Robust Estimation and Inference for Time-varying Unconditional Volatility[1]

Rickard Sandberg[2] and Genaro Sucarrat[3]

This version: 1st September 2022

Abstract

The unconditional volatility of financial return is often time-varying. To model this, a common approach is to decompose the volatility σ_t^2 multiplicatively into a non-stochastic process g_t, and a de-volatilitised stochastic process h_t: $\sigma_t^2 = g_t h_t$. We prove the consistency and asymptotic normality of the single-step Quasi Maximum Likelihood Estimator (QMLE) of the parameters of g_t for a large class of specifications of g_t. Next, we derive a simple but robust and consistent estimator of the asymptotic coefficient covariance. The exact specification of h_t need not be estimated or known, and h_t can even be non-stationary in the distribution. This is important in empirical applications, since financial returns are frequently characterised by a non-stationary zero-process. Next, we derive a period-by-period estimator of time-varying periodic unconditional volatility. Due to the assumptions we rely upon, our results extend directly to the Multiplicative Error Model (MEM) interpretation of volatility models. So our results can also be applied to the modelling of the time-varying unconditional mean of non-negative processes (e.g. volume, duration, realised volatility, dividends and unemployment). Three applications illustrate our results.

JEL Classification: C01, C13, C14, C22, C58
Keywords: Volatility, time-varying unconditional volatility, ARCH, GARCH, MEM, robust estimation

[1]We are grateful to Steffen Grønneberg, Fabian Harang, participants at the ISNPS 2022 conference (June, Paphos), QFFE 2022 workshop (June, Marseille) and the internal ECON seminar at BI Norwegian Business School (June 2022) for their helpful comments, suggestions and questions.

[2]Center for Data Analytics, Stockholm School of Economics.

[3]Corresponding author: Department of Economics, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway. Webpage: https://www.sucarrat.net/.

1 Introduction

Financial returns are frequently characterised by a time-varying unconditional volatility. This has important implications for statistical inference and economic decision making. Lamoureux and Lastrapes (1990), Mikosch and Starica (2004), and Hillebrand (2005), for example, document that ignoring changes in the unconditional volatility can lead to spurious persistence and long-memory effects. In turn, the distortions induced by faulty estimates and inference, affect quantities that are key in economic decision making. Examples include risk estimation (e.g. Andreou and Ghysels, 2008), asset allocation (e.g. Pettenuzzo and Timmermann, 2011), the equity premium (e.g. Pastor and Stambaugh, 2001) and the shape of the option volatility smile (e.g. Bates, 2000).

Let ϵ_t denote an observed financial return (possibly de-meaned), whose variability ϵ_t^2 is governed by

$$\epsilon_t^2 = \sigma_t^2 \eta_t^2, \qquad \sigma_t^2 > 0 \ \ a.s., \qquad \eta_t^2 \geq 0 \ \ a.s., \qquad t = 1, \ldots, T.$$

The σ_t^2 is the volatility, η_t^2 is the squared innovation and T is the sample size. To model changes in the unconditional volatility $E(\sigma_t^2)$, it is common to decompose σ_t^2 multiplicatively as

$$\sigma_t^2 = g_t h_t, \tag{1}$$

where g_t is a non-stochastic "long-term" component, and h_t is a stochastic "short-term" component. An example of h_t is the scaled version of the GARCH(1,1) model of Bollerslev (1986):

$$h_t = \omega + \alpha \phi_t^2 + \beta h_{t-1}, \qquad \phi_t^2 = \epsilon_t^2 / g_t. \tag{2}$$

Other examples of h_t include scaled versions of Stochastic Volatility (SV) models, and scaled versions of Dynamic Conditional Score (DCS) models. See Amado et al. (2019) for a survey of multiplicative decompositions of volatility.

Broadly, there are two approaches to the specification and estimation of time-varying unconditional volatility g_t. In the first approach, estimation of g_t is nonparametric. Examples include Feng (2004), the "Lip" specification in Van Bellegem and Von Sachs (2004), Feng and McNeil (2008), Hafner and Linton (2010), Kim and Kim (2016), and Jiang et al. (2021). In the second approach, g_t is parametrised by a parameter $\boldsymbol{\theta}$. An early example is the piecewise constant specification in Van Bellegem and Von Sachs (2004). For estimation, they proposed the sample variance of each constant period under the assumption that break-locations are known. However, asymptotic methods for the joint estimation and inference of multiple break-sizes were not considered. Engle and Rangel (2008) (without regressors), and Brownlees and Gallo (2010), specify g_t as a deterministic spline function. The former use Gaussian Maximum Likelihood (ML) for estimation, whereas the latter employ penalised ML. However, no asymptotic results are established in either work. Zhang et al. (2020) derive asymptotic results for a least squares estimator of B-splines. In a series of papers, see e.g. Amado and Teräsvirta (2013), Amado and Teräsvirta (2014), and Amado and Teräsvirta (2017), g_t is specified as a smooth transition function. In all of these papers the Gaussian Quasi ML Estimator (QMLE) is used to estimate the parameter $\boldsymbol{\theta}$ in the first step of an iterative estimation algorithm. However, consistency of the first step Gaussian QMLE is only proved under the restrictive and unrealistic assumption that $\phi_t^2 \equiv \epsilon_t^2/g_t$ is *iid*, see the assumption that $h_t = 1$ for all t in Theorem 1 of Amado and Teräsvirta (2013, p. 145). Also, neither consistency nor asymptotic normality of their iterative estimator is established formally (their Theorem 2 on p. 146 is on the infeasible two-step estimator where the parameter values from the first step are known). To accommodate the possibility of cyclical patterns, Andersen and Bollerslev (1997), and Mazur and Pipien (2012), specify (3) as a Fourier Flexible Form (FFF). In the former estimation is by a least squares procedure (see their Appendix B), and in the latter Bayesian methods are used. No asymptotic results are derived in either work. Escribano and Sucarrat (2018) propose a log-linear version of g_t, and use least squares methods to estimate the parameter $\boldsymbol{\theta}$. However, they obtain no asymptotic results.

Here, in this paper, g_t is parametrised by a finite dimensional parameter $\boldsymbol{\theta}$, and re-scaled time $t/T \in [0,1]$. Formally,

$$g_t = g(\boldsymbol{\theta}, t/T). \tag{3}$$

We prove that the Gaussian QMLE provides Consistent and Asymptotically Normal (CAN) estimates of $\boldsymbol{\theta}$ for a large class of specifications contained in (3). In particular, most of the parametric specifications of the literature review above are contained in (3) (see Section 3 for specific examples). For consistency (Theorem 1), $g(\boldsymbol{\theta}, t/T)$ is assumed to be strictly non-negative and bounded, and continuously first order differentiable in $\boldsymbol{\theta}$ for all $t/T \in [0,1]$, see assumption A 2. For asymptotic normality (Theorem 2), we also rely on three times continous differentiability in $\boldsymbol{\theta}$, see assumption A 7. Next, we derive a simple but robust and consistent positive definite estimator of the asymptotic coefficient covariance (Corollary 1 and Theorem 3). Our results are characterised by several attractive properties. First, there is no need to specify – or know – the exact specification of the stochastic component $\phi_t^2 \equiv \epsilon_t^2/g_t$. We only rely on fairly mild mixing assumptions. So our results hold for a large class of specifications of ϕ_t^2, including GARCH models, Stochastic Volatility (SV) models and Dynamic

Conditional Score (DCS) models. This contrasts with several previous results, where the expression of the coefficient covariance is derived under specific assumptions on the structure of ϕ_t^2. Nevertheless, in empirical applications our results can also be used in multi-step estimation procedures, where the parameters of h_t – e.g. a GARCH model – are estimated in a second step (Section 4.1 contains a numerical illustration). Second, a novel and attractive property of our results is that we do not rely on the stochastic component $\{\phi_t^2\}$ being strictly stationary. This is important empirically, since recent studies reveal the zero-process of financial returns – both daily and intradaily – is frequently nonstationary, see e.g. Kolokolov et al. (2020), Sucarrat and Grønneberg (2020), and Francq and Sucarrat (2021). In other words, the common practice of scaling ϵ_t^2 by $E(\epsilon_t^2)$ will *not* result in $\epsilon_t^2/E(\epsilon_t^2)$ being strictly stationary (Section 4.2 contains an empirical illustration). Third, in analogy to equation-by-equation estimation of multivariate volatility models – see Francq and Zakoïan (2016), we use our results to derive a period-by-period estimator of time-varying cyclical volatility (Section 2.4). This means our results can be used to both model and test whether periodic volatility is time-varying (Section 4.3 contains an empirical illustration). Fourth, another attractive property of our results, due to the assumptions we rely upon, is that the Multiplicative Error Model (MEM) interpretation of volatility models holds straightforwardly. The reason is that our assumptions are on ϵ_t^2, not on ϵ_t. Accordingly, our results can also be used to model the time-varying unconditional mean of non-negative processes like volume, duration, realised volatility, dividends, unemployment, and so on by simply interpreting ϵ_t^2 as the non-negative variable in question. Finally, a possible drawback with our estimator is that it may be less efficient asymptotically than alternatives that estimate the parameters of both the deterministic and stochastic component. An example is the iterative estimator proposed by Amado and Teräsvirta (2013). However, our simulations (see Section 4.1) do not support the conjecture that the iterative estimator is more efficient.

The rest of the paper is organised as follows. The next section, Section 2, contains our main results, and a presentation and discussion of the assumptions they rely on. Section 3 gives examples of g_t specifications that are contained in (3). Section 4 contains numerical illustrations of our results, whereas Section 5 concludes. The appendix contains the proofs of our main results.

2 Consistency and Asymptotic Normality

2.1 Consistency

Our objective function, the normal log-likelihood, is given by

$$L_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} l_t(\boldsymbol{\theta}, \epsilon_t^2), \qquad l_t(\boldsymbol{\theta}, \epsilon_t^2) = \ln g_t(\boldsymbol{\theta}) + \frac{\epsilon_t^2}{g_t(\boldsymbol{\theta})},$$

and minimisation of this function leads to the Quasi Maximum Likelihood Estimator (QMLE):

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_T(\boldsymbol{\theta}). \tag{4}$$

Let $\boldsymbol{\theta}_0$ denote the true parameter value. Our proof of consistency relies on the following assumptions.

A 1 $\boldsymbol{\Theta}$ *is compact.*

A 2 *Let* $\boldsymbol{\Theta}^*$ *be an open and convex set that contains* $\boldsymbol{\Theta}$*, that is,* $\boldsymbol{\Theta} \subsetneq \boldsymbol{\Theta}^*$*, and let* $\{g_t\}$ *be a non-stochastic process,* $g_t = g(\boldsymbol{\theta}, t/T)$*,* $g : \boldsymbol{\Theta}^* \times [0,1] \to \mathbb{R}$*. For all* $t/T \in [0,1]$*:*

a) g_t *is bounded and strictly positive:* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}^*} g_t < \infty$ *and* $\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}^*} g_t > 0$*;*

b) $g_t(\boldsymbol{\theta})$ *is continuously differentiable on* $\boldsymbol{\Theta}^*$*.*

A 3 *Let* $\{\epsilon_t^2\}$ *be a stochastic process that is* ϕ*-mixing of size* $-r/(2r-1), r \geq 1$*, or* α*-mixing of size* $-r/(r-1), r > 1$*, and let* $\epsilon_t^2 \geq 0$ *a.s. for all* t*.*

A 4 *Let* $\phi_t^2 \equiv \epsilon_t^2/g_t(\boldsymbol{\theta}_0)$ *be a non-degenerate random variable such that:*

a) $E(\phi_t^2) = 1$ *for all* t*;*

b) $E|\phi_t^2|^{r+\varepsilon} < \infty$ *for all* t*, where* $r > 1$ *and* $\varepsilon > 0$*.*

A 5 $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\big(l_t(\boldsymbol{\theta}, \epsilon_t^2)\big)$ *attains a unique minimum at* $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$*.*

A 1 is a standard assumption. A 2 defines the class of g_t functions we consider, which is very general. Section 3 gives some specific examples. The open and convex set $\boldsymbol{\Theta}^*$, and the differentiability assumed in A 2b), are needed because the score is part of the dominating function we use in our proof of uniform convergence. A 3 is a fairly mild dependence assumption. In particular, it is substantially milder than the assumptions used by Amado and Teräsvirta (2013) in their derivations, since they rely on $\{\phi_t^2\}$

being *iid*, see their Theorem 1 on p. 145 just below equation (15). Here, by contrast, A 3 can be compatible with any volatility model of ϕ_t^2, stationary or not, as long as the mixing conditions are satisfied. This means our results apply not only to standard models within the ARCH, GAS and SV classes, but also to semi-strong volatility models, see e.g. Escanciano (2009) and Francq and Thieu (2018), and to models that are only weakly identified as models of the variance (e.g. intraday high-frequency measures of volatility), see Sucarrat (2021). The series $\{\epsilon_t^2\}$ being α- or ϕ-mixing of size $-a$ means $\alpha(m) = O(m^{-a-\varepsilon})$ for some $\varepsilon > 0$.[4] For both α- and ϕ-mixing, the greater the r, the greater the dependence.

A 4a) is a very mild identification assumption. The reason the assumption is mild is that almost all volatility models are invariant to scale-tranformations in the sense that there exists a finite constant $c > 0$ such that the stochastic process $\{\phi_t^{2*}\}$ with $E(\phi_t^{2*}) = \mu$ for all t satisfies $E(c\phi_t^{2*}) = cE(\phi_t^{2*}) = 1$ for all t. For volatility models that are not invariant to scale transformations in this sense, in particular those whose stability conditions are affected by scaling (e.g. the Dynamic Conditional Score (DCS) model of Harvey and Sucarrat (2014)), the condition $E(\phi_t^2) = 1$ may be restrictive. It should also be noted that A 4a) is compatible with $\{\phi_t^2\}$ being nonstationary. A case in point is the common situation where the zero-process of a financial return is nonstationary, see e.g. Sucarrat and Grønneberg (2020), and Francq and Sucarrat (2021). In particular, part ii) of Proposition 2.1 in Sucarrat and Grønneberg (2020) implies that $E(\phi_t^2)$ can be constant over time even though the zero-process of a financial return is nonstationary. Another implication of A 4a) is that $E(\epsilon_t^2) = g_t(\boldsymbol{\theta}_0)$. A 4b) is also a fairly mild moment assumption. For example, it holds when $\{h_t\}$ is a stationary de-volatilitised GARCH(1,1), as in (2), with finite $E(\phi_t^4)$. A finite fourth moment is needed for standard inference on the parameters, see Francq and Zakoïan (2019). A finite fourth moment is, however, more restrictive than the usual second moment requirement for consistency in the standard case.

Finally, A 5 is a standard regularity assumption. Note that our proof of consistency shows that $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\big(l_t(\boldsymbol{\theta}, \epsilon_t^2)\big)$ exists for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. So A 5 simply ensures there exists a unique minimiser $\boldsymbol{\theta}_0$.

Theorem 1 (consistency) *Suppose A 1 – A 5 hold. Then* $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$.

Proof: See Appendix A.1.

[4] That is, there exists a $\Delta > 0$ and a finite integer N such that $\left|\frac{\alpha(m)}{m^{-a-\varepsilon}}\right| < \Delta$ for all $m \geq N$, see White (2001, Definition 3.45 on p. 49).

2.2 Asymptotic normality

Let int($\mathbf{\Theta}$) denote the interior of $\mathbf{\Theta}$. For Asymptotic Normality (AN), we rely on the following additional assumptions:

A 6 *The true parameter $\boldsymbol{\theta}_0$ lies in int($\mathbf{\Theta}$).*

A 7 *Let g_t be as in A 2. For all t and $t/T \in [0,1]$: $g_t(\boldsymbol{\theta})$ is three times continuously differentiable on $\mathbf{\Theta}^*$.*

A 8 *Let $\boldsymbol{H}_t(\boldsymbol{\theta}, \epsilon_t^2)\big) \equiv \partial^2 l_t(\boldsymbol{\theta}, \epsilon_t^2)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$. The limit $\boldsymbol{A} = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\big(\boldsymbol{H}_t(\boldsymbol{\theta}_0, \epsilon_t^2)\big)$ is positive definite.*

A 9 *$E|\phi_t|^{4+\nu} < \infty$ for all t with $\nu > 2(r-1)$, where $r > 1$ determines the α-mixing size in A 3.*

A 10 *The limit $\boldsymbol{B} = \lim_{T\to\infty} Var\big(T^{-1/2}\sum_{t=1}^{T} \boldsymbol{s}_t(\boldsymbol{\theta}_0)\big)$ is positive definite.*

Assumptions A 6 is standard. In A 7, the third order differentiability requirement is needed, since the dominating function that we use to show that each entry of the Hessian satisfies the Uniform Law of Large Numbers (UWLLN) contains the partial derivatives of the Hessian. In the course of proving UWLLN, we show that the limit in A 8 exists for all $\boldsymbol{\theta} \in \mathbf{\Theta}$. So A 8 simply ensures the limit is positive definite. In A 9, the value $r > 1$ determines the α-mixing size in A 3, so $\nu > 0$. This means a moment higher than the 4th. must exist as a minimum, and that the number of existing moments that are required depends on the α-mixing size: The more dependence (i.e. the higher r is), the more moments are required. The limit in A 10 exists due to the preceding assumptions (see the proof of Theorem 2). So A 10 simply ensures the limit is positive definite.

Theorem 2 (asymptotic normality) *Suppose assumptions 6 – 10 hold in addition to the assumptions of Theorem 1. Then $\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N\big(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\big)$.*

Proof: See Appendix A.2.

Note that a direct consequence of Theorem 2 is that

$$\widehat{\boldsymbol{A}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{H}_t(\widehat{\boldsymbol{\theta}}, \epsilon_t^2)$$

provides a consistent estimate of $\boldsymbol{A} = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\big(\boldsymbol{H}_t(\boldsymbol{\theta}_0, \epsilon_t^2)\big)$.

Corollary 1 (Consistent estimation of $\boldsymbol{A}$) *Suppose assumptions A 6 – A 7 hold in addition to those of Theorem 1. Then $\widehat{\boldsymbol{A}} \xrightarrow{p} \boldsymbol{A}$ as $T \to \infty$.*

Proof. In the proof of Theorem 2, assumptions A 6 – A 7 in addition to those of Theorem 1 are used to show that $\{\boldsymbol{H}_t(\boldsymbol{\theta}, \epsilon_t^2)\}$ satisfies the UWLLN (Theorem 4.2 of Wooldridge, 1994). By Lemma A.1 in Wooldridge (1994, p. 2727) it thus follows that $\widehat{\boldsymbol{A}} \xrightarrow{p} \boldsymbol{A}$. □

2.3 Estimation of $\boldsymbol{B}$

The general form of our estimator is given by

$$
\begin{aligned}
\widehat{\boldsymbol{B}} &= \sum_{j=-T}^{T} k(j/S_T)\widehat{\boldsymbol{\Gamma}}(j), \\
\widehat{\boldsymbol{\Gamma}}(j) &= \frac{1}{T}\sum_{t=1}^{T-j} \widehat{\boldsymbol{s}}_t \widehat{\boldsymbol{s}}'_{t+j}, \qquad j \geq 0, \\
\widehat{\boldsymbol{\Gamma}}(j) &= \widehat{\boldsymbol{\Gamma}}(-j)', \qquad j < 0,
\end{aligned}
$$

where the $k(\cdot)$'s are kernel weights, and S_T is the bandwidth. For consistent estimation of $\boldsymbol{B}$, we rely on the following additional assumptions:

A 11 (Kernel) *For all $x \in \mathbb{R}$, $|k(x)| \leq 1$ and $k(x) = k(-x)$; $k(0) = 1$; $k(x)$ is continuous at zero and for almost all $x \in \mathbb{R}$; $\int_{\mathbb{R}} |k(x)|dx < \infty$.*

A 12 (Bandwidth) *$S_T \to \infty$, and for some $q \in (1/2, \infty)$, $S_T^{1+2q}/T = O(1)$.*

A 13 *For some $u \in (2, 4]$ such that $u > 2 + 1/q$, and some $p > u$:*

a) $2(1/u - 1/p) > (r-1)/r$, $\quad r > 1$, *where $-r/(r-1)$ is the α-mixing size in A 3;*

b) $E\big(|\phi_t^2|^p\big) < \infty$.

Most kernels considered in the literature satisfy A 11. This includes, amongst other, the Bartlett kernel (set $k(x) = 1 - |x|$ for $x \leq 1$), the Parzen kernel and the Quadratic Spectral kernel. A 13a) introduces mixing requirements beyond those of A 3 related to the bandwidth parameter q: The greater the dependence (i.e. the greater the r), the greater q must be for the conditions in A 13a) to hold. In A 13b) there is a trade-off between q and the number of moments required for ϕ_t^2, since $p > u$: The lower q is, the more moments are required.

Theorem 3 (Consistent estimation of $\boldsymbol{B}$) *Suppose assumptions A 11 – A 13 hold in addition to those of Theorems 1 and 2. Then* $\widehat{\boldsymbol{B}} \xrightarrow{p} \boldsymbol{B}$.

Proof: See Appendix A.3.

2.4 Time-varying periodic unconditional volatility

To model intraday periodic volatility, Andersen and Bollerslev (1997), and Mazur and Pipien (2012) specify $g_t(\boldsymbol{\theta})$ as a Fourier Flexible Form (FFF) in terms of nominal time t. No asymptotic results are established in either work. While it may be possible to extend our results in this direction, an easier and more general approach is to derive a period-by-period estimator. Period-by-period estimation can be viewed as the periodic analog of equation-by-equation estimation of multivariate volatility, see e.g. Francq and Zakoïan (2016). An example is Escribano and Sucarrat (2018). Also, the common practice of estimating the intraday unconditional volatilities with cross-day averages of squared return is a special case of period-by-period estimation.

Let $\epsilon_{m,t}$ denote the period m return at time t. For example, if $\epsilon_{m,t}$ is the hour m return in day t of an exchange rate that is traded 24-hours a day, then $m = 1, \ldots, 24$. Let M denote the number of periods at time t, and let

$$\epsilon_{m,t}^2 = g_{m,t}(\boldsymbol{\theta}_m)\phi_m^2, \qquad t = 1, 2, \ldots, \qquad \text{for} \qquad m = 1, \ldots, M.$$

The process $\{\boldsymbol{\epsilon}_t\}_{t=1}^{\infty}$ with $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \ldots, \epsilon_{M,t})'$ is thus an M-dimensional multivariate process. Let

$$\widehat{\boldsymbol{\theta}}_m = \underset{\boldsymbol{\theta}_m \in \boldsymbol{\Theta}_m}{\arg\min}\ L_T(\boldsymbol{\theta}_m), \qquad L_T(\boldsymbol{\theta}_m) = \frac{1}{T}\sum_{t=1}^{T} l_t(\boldsymbol{\theta}_m, \epsilon_{m,t}^2),$$

denote the period m estimator, where $l_{m,t}(\boldsymbol{\theta}_m, \epsilon_{m,t}^2) = \ln g_{m,t}(\boldsymbol{\theta}_m) + \epsilon_{m,t}^2 / g_{m,t}(\boldsymbol{\theta}_m)$. The following corollary is a direct consequence of Theorems 1 – 3, and Corollary 1.

Corollary 2 (period-by-period estimation) *Suppose* $\boldsymbol{\theta}_{m0}$, $\boldsymbol{\Theta}_m$, $\{g_{mt}\}$ *and* $\{\epsilon_{m,t}^2\}$ *satisfy:*

a) A 1 – A 5 for each $m = 1, \ldots, M$. *Then* $\widehat{\boldsymbol{\theta}}_m \xrightarrow{p} \boldsymbol{\theta}_{m0}$, $m = 1, \ldots, M$.

b) A 1 – A 10 for each $m = 1, \ldots, M$. *Then*

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m0}) \xrightarrow{d} N\big(\boldsymbol{0}, \boldsymbol{A}_m^{-1}\boldsymbol{B}_m\boldsymbol{A}_m^{-1}\big), \qquad m = 1, \ldots, M.$$

c) A 1 – A 7 for each $m = 1, \dots, M$*. Then* $\widehat{\boldsymbol{A}}_m \xrightarrow{p} \boldsymbol{A}_m$*,* $m = 1, \dots, M$*.*

d) A 1 – A 13 for each $m = 1, \dots, M$*. Then* $\widehat{\boldsymbol{B}}_m \xrightarrow{p} \boldsymbol{B}_m$*,* $m = 1, \dots, M$*.*

The period-by-period estimator is thus the generalisation of an estimator that is frequently used in empirical applications. Consider, for example, the intraday returns $\epsilon_{m,t}$, $m = 1, \dots, M$ of day t. Often, the sample averages $T^{-1}\sum_{t=1}^{T}\epsilon_{m,t}^2$, $m = 1, \dots, M$, are used to estimate the intraday unconditional volatilities $E(\epsilon_{1,t}^2), \dots, E(\epsilon_{M,t}^2)$. This collection of sample averages is a special case of the period-by-period estimator. However, it is only consistent in the special case where the unconditional intraday volatilities are constant across days, i.e. for each $m = 1, \dots, M$ we have $E(\epsilon_{m,t_1}^2) = E(\epsilon_{m,t_2}^2)$ for all t_1, t_2. By contrast, the period-by-period estimator above can also be used to estimate unconditional intraday volatilities that vary across days (Section 4.3 contains an empirical illustration).

3 Examples of $g_t(\boldsymbol{\theta})$

Here we provide some examples of $g_t(\boldsymbol{\theta})$. The main focus is on whether the key assumptions A 2 and A 7 hold.

3.1 Smooth transition models

A variety of smooth transition models have been considered, see Amado and Teräsvirta (2013) for a survey. Amado and Teräsvirta (2013) consider the following in more detail:

$$g(\boldsymbol{\theta}, t/T) = \delta_0 + \sum_{l=1}^{s}\delta_l G_l(\gamma_l, c_l, t/T), \quad G_l(\gamma_l, c_l, t/T) = \frac{1}{1 + \exp\big(-\gamma_l(t/T - c_l)\big)}, \tag{5}$$

where $\boldsymbol{\theta} = (\boldsymbol{\delta}', \boldsymbol{\gamma}', \boldsymbol{c}')'$ with $\boldsymbol{\delta} = (\delta_0, \delta_1, \dots, \delta_s)'$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_s)'$ and $\boldsymbol{c} = (c_1, \dots, c_s)'$. The δ_l is the total size of break l, γ_l is the speed of transition of break l, c_l is the location of break l and s is the number of breaks. If $\delta_1 = \cdots = \delta_s = 0$, then there are no breaks. A 1 entails that the parameter space $\boldsymbol{\Theta} = \{\Delta \times \Gamma \times C\}$ is compact, where $\boldsymbol{\delta} \in \Delta, \boldsymbol{\gamma} \in \Gamma$ and $\boldsymbol{c} \in C$. Lemma A.1 in Amado and Teräsvirta (2013, p. 150) contains the derivatives of g_t with respect to $\boldsymbol{\theta}$. If the open and convex set $\boldsymbol{\Theta}^*$ in A 2 is suitably bounded, then all the characteristics of A 2 hold. Lemma A.2 in Amado and Teräsvirta (2013, p. 150) contains the second derivatives of g_t with respect to $\boldsymbol{\theta}$. It is straightforward to verify that also the third order derivatives are continuous on a suitably bounded open and convex set $\boldsymbol{\Theta}^*$. Accordingly, A 7 also hold.

3.2 Piecewise constant models

Van Bellegem and Von Sachs (2004) specify $g_t(\boldsymbol{\theta})$ as piecewise constant. This amounts to

$$g_t(\boldsymbol{\theta}) = \delta_0 + \sum_{l=1}^{s} \delta_l I(t/T \geq c_l), \qquad \boldsymbol{\theta} = (\delta_0, \delta_1, \dots, \delta_s)', \tag{6}$$

where $I(A)$ is an indicator function equal to 1 if A holds and 0 otherwise. The values of the possible break-locations $c_1, \dots, c_s$ are known. For estimation, Van Bellegem and Von Sachs (2004) proposed the sample variance of each constant period. They did not consider asymptotic methods for the joint estimation and inference of multiple break-sizes. Our results do not permit the c_i's to be estimated in (6), since g_t is not continuously differentiable with respect to the c_i's. It is straightforward to verify that, for all $t/T \in [0,1]$, $g_t(\boldsymbol{\theta})$ is continuously first, second and third order differentible with respect to $\boldsymbol{\theta}$ on $\mathbb{R}^K$. So A 2 and A 7 hold.

In Escribano and Sucarrat (2018), $g_t(\boldsymbol{\theta})$ is specified as a generic log-linear function. Least Squares (LS) methods are used for estimation, but no asymptotic results are established. The log-linear version of a piecewise constant specification is an example of a model contained in their class of models:

$$g_t(\boldsymbol{\theta}) = \exp\left(\delta_0 + \sum_{l=1}^{s} \delta_l \cdot I(t/T \geq c_l)\right), \qquad \boldsymbol{\theta} = (\delta_0, \delta_1, \dots, \delta_s)'. \tag{7}$$

The advantage of this specification over (6) is that non-negativity constraints on $\boldsymbol{\theta}$ are not needed. Again, it is straightforward to verify that, for all $t/T \in [0,1]$, $g_t(\boldsymbol{\theta})$ is continuously first, second and third order differentible with respect to $\boldsymbol{\theta}$ on a suitably bounded open and convex set $\boldsymbol{\Theta}^*$. So A 2 and A 7 hold.

3.3 Splines

Engle and Rangel (2008), and Brownlees and Gallo (2010), specify $g_t(\boldsymbol{\theta})$ as a deterministic spline. The former use Gaussian ML for estimation, whereas the latter employs penalised ML. However, no asymptotic results are established in either work. Zhang et al. (2020) derive asymptotic results for a least squares estimator of B-splines.

Splines that are suitably expressed in terms of re-scaled time can satisfy A 2 and A 7. An example is the exponential quadratic spline function considered by Engle and Rangel (2008) (without regressors).

If we remove the trend and replace nominal time with re-scaled time, then we obtain

$$g_t(\boldsymbol{\theta}) = \exp\left(\delta_0 + \sum_{l=1}^{s} \delta_l (t/T - c_l)^2 I(t/T \geq c_l)\right), \qquad \boldsymbol{\theta} = (\delta_0, \delta_1, \ldots, \delta_s)',$$

where $I(A)$ is an indicator function equal to 1 if A holds and 0 otherwise, and the c_l's are given knot-locations. Typically, to facilitate estimation, the c_l's are assumed to be equidistant from each other. However, this is not a requirement. The value s is the number of knots, and $\delta_1, \ldots, \delta_s$ are the knot-coefficients. Large values of s imply more frequent cycles, and the sharpness of each cycle is governed by the knot-coefficients. Let $\tau(t/T, c) = (t/T - c)^2 I(t/T \geq c)$. Note that $\tau(t/T, c) \geq 0$ for all $t/T, c \in [0, 1]$. For all $t/T \in [0, 1]$, $g_t(\boldsymbol{\theta})$ is continuously first, second and third order differentiable with respect to $\boldsymbol{\theta}$ on a suitably bounded open and convex set $\boldsymbol{\Theta}^*$. So A 2 and A 7 hold.

4 Numerical illustrations

4.1 Comparison with the iterative estimator of Amado and Teräsvirta (2013)

Our estimator coincides with the first step of the iterative estimator proposed by Amado and Teräsvirta (2013). Their iterative estimator is an adaption of the maximisation by part algorithm developed by Song et al. (2005) for independent data. A reasonable conjecture is that the iterative estimator of Amado and Teräsvirta (2013) is more efficient asymptotically than our estimator. However, the iterative estimator is also likely to be more fragile numerically in finite samples, and substantially slower in relative terms. To shed light on this, we run a simulation experiment where the Data Generating Process (DGP) is

$$\begin{aligned}
\epsilon_t &= \sigma_t \eta_t, \qquad \sigma_t^2 = g_t h_t, \qquad \eta_t \overset{iid}{\sim} N(0,1), \qquad t = 1, \ldots, T, \\
g_t &= \delta_0 + \frac{\delta_1}{1 + \exp\big(-\gamma(t/T - c)\big)}, \qquad (\delta_0, \delta_1, \gamma, c) = (1, 1, 10, 0.5), \\
h_t &= \omega + \alpha \phi_{t-1}^2 + \beta h_{t-1}, \qquad \phi_t^2 = h_t \eta_t^2, \qquad (\omega, \alpha, \beta) = (0.1, 0.1, 0.8).
\end{aligned}$$

Thus, the time-varying unconditional volatility g_t is governed by a single break centered about the middle of the sample, i.e. $c = 0.5$, and the stochastic component ϕ_t^2 is governed by a strictly stationary and ergodic scaled GARCH(1,1).

Table 1 contains the comparison of the g_t parameters. As expected, both the bias and standard deviation of the estimates fall as the sample size T increases. However, the results do not suggest the

Iterative estimator is asymptotically more efficient. That is, as the sample size becomes large (i.e. $T \geq 10000$), the evidence does not suggest it becomes more efficient at T increases. For the smallest sample size we investigate ($T = 1000$), the results suggest the Iterative estimator is substantially less efficient. For the intermediate sample sizes, the results are mixed. In sum, we do not find clear support of the hypothesis that the Iterative estimator is more efficient asymptotically for the g_t parameters.

Table 2 contains the comparison of the h_t parameters. In our case, these estimates are obtained via a two-step estimation procedure. Again, as expected, the bias and standard deviation both fall as the sample size increases. When the sample size becomes large (i.e. $T \geq 10000$), the evidence does not suggest – in a clear way – that it becomes more efficient at T increases. Instead, if anything, the evidence suggest the estimators are equally efficient asymptotically, since the differences are so small that numerical simulation error cannot be ruled out. For the smaller and intermediate sample sizes that we investigate ($T = 1000$ to $T = 5000$), the results are more mixed. In sum, we do not find clear support of the hypothesis that the Iterative estimator is more efficient asymptotically for the h_t parameters than a two-step estimator.

4.2 Daily return with a non-stationary zero-process

An attractive feature of our estimator is that the stochastic component ϕ_t^2 need not be stationary. To illustrate this, we revisit one of the daily stock returns investigated by Sucarrat and Grønneberg (2020). Eros International plc., whose market ticker is EROS, was an Indian multinational global mass media conglomerate (a "Bollywood" company) that merged with the US company STX Entertainment in April 2020. The left graph of Figure 1 depicts the daily returns at the New York Stock Exchange (NYSE) from 21 December 2009 to 4 February 2019 ($T = 2295$). The datasource is Bloomberg. In the beginning of the period the primary listing of the stock was in India. This explains all the zeros in the series until November 2013. Thereafter, there are few zeros. The return series thus exhibits a clear break in the unconditional zero-probability, and so its zero-process is non-stationary. Accordingly, the return process and the transformation $\phi_t^2 = \epsilon_t^2/E(\epsilon_t^2)$ are therefore also non-stationary.

Interestingly, the 500-day moving average of squared return in the right graph of Figure 1 does not suggest in a clear way that there is a break in the unconditional volatility in November 2013. Instead, the graph suggests the break or breaks occur later, namely in October 2015 and in October 2017. To illustrate the estimation of a piecewise constant log-linear specification g_t, we will use it to investigate whether there are breaks at the aforementioned points. The advantage of piecewise constant specifications over smooth transition models is that the latter are not identified if the number of transition terms exceeds the number of breaks. Moreover, closer inspection reveals that the possible break locations can be identified with a fairly high degree of precision, i.e. the breaks are quite abrupt. This justifies the usage of a piecewise constant specification. Specifically, the data suggest the possible break-locations are 11 November 2013, 14 October 2015 and 6 October 2017, respectively. In terms of re-scaled time these correspond to $(c_1, c_2, c_3)' = (0.427, 0.638, 0.855)$, and the estimated model is

$$\widehat{\ln g_t} = \underset{(0.4174)}{1.795} + \underset{(0.4355)}{0.351} I(t/T \geq c_1) + \underset{(0.2342)}{1.215} I(t/T \geq c_2) - \underset{(0.2545)}{0.912} I(t/T \geq c_3).$$

The numbers in parentheses are the standard errors of the estimates. These are computed as the square root of the diagonal of $\widehat{\boldsymbol{\Sigma}}/T$, where $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{A}}^{-1}$ is the estimate of the asymptotic coefficient covariance. A Bartlett kernel is used in the computation of $\widehat{\boldsymbol{B}}$, and the bandwidth is obtained as the integer part of $4(T/100)^{(2/9)}$. The t-ratios of the break-size estimates are $0.351/0.4355 = 0.806$, $1.215/0.2345 = 5.181$ and $-0.912/0.2545 = -3.583$, respectively. So two-sided t-tests at common significance levels (i.e. 10%, 5% and 1%) suggest there are breaks at c_2 and c_3, but not at c_1.

4.3 Time-varying intraday periodic volatility

The period-by-period estimator of Section 2.4 can be used to estimate time-varying periodic volatility. Here, we illustrate this for intraday hourly USD/EUR exchange rate volatility. Let $S_{m,t}$ denote the exchange rate at the end of hour m in day t, and let $\epsilon_{m,t} = 100^2 \cdot (\ln S_{m,t} - \ln S_{m-1,t})$ denote the hour m log-return denominated in basis points. The left graph of Figure 2 plots the hourly returns at Forexite, a currency trading platform, from 2 January 2017 to 31 December 2018 (12 184 hourly returns). Only trading days are included in the sample (i.e. weekends are excluded), and a trading day contains $M = 24$ returns. The first return of a trading day covers the interval from 00:00 CET to 01:00 CET, whereas the last covers 23:00 CET to 00:00 CET. The right graph of Figure 2 contains the sample averages of squared returns across days, i.e. $T_m^{-1}\sum_{t=1}^{T_m}\epsilon_{m,t}^2$, where T_m is the number of observations available for period m. If the period m unconditional volatilities are constant across days, then the sample averages are consistent. As is clear from the graph, the intraday hourly unconditional volatility is time-varying. It is at its lowest at the end of the day at 24h CET, and it is at its highest at 15h CET.

To shed light on whether the intraday unconditional volatilites are constant across days, we estimate a quadratic spline function a la Engle and Rangel (2008) with re-scaled time and four knots at equidistant locations, i.e.

$$\widehat{\ln g}_{m,t} = \delta_{m,0} + \sum_{l=1}^{4} \delta_{m,l}(t/T - c_l)^2 I(t/T \geq c_l), \qquad (c_1, c_2, c_3, c_4) = (0.2, 0.4, 0.6, 0.8),$$

for each period $m = 1, \ldots, M$. Table 3 contains the estimation results together with a Wald-test of $H_0 : \delta_1 = \cdots = \delta_4 = 0$. So under the null the unconditional volatility of period m is constant and equal to $g_{m,t} = \exp(\delta_{m,0})$ for all t. The p-values of the test are contained in the square brackets of the last column. Out of the 24 tests, 8 reject the null at the 5% significance level, and 4 reject the null at 1%. Without time-varying period m volatilities, we should on average expect 1.2 rejections at 5%, and 0.24 rejections at 1%. Accordingly, the results support the hypothesis that some of the unconditional intraday volatilites are time-varying across days.

5 Conclusions

The unconditional volatility of financial return is frequently time-varying. Let $g_t(\boldsymbol{\theta})$ denote the time-varying unconditional volatility as a function of a finite-dimensional parameter $\boldsymbol{\theta}$. We establish the Consistency and Asymptotic Normality (CAN) of the (Gaussian) QMLE for a large class of specifications g_t. For consistency, continuous first order differentiability in $\boldsymbol{\theta}$ is the main requirement for our results to be applicable, whereas for asymptotic normality continuous third order differentiability is the main requirement. We also derive a simple but robust and consistent positive definite estimator of the asymptotic coefficient covariance. Our results are characterised by several attractive properties. First, there is no need to specify – or know – the exact specification of the stochastic component (we only rely on fairly mild mixing assumptions). So our results hold for a large class of specifications of the stochastic component, including GARCH models, Stochastic Volatility (SV) models and Dynamic Conditional Score (DCS) models. Nevertheless, in empirical applications our results can also be used in multi-step estimation procedures, where the parameters of the stochastic component – e.g. a GARCH model – are estimated in a second step. Another attractive property of our results is that the stochastic component does not have to be strictly stationary. This is important empirically, since recent studies reveal the zero-process of financial returns – both daily and intradaily – is frequently nonstationary. Third, we use our results to derive a period-by-period estimator of time-varying periodic volatility. A fourth attractive property of our results, due to the assumptions we rely upon, is that the Multiplicative Error Model (MEM) interpretation of volatility models holds straightforwardly. Accordingly, our results can also be used to model the time-varying unconditional mean of non-negative processes like volume, duration, realised volatility, dividends, unemployment, and so on. Finally, a possible drawback with our estimator is that it may be less efficient asymptotically than alternatives that estimate the parameters of both the deterministic and stochastic compoments. An example is the iterative estimator proposed by Amado and Teräsvirta (2013). However, our simulations do not provide support of the conjecture that the iterative estimator is more efficient.

References

Amado, C., A. Silvennoinen, and T. Teräsvirta (2019). Models with Multiplicative Decomposition of Conditional Variances and Correlations. In J. Chevallier, S. Goutte, D. Guerreiro, S. Saglio and B. Sanhadji (eds.): *Financial Mathematics, Volatility and Covariance Modelling, Volume 2*.

Amado, C. and T. Teräsvirta (2013). Modelling volatility by variance decomposition. *Journal of Econometrics 175*, 142–153.

Amado, C. and T. Teräsvirta (2014). Modelling Changes in the unconditional variance of long stock return series. *Journal of Empirical Finance 25*, 15–35.

Amado, C. and T. Teräsvirta (2017). Specification and testing of multiplicative time-varying GARCH models with applications. *Econometric Reviews 36*, 421–446.

Andersen, T. G. and T. Bollerslev (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance 4*, 115–158.

Andreou, E. and E. Ghysels (2008). Quality control for structural credit risk models. *Journal of Econometrics 146*, 364–375.

Bates, D. S. (2000). Post-'87 crash fears in the S& P 500 futures option market. *Journal of Econometrics 94*, 181–238.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics 31*, 307–327.

Brownlees, C. and G. Gallo (2010). Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics 8*, 29–56.

Campos-Martins, S. and G. Sucarrat (2021). *tvgarch: Time Varying GARCH Modelling*. R package version 2.2.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.

Engle, R. F. and J. G. Rangel (2008). The Spline GARCH Model for Low Frequency Volatility and its Global Macroeconomic Causes. *Review of Financial Studies 21*, 1187–1222.

Escanciano, J. C. (2009). Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory 25*, 561–570.

Escribano, Á. and G. Sucarrat (2018). Equation-by-Equation Estimation of Multivariate Periodic Electricity Price Volatility. *Energy Economics 74*, 287–298.

Feng, Y. (2004). Simultaneously modeling the conditional heteroscedasticity and scale change. *Econometric Theory 20*, 563–596.

Feng, Y. and A. J. McNeil (2008). Modeling of scale change, periodicity and conditional heteroskedasticity in return volatility. *Economic Modeling 25*, 850–867.

Francq, C. and G. Sucarrat (2021). Volatility Estimation when the Zero-Process is Nonstationary. *Journal of Business and Economic Statistics*. In press. DOI: https://doi.org/10.1080/07350015.2021.1999821.

Francq, C. and L. Q. Thieu (2018). Qml inference for volatility models with covariates. *Econometric Theory*. https://doi.org/10.1017/S0266466617000512.

Francq, C. and J.-M. Zakoïan (2016). Estimating multivariate volatility models equation by equation. *The Journal of the Royal Statistical Society. Series B 78*, 613–635. Working paper version: MPRA Paper No. 54250: http://mpra.ub.uni-muenchen.de/54250/.

Francq, C. and J.-M. Zakoïan (2019). *GARCH Models*. New York: Wiley. 2nd. Edition.

Hafner, C. and O. Linton (2010). Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics 159*, 55–73.

Hansen, B. E. (1992). Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes. *Econometrica 60*, 697–972.

Hansen, B. E. (2021). *Introduction to Econometrics.* Wisconsin. 9 March 2021. https://www.ssc.wisc.edu/~bhansen/.

Harvey, A. C. and G. Sucarrat (2014). EGARCH models with fat tails, skewness and leverage. *Computational Statistics and Data Analysis 76*, 320–338.

Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *Annals of Probability 12*, 141–153.

Hillebrand, E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics 129*, 121–138.

Jiang, F., D. Li, and K. Zhu (2021). Adaptive inference for a semiparametric generalized autoregressive conditional heteroskedasticity model. *Journal of Econometrics 224*, 306–329.

Kim, K. H. and T. Kim (2016). Capital asset pricing model: A time-varying volatility approach. *Journal of Empirical Finance 37*, 268–281.

Kolokolov, A., G. Livieri, and D. Pirino (2020). Statistical inferences for price staleness. *Journal of Econometrics 218*, 32–81.

Lamoureux, C. G. and W. D. Lastrapes (1990). Persistence in Variance, Sructural Breaks, and the GARCH Model. *Journal of Business and Economic Statistics 8*, 225–234.

Mazur, B. and M. Pipien (2012). On the Empirical Importance of Periodicity in the Volatility of Financial Returns – Time Varying GARCH as a Second Order APC(2) Process. *Central European Journal of Economic Modelling and Econometrics 4*, 95–116.

Mikosch, T. and C. Starica (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *The Review of Economics and Statistics 86*, 378–390.

Pastor, L. and R. F. Stambaugh (2001). The Equity Premium and Structural Breaks. *The Journal of Finance 56*, 1207–1239.

Pettenuzzo, D. and A. Timmermann (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics 164*, 60–78.

R Core Team (2021). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Song, P. X.-K., Y. Fan, and J. D. Kalbfleisch (2005). Maximization by parts in likelihood inference. *Journal of the American Statistical Association 100*, 1145–1158.

Sucarrat, G. (2021). Identification of Volatility Proxies as Expectation of Squared Financial Return. *Forthcoming in International Journal of Forecasting*. DOI: https://doi.org/10.1016/j.ijforecast.2021.03.008.

Sucarrat, G. and S. Grønneberg (2020). Risk Estimation with a Time Varying Probability of Zero Returns. *Journal of Financial Econometrics*. Forthcoming. DOI: https://doi.org/10.1093/jjfinec/nbaa014.

Van Bellegem, S. and R. Von Sachs (2004). Forecasting economic time-series with unconditional time-varying variance. *International Journal of Forecasting 20*, 611–627.

White, H. (2001). *Asymptotic Theory for Econometricians.* Bingley UK: Emerald Group Publishing Ltd. Revised Edition.

Wooldridge, J. M. (1994). Estimation and inference for dependent processes. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume IV.* New York: North-Holland.

Zhang, Y., R. Liu, Q. Shao, and L. Yang (2020). Two-step estimation for time varying ARCH models. *Journal of Time Series Analysis 41*, 551–570.

A Proofs of main results

A.1 Proof of Theorem 1

To establish consistency, we use two theorems from Wooldridge (1994), which rely on the following two definitions:

> **Definition 4.1 (Wooldridge, 1994, p. 2651): WLLN**. A sequence of random variables $\{z_t\}$ satisfies the *weak law of large numbers* (WLLN) if
>
> i) $E(|z_t|) < \infty$ for $t = 1, 2, \ldots$;
>
> ii) $\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(z_t)$ exists;
>
> iii) $T^{-1} \sum_{t=1}^{T} \big(z_t - E(z_t)\big) \xrightarrow{p} 0$.

Definition 4.2 (Wooldridge, 1994, p. 2651): UWLLN. Let $\boldsymbol{\Theta} \subset \mathbb{R}^P$, let $\{\boldsymbol{w}_t : t = 1, 2, \ldots\}$ be a sequence of random vectors with $\boldsymbol{w}_t \in \mathcal{W}_t$ and let $\{q_t : \boldsymbol{\Theta} \times \mathcal{W}_t \to \mathbb{R}, t = 1, 2, \ldots\}$ be a sequence of real-valued functions. Assume that

i) $\boldsymbol{\Theta}$ is compact;

ii) $q_t(\boldsymbol{\theta}, \boldsymbol{w}_t)$ satisfies the following measurability and continuitiy requirements on $\boldsymbol{\Theta} \times \mathcal{W}_t$:

a) for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $q_t(\boldsymbol{\theta}, \cdot)$ is measurable;

b) for each $\boldsymbol{w}_t \in \mathcal{W}_t$, $q_t(\cdot, \boldsymbol{w}_t)$ is continuous on $\boldsymbol{\Theta}$;

iii) $E\big(|q_t(\boldsymbol{\theta}, \boldsymbol{w}_t)|\big) < \infty$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $t = 1, 2, \ldots$;

iv) $\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E\big(q_t(\boldsymbol{\theta}, \boldsymbol{w}_t)\big)$ exists for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$;

v) $\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|T^{-1} \sum_{t=1}^{T} q_t(\boldsymbol{\theta}, \boldsymbol{w}_t) - E\big(q_t(\boldsymbol{\theta}, \boldsymbol{w}_t)\big)\big| \xrightarrow{p} 0$.

Then $\{q_t(\boldsymbol{\theta}, \boldsymbol{w}_t)\}$ is said to satisfy the *uniform weak law of large numbers* (UWLLN) on $\boldsymbol{\Theta}$.

The first theorem from Wooldridge (1994) that we rely upon ensures UWLLN holds:

Theorem 4.2 (Wooldridge, 1994, p. 2652): UWLLN for the heterogeneous case. Let $\boldsymbol{\Theta} \subset \mathbb{R}^P$ and $\{\boldsymbol{w}_t : t = 1, 2, \ldots\}$ be as in Definition 4.2. Assume that

i) $\boldsymbol{\Theta}$ is compact;

ii) $q_t(\boldsymbol{\theta}, \boldsymbol{w}_t)$ satisfies the standard measurability and continuitiy requirements on $\boldsymbol{\Theta} \times \mathcal{W}_t$, see ii) in Definition 4.2 above;

iii) for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\{q_t(\boldsymbol{\theta}, \boldsymbol{w}_t)\}$ satisfies the Weak Law of Large Numbers (WLLN);

iv) there exists a function $c_t(\boldsymbol{w}_t) \geq 0$ such that

(a) for all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$, $\big|q_t(\boldsymbol{\theta}_1, \boldsymbol{w}_t) - q_t(\boldsymbol{\theta}_2, \boldsymbol{w}_t)\big| \leq c_t(\boldsymbol{w}_t) \parallel \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 \parallel$;

(b) $\{c_t(\boldsymbol{w}_t)\}$ satisfies the WLLN.

Then $\{q_t(\boldsymbol{\theta}, \boldsymbol{w}_t)\}$ satisfies the UWLLN on $\boldsymbol{\Theta}$.

We now verify each condition of this theorem for $l_t(\boldsymbol{\theta}, \epsilon_t^2)$, so $\epsilon_t^2 = \boldsymbol{w}_t$ and $l_t(\boldsymbol{\theta}, \epsilon_t^2) = q_t(\boldsymbol{\theta}, \boldsymbol{w}_t)$.

i) This holds due to A 1.

ii) $g_t(\boldsymbol{\theta})$ is bounded (due to A 2a)) and continuous (due to A 2b)) on $\boldsymbol{\theta} \in \boldsymbol{\Theta}^*$. Accordingly, $l_t(\cdot, \epsilon_t^2)$ is continuous on $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ for each $\epsilon_t^2 \in [0, \infty)$, and $l_t(\boldsymbol{\theta}, \cdot)$ is measurable for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. So ii) holds.

iii) $\{l_t(\boldsymbol{\theta}, \epsilon_t^2)\}$ satisfies the WLLN:

i) A 2a) ensures $0 < g_t(\boldsymbol{\theta}) < \infty$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and A 4a) implies that $\epsilon_t^2 = g_t(\boldsymbol{\theta}_0)\phi_t^2$ and $E(\phi_t^2) = 1 < \infty$. Hence, by the triangle inequality, $E\big|l_t(\boldsymbol{\theta}, \epsilon_t^2)\big| \leq |\ln g_t(\boldsymbol{\theta})| + \frac{g_t(\boldsymbol{\theta}_0)}{g_t(\boldsymbol{\theta})} E(\phi_t^2) < \infty$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $t = 1, 2, \ldots$.

ii) Recall that $g_t(\boldsymbol{\theta}) = g(\boldsymbol{\theta}, t/T)$. We have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} E\big(l_t(\boldsymbol{\theta}, \epsilon_t^2)\big) &= \frac{1}{T}\sum_{t=1}^{T}\left(\ln g_t(\boldsymbol{\theta}) + \frac{E(\epsilon_t^2)}{g_t(\boldsymbol{\theta})}\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\ln g_t(\boldsymbol{\theta}) + \frac{g_t(\boldsymbol{\theta}_0)}{g_t(\boldsymbol{\theta})} \cdot 1\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\ln g(\boldsymbol{\theta}, t/T) + \frac{g(\boldsymbol{\theta}_0, t/T)}{g(\boldsymbol{\theta}, t/T)}\right)\end{aligned}$$

for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. That $E(\phi_t^2) = 1$ follows from A 4a). Next, the compactness of $\boldsymbol{\Theta}$ (A 1), the continuity of $g_t(\boldsymbol{\theta})$ in $\boldsymbol{\theta}$ for all $t/T \in [0, 1]$ and the boundedness of $g_t(\boldsymbol{\theta})$ for all $t/T \in [0, 1]$ (assumed in A 2), imply there exists a strictly positive constant $C < \infty$, such that

$$\left|\ln g(\boldsymbol{\theta}, t/T) + \frac{g(\boldsymbol{\theta}_0, t/T)}{g(\boldsymbol{\theta}, t/T)}\right| \leq C$$

for all t. Hence, we also have $|T^{-1}E\big(l_t(\boldsymbol{\theta}, \epsilon_t^2)\big)| \leq T^{-1}C$ for all t, and $\lim_{T\to\infty}\sum_{t=1}^{T} T^{-1}C = C$. From a comparison test it thus follows that also $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E\big(l_t(\boldsymbol{\theta}, \epsilon_t^2)\big)$ exists.

iii) From the measurability of $l_t(\boldsymbol{\theta}, \epsilon_t^2)$, it follows from Theorem 14.1 in Davidson (1994, p. 210) that $\{l_t(\boldsymbol{\theta}, \epsilon_t^2)\}$ inherits the α- and ϕ-mixing properties of $\{\epsilon_t^2\}$ in A 3. Next, A 4b) implies that $E|\epsilon_t^2|^{r+\varepsilon} < \infty$ for some $r > 1, \varepsilon > 0$. From Corollary 3.48 in White (2001, p. 49) it thus follows that $\left|T^{-1}\sum_{t=1}^{T} l_t(\boldsymbol{\theta}, \epsilon_t^2) - E\big(l_t(\boldsymbol{\theta}, \epsilon_t^2)\big)\right| \xrightarrow{p} 0$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

iv) Let $c_t(\epsilon_t^2) = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} ||\boldsymbol{s}_t(\boldsymbol{\theta})||$, where $\boldsymbol{s}_t(\boldsymbol{\theta})$ is the score at t:

$$\boldsymbol{s}_t(\boldsymbol{\theta}) = \frac{\partial l_t(\boldsymbol{\theta}, \epsilon_t^2)}{\partial \boldsymbol{\theta}} = \frac{\partial g_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \cdot \left(\frac{1}{g_t(\boldsymbol{\theta})} - \frac{\epsilon_t^2}{g_t(\boldsymbol{\theta})^2}\right).$$

A 2b) ensures $\partial g_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ exists on $\boldsymbol{\Theta}^*$, and therefore also on $\boldsymbol{\Theta}$.

a) By a mean value expansion, there exists a $\boldsymbol{\theta} \in \boldsymbol{\Theta}^*$ such that

$$\begin{aligned} l_t(\boldsymbol{\theta}_1, \epsilon_t^2) - l_t(\boldsymbol{\theta}_2, \epsilon_t^2) &= \boldsymbol{s}_t(\boldsymbol{\theta})'(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \\ \Rightarrow \left| l_t(\boldsymbol{\theta}_1, \epsilon_t^2) - l_t(\boldsymbol{\theta}_2, \epsilon_t^2) \right| &= \left| \boldsymbol{s}_t(\boldsymbol{\theta})'(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \right|. \end{aligned}$$

Let $||\boldsymbol{x}|| = \sqrt{x_1^2 + \cdots + x_K^2}$. From the Cauchy-Schwarz inequality we obtain

$$\begin{aligned} \left| \boldsymbol{s}_t(\boldsymbol{\theta})'(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \right| &\leq \| \boldsymbol{s}_t(\boldsymbol{\theta}) \| \| \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 \| \\ &\leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \| \boldsymbol{s}_t(\boldsymbol{\theta}) \| \| \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 \| \end{aligned}$$

for all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$. So a) holds.

b) $\{c_t(\epsilon_t^2)\}$ satisfies the WLLN:

i) Let $\partial g_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = (\dot{g}_{1t}, \ldots, \dot{g}_{Kt})'$ and $y_t = \left(\frac{1}{g_t(\boldsymbol{\theta})} - \frac{\epsilon_t^2}{g_t(\boldsymbol{\theta})^2} \right)$, and recall that $\epsilon_t^2 = g_t(\boldsymbol{\theta}_0)\phi_t^2$ (due to A 4). We have:

$$\begin{aligned} ||\boldsymbol{s}_t(\boldsymbol{\theta})|| = \sqrt{\dot{g}_{1t}^2 y_t^2 + \cdots + \dot{g}_{Kt}^2 y_t^2} &= \sqrt{\dot{g}_{1t}^2 + \cdots + \dot{g}_{Kt}^2} \cdot |y_t| \\ &= \sqrt{\dot{g}_{1t}^2 + \cdots + \dot{g}_{Kt}^2} \cdot \left| \frac{1}{g_t(\boldsymbol{\theta})} - \frac{g_t(\boldsymbol{\theta}_0)}{g_t(\boldsymbol{\theta})^2}\phi_t^2 \right| \\ &= \left| v_t(\boldsymbol{\theta}) - w_t(\boldsymbol{\theta}) \right|, \end{aligned}$$

where

$$v_t(\boldsymbol{\theta}) = \sqrt{\dot{g}_{1t}^2 + \cdots + \dot{g}_{Kt}^2} \cdot \frac{1}{g_t(\boldsymbol{\theta})} \qquad \text{and} \qquad w_t(\boldsymbol{\theta}) = \sqrt{\dot{g}_{1t}^2 + \cdots + \dot{g}_{Kt}^2} \cdot \frac{g_t(\boldsymbol{\theta}_0)}{g_t(\boldsymbol{\theta})^2}\phi_t^2.$$

By the compactness of $\boldsymbol{\Theta}$ (A 1) and the continuity of $g_t(\boldsymbol{\theta})$ and $\partial g_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ in $\boldsymbol{\theta}$ (A 2b)), there exists constants $C_1, C_2 > 0$ such that

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} ||\boldsymbol{s}_t(\boldsymbol{\theta})|| = \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |v_t(\boldsymbol{\theta}) - w_t(\boldsymbol{\theta})| &\leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \Big(|v_t(\boldsymbol{\theta})| + |w_t(\boldsymbol{\theta})| \Big) \\ &\leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \Big(v_t(\boldsymbol{\theta}) + w_t(\boldsymbol{\theta}) \Big) \\ &\leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} v_t(\boldsymbol{\theta}) + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} w_t(\boldsymbol{\theta}) \\ &\leq C_1 + C_2\phi_t^2 \end{aligned}$$

for all t. Using that $E(\phi_t^2) = 1$ from A 4a), we obtain

$$E(|c_t(\epsilon_t^2)|) = E(c_t(\epsilon_t^2)) \leq E(C_1 + C_2\phi_t^2) = C_1 + C_2 < \infty.$$

This holds for all t, so i) holds.

ii) What we have just shown implies that there exists a positive constant C such that $|E(c_t(\epsilon_t^2))| \leq C$ for all t. This implies that also $|T^{-1}E(c_t(\epsilon_t^2))| \leq T^{-1}C$ for all t. Since $\lim_{T\to\infty}\sum_{t=1}^{T} T^{-1}C = C$, the comparison test implies that also $\lim_{T\to\infty}\sum_{t=1}^{T} T^{-1}E(c_t(\epsilon_t^2))$ exists. So ii) holds.

iii) From the measurability of $c_t(\epsilon_t^2)$, it follows from Theorem 14.1 in Davidson (1994, p. 210) that $\{c_t(\epsilon_t^2)\}$ inherits the α- and ϕ-mixing properties of $\{\epsilon_t^2\}$ in A 3. Next, using that $c_t(\epsilon_t^2) \leq C_1 + C_2\phi_t^2$ from above with $C_1, C_2 > 0$, and setting $r > 1$ and $\varepsilon > 0$, we obtain

$$E(|c_t(\epsilon_t^2)|^{r+\varepsilon}) \leq E|C_1 + C_2\phi_t^2|^{r+\varepsilon} \leq C^{r+\varepsilon} \cdot E|1 + \phi_t^2|^{r+\varepsilon} \qquad \text{for all } t,$$

where $C = \max\{C_1, C_2\}$. This yields

$$\begin{aligned} E|1+\phi_t^2|^{r+\varepsilon} &= \int_{\phi_t^2<1} |1+\phi_t^2|^{r+\varepsilon} dP + \int_{\phi_t^2\geq 1} |1+\phi_t^2|^{r+\varepsilon} dP \\ &\leq \int_{\phi_t^2<1} |1+\phi_t^2|^{r+\varepsilon} dP + \int_{\phi_t^2\geq 1} |2\phi_t^2|^{r+\varepsilon} dP \\ &\leq \int_{\phi_t^2<1} |1+\phi_t^2|^{r+\varepsilon} dP + 2^{r+\varepsilon} E|\phi_t^2|^{r+\varepsilon}. \end{aligned}$$

The first expression on the right hand side is finite due to the measurability of $|1+\phi_t^2|^{r+\varepsilon}$, and due to its boundedness on $[0, 1]$. The second expression is finite due to the assumption that $E|\phi_t^2|^{r+\varepsilon} < \infty$ in A 4b). Accordingly, we have shown that $E|c_t(\epsilon_t^2)|^{r+\varepsilon} < \infty$ for all t. From Corollary 3.48 in White (2001, p. 49) it thus follows that $\left|T^{-1}\sum_{t=1}^{T} c_t(\epsilon_t^2) - E(c_t(\epsilon_t^2))\right| \xrightarrow{p} 0$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

As a consequence, $\{l_t(\boldsymbol{\theta}, \epsilon_t^2)\}$ satisfies the UWLLN.

To complete the proof of consistency, we use the following theorem from Wooldridge (1994):

Theorem 4.3 (Wooldridge, 1994, p. 2653): Weak consistency of M-estimators. Let $\boldsymbol{\Theta} \subset \mathbb{R}^P$, $\Gamma \subset \mathbb{R}^R$, $\{\boldsymbol{w}_t \in \mathcal{W}_t : t = 1, 2, \ldots\}$ be a sequence of random vectors, and let $\{q_t : \boldsymbol{\Theta} \times \mathcal{W}_t \times \Gamma \to \mathbb{R} : t = 1, 2, \ldots\}$ be a sequence of objective functions. Assume that

M.1 i) $\boldsymbol{\Theta}$ and Γ^* are compact;

ii) $\widehat{\gamma}^* \xrightarrow{p} \gamma^* \in \Gamma^*$;

iii) q_t satisfies the standard measurability and continuity requirements, see ii) in Definition 4.2;

M.2 $\{q_t(\boldsymbol{\theta}, \boldsymbol{w}_t, \widehat{\gamma}^*) : t = 1, 2, \ldots\}$ satisfies the UWLLN on $\boldsymbol{\Theta} \times \Gamma^*$;

M.3 $\boldsymbol{\theta}_0$ is the unique minimiser of

$$\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E\big[q_t(\boldsymbol{\theta}, \boldsymbol{w}_t, \gamma^*)\big] \quad \text{on} \quad \boldsymbol{\Theta}.$$

Then a random vector $\widehat{\boldsymbol{\theta}}$ exists that solves (4) and $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$.

Again, $\epsilon_t^2 = \boldsymbol{w}_t$ and $l_t = q_t$ in our proof. The theorem accommodates the presence of a prior estimator $\widehat{\gamma}^*$ of a nuisance parameter γ^*, which we do not have here. So the conditions involving γ^* hold trivially. The compactness of $\boldsymbol{\Theta}$ is assumed in A 1, whereas M.1 iii) and M.2 were verified above. Finally, condition M.3 is assumed in A 5. This completes the proof of consistency. □

A.2 Proof of Theorem 2

To prove asymptotic normality, we use Theorem 4.4 in Wooldridge (1994):

Theorem 4.4 (Wooldridge, 1994, p. 2655): Asymptotic Normality of M-estimators. Let $\boldsymbol{\Theta}$, Γ, $\{\boldsymbol{w}_t : t = 1, 2, \ldots\}$ and $\{q_t : \boldsymbol{\Theta} \times \mathcal{W}_t \times \Gamma \to \mathbb{R} : t = 1, 2, \ldots\}$ be as in Theorem 4.3. In addition to M.1 – M.3, assume

M.4 i) $\boldsymbol{\theta}_0$ is interior to $\boldsymbol{\Theta}$;

ii) γ^* is interior to Γ^*;

iii) $\sqrt{T}(\widehat{\gamma}^* - \gamma^*) = O_p(1)$;

iv) For each $\gamma^* \in \Gamma$, q_t satisfies the following measurability and differentiability requirements on $\boldsymbol{\Theta} \times \mathcal{W}_t$:

a) for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $q_t(\boldsymbol{\theta}, \cdot, \gamma^*)$ is measurable;

b) for each $\boldsymbol{w}_t \in \mathcal{W}_t$, $q_t(\cdot, \boldsymbol{w}_t, \gamma^*)$ is twice continuously differentiable on $\text{int}(\boldsymbol{\Theta})$;

Define the $P \times 1$ score vector $\boldsymbol{s}_t(\boldsymbol{\theta}; \gamma^*) = \partial q_t(\boldsymbol{\theta}, \boldsymbol{w}_t, \gamma^*)/\partial \boldsymbol{\theta}$ and the $P \times P$ Hessian matrix $\boldsymbol{H}_t(\boldsymbol{\theta}, \boldsymbol{w}_t, \gamma^*) = \partial^2 q_t(\boldsymbol{\theta}, \boldsymbol{w}_t, \gamma^*)/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$.

v) For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\boldsymbol{s}_t(\boldsymbol{\theta})$ is continuously differentiable on $\text{int}(\Gamma^*)$;

M.5 i) $\{\boldsymbol{H}_t(\boldsymbol{\theta}, \epsilon_t^2; \gamma^*)\}$ satisfies the UWLLN on $\boldsymbol{\Theta} \times \Gamma^*$;

ii) $\boldsymbol{A} = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E\big(\boldsymbol{H}_t(\boldsymbol{\theta}_0, \epsilon_t^2; \gamma^*)\big)$ is positive definite;

iii) $\{\partial \boldsymbol{s}_t(\boldsymbol{\theta}, \gamma^*)/\partial\gamma^*\}$ satisfies the UWLLN on $\boldsymbol{\Theta} \times \Gamma^*$;

M.6 At the true value $\boldsymbol{\theta}_0$, $\{\boldsymbol{s}_t(\boldsymbol{\theta}_0, \gamma^*)\}$ satisfies

i) $E(\boldsymbol{s}_t'\boldsymbol{s}_t) < \infty$ for all t;

ii) $T^{-1/2} \sum_{t=1}^{T} E(\boldsymbol{s}_t) \to \boldsymbol{0}$ as $T \to \infty$;

iii) $T^{-1/2} \sum_{t=1}^{T} \boldsymbol{s}_t \xrightarrow{d} N\big(\boldsymbol{0}, \boldsymbol{B}\big)$, where $\boldsymbol{B} = \lim_{T\to\infty} Var\left(T^{-1/2} \sum_{t=1}^{T} \boldsymbol{s}_t\right)$ is positive definite.

M.7 $E\big(\partial \boldsymbol{s}_t(\boldsymbol{\theta}, \gamma^*)/\partial\gamma^*\big) = \boldsymbol{0}$ for all t.

Then $\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N\big(\boldsymbol{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\big)$.

The theorem accommodates the presence of a prior estimator $\widehat{\gamma}^*$ of a "nuisance" parameter γ^*, which we do not have here. Accordingly, conditions M.4 ii)–iii), M.4 v), M.5 iii) and M.7 hold trivially. Next, M.4 i) holds due to A 6. M.4 iv) a) holds due to the measurability of l_t. To verify M.4 iv) b), we need to show that a generic entry of the Hessian $\boldsymbol{H}_t(\boldsymbol{\theta}, \epsilon_t^2)$ exists and is continuous on int$(\boldsymbol{\Theta})$. The Hessian matrix at t is

$$\begin{aligned} \boldsymbol{H}_t(\boldsymbol{\theta}, \epsilon_t^2) &= \frac{\partial^2 l_t(\boldsymbol{\theta}, \epsilon_t^2)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \cdot a_t(\boldsymbol{\theta}) + \frac{\partial g_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \cdot \left(\frac{\partial a_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)' \qquad \text{where} \\ a_t(\boldsymbol{\theta}) &= \left(\frac{1}{g_t(\boldsymbol{\theta})} - \frac{\epsilon_t^2}{g_t(\boldsymbol{\theta})^2}\right). \end{aligned}$$

The generic (i, j)th. entry of $\boldsymbol{H}_t(\boldsymbol{\theta}, \epsilon_t^2)$ can be written as

$$m_{ij,t}(\boldsymbol{\theta}, \epsilon_t^2) = \frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_i}\left(\frac{1}{g_t(\boldsymbol{\theta})} - \frac{\epsilon_t^2}{g_t(\boldsymbol{\theta})^2}\right) - \frac{\partial g_t(\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial g_t(\boldsymbol{\theta})}{\partial\theta_j}\left(\frac{1}{g_t(\boldsymbol{\theta})^2} - \frac{2\epsilon_t^2}{g_t(\boldsymbol{\theta})^3}\right). \tag{8}$$

A 2a) ensures $g_t(\boldsymbol{\theta})$ is non-zero, positive and finite, and A 7 ensures the first and second order partial derivatives of $g_t(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ exist. Accordingly, $l_t(\boldsymbol{\theta}, \epsilon_t^2)$ is twice continuously differentiable on int$(\boldsymbol{\Theta})$ for each $\epsilon_t^2 \in [0, \infty)$, so M.4 iv) b) holds.

To verify condition M.5 i), we need to show that the generic (i, j)th. entry of the Hessian satisfies the UWLLN. To this end, we verify the conditions of Theorem 4.2:

i) This holds due to A 1.

ii) $g_t(\boldsymbol{\theta})$, $\partial g_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ are all continuous and bounded on $\boldsymbol{\Theta}$ due to A 1, A 2a)–b) and A 7. Accordingly, $m_{ij,t}(\cdot,\epsilon_t^2)$ is continuous in $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ for each $\epsilon_t^2\in[0,\infty)$, and $m_{ij,t}(\boldsymbol{\theta},\cdot)$ is thus measurable for each $\boldsymbol{\theta}\in\boldsymbol{\Theta}$. So ii) holds.

iii) $\{m_{ij,t}(\boldsymbol{\theta},\epsilon_t^2): t=1,2,\ldots\}$ satisfies the WLLN:

i) Using that $\epsilon_t^2 = g_t(\boldsymbol{\theta}_0)\phi_t^2$ (recall A 4a)), the (i,j)th entry of the Hessian at t can be written as

$$
\begin{aligned}
m_{ij,t}(\boldsymbol{\theta},\epsilon_t^2) \;=\; & \frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_i}\frac{1}{g_t(\boldsymbol{\theta})} - \frac{\partial g_t(\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial g_t(\boldsymbol{\theta})}{\partial\theta_j}\frac{1}{g_t(\boldsymbol{\theta})^2} \\
& - \left(\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_i}\frac{1}{g_t(\boldsymbol{\theta})} + \frac{\partial g_t(\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial g_t(\boldsymbol{\theta})}{\partial\theta_j}\frac{2}{g_t(\boldsymbol{\theta})^2}\right)\frac{g_t(\boldsymbol{\theta}_0)\phi_t^2}{g_t(\boldsymbol{\theta})}.
\end{aligned}
$$

A 2a) ensures $0 < g_t(\boldsymbol{\theta}) < \infty$ for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$. A 1, A 2b) and A 7 ensure $\partial g_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ are bounded on $\boldsymbol{\Theta}$, and A 4a) implies $E(\phi_t^2)=1<\infty$. Hence, by the triangle inequality, since we can write $m_{ij,t}(\boldsymbol{\theta},\epsilon_t^2) = v_{ij,t}(\boldsymbol{\theta}) + w_{ij,t}(\boldsymbol{\theta})\phi_t^2$ where $v_{ij,t}(\boldsymbol{\theta})$ is equal to the first term of $m_{ij,t}$ and $w_{ij,t}(\boldsymbol{\theta})\phi_t^2$ is equal to the second, the compactness of $\boldsymbol{\Theta}$ entails that there exists constants $C_1, C_2 > 0$ such that

$$
E\big|m_{ij,t}(\boldsymbol{\theta},\epsilon_t^2)\big| \;\leq\; C_1 + C_2 E(\phi_t^2) \;=\; C_1 + C_2 < \infty \quad \text{for all } \boldsymbol{\theta}\in\boldsymbol{\Theta}, \quad t=1,2,\ldots
$$

ii) Since A 4a) implies $E(\epsilon_t^2) = g_t(\boldsymbol{\theta}_0)$, there exists, for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, a constant $C>0$ such that $\big|E\big(m_{ij,t}(\boldsymbol{\theta},\epsilon_t^2)\big)\big| \leq C$ for all t, and hence also that

$$
\big|T^{-1}E\big(m_{ij,t}(\boldsymbol{\theta},\epsilon_t^2)\big)\big| \;\leq\; T^{-1}C
$$

for $t=1,2,\ldots,T$. Since $\lim_{T\to\infty}\sum_{t=1}^T T^{-1}C = C$, the comparison test implies that also $\lim_{T\to\infty}\sum_{t=1}^T T^{-1}E\big(m_{ij,t}(\boldsymbol{\theta},\epsilon_t^2)\big)$ exists.

iii) From the measurability of $m_{ij,t}(\boldsymbol{\theta},\epsilon_t^2)$, it follows from Theorem 14.1 in Davidson (1994, p. 210) that $\{m_{ij,t}(\boldsymbol{\theta},\epsilon_t^2)\}$ inherits the α- and ϕ-mixing properties of $\{\epsilon_t^2\}$ in A 3. Next, A 4b) implies that $E|\epsilon_t^2|^{r+\varepsilon}<\infty$ for some $r>1, \varepsilon>0$. From Corollary 3.48 in White (2001, p. 49) it thus follows that $\Big|T^{-1}\sum_{t=1}^T m_{ij,t}(\boldsymbol{\theta},\epsilon_t^2) - E\big(m_{ij,t}(\boldsymbol{\theta},\epsilon_t^2)\big)\Big| \xrightarrow{p} 0$ for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$.

iv) Let $c_t(\epsilon_t) = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} ||\boldsymbol{n}_t(\boldsymbol{\theta})||$, where $\boldsymbol{n}_t(\boldsymbol{\theta}) = \partial m_{ij,t}(\boldsymbol{\theta},\epsilon_t^2)/\partial\boldsymbol{\theta}$. The existence of $\boldsymbol{n}_t(\boldsymbol{\theta})$ on $\boldsymbol{\Theta}^*$ is ensured by A 2 and A 7.

a) By a mean value expansion, there exists a $\boldsymbol{\theta} \in \boldsymbol{\Theta}^*$ such that

$$\begin{aligned} m_{ij,t}(\boldsymbol{\theta}_1, \epsilon_t^2) - m_{ij,t}(\boldsymbol{\theta}_2, \epsilon_t^2) &= \boldsymbol{n}_t(\boldsymbol{\theta})'(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \\ \Rightarrow \left| m_{ij,t}(\boldsymbol{\theta}_1, \epsilon_t^2) - m_{ij,t}(\boldsymbol{\theta}_2, \epsilon_t^2) \right| &= \left| \boldsymbol{n}_t(\boldsymbol{\theta})'(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \right|. \end{aligned}$$

The right-hand side satisfies

$$\begin{aligned} \left| \boldsymbol{n}_t(\boldsymbol{\theta})'(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \right| &\leq \| \boldsymbol{n}_t(\boldsymbol{\theta}) \| \| \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 \| \\ &\leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \| \boldsymbol{n}_t(\boldsymbol{\theta}) \| \| \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 \| \end{aligned}$$

for all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$. So a) holds.

b) $\{c_t(\epsilon_t^2)\}$ satisfies the WLLN:

i) Let $\boldsymbol{n}_t(\boldsymbol{\theta}) = n_{1t}(\boldsymbol{\theta}) + \cdots + n_{Kt}(\boldsymbol{\theta})$. By Loève's inequality, see e.g. Theorem 2.14 in Hansen (2021), we have

$$||\boldsymbol{n}_t(\boldsymbol{\theta})|| = \sqrt{n_{1t}(\boldsymbol{\theta})^2 + \cdots + n_{Kt}(\boldsymbol{\theta})^2} \leq |n_{1t}(\boldsymbol{\theta})| + \cdots + |n_{Kt}(\boldsymbol{\theta})|$$

for all $\boldsymbol{\theta}$. Accordingly,

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} ||\boldsymbol{n}_t(\boldsymbol{\theta})|| \leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |n_{1t}(\boldsymbol{\theta})| + \cdots + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |n_{Kt}(\boldsymbol{\theta})|.$$

Write $m_{ij,t}(\boldsymbol{\theta}, \epsilon_t^2) = v_{ij,t}(\boldsymbol{\theta}) + w_{ij,t}(\boldsymbol{\theta})\phi_t^2$ so that

$$\boldsymbol{n}_t(\boldsymbol{\theta}) = \frac{\partial m_{ij,t}(\boldsymbol{\theta}, \epsilon_t^2)}{\partial \boldsymbol{\theta}} = \frac{v_{ij,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{w_{ij,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\phi_t^2, \qquad n_{kt}(\boldsymbol{\theta}) = \frac{v_{ij,t}(\boldsymbol{\theta})}{\partial \theta_k} + \frac{w_{ij,t}(\boldsymbol{\theta})}{\partial \theta_k}\phi_t^2.$$

For each $k = 1, \ldots, K$, there exists constants $C_{1,k}, C_{2,k} > 0$ such that

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |n_{kt}(\boldsymbol{\theta})| = \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{v_{ij,t}(\boldsymbol{\theta})}{\partial \theta_k} + \frac{w_{ij,t}(\boldsymbol{\theta})}{\partial \theta_k}\phi_t^2 \right| &\leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{v_{ij,t}(\boldsymbol{\theta})}{\partial \theta_k} \right| + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{w_{ij,t}(\boldsymbol{\theta})}{\partial \theta_k}\phi_t^2 \right| \\ &\leq C_{1,k} + C_{2,k}\phi_t^2. \end{aligned}$$

This means

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} ||\boldsymbol{n}_t(\boldsymbol{\theta})|| \leq \overline{C}_1 + \overline{C}_2\phi_t^2,$$

where $\overline{C}_1 = C_{1,k} + \cdots + C_{1,K}$ and $\overline{C}_2 = C_{2,1} + \cdots + C_{2,K}$. This gives

$$E\big(|c_t(\epsilon_t^2)|\big) \;=\; E\big(c_t(\epsilon_t^2)\big) \;\leq\; E\big(\overline{C}_1 + \overline{C}_2\phi_t^2\big) \;=\; \overline{C}_1 + \overline{C}_2 < \infty.$$

Since there exists suitable constants $\overline{C}_1, \overline{C}_2$ such that this holds for all t, it follows that i) holds.

ii) Since $c_t(\epsilon_t^2) \geq 0$ for all t, there exists a suitable constant $C > 0$ such that $\big|E\big(c_t(\epsilon_t^2)\big)\big| \leq C$ for all t. This implies that $\big|T^{-1}E\big(c_t(\epsilon_t^2)\big)\big| \leq T^{-1}C$ for all t. Since $\lim_{T\to\infty}\sum_{t=1}^{T} T^{-1}C = C$, the comparison test implies that also $\lim_{T\to\infty}\sum_{t=1}^{T} T^{-1}E\big(c_t(\epsilon_t^2)\big)$ exists.

iii) From the measurability of $c_t(\epsilon_t^2)$, it follows from Theorem 14.1 in Davidson (1994, p. 210) that $\{c_t(\epsilon_t^2)\}$ inherits the α- and ϕ-mixing properties of $\{\epsilon_t^2\}$ in A 3. Next, by an argument similar to as in b)i) above with $r > 1$ and $\varepsilon > 0$,

$$E\big(|c_t(\epsilon_t^2)|^{r+\varepsilon}\big) \leq E\big|C_1 + C_2\phi_t^2\big|^{r+\varepsilon} \leq C \cdot E\big|1 + \phi_t^2\big|^{r+\varepsilon} \qquad \text{for all } t,$$

where $C_1, C_2 > 0$ are constants, and $C = \max\{C_1, C_2\}$. Next:

$$\begin{aligned} E\big|1+\phi_t^2\big|^{r+\varepsilon} &= \int_{\phi_t^2<1}\big|1+\phi_t^2\big|^{r+\varepsilon}dP + \int_{\phi_t^2\geq 1}\big|1+\phi_t^2\big|^{r+\varepsilon}dP \\ &\leq \int_{\phi_t^2<1}\big|1+\phi_t^2\big|^{r+\varepsilon}dP + \int_{\phi_t^2\geq 1}\big|2\phi_t^2\big|^{r+\varepsilon}dP \\ &\leq \int_{\phi_t^2<1}\big|1+\phi_t^2\big|^{r+\varepsilon}dP + 2^{r+\varepsilon}E\big|\phi_t^2\big|^{r+\varepsilon}. \end{aligned}$$

The first expression on the right hand side is finite due to the measurability of $\big|1+\phi_t^2\big|^{r+\varepsilon}$, and due to its boundedness on $[0,1]$. The second expression is finite due to the assumption that $E\big|\phi_t^2\big|^{r+\varepsilon} < \infty$ in A 4b). Accordingly, we have shown that $E\big|c_t(\epsilon_t^2)\big|^{r+\varepsilon} < \infty$ for all t. From Corollary 3.48 in White (2001, p. 49) it thus follows that $\Big|T^{-1}\sum_{t=1}^{T} c_t(\epsilon_t^2) - E\big(c_t(\epsilon_t^2)\big)\Big| \xrightarrow{p} 0$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

As a consequence, $\{m_{ij,t}(\boldsymbol{\theta}, \epsilon_t^2)\}$ satisfies the UWLLN.

We now turn to M.5 ii). Above, when verifying that the generic (i,j)th. entry of the Hessian $m_{ij,t}(\boldsymbol{\theta}, \epsilon_t^2)$ satisfies the UWLLN, we showed that $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\big(m_{ij,t}(\boldsymbol{\theta}, \epsilon_t^2)\big)$ exists for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. A 8 asserts the limit is positive definite, so M.5 ii) holds.

We now turn to M.6:

i) At the true value $\boldsymbol{\theta}_0$, the score is

$$\boldsymbol{s}_t = \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \cdot \left(\frac{1}{g_t(\boldsymbol{\theta}_0)} - \frac{\epsilon_t^2}{g_t(\boldsymbol{\theta}_0)^2} \right) = \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \cdot \left(\frac{1}{g_t(\boldsymbol{\theta}_0)} - \frac{\phi_t^2}{g_t(\boldsymbol{\theta}_0)} \right)$$
$$= \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \cdot \frac{1}{g_t(\boldsymbol{\theta}_0)} (1 - \phi_t^2).$$

From A 2 we have that $g_t(\boldsymbol{\theta}_0)$ is strictly positive, and that $g_t(\boldsymbol{\theta}_0)$ and $\partial g_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}$ are bounded. Accordingly, $E(\boldsymbol{s}_t' \boldsymbol{s}_t) < \infty$ if $E(\phi_t^4) < \infty$ for each t, which is a consequence of A 9. So condition i) holds.

ii) From A 4a) it follows that $E(1 - \phi_t^2) = 1 - E(\phi_t^2) = 0$ for all t, so condition ii) holds.

iii) Under suitable α-mixing we can use the CLT by Herrndorf (1984). Given a series $\{X_t\}$, define the mixing coefficient between two σ-fields as

$$\alpha_X(h) = \sup_t \alpha\{\sigma(X_u, u \leq t), \sigma(X_u, u \geq t + h)\}.$$

If $\alpha_X(h) \to 0$ as $h \to \infty$, then $\{X_t\}$ is said to be α-mixing. In Francq and Zakoïan (2019, p. 375), Corollary 1 in Herrndorf (1984, p. 42) is written as:

Theorem A.4 (Herrndorf, 1984): CLT for α-mixing processes. Let $\{X_t\}$ be a centred process (i.e. $E(X_t) = 0$ for all t) such that

$$\sup_t ||X_t||_{2+\nu} < \infty, \qquad \sum_{h=0}^{\infty} \alpha_X(h)^{\nu/(2+\nu)} < \infty, \quad \text{for some} \quad \nu > 0,$$

where $||X||_b = E^{1/b}|X|^b$. If $\sigma_X^2 = \lim_{T\to\infty} Var(T^{-1/2} \sum_{t=1}^T X_t)$ exists and is not zero, then

$$T^{-1/2} \sum_{t=1}^{T} X_t \xrightarrow{d} N(0, \sigma_X^2).$$

Let $\boldsymbol{\lambda}$ be a $(K \times 1)$ vector of finite scalars such that $\boldsymbol{\lambda}'\boldsymbol{\lambda} = 1$, and let

$$X_t = \boldsymbol{\lambda}' \boldsymbol{s}_t = g_t^*(\boldsymbol{\theta}_0) \cdot (1 - \phi_t^2), \qquad \text{where} \qquad g_t^*(\boldsymbol{\theta}_0) = \boldsymbol{\lambda}' \cdot \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \cdot \frac{1}{g_t(\boldsymbol{\theta}_0)}.$$

A 1, A 2 and the finiteness of $\boldsymbol{\lambda}$ imply that $g_t^*(\boldsymbol{\theta}_0)$ is bounded for all t, and A 4a) implies that $E(X_t) = 0$ for all t. From Theorem 14.1 in Davidson (1994, p. 210) it follows that $\{X_t\}$ inherits the α-mixing properties of $\{\epsilon_t^2\}$ in A 3, since $\phi_t^2 \equiv \epsilon_t^2 / g_t(\boldsymbol{\theta}_0)$. So $\{X_t\}$ is α-mixing of size $-r/(r-1), r>1$. Next, for some $\nu > 0$,

$$\begin{aligned} E\big(|X_t|^{2+\nu}\big) &= \big|g_t^*(\boldsymbol{\theta}_0)\big|^{2+\nu} \cdot E\big|1-\phi_t^2\big|^{2+\nu}, \\ &\leq \big|g_t^*(\boldsymbol{\theta}_0)\big|^{2+\nu} \cdot E\big|1+\phi_t^2\big|^{2+\nu}, \end{aligned}$$

and

$$\begin{aligned} E\big|1+\phi_t^2\big|^{2+\nu} &= \int_{\phi_t^2<1} \big|1+\phi_t^2\big|^{2+\nu} dP + \int_{\phi_t^2\geq 1} \big|1+\phi_t^2\big|^{2+\nu} dP \\ &\leq \int_{\phi_t^2<1} \big|1+\phi_t^2\big|^{2+\nu} dP + \int_{\phi_t^2\geq 1} \big|2\phi_t^2\big|^{2+\nu} dP \\ &\leq \int_{\phi_t^2<1} \big|1+\phi_t^2\big|^{2+\nu} dP + 2^{2+\nu} E\big|\phi_t^2\big|^{2+\nu}. \end{aligned}$$

The first expression on the right hand side is finite due to the measurability of $\big|1+\phi_t^2\big|^{2+\nu}$, and due to its boundedness on $[0,1]$. The second expression is finite due to the assumption that $E|\phi_t^2|^{2+\nu} < \infty$ in A 9. Accordingly, we have shown that $E\big(|X_t|^{2+\nu}\big) < \infty$ for all t, and so $\sup_t ||X_t||_{2+\nu} < \infty$.

We now show that $\sum_{h=0}^{\infty} \alpha(h)^{\nu/(2+\nu)} < \infty$. Since $\{X_t\}$ is α-mixing of size $-r/(r-1)$ with $r>1$, there exists an integer N such that $\frac{\alpha(h)}{h^{-a}} < \Delta$ for $T \geq N$, where $a = r/(r-1)+\varepsilon$ for some $\Delta, \varepsilon > 0$. Set $b = \nu/(2+\nu)$, and without loss of generality set $\Delta = 1$. We have:

$$\begin{aligned} \sum_{h=0}^{\infty} \alpha(h)^b = \sum_{h=0}^{\infty} h^{-ab} \left(\frac{\alpha(h)}{h^{-a}}\right)^b &= \sum_{h=0}^{N-1} h^{-ab} \left(\frac{\alpha(h)}{h^{-a}}\right)^b + \sum_{h=N}^{\infty} h^{-ab} \left(\frac{\alpha(h)}{h^{-a}}\right)^b \\ &\leq \sum_{h=0}^{N-1} h^{-ab} \left(\frac{\alpha(h)}{h^{-a}}\right)^b + \sum_{h=N}^{\infty} h^{-ab} \Delta^b \quad \left(\text{since } \frac{\alpha(h)}{h^{-a}} < \Delta\right) \\ &\leq \sum_{h=0}^{N-1} h^{-ab} \left(\frac{\alpha(h)}{h^{-a}}\right)^b + \sum_{h=N}^{\infty} h^{-ab}. \end{aligned}$$

The first term on the right hand side is finite, and the second term is finite if $ab > 1$, i.e. if

$$\nu > \frac{2(r-1)}{(r-1)\varepsilon + 1}$$

for some $\varepsilon > 0$. This holds for all $\varepsilon > 0$ when $\nu > 2(r-1)$, which is assumed in A 9. So $\sum_{h=0}^{\infty} \alpha(h)^{\nu/(2+\nu)} < \infty$.

We now turn to the variance. We have $Var(T^{-1/2}\sum_{t=1}^{T} X_t) = T^{-1}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T} E(X_{t_1}X_{t_2})$.

A 1, A 2 and the finiteness of $\boldsymbol{\lambda}$ imply that $g_t^*(\boldsymbol{\theta}_0)$ is bounded for all t, and A 9 implies $E(\phi_{t_1}^2 \phi_{t_2}^2) < \infty$ for all pairs t_1, t_2. Accordingly, $|E(X_{t_1} X_{t_2})| < \infty$ for all pairs t_1, t_2, and so $Var(T^{-1/2} \sum_{t=1}^T X_t) < \infty$ for finite T. Next, write $Var(T^{-1/2} \sum_{t=1}^T X_t) = T \cdot E\big[(T^{-1} \sum_{t=1}^T X_t)^2\big] = T \cdot E(\overline{X}^2)$. Since $\sup_t ||X_t||_{2+\nu} < \infty$ and $\sum_{h=0}^{\infty} \alpha_X(h)^{\nu/(2+\nu)} < \infty$, Corollary A.2 in Francq and Zakoïan (2019, p. 373) implies that $E(\overline{X}^2) = O(T^{-1})$. So $T \cdot E(\overline{X}^2) = O(1)$. Accordingly, $\lim_{T\to\infty} Var(T^{-1/2} \sum_{t=1}^T X_t) < \infty$. The non-degenerateness of ϕ_t^2 assumed in A 4 ensures the limit is non-zero.

Thus far, we have shown that $T^{-1/2} \sum_{t=1}^T X_t \xrightarrow{d} N(0, \sigma_X^2)$ for all $\boldsymbol{\lambda}$ that satisfies $\boldsymbol{\lambda}'\boldsymbol{\lambda} = 1$, where $X_t = \boldsymbol{\lambda}' \boldsymbol{s}_t$. If a stochastic vector $\boldsymbol{s}$ satisfies $\boldsymbol{s} \sim N\big(\mathbf{0}, \boldsymbol{B}\big)$, then $\boldsymbol{\lambda}'\boldsymbol{s} \sim N\big(\mathbf{0}, \boldsymbol{\lambda}'\boldsymbol{B}\boldsymbol{\lambda}\big)$. From the Cramér-Wold theorem, see e.g. Hansen (2021, Theorem 8.4), it thus follows that $\sqrt{T} \sum_{t=1}^T \boldsymbol{s}_t \xrightarrow{d} N\big(\mathbf{0}, \boldsymbol{B}\big)$. The existence of $\boldsymbol{B}$ follows from $\lim_{T\to\infty} Var(T^{-1/2} \sum_{t=1}^T X_t) < \infty$ (shown above), and its positive definiteness is assumed in A 10.

This completes the proof of asymptotic normality. □

A.3 Proof of Theorem 3

The result follows from Theorem 2 in Hansen (1992):

Theorem 2 (Hansen, 1992, p. 969): Assume

(K) (Kernel) For all $x \in \mathbb{R}$, $|k(x)| \leq 1$ and $k(x) = k(-x)$; $k(0) = 1$; $k(x)$ is continuous at zero and for almost all $x \in \mathbb{R}$; $\int_{\mathbb{R}} |k(x)| dx < \infty$.

(S) (Bandwidth) $S_T \to \infty$, and for some $q \in (1/2, \infty)$, $S_T^{1+2q}/T = O(1)$;

(V1) For some $u \in (2, 4]$ such that $u > 2 + 1/q$, and some $p > u$,

- (i) $12 \sum_{h=1}^{\infty} \alpha_h^{2(1/u - 1/p)} < \infty$ or $4 \sum_{h=1}^{\infty} \phi_h^{1-2/p} < \infty$;
- (ii) $\sup_{t \geq 1} ||\boldsymbol{s}_t(\boldsymbol{\theta}_0)||_p < \infty$, where $||\boldsymbol{s}_t(\boldsymbol{\theta}_0)||_p = \big(\sum_j E|s_{jt}(\boldsymbol{\theta}_0)|^p\big)^{1/p}$;

where $\{\alpha_h\}_{h=1}^{\infty}$ and $\{\phi_h\}_{h=1}^{\infty}$ denote the α-mixing and ϕ-mixing coefficients, respectively, for $\{\boldsymbol{s}_t(\boldsymbol{\theta}_0)\}_{t=1}^{\infty}$;

(V2) Let $\mathcal{N}$ denote some neighbourhood of $\boldsymbol{\theta}_0$, and let $||\cdot||$ denote the Euclidean norm:

- (i) $\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p(1)$;
- (ii) $\sup_{t \geq 1} E\left(\sup_{\boldsymbol{\theta} \in \mathcal{N}} ||\boldsymbol{s}_t(\boldsymbol{\theta})||^2\right) < \infty$;
- (iii) $\sup_{t \geq 1} E\left(\sup_{\boldsymbol{\theta} \in \mathcal{N}} ||\frac{\partial}{\partial \boldsymbol{\theta}'} \boldsymbol{s}_t(\boldsymbol{\theta})||^2\right) < \infty$.

Then $\widehat{\boldsymbol{B}} \xrightarrow{p} \boldsymbol{B}$.

Conditions (K) and (S) hold by assumption (i.e. A 11 and A 12).

To verify condition (V1)(i) for α-mixing, we use the same approach as in our proof of Theorem 2 when we verified $\sum_{h=0}^{\infty} \alpha(h)^{\nu/(2+\nu)} < \infty$ in Herrndorf's (1984) theorem. Set $a = r/(r-1) + \varepsilon$ and $b = 2(1/u - 1/p)$. Using the same reasoning, $\sum_{h=0}^{\infty} \alpha(h)^b < \infty$ if $ab > 1$ for some $\varepsilon > 0$. This holds for all $\varepsilon > 0$ when $2(1/u - 1/p) > (r-1)/r$, which is assumed in A 13a). So condition (V1)(i) holds.

To verify condition (V1)(ii), note that the score at $\boldsymbol{\theta}_0$ can be written as

$$\begin{aligned}
\boldsymbol{s}_t(\boldsymbol{\theta}_0) = \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \cdot \left(\frac{1}{g_t(\boldsymbol{\theta}_0)} - \frac{\epsilon_t^2}{g_t(\boldsymbol{\theta}_0)^2} \right) &= \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \cdot \left(\frac{1}{g_t(\boldsymbol{\theta}_0)} - \frac{\phi_t^2}{g_t(\boldsymbol{\theta}_0)} \right) \\
&= \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \cdot \frac{1}{g_t(\boldsymbol{\theta}_0)} \left(1 - \phi_t^2\right) \\
&= \boldsymbol{k}_t(\boldsymbol{\theta}_0)(1 - \phi_t^2).
\end{aligned}$$

From A 2 we have that $g_t(\boldsymbol{\theta}_0)$ is strictly positive, and that $g_t(\boldsymbol{\theta}_0)$ and $\partial g_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}_0$ are bounded. Accordingly, there exists a constant $C > 0$ such that

$$\begin{aligned}
||\boldsymbol{s}_t(\boldsymbol{\theta}_0)||_p = \left(\sum_j E\left| k_{jt}(\boldsymbol{\theta}_0)(1 - \phi_t^2) \right|^p \right)^{1/p} &\leq \left(\sum_j E|k_{jt}(\boldsymbol{\theta}_0)|^p |(1 - \phi_t^2)|^p \right)^{1/p} \\
&\leq \left(\sum_j |k_{jt}(\boldsymbol{\theta}_0)|^p \right)^{1/p} \cdot \left(E|(1 - \phi_t^2)|^p \right)^{1/p} \\
&\leq C \cdot \left(E|(1 + \phi_t^2)|^p \right)^{1/p},
\end{aligned}$$

where

$$\begin{aligned}
E\left|1 + \phi_t^2\right|^p &= \int_{\phi_t^2 < 1} \left|1 + \phi_t^2\right|^p dP + \int_{\phi_t^2 \geq 1} \left|1 + \phi_t^2\right|^p dP \\
&\leq \int_{\phi_t^2 < 1} \left|1 + \phi_t^2\right|^p dP + \int_{\phi_t^2 \geq 1} \left|2\phi_t^2\right|^p dP \\
&\leq \int_{\phi_t^2 < 1} \left|1 + \phi_t^2\right|^p dP + 2^p E\left|\phi_t^2\right|^p.
\end{aligned}$$

The first expression on the right hand side is finite due to the measurability of $\left|1 + \phi_t^2\right|^p$, and due to its boundedness on $[0, 1]$. The second expression is finite due to the assumption that $E|\phi_t^2|^p < \infty$ for $p > u > 2 + 1/q$ in A 13. Finally, since this holds for all t, it follows that condition (V1)(ii) holds.

In verifying condition (V2)(i), set $\mathcal{N} = \text{int}(\boldsymbol{\Theta})$. Then condition (V2)(i) holds, since $\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d}$

$N(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1})$ by Theorem 2.

To verify condition (V2)(ii), we use an argument similar to the one we used to verify that $c_t(\epsilon_t^2)$ satisfies the WLLN in the proof of Theorem 1. Note that $\mathcal{N} \subset \boldsymbol{\Theta}$. Let $\partial g_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = (\dot{g}_{1t}, \ldots, \dot{g}_{Kt})'$. Using the triangle inequality, and that $\epsilon_t^2 = g_t(\boldsymbol{\theta}_0)\phi_t^2$ (due to A 4) and $g_t > 0$ (due to A 2a)), we obtain

$$\begin{aligned}
||\boldsymbol{s}_t(\boldsymbol{\theta})||^2 &= \left(\dot{g}_{1t}^2 + \cdots + \dot{g}_{Kt}^2\right) \cdot \left(\frac{1}{g_t(\boldsymbol{\theta})} - \frac{g_t(\boldsymbol{\theta}_0)}{g_t(\boldsymbol{\theta})^2}\phi_t^2\right)^2 \\
&= \left(\dot{g}_{1t}^2 + \cdots + \dot{g}_{Kt}^2\right) \cdot \left(\frac{1}{g_t(\boldsymbol{\theta})^2} - 2\frac{g_t(\boldsymbol{\theta}_0)}{g_t(\boldsymbol{\theta})^3}\phi_t^2 + \frac{g_t(\boldsymbol{\theta}_0)^2}{g_t(\boldsymbol{\theta})^4}\phi_t^4\right) \\
&\leq u_t(\boldsymbol{\theta}) + v_t(\boldsymbol{\theta})\phi_t^2 + w_t(\boldsymbol{\theta})\phi_t^4,
\end{aligned}$$

where $u_t(\boldsymbol{\theta}), v_t(\boldsymbol{\theta}), w_t(\boldsymbol{\theta}) \geq 0$ for all t and all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. By A 1 and A 2b), $\partial g_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is bounded for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. So there exists strictly positive constants $C_1, C_2, C_3 > 0$ such that

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} ||\boldsymbol{s}_t(\boldsymbol{\theta})||^2 &\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} u_t(\boldsymbol{\theta}) + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} v_t(\boldsymbol{\theta})\phi_t^2 + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} w_t(\boldsymbol{\theta})\phi_t^4 \\
&\leq C_1 + C_2\phi_t^2 + C_3\phi_t^4.
\end{aligned}$$

A 4 and A 9 ensure $E(\phi_t^2)$ and $E(\phi_t^4)$ are finite for all t, so

$$E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} ||\boldsymbol{s}_t(\boldsymbol{\theta})||^2\right) \leq E\big(C_1 + C_2\phi_t^2 + C_3\phi_t^4\big) = C_1 + C_2 + C_3E(\phi_t^4) < \infty.$$

This holds for at each t for suitable constants C_1, C_2, C_3. Finally, since $\mathcal{N} \subset \boldsymbol{\Theta}$, it follows that

$$\sup_{t\geq 1} E\left(\sup_{\boldsymbol{\theta}\in\mathcal{N}} ||\boldsymbol{s}_t(\boldsymbol{\theta})||^2\right) < \infty.$$

So condition (V2)(ii) holds.

In condition (V2)(iii), the term inside $||\cdot||^2$ is the Hessian. In the proof of Theorem 2 (when proving that each entry in the Hessian satisfies the UWLLN), the (i,j)th entry of the Hessian at t was written as $m_{ij,t}(\boldsymbol{\theta}, \epsilon_t^2) = v_{ij,t}(\boldsymbol{\theta}) + w_{ij,t}(\boldsymbol{\theta})\phi_t^2$. This means, by Loève's inequality (see e.g. Theorem

2.14 in Hansen (2021)), that

$$\begin{aligned}\left|\left|\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{s}_t(\boldsymbol{\theta})\right|\right|^2 &= \sum_i\sum_j\big(v_{ij,t}(\boldsymbol{\theta}) + w_{ij,t}(\boldsymbol{\theta})\phi_t^2\big)^2 \\ &= \sum_i\sum_j\big(u_{ij,t}^{(1)}(\boldsymbol{\theta}) + u_{ij,t}^{(2)}(\boldsymbol{\theta})\phi_t^2 + u_{ij,t}^{(3)}(\boldsymbol{\theta})\phi_t^4\big) \\ &\leq \sum_i\sum_j |u_{ij,t}^{(1)}(\boldsymbol{\theta})| + |u_{ij,t}^{(2)}(\boldsymbol{\theta})|\phi_t^2 + |u_{ij,t}^{(3)}(\boldsymbol{\theta})|\phi_t^4,\end{aligned}$$

where $u_{ij,t}^{(1)}$, $u_{ij,t}^{(2)}$ and $u_{ij,t}^{(3)}$ are introduced to reduce the notational burden. Due to the compactness of $\boldsymbol{\Theta}$, A 2 and A 7, there exists constants $C_{ij}^{(1)}, C_{ij}^{(2)}, C_{ij}^{(3)} > 0$ such that

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\left|\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{s}_t(\boldsymbol{\theta})\right|\right|^2 &\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left(\sum_i\sum_j |u_{ij,t}^{(1)}(\boldsymbol{\theta})| + |u_{ij,t}^{(2)}(\boldsymbol{\theta})|\phi_t^2 + |u_{ij,t}^{(3)}(\boldsymbol{\theta})|\phi_t^4\right) \\ &\leq \sum_i\sum_j \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|u_{ij,t}^{(1)}(\boldsymbol{\theta})| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|u_{ij,t}^{(2)}(\boldsymbol{\theta})|\phi_t^2 + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|u_{ij,t}^{(3)}(\boldsymbol{\theta})|\phi_t^4 \\ &\leq \sum_i\sum_j C_{ij}^{(1)} + C_{ij}^{(2)}\phi_t^2 + C_{ij}^{(3)}\phi_t^4 \\ &\leq \overline{C}^{(1)} + \overline{C}^{(2)}\phi_t^2 + \overline{C}^{(3)}\phi_t^4,\end{aligned}$$

where $\overline{C}^{(1)}$, $\overline{C}^{(2)}$ and $\overline{C}^{(3)}$ are suitable sums of $C_{ij}^{(\cdot)}$'s. A 9 ensure $E(\phi_t^4)$ is finite for all t, so

$$E\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\frac{\partial}{\partial\boldsymbol{\theta}'}\boldsymbol{s}_t(\boldsymbol{\theta})\|^2\right) \leq \overline{C}^{(1)} + \overline{C}^{(2)}E(\phi_t^2) + \overline{C}^{(3)}E(\phi_t^4) \quad < \quad \infty,$$

This holds for all t. And since $\mathcal{N} \subset \boldsymbol{\Theta}$, condition (V2)(iii) is satisfied.

This completes the proof of $\widehat{\boldsymbol{B}} \xrightarrow{p} \boldsymbol{B}$. □

Table 1: Comparison of estimators of the g_t parameters (see Section 4.1)

T	$m(\widehat{\delta}_0)$	$se(\widehat{\delta}_0)$	$m(\widehat{\delta}_1)$	$se(\widehat{\delta}_1)$	$m(\widehat{\gamma})$	$se(\widehat{\gamma})$	$m(\widehat{c})$	$se(\widehat{c})$
Our estimator:								
1000	−0.156	0.375	0.575	1.182	35.108	42.987	0.040	0.246
2000	−0.138	0.340	0.489	1.077	24.069	38.212	0.037	0.217
5000	−0.093	0.272	0.295	0.759	12.852	28.732	0.024	0.161
10000	−0.053	0.187	0.151	0.482	5.430	17.562	0.011	0.102
20000	−0.024	0.106	0.072	0.288	1.795	9.432	0.007	0.062
40000	−0.009	0.049	0.022	0.114	0.447	2.951	0.001	0.029
Iterative estimator (Amado and Teräsvirta, 2013):								
1000	−0.177	0.395	77.75	1469.7	82.010	109.20	0.030	0.245
2000	−0.146	0.349	0.856	4.116	61.671	99.219	0.018	0.200
5000	−0.094	0.262	0.292	0.893	27.127	68.009	0.016	0.153
10000	−0.045	0.173	0.143	0.529	9.768	39.112	0.012	0.096
20000	−0.026	0.131	0.073	0.332	2.314	15.693	0.007	0.061
40000	−0.013	0.090	0.036	0.246	0.552	3.437	0.003	0.039

Relative efficiency (Our/Iterative):

T	$se(\widehat{\delta}_0)$	$se(\widehat{\delta}_1)$	$se(\widehat{\gamma})$	$se(\widehat{c})$
1000	0.949	0.001	0.394	1.008
2000	0.974	0.262	0.385	1.086
5000	1.038	0.850	0.422	1.048
10000	1.080	0.910	0.449	1.063
20000	0.809	0.869	0.601	1.023
40000	0.537	0.466	0.858	0.764

T, sample size. $m(\widehat{x})$, average bias of estimate $\widehat{x}$ across replications (no. of replications = 1000). $se(\widehat{x})$, sample standard deviation of estimate $\widehat{x}$ across replications. All computations in R (R Core Team, 2021). Our estimator is implemented with own code. The Iterative estimator is implemented with the `tvgarch()` function of the CRAN package **tvgarch** (Campos-Martins and Sucarrat, 2021).

Table 2: Comparison of estimators of the h_t parameters (see Section 4.1)

T	$m(\widehat{\omega})$	$se(\widehat{\omega})$	$m(\widehat{\alpha})$	$se(\widehat{\alpha})$	$m(\widehat{\beta})$	$se(\widehat{\beta})$
Our estimator:						
1000	0.044	0.082	−0.002	0.030	−0.042	0.096
2000	0.019	0.039	0.000	0.020	−0.019	0.050
5000	0.007	0.020	0.000	0.013	−0.007	0.028
10000	0.003	0.015	0.000	0.009	−0.003	0.020
20000	0.001	0.010	0.000	0.006	−0.001	0.014
40000	0.001	0.007	0.000	0.004	−0.001	0.009
Iterative estimator (Amado and Teräsvirta, 2013):						
1000	0.078	0.278	−0.001	0.029	−0.042	0.095
2000	0.024	0.066	0.000	0.020	−0.020	0.051
5000	0.008	0.024	−0.001	0.013	−0.007	0.029
10000	0.003	0.015	0.000	0.009	−0.004	0.020
20000	0.002	0.010	0.000	0.006	−0.002	0.014
40000	0.001	0.007	0.000	0.004	0.000	0.009

Relative efficiency (Our/Iterative):

T	$se(\widehat{\omega})$	$se(\widehat{\alpha})$	$se(\widehat{\beta})$
1000	0.293	1.048	1.009
2000	0.597	0.990	0.979
5000	0.826	0.980	0.977
10000	0.995	1.009	1.019
20000	1.023	1.027	1.033
40000	0.990	0.947	0.971

T, sample size. $m(\widehat{x})$, average bias of estimate $\widehat{x}$ across replications (no. of replications = 1000). $se(\widehat{x})$, sample standard deviation of estimate $\widehat{x}$ across replications. All computations in R (R Core Team, 2021). Our estimator is implemented with own code. The Iterative estimator is implemented with the `tvgarch()` function of the CRAN package **tvgarch** (Campos-Martins and Sucarrat, 2021).

Table 3: Spline estimates of intraday hourly volatility (see Section 4.3)

m	$\widehat{\delta}_{m,0}$ (*s.e.*)	$\widehat{\delta}_{m,1}$ (*s.e.*)	$\widehat{\delta}_{m,2}$ (*s.e.*)	$\widehat{\delta}_{m,3}$ (*s.e.*)	$\widehat{\delta}_{m,4}$ (*s.e.*)	T_m	$\chi^2(4)$ [p-value]
1	5.299 (0.7500)	−15.810 (13.4499)	21.374 (35.3870)	−6.012 (42.0694)	53.597 (52.1365)	515	21.853 [0.0002]
2	3.775 (0.2522)	1.133 (5.5682)	−11.494 (17.2066)	9.289 (27.5935)	53.769 (44.4830)	516	13.659 [0.0085]
3	3.643 (0.1799)	2.215 (5.6775)	−8.954 (18.0865)	18.192 (29.0815)	−22.852 (51.8127)	516	0.518 [0.9717]
4	3.881 (0.2042)	−1.513 (5.8468)	−0.526 (17.8306)	10.394 (24.7749)	−31.197 (36.2192)	516	3.739 [0.4425]
5	3.374 (0.2480)	−3.385 (5.6154)	17.916 (25.1045)	−22.709 (50.8933)	−58.255 (65.4091)	516	11.252 [0.0239]
6	2.735 (0.1639)	4.456 (4.8266)	−13.627 (14.8996)	17.628 (21.9418)	−6.619 (35.6272)	516	2.678 [0.6131]
7	2.828 (0.1880)	2.119 (4.8304)	2.422 (14.8860)	−18.812 (23.0798)	24.475 (41.8866)	516	4.935 [0.2940]
8	3.857 (0.1761)	0.895 (4.7520)	−5.108 (15.6553)	27.629 (27.3967)	−85.692 (50.4111)	515	4.605 [0.3303]
9	4.521 (0.1046)	1.519 (3.3636)	1.150 (11.0443)	−10.082 (17.3552)	−6.229 (26.6340)	517	9.899 [0.0422]
10	5.017 (0.1083)	−4.345 (4.1078)	20.107 (14.1020)	−44.509 (23.5988)	86.918 (39.2159)	517	5.067 [0.2805]
11	4.639 (0.1883)	−6.359 (4.2830)	31.118 (13.0861)	−61.819 (19.7580)	82.749 (32.9960)	517	12.851 [0.0120]
12	4.270 (0.1340)	3.207 (6.0619)	−9.070 (19.7662)	7.739 (27.8525)	8.046 (31.5353)	517	1.285 [0.8638]
13	4.557 (0.2465)	−1.464 (5.8333)	4.095 (16.2240)	−6.101 (20.1852)	−3.786 (27.0234)	517	6.345 [0.1748]
14	4.543 (0.1308)	−2.035 (5.9661)	19.024 (20.3573)	−47.267 (30.2697)	52.408 (33.8313)	517	6.326 [0.1761]
15	5.459 (0.1745)	−2.838 (6.1910)	10.719 (19.1027)	−19.411 (26.1795)	11.935 (39.5819)	517	4.168 [0.3838]
16	5.119 (0.1370)	2.594 (3.5311)	−9.415 (11.2707)	11.557 (18.2838)	10.993 (29.4810)	517	4.806 [0.3078]
17	5.409 (0.2084)	−4.002 (6.1471)	10.676 (18.9852)	−19.079 (26.0643)	51.860 (33.0375)	517	6.545 [0.1620]
18	4.525 (0.1408)	2.129 (4.2691)	−5.967 (14.9245)	8.191 (25.3191)	−22.307 (35.8562)	517	1.757 [0.7803]
19	4.538 (0.2072)	−8.461 (4.6763)	28.788 (14.3695)	−45.468 (22.1033)	89.723 (37.5707)	517	11.391 [0.0225]
20	4.812 (0.2824)	−12.499 (6.1909)	42.377 (18.0694)	−66.247 (25.3022)	78.278 (37.1935)	515	7.316 [0.1201]
21	4.762 (0.2543)	12.607 (8.3225)	−51.998 (27.7942)	68.146 (41.7357)	13.428 (60.2765)	516	13.748 [0.0081]
22	4.247 (0.2998)	−9.138 (6.8886)	26.477 (22.1287)	−37.022 (36.0341)	50.927 (50.1771)	516	3.522 [0.4746]
23	3.192 (0.1530)	−0.677 (6.5807)	−1.918 (21.7801)	17.689 (34.8831)	−34.275 (58.9980)	412	3.196 [0.5257]
24	2.203 (0.1849)	20.118 (4.9577)	−68.402 (15.6429)	80.915 (24.5480)	−18.442 (37.1555)	413	34.318 [0.0000]

The estimated model is $\widehat{\ln g_{m,t}} = \delta_{m,0} + \sum_{l=1}^{4} \delta_{m,l}(t/T - c_l)^2 I(t/T \geq c_l)$ with $(c_1, c_2, c_3, c_4)' = (0.2, 0.4, 0.6, 0.8)$. m, intraday period/hour. *s.e.*, standard error of estimate. T, number of observations. $\chi^2(4)$, the test statistic of a Wald-test with $H_0 : \delta_{m,1} = \cdots = \delta_{m,4} = 0$ (p-value in square brackets).

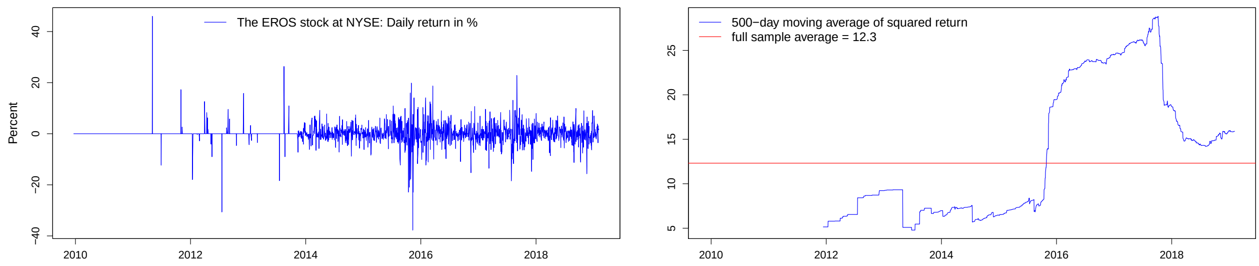

Figure 1: Daily log-returns in % of the EROS stock at NYSE (left) and 500-day moving average of squared returns (right), 21 December 2009 – 4 February 2021 (see Section 4.2). Datasource: Bloomberg

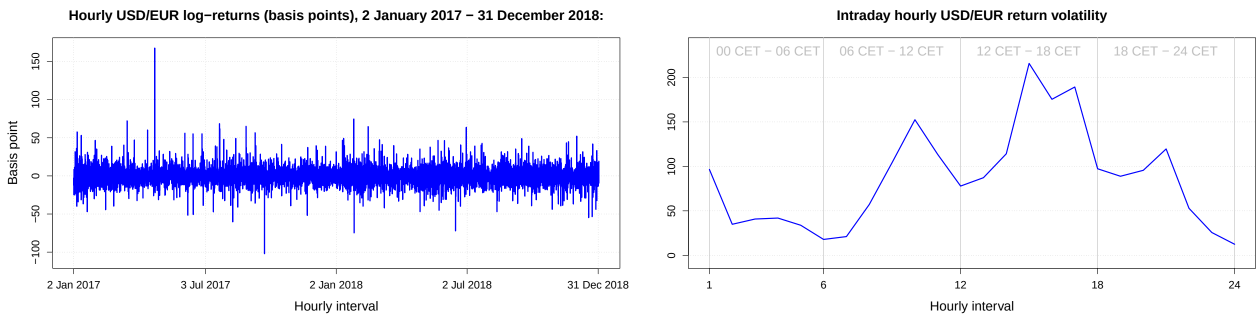

Figure 2: Hourly log-returns in basis points of the USD/EUR exchange rate (left) and estimates of its intraday hourly volatility (right), 2 January 2017 – 31 December 2018 (see Section 4.3). Datasource: Forexite