applied Econometrics*, 8:S85–S118, 1993.
- [13] C. Gouriéroux, E. Renault, and N. Touzi. Calibration by simulation for small sample bias correction. In R. Mariano, T. Schuermann, and M. J. Weeks, editors, *Simulation-based Inference in Econometrics: Methods and Applications*, pages 328–358. Cambridge University Press, Cambridge, 2000.

- [14] C. Gouriéroux, P. C. B. Phillips, and J. Yu. Indirect Inference for dynamic panel models. *Journal of Econometrics*, 157(1):68–77, 2010. ISSN 03044076. doi: 10.1016/j.jeconom.2009.10.024.
- [15] S. Haug, C. Klüppelberg, A. Lindner, and M. Zapp. Method of moment estimation in the COGARCH(1,1) model. *Econometrics Journal*, 10(2):320–341, 2007.
- [16] J. Hutton and P. Nelson. Interchanging the order of stochastic integration and ordinary differentiation. *Stochastic Processes and their Applications*, 18:371–377, 1984.
- [17] S. Iacus, L. Mercuri, and E. Rroji. Cogarch (p, q): Simulation and inference with the yuima package. *Journal of Statistical Software*, 80(4), 2017.
- [18] M. Iannace. *COGARCH Processes: Theory and Asymptotics for the Pseudo-Maximum Likelihood Estimator*. PhD thesis, University of Milano-Bicocca, 2014.
- [19] I. Ibragimov and Y. Linnik. *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen, 1971.
- [20] G. J. Jiang. Estimation of jump-diffusion processes based on Indirect Inference. *IFAC Proceedings Volumes*, 31(16):385–390, 1998.
- [21] D. Khoshnevisan. *Multiparameter Processes: An Introduction to Random Fields*. Springer, New York, 2002.
- [22] M. Kim and S. Lee. On the maximum likelihood estimator for irregularly observed time series data from COGARCH(1,1) models. *Revstat Statistical Journal*, 11(2):135–168, 2013.
- [23] C. Klüppelberg, A. Lindner, and R. Maller. A continuous-time GARCH process driven by a Lévy process: stationarity and second-order behaviour. *Journal of Applied Probability*, 41(3):601–622, 2004.
- [24] C. Klüppelberg, A. Lindner, and R. Maller. Continuous time volatility modelling: COGARCH versus Ornstein-Uhlenbeck models. In Y. Kabanov, R. Liptser, and J. Stoyanov, editors, *The Shiryaev Festschrift: From Stochastic Calculus to Mathematical Finance*, pages 393–419. Springer, Berlin, 2006.
- [25] C. Klüppelberg, R. Maller, and A. Szimayer. The COGARCH: a review, with news on option pricing and statistical inference. In J. Błath, P. Imkeller, and S. Roelly, editors, *Survey in Stochastic Processes. Proc. of the 33rd SPA Conference in Berlin*, pages 29–50. EMS Series of Congress Reports, EMS Publishing House, Zürich, 2011.
- [26] K. Königsberger. *Analysis 1*. Springer, Berlin, 6 edition, 2004.
- [27] U. Krengel. *Ergodic Theorems*. Walter de Gruyter, Berlin, 1985.
- [28] M. P. Laurini and L. K. Hotta. Indirect inference in fractional short-term interest rate diffusions. *Mathematics and Computers in Simulation*, 94:109–126, 2013.
- [29] A. Lindner and R. Maller. Lévy integrals and the stationarity of generalised Ornstein-Uhlenbeck processes. *Stochastic Processes and their Applications*, 115(10):1701–1722, 2005.
- [30] M. J. Lombardi and G. Calzolari. Indirect estimation of alpha-stable distributions and processes. *Econometrics Journal*, 11(1):193–208, 2008.

- [31] R. A. Maller, G. Müller, and A. Szimayer. GARCH modelling in continuous time for irregularly spaced time series data. *Bernoulli*, 14(2):519–542, 2008.
- [32] G. Müller. MCMC estimation of the COGARCH(1,1) model. *Journal of Financial Econometrics*, 8(4):481–510, 2010.
- [33] W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245, 1994.
- [34] D. Pollard. *Empirical Processes: Theory and Applications*, volume 2. Institute of Mathematical Statistics, Hayward, CA, 1990.
- [35] P. Protter. *Stochastic Integration and Differential Equations*. Springer, New York, 2 edition, 2005.
- [36] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL <https://www.R-project.org/>.
- [37] A. Raknerud and Ø. Skare. Indirect inference methods for stochastic volatility models based on non-Gaussian Ornstein-Uhlenbeck processes. *Computational Statistics & Data Analysis*, 56(11):3260–3275, 2012.
- [38] R. L. Schilling and L. Partzsch. *Brownian Motion: An Introduction to Stochastic Processes*. De Gruyter, Berlin/Boston, 2014.
- [39] P. Shaman and R. A. Stine. The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association*, 83(403):842–848, 1988.
- [40] A. A. Smith. Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8:S63–S84, 1993.
- [41] M. Sørensen. Prediction-based estimating functions. *Econometrics Journal*, 3:123–147, 2000.
- [42] D. Straumann and T. Mikosch. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics*, 34(5):2449–2495, 2006.
- [43] B. Wahlberg, J. Welsh, and L. Ljung. Identification of stochastic Wiener systems using Indirect Inference. *IFAC-PapersOnLine*, 48(28):620–625, 2015.