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

Adam Lee,[2], Rickard Sandberg[3] and Genaro Sucarrat[4]

This version: 11 May 2024

Abstract

We derive a unified and general framework for estimation and inference in a large class of parametric models of time-varying unconditional volatility of financial return, both univariate and multivariate. A large number of well-known and widely used specifications, for many of which asymptotic results have not been specifically established, are contained in the class. Our framework is based on the multivariate equation-by-equation version of the Gaussian Quasi Maximum Likelihood Estimator (QMLE). An attractive property of the estimator is its ease of implementation, since the equation-by-equation nature reduces the curse of dimensionality associated with multivariate methods. Another attractive property is that the exact specification of the conditional volatility dynamics need not be known or estimated. Nevertheless, a model of conditional volatility can be estimated in a second step. In particular, we show that the (scaled) GARCH(1,1) specification is well-defined under both correct and incorrect specification within our framework. 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 other non-negative processes like volume, duration, realised volatility, dividends, unemployment and so on. Finally, three numerical 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 and Ovidijus Stauskas at BI, Dennis Kristensen, and participants at Forskermøtet (January 2024), the economics seminar at the University of Lund (November 2023), ICEEE 2023 congress (May 2023), Zaragoza Workshop on Time Series Econometrics (March 2023), CFE 2022 conference (December, London), statistics seminar at the University of Padova (September 2022), economics seminar at the University of Verona (September 2022), ISNPS 2022 conference (June, Paphos), QFFE 2022 workshop (June, Marseille) and the internal economics seminar at BI Norwegian Business School (June 2022) for their helpful comments, suggestions and questions.

[2]Department of Data Science and Analytics, BI Norwegian Business School.

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

[4]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, and it has long been known that 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), to name but a few.

Let ϵ_t denote an observed financial return at t. If $0 < E(\epsilon_t^2) < \infty$ for all t, then ϵ_t^2 can be decomposes multiplicatively as

$$\epsilon_t^2 = g_t \phi_t^2 \qquad \text{with} \qquad g_t := E(\epsilon_t^2) \quad \text{and} \quad \phi_t^2 := \epsilon_t^2 / E(\epsilon_t^2) \tag{1}$$

for all t. Henceforth, we refer to g_t as the unconditional volatility at t. The decomposition in (1) implies $E(\phi_t^2) = 1$ for all t. For the implications of this, see our discussion in relation with Assumption 4 further below. For a complete characterisation of the conditional volatility dynamics, a model of the stochastic part ϕ_t^2 is required. A leading example is the scaled version of the stationary GARCH(1,1) model of Bollerslev (1986),

$$\phi_t^2 = h_t \eta_t^2, \qquad \eta_t \sim iid(0,1), \qquad h_t = \omega + \alpha \phi_t^2 + \beta h_{t-1}, \tag{2}$$

in which the conditional volatility or variance is $\sigma_t^2 = g_t h_t$, and the unconditional volatility at t is $E(\sigma_t^2) = g_t$. (Henceforth, to simplify the exposition, we refer to both σ_t and

σ_t^2 as conditional volatility, since one is obtained from the other via a straightforward transformation.) Other examples of h_t include scaled versions of Stochastic Volatility (SV) models (in which σ_t^2 need not equal the conditional variance), and scaled versions of Dynamic Conditional Score (DCS) models. Amado et al. (2019) contains 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), Koo and Linton (2015), Kim and Kim (2016), and Jiang et al. (2021). In the second approach, which we follow here, 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 in 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), in their specification 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. No asymptotic results are established in either work, but in later 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, 2014, 2017), and Silvennoinen and Teräsvirta (2021), g_t is specified as a smooth transition function, and ϕ_t^2 is governed by a first-order GARCH model. In 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, in the former consistency of the first step Gaussian QMLE is proved under the restrictive and unrealistic assumption that ϕ_t^2 is *iid*, and in the latter the standardised error is assumed to be *iid*.[5] Next, Consistency and Asymptotic Normality (CAN) of the parameters of a GARCH model that governs ϕ_t^2 is established in the infeasible case where $\boldsymbol{\theta}$ is known from the first step.[6] Theorem 7 in Silvennoinen and Teräsvirta (2021) uses Theorem 3 in Song et al. (2005) to establish joint CAN of all the parameters of the multivariate model, but the proof of Theorem 7 appears to be incomplete.[7] To accommodate the possibility of cyclical patterns in volatility, which is a common feature of intraday financial returns, 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 established 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 do not establish any asymptotic results. In He et al. (2019), g_t is specified as a seasonal smooth transition function, and CAN is established for a likelihood-based estimator under the assumption that $\phi_t \overset{iid}{\sim} N(0,1)$ (see their assumption A5 in Section 5.2). This is generalised to the multivariate case in He et al. (in press), but ϕ_t is still required to be $\overset{iid}{\sim} N(0,1)$ for all t conditional on the past (see assumption AV6 in He et al., 2023b).

In this paper, g_t is parametrised by a finite dimensional parameter $\boldsymbol{\theta}$ and the sample size T. Specifically,

$$g_t = g_{t,T}(\boldsymbol{\theta}), \tag{3}$$

so $\{g_{t,T} : T \in \mathbb{N},\ 1 \leq t \leq T\}$ forms a triangular array of functions from $\boldsymbol{\Theta}$ to $(0,\infty)$. We

[5]See the assumption that $h_t = 1$ for all t in Theorem 1 of (Amado and Teräsvirta, 2013, p. 145), and Theorem 3 in Silvennoinen and Teräsvirta (2021).

[6]See Amado and Teräsvirta (2013, Theorem 2), and Silvennoinen and Teräsvirta (2021, Theorem 6).

[7]Theorem 3 in Song et al. (2005) assumes a set of unstated regularity conditions hold, and that consistency of the first and second step estimators have been established, see the proof of Theorem 3 on p. 1156 in Song et al. (2005). What the unstated regularity conditions are is particularly important in the current context due to the triangular nature of the sequence of $g_{t,T}$'s, and due to how this may affect invertibility (i.e. the asymptotic irrelevance of the initial values of the GARCH recursion at the true parameter value) and estimation error in the in the second step estimation of the parameters of ϕ_t^2 (cf. Francq and Zakoïan, 2019, p. 190, Francq et al., 2011, and Francq et al., 2016).

prove that the equation-by-equation Gaussian QMLE provides Consistent and Asymptotically Normal (CAN) estimates of $\boldsymbol{\theta}$ for a large number and widely used specifications in (3), both univariate and multivariate versions. In particular, most of the parametric specifications in the literature reviewed above are covered by our theory, since we allow the g_t functions to change with T. A sub-class of special interest contained in (3) is $g_{t,T}(\boldsymbol{\theta}) = g(\boldsymbol{\theta}, t/T)$, $g : \Theta \times [0, 1] \to (0, \infty)$, where time enters in the re-scaled form t/T.

Our results are characterised by several attractive properties. First, there is no need to specify – or know – the exact specification of the stochastic component ϕ_t^2 in the estimation of $\boldsymbol{\theta}$. Also, the ϕ_t^2's can be dependent over time. Our results thus hold for a large class of specifications of ϕ_t^2, including the most common GARCH and Stochastic Volatility (SV) models, both univariate and multivariate. This contrasts with previous results, which rely on specific and often restrictive assumptions on ϕ_t^2. Second, while our results do not require the estimation and explicit specification of a model of ϕ_t^2, a model can nevertheless be estimated in a second step. Our focus is on the GARCH(1,1) model, arguably the most common conditional volatility model in empirical practice. A particularly interesting outcome of our results is that the (scaled) GARCH(1,1) prediction is well-defined under both correct and incorrect specification within our framework, including under certain types of non-stationarities of the stochastic component ϕ_t^2. This is very useful in practice, since it means the user is not required to know the exact DGP of the conditional volatility dynamics, or to rely on restrictive assumptions like strict stationarity of $\{\phi_t^2\}$ or that the scaled error ϕ_t^2/h_t is *iid*. The latter is important, since recent studies reveal that the zero-process of financial returns – both daily and intradaily – is frequently non-stationary, see e.g. Kolokolov et al. (2020), Sucarrat and Grønneberg (2022), Francq and Sucarrat (2023), and Stauskas and Sucarrat (2023). Sections 5.2 and 5.3 provide illustrations. A third attractive property of our estimator is its equation-by-equation nature (cf. Francq and Zakoïan, 2016). This reduces the numerical challenges ("the curse of dimensionality") typically associated with multivariate models. A fourth attractive property pertains to the challenge of modelling non-stationary periodic volatility (e.g. as in intraday returns). Standard ways of describing periodicity do not readily lend themselves to tractable re-formulations in terms of re-scaled time. By instead approaching the problem in terms of the vector-of-seasons representation, this problem is side-stepped (Section 5.3 provides an empirical illustration). Fifth, for parameter identification, previous theoretical results either rely on the high-level assumption that the true parameter is the unique optimiser, see e.g. Amado and Teräsvirta (2013, Assumption AG2 on p. 145), or on restrictive density and *iid* assumptions on the scaled error ϕ_t^2/h_t, see e.g. He et al. (2019), He et al. (2023a), and Silvennoinen and Teräsvirta (2021). Here, we establish milder, more primitive and verifiable sufficient conditions for important sub-classes of g_t, see Section 3. This is possible due to the nature of our estimator. Finally, due to the assumptions we rely on, the Multiplicative Error Model (MEM) interpretation of volatility models holds straightforwardly. The reason is that our assumptions are on ϵ_t^2 and ϕ_t^2, not on ϵ_t and ϕ_t. Accordingly, our results also apply to models of 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.

The rest of the paper is organised as follows. The next section, Section 2, contains our main theoretical results and the assumptions they rely on. Section 3 gives examples of $g_{t,T}$ specifications contained in (3), and derive primitive sufficient conditions for a unique optimiser for three important sub-classes of g_t. Section 4 outlines how a GARCH(1,1) specification can be used to estimate the conditional volatility dynamics in a second step under both correct and incorrect specification. Section 5 contains numerical illustrations of our results, whereas Section 6 concludes. The proofs of our results are contained in the supplemental Appendix.

2 Consistency and Asymptotic Normality

2.1 Consistency

Let $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \ldots, \epsilon_{M,t})'$ denote an M-dimensional multivariate return at t with $M \in \mathbb{N}$, and let

$$\epsilon_{m,t,T}^2 = g_{m,t,T}(\boldsymbol{\theta}_m)\phi_{m,t,T}^2, \qquad m = 1, \ldots, M, \quad 1 \leq t \leq T, \quad T \in \mathbb{N},$$

with $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_M')'$. Our estimator of $\boldsymbol{\theta}$ is derived from the objective function

$$L_T(\boldsymbol{\theta}) = \sum_{m=1}^{M} L_{m,T}(\boldsymbol{\theta}_m) \qquad \text{with} \qquad L_{m,T}(\boldsymbol{\theta}_m) = \frac{1}{T}\sum_{t=1}^{T} l_{m,t,T}(\boldsymbol{\theta}_m, \epsilon_{m,t,T}^2), \tag{4}$$

where

$$l_{m,t,T}(\boldsymbol{\theta}_m, \epsilon_{m,t,T}^2) = \ln g_{m,t,T}(\boldsymbol{\theta}_m) + \frac{\epsilon_{m,t,T}^2}{g_{m,t,T}(\boldsymbol{\theta}_m)}, \qquad m = 1, \ldots, M.$$

Hence, minimisation of (4) leads to the Equation-by-Equation (EBE) Quasi Maximum Likelihood Estimator (QMLE):

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L_T(\boldsymbol{\theta}) = (\widehat{\boldsymbol{\theta}}_1', \ldots, \widehat{\boldsymbol{\theta}}_M')', \quad \widehat{\boldsymbol{\theta}}_m = \arg\min_{\boldsymbol{\theta}_m \in \boldsymbol{\Theta}_m} L_{m,T}(\boldsymbol{\theta}_m), \quad m = 1, \ldots, M. \tag{5}$$

This is an EBE estimator, since the parameters of equation m, i.e. $\boldsymbol{\theta}_m$, can be estimated separately from the parameters of the other equations.

Let $\boldsymbol{\theta}^\star = (\boldsymbol{\theta}_1^{\star\prime}, \ldots, \boldsymbol{\theta}_M^{\star\prime})' \in \prod_{m=1}^{M} \boldsymbol{\Theta}_m = \boldsymbol{\Theta}$ denote the true parameter value. In establishing consistency of the EBE-QMLE, we rely on the following assumptions.

Assumption 1. $\boldsymbol{\Theta} \subset \mathbb{R}^{d_{\boldsymbol{\theta}}}$ *is compact.*

Assumption 2. *For each* $m = 1, \ldots, M$*, let* $\boldsymbol{\Theta}_m^*$ *be an open, convex set containing* $\boldsymbol{\Theta}_m$*. For all* $1 \leq t \leq T$*,* $T \in \mathbb{N}$*,* *(i)* $g_{m,t,T}(\boldsymbol{\theta}_m)$ *is bounded away from zero and infinity, i.e.*

$$0 < \inf_{\boldsymbol{\theta}_m \in \boldsymbol{\Theta}_m^*,\, 1 \le t \le T,\, T \in \mathbb{N}} g_{m,t,T}(\boldsymbol{\theta}_m) \le \sup_{\boldsymbol{\theta}_m \in \boldsymbol{\Theta}_m^*,\, 1 \le t \le T,\, T \in \mathbb{N}} g_{m,t,T}(\boldsymbol{\theta}_m) < \infty.$$

(ii) $\boldsymbol{\theta}_m \mapsto g_{m,t,T}(\boldsymbol{\theta}_m)$ *is continuously differentiable on* $\boldsymbol{\Theta}_m^*$ *and the derivatives* $\dot{\boldsymbol{g}}_{m,t,T}$ *are uniformly bounded:*

$$\sup_{\boldsymbol{\theta}_m \in \boldsymbol{\Theta}_m^*,\, 1 \le t \le T,\, T \in \mathbb{N}} \|\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)\| < \infty.$$

Assumption 3. *For each* $m = 1, \ldots, M$, $\{\epsilon^2_{m,t,T} : 1 \le t \le T,\, T \in \mathbb{N}\}$ *forms a triangular array of a.s. non-negative random variables. Let* $\alpha_{m,T}(k)$ *be the* α*-mixing coefficients corresponding to* $\{\epsilon^2_{m,t,T} : 1 \le t \le T\}$ *and suppose that as* $k \to \infty$,

$$\sup_{T \in \mathbb{N}} \alpha_{m,T}(k) \to 0.$$

Assumption 4. *For each* $m = 1, \ldots, M$, $\phi^2_{m,t,T} := \epsilon^2_{m,t,T}/g_{m,t,T}(\boldsymbol{\theta}^\star_m)$ *is a non-degenerate random variable such that:*

(i) $E(\phi^2_{m,t,T}) = 1$ *for all* $1 \le t \le T$, $T \in \mathbb{N}$*;*

(ii) $\sup_{1 \le t \le T,\, T \in \mathbb{N}} E|\phi^2_{m,t,T}|^{1+\delta} < \infty$ *for some* $\delta > 0$.

Assumption 5. *For each* $m = 1, \ldots, M$, $L_m(\boldsymbol{\theta}_m) := \lim_{T \to \infty} T^{-1} \sum_{t=1}^T E\left(l_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})\right)$ *exists and attains a unique minimum at* $\boldsymbol{\theta}^\star_m \in \boldsymbol{\Theta}_m$.

Assumption 1 is standard. Assumption 2 defines the general class of g_m functions that we consider. Section 3 gives specific examples. Assumption 3 is a mild dependence assumption. In particular, it is substantially milder than the assumptions used by Amado and Teräsvirta (2013) in their univariate derivations, since they rely – in our notation – on $\{\phi^2_{m,t,T}\}$ being *iid*, see their Theorem 1 on p. 145 just below equation (15). Here, by contrast, Assumption 3 is compatible with any volatility model of $\phi^2_{m,t,T}$, stationary or not, as long strong mixing holds. 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 (2019), and to models that are only weakly identified as models of the variance (e.g. intraday high-frequency measures of volatility), see Sucarrat (2021b). Specific examples of GARCH and SV models that are compatible with Assumption 3 are studied in Carrasco and Chen (2002), Lindner (2009), Davis and Mikosch (2009), and Francq and Zakoïan (2019, Ch. 3). Note that, in the definition of mixing size, the underlying mixing coefficients are defined across σ-fields generated by the $\epsilon^2_{m,t,T}$'s and *not* the $\boldsymbol{\epsilon}^2_{t,T}$'s. Since the σ-fields generated by the former are contained in those of generated by the latter, the dependence as measured by mixing is stronger for the latter than for the former (cf. the discussion of Assumption 9).

Assumption 4(i) is a very mild identification assumption. The reason is that almost all volatility models are invariant to scale-transformations in the sense that there exists a finite constant $c > 0$ such that the stochastic process $\{\phi^{2*}_{m,t,T}\}$ with $E(\phi^{2*}_{m,t,T}) = \mu$ for all t, T satisfies $E(c\phi^{2*}_{m,t,T}) = cE(\phi^{2*}_{m,t,T}) = 1$ for all t, 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 $E(\phi^2_{m,t,T}) = 1$ may be restrictive. It should also be noted that Assumption 4(i) is compatible with $\{\phi^2_{m,t,T}\}$ being non-stationary. A case in point is the common situation where the zero-process of a financial return is non-stationary, see e.g. Sucarrat and Grønneberg (2022), and Francq and Sucarrat (2023). In particular, Proposition 2.1(ii) in Sucarrat and Grønneberg (2022) implies $E(\phi^2_{m,t,T})$ can be constant over time even though the zero-process of a financial return is non-stationary. Another implication of Assumption 4(i) is that $E(\epsilon^2_{m,t,T}) = g_{m,t,T}(\boldsymbol{\theta}^\star_m)$. This facilitates interpretation. Assumption 4(ii) is also a fairly mild moment assumption. For example, it holds when $\{\phi^2_{m,t,T}\}$ is governed by a stationary GARCH(1,1), as in (2), with finite $E(\phi^4_{m,t,T})$.[8] Finally, Assumption 5 is a standard regularity condition.

These assumptions are sufficient for consistency of the $\widehat{\boldsymbol{\theta}}_m$ estimators as defined in (5).

Theorem 1 (Consistency)**.** *Suppose Assumptions 1 – 5 hold. Then* $\widehat{\boldsymbol{\theta}}_m \xrightarrow{P} \boldsymbol{\theta}^\star_m$ *for each* $m = 1, \ldots, M$.

2.2 Asymptotic Normality Equation-by-Equation

We now establish asymptotic normality of $\widehat{\boldsymbol{\theta}}_m$, for each $m = 1, \ldots, M$ separately. For this we have to strengthen the imposed conditions.

Assumption 6. $\boldsymbol{\theta}^\star \in \operatorname{int}(\boldsymbol{\Theta})$.

Assumption 7. *For each* $m = 1, \ldots, M$, $\sup_{1 \le t \le T,\, T \in \mathbb{N}} E|\phi^2_{m,t,T}|^{2+\delta_m} < \infty$ *for some* $\delta_m > 0$.

Assumption 8. *For each* $m = 1, \ldots, M$, $1 \le t \le T$, $T \in \mathbb{N}$,

(i) $\boldsymbol{\theta}_m \mapsto g_{m,t,T}(\boldsymbol{\theta}_m)$ *is twice continuously differentiable on a neighbourhood* $\mathcal{V}_m$ *of* $\boldsymbol{\theta}^\star_m$ *in* $\boldsymbol{\Theta}_m$.

(ii) *On* $\mathcal{V}_m$, *define*

$$\boldsymbol{S}_{m,T}(\boldsymbol{\theta}_m) := \frac{1}{T}\sum_{t=1}^{T} \dot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T}),$$

and

$$\hat{\boldsymbol{A}}_{m,T}(\boldsymbol{\theta}_m) := \frac{1}{T}\sum_{t=1}^{T} \ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T}),$$

where $\dot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})$ *and* $\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})$ *are respectively the first and second derivative of* $l_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})$ *with respect to* $\boldsymbol{\theta}_m$.

[8] 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.

(iii) There are deterministic functions $\varphi_{m,t,T} : \mathcal{V}_m \to \mathbb{R}$ and random variables $\upsilon_{m,t,T}$ such that,

$$\|\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})\| \leq \varphi_{m,t,T}(\boldsymbol{\theta}_m)\upsilon_{m,t,T}, \qquad \boldsymbol{\theta}_m \in \mathcal{V}_m,$$

where

$$\sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \sup_{1 \leq t \leq T, T \in \mathbb{N}} \varphi_{m,t,T}(\boldsymbol{\theta}_m) < \infty, \qquad \sup_{1 \leq t \leq T, t \in \mathbb{N}} E\upsilon^2_{m,t,T} < \infty.$$

(iv) There exist random variables $\psi_{m,t,T}$ such that for $\boldsymbol{\theta}_m, \boldsymbol{\theta}'_m \in \mathcal{V}_m$,

$$\|\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T}) - \ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}'_m, \epsilon^2_{m,t,T})\| \leq \psi_{m,t,T}\|\boldsymbol{\theta}_m - \boldsymbol{\theta}'_m\|,$$

where

$$\sup_{1 \leq t \leq T, T \in \mathbb{N}} E\left|\psi_{m,t,T}\right| < \infty.$$

Assumption 9. *For each $m = 1, \ldots, M$, the strong mixing coefficients $\alpha_{m,T}(k)$ satisfy*

$$\sup_{T \in \mathbb{N}} \alpha_{m,T}(k) = O(k^{-\rho_m - \varepsilon}),$$

for some $\varepsilon > 0$, where $\rho_m := r_m/(r_m - 2)$, $r_m = 2 + \delta_m$ with $\delta_m > 0$ as in Assumption 7.

Assumption 10. *For each $m = 1, \ldots, M$, as $T \to \infty$*

$$\boldsymbol{B}_{m,T} := \operatorname{Var}\left(T^{-1/2}\sum_{t=1}^{T} \dot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}^\star_m, \epsilon^2_{m,t,T})\right) \to \boldsymbol{B}^\star_m,$$

with $\boldsymbol{B}^\star_m$ positive definite.

Assumption 11. *For each $m = 1, \ldots, M$, as $T \to \infty$,*

$$\boldsymbol{A}_{m,T}(\boldsymbol{\theta}_m) := \frac{1}{T}\sum_{t=1}^{T} E\left[\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})\right] \to \boldsymbol{A}_m(\boldsymbol{\theta}_m), \quad \boldsymbol{\theta}_m \in \mathcal{V}_m,$$

where $\mathcal{V}_m$ is as in assumption 8. $\boldsymbol{A}^\star_m := \boldsymbol{A}_m(\boldsymbol{\theta}^\star_m)$ is positive definite.

Assumption 6 is standard. Assumption 7 is a strengthened version of Assumption 4(ii). Assumption 8 imposes twice continuous differentiability of each $g_{m,t,T}$ in a neighbourhood of the true parameter and assumes that the second derivative satisfies (iii) a domination condition and (iv) a Lipschitz-type condition. If the second derivative matrix of $g_{m,t,T}$ is bounded on $\mathcal{V}_m$, uniformly in t, T, (iii) holds (see Lemma 1). A sufficient condition for Assumption 8 (given Assumptions 2 & 4) is that $g_{m,t,T}$ is three-times differentiable on $\mathcal{V}_m$ with its second and third derivatives uniformly bounded (over t, T and $\mathcal{V}_m$); see Lemma 7. In Assumption 9, the mixing size $r_m = 2 + \delta_m$ is connected to the moments requirements in Assumption 7. The more dependence (i.e. the higher r_m is), the more moments are required. Assumptions 2, 4 and 9 are sufficient for $\boldsymbol{B}_{m,T} = O(1)$ (cf. Lemma 2); Assumption 10 further ensures that $\boldsymbol{B}_{m,T}$ converges to a positive definite limit. Similarly Assumptions 2, 4 and 8 suffice that each $A_{m,T}(\boldsymbol{\theta}_m) = O(1)$ (cf. Lemma 3); existence of the limit is assumed in Assumption 11.

These assumptions are sufficient for marginal asymptotic normality of each $\hat{\boldsymbol{\theta}}_m$ and that $\hat{\boldsymbol{A}}_{m,T}(\hat{\boldsymbol{\theta}}_m)$ is consistent for $\boldsymbol{A}^\star_m$.

Theorem 2. *Suppose Assumptions 1 – 11 hold. Then $\sqrt{T}(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^\star_m) \xrightarrow{D} \mathcal{N}(\mathbf{0}, [\boldsymbol{A}^\star_m]^{-1}\boldsymbol{B}^\star_m[\boldsymbol{A}^\star_m]^{-1})$ for $m = 1, \ldots, M$.*

Corollary 1. *Suppose Assumptions 1 – 11 hold. Then $\widehat{\boldsymbol{A}}_{m,T}(\widehat{\boldsymbol{\theta}}_m) \xrightarrow{P} \boldsymbol{A}^\star_m$ for $m = 1, \ldots, M$.*

2.2.1 Variance estimation

In order to operationalise inference based on the asymptotic approximation of Theorem 2, beyond Corollary 1 we require a consistent estimator of $\boldsymbol{B}^\star_m$. We can consistently estimate this matrix using kernel weighted sample autocovariances. The general form of our estimator is

$$
\begin{aligned}
\hat{\boldsymbol{B}}_{m,T} &\coloneqq \sum_{j=-T}^{T} k_m(j/\kappa_{m,T})\hat{\boldsymbol{\Gamma}}_{m,T}(j), \\
\hat{\boldsymbol{\Gamma}}_{m,T}(j) &\coloneqq \frac{1}{T}\sum_{t=1}^{T-j} \boldsymbol{i}_{m,t,T}(\hat{\boldsymbol{\theta}}_m, \epsilon^2_{m,t,T})\boldsymbol{i}_{m,t,T}(\hat{\boldsymbol{\theta}}_m, \epsilon^2_{m,t+j,T})' \qquad (j \geq 0), \\
\hat{\boldsymbol{\Gamma}}_{m,T}(j) &\coloneqq \hat{\boldsymbol{\Gamma}}_{m,T}(-j)' \qquad (j < 0).
\end{aligned} \tag{6}
$$

where the $k_m(\cdot)$'s are kernel weights, and $\kappa_{m,T}$ is the bandwidth. The permitted kernel functions are those which belong to the class $\mathscr{K}$ of de Jong and Davidson (2000, p. 409), defined as:

$$
\mathscr{K} \coloneqq \Bigg\{ k : \mathbb{R} \to [-1,1] : k(0) = 1,\ k(x) = k(-x),\ \int |k(x)|\,\mathrm{d}x < \infty,\ \int |\phi(\xi)|\,\mathrm{d}\xi < \infty, \\
k \text{ is continuous at } 0 \text{ and at all but a finite number of points} \Bigg\},
$$

where $\phi(\xi) \coloneqq \frac{1}{2\pi}\int k(x)e^{i\xi x}\,\mathrm{d}x$.

Assumption 12 (Kernel)**.** *For each* $m = 1, \ldots, M$, $k_m \in \mathscr{K}$.

Assumption 13 (Bandwidth)**.** $\kappa_{m,T} \to \infty$ *and* $\kappa_{m,T} = o(T^{1/2})$ *for each* $m = 1, \ldots, M$.

Most kernels considered in the literature satisfy Assumption 12. This includes, amongst other, the Bartlett, Parzen and Quadratic Spectral kernels. Assumption 13 governs the divergence rate of the bandwidth.

The following Proposition is proven by verifying the conditions of Theorem 2.2 of de Jong and Davidson (2000) and demonstrates that – under our Assumptions – $\hat{\boldsymbol{B}}_{m,T}$ is consistent for $\boldsymbol{B}^\star_m$.

Proposition 1. *Suppose Assumptions 1 – 13 hold. Then* $\widehat{\boldsymbol{B}}_{m,T} \xrightarrow{P} \boldsymbol{B}^\star_m$ *for* $m = 1, \ldots, M$.

2.3 Joint Asymptotic Normality

We next establish the joint asymptotic normality of $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1', \ldots, \hat{\boldsymbol{\theta}}_M')'$. Let $\boldsymbol{\epsilon}_{t,T}^2 := (\epsilon_{1,t,T}^2, \ldots, \epsilon_{M,t,T}^2)'$. We can re-write the objective function in (4) as

$$L_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} l_{t,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}_{t,T}^2), \qquad l_{t,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}_{t,T}^2) := \sum_{m=1}^{M} l_{m,t,T}(\boldsymbol{\theta}_m, \epsilon_{m,t,T}^2). \tag{7}$$

Note that (under Assumption 2) the first and second derivatives of $\boldsymbol{\theta} \mapsto l_{t,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}_{t,T}^2)$ are

$$\begin{aligned}
\dot{\boldsymbol{l}}_{t,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}_{t,T}^2) &= \left(\dot{\boldsymbol{l}}_{1,t,T}(\boldsymbol{\theta}_1, \epsilon_{1,t,T}^2)', \ldots, \dot{\boldsymbol{l}}_{M,t,T}(\boldsymbol{\theta}_M, \epsilon_{M,t,T}^2)'\right)', \\
\ddot{\boldsymbol{l}}_{t,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}_{t,T}^2) &= \operatorname{diag}\left(\ddot{\boldsymbol{l}}_{1,t,T}(\boldsymbol{\theta}_1, \epsilon_{1,t,T}^2), \ldots, \ddot{\boldsymbol{l}}_{M,t,T}(\boldsymbol{\theta}_M, \epsilon_{M,t,T}^2)\right).
\end{aligned}$$

Note that under Assumption 11,

$$\boldsymbol{A}_T(\boldsymbol{\theta}) := \frac{1}{T}\sum_{t=1}^{T} E\left[\ddot{\boldsymbol{l}}_{t,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}_{t,T}^2)\right] \to \boldsymbol{A}(\boldsymbol{\theta}) := \operatorname{diag}\left(\boldsymbol{A}_1(\boldsymbol{\theta}_1), \ldots, \boldsymbol{A}_M(\boldsymbol{\theta}_M)\right), \quad \boldsymbol{\theta} \in \mathcal{V}, \tag{8}$$

where $\mathcal{V} := \prod_{m=1}^{M} \mathcal{V}_m$ and $\boldsymbol{A}^\star := \boldsymbol{A}(\boldsymbol{\theta}^\star)$ is positive definite. Define

$$\hat{\boldsymbol{A}}_T(\boldsymbol{\theta}) := \operatorname{diag}\left(\hat{\boldsymbol{A}}_{1,T}(\boldsymbol{\theta}_1), \ldots, \hat{\boldsymbol{A}}_{M,T}(\boldsymbol{\theta}_M)\right). \tag{9}$$

To establish joint asymptotic normality we need to strengthen Assumptions 9 and 10 to (respectively) Assumptions 14 and 15 below.

Assumption 14. *If $\alpha_T(k)$ are the strong mixing coefficients corresponding to $\{\boldsymbol{\epsilon}_{t,T}^2 : 1 \leq t \leq T,\ T \in \mathbb{N}\}$, then*

$$\sup_{T\in\mathbb{N}} \alpha_T(k) = O(k^{-\rho-\varepsilon}),$$

for some $\varepsilon > 0$, where $\rho := \frac{r}{r-2}$, $r := 2 + \min\{\delta_1, \ldots, \delta_M\}$ with δ_m as in Assumption 7.

Assumption 15. *As $T \to \infty$*

$$\boldsymbol{B}_T \coloneqq \mathrm{Var}\left(T^{-1/2}\sum_{t=1}^{T}\dot{\boldsymbol{l}}_{t,T}(\boldsymbol{\theta}^\star, \boldsymbol{\epsilon}_{t,T}^2)\right) \to \boldsymbol{B}^\star,$$

with $\boldsymbol{B}^\star$ positive definite.

Assumptions 2, 4 and 14 are sufficient for $\boldsymbol{B}_T = O(1)$ (cf. Lemma 4); Assumption 15 further ensures that $\boldsymbol{B}_T$ converges to a positive definite limit.

Theorem 3. *Suppose Assumptions 1 – 8, 11, 14 and 15 hold. Then $\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) \xrightarrow{D} \mathcal{N}(\mathbf{0}, [\boldsymbol{A}^\star]^{-1}\boldsymbol{B}^\star[\boldsymbol{A}^\star]^{-1})$.*

2.3.1 Variance estimation

We can consistently estimate $\boldsymbol{B}^\star$ in the same manner as $\boldsymbol{B}_m^\star$. Let

$$\begin{aligned}
\hat{\boldsymbol{B}}_T &\coloneqq \sum_{j=-T}^{T} k(j/\kappa_T)\hat{\boldsymbol{\Gamma}}_T(j), \\
\hat{\boldsymbol{\Gamma}}_T(j) &\coloneqq \frac{1}{T}\sum_{t=1}^{T-j}\dot{\boldsymbol{l}}_{t,T}(\hat{\boldsymbol{\theta}}, \boldsymbol{\epsilon}_{t,T}^2)\dot{\boldsymbol{l}}_{t,T}(\hat{\boldsymbol{\theta}}, \boldsymbol{\epsilon}_{t+j,T}^2)' \quad (j \geq 0), \\
\hat{\boldsymbol{\Gamma}}_T(j) &\coloneqq \hat{\boldsymbol{\Gamma}}_T(-j)' \quad (j < 0).
\end{aligned} \tag{10}$$

where the $k(\cdot)$'s are kernel weights, and κ_T is the bandwidth. We replace Assumptions 12 and 13 by Assumptions 16 and 17 below.

Assumption 16. *$k \in \mathscr{K}$, with $\mathscr{K}$ defined as in Assumption 12.*

Assumption 17. *$\kappa_T \to \infty$ and $\kappa_T = o(T^{1/2})$.*

Proposition 2. *Suppose Assumptions 1 – 8, 11 and 14 – 17 hold. Then $\widehat{\boldsymbol{B}}_T \xrightarrow{P} \boldsymbol{B}^\star$.*

3 Examples of $g_{t,T}$

Here we provide examples of $g_{m,t,T}(\boldsymbol{\theta}_m)$ and derive verifiable conditions that ensure the high-level Assumptions 5 and 8 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:

$$g_{m,t,T}(\boldsymbol{\theta}_m) = \delta_{m,0} + \sum_{l=1}^{s_m} \frac{\delta_{m,l}}{1 + \exp\big(- \gamma_{m,l}(t/T - c_{m,l})\big)}, \tag{11}$$

where $\boldsymbol{\theta}_m = (\boldsymbol{\delta}'_m, \boldsymbol{\gamma}'_m, \boldsymbol{c}'_m)'$ with $\boldsymbol{\delta}_m = (\delta_{m,0}, \delta_{m,1}, \ldots, \delta_{m,s_m})'$, $\boldsymbol{\gamma}_m = (\gamma_{m,1}, \ldots, \gamma_{m,s_m})'$ and $\boldsymbol{c}_m = (c_{m,1}, \ldots, c_{m,s_m})'$. For $l = 1, \ldots, s_m$, the $\delta_{m,l}$ is the total size of break l, $\gamma_{m,l}$ is the speed of transition of break l, c_l is the centre of break location l and s_m is the number of breaks. There are no breaks if $\delta_{m,1} = \cdots = \delta_{m,s_m} = 0$. Note that, for Assumption 5 to hold, the $\delta_{m,l}$'s and $\gamma_{m,l}$'s must all differ from zero. The following result ensures that the high-level Assumptions 5 and 8 hold.

Proposition 3. *Suppose $g_{m,t,T}(\boldsymbol{\theta}_m)$ is given by (11) with $\delta_{m,1} \neq 0, \ldots, \delta_{m,s_m} \neq 0$, with $\gamma_{m,1} \neq 0, \ldots, \delta_{m,s_m} \neq 0$, and with $c_{m,0} < c_{m,1} < \cdots < c_{m,s_m} < c_{m,s_m+1}$ where $c_{m,0} = 0$ and $c_{m,s_m+1} = 1$. Suppose further that Assumption 1 hold, that $\boldsymbol{\theta}^\star_m \in \boldsymbol{\Theta}_m$, that $\boldsymbol{\Theta}^*_m$ is an open, bounded and convex set that contains $\boldsymbol{\Theta}_m$, that Assumption 4 hold, and that $\mathcal{V}_m$ in Assumption 8 is contained in $\boldsymbol{\Theta}_m$. Then Assumptions 2 and 8 hold and the limit $L_m(\boldsymbol{\theta}_m)$ in Assumption 5 exists. Moreover, if the Hessian $\ddot{\boldsymbol{L}}_m(\boldsymbol{\theta}_m)$ is positive definite on $\boldsymbol{\Theta}^*_m$, $L_m(\boldsymbol{\theta}_m)$ attains a unique minimum at $\boldsymbol{\theta}^\star_m \in \boldsymbol{\Theta}^*_m$.*

Proof: See Section C.2 in the appendix.

Establishing conditions under which $\ddot{\boldsymbol{L}}_m(\boldsymbol{\theta}_m)$ is positive definite is tedious, even when there is only one transition ($s_m = 1$). However, numerical verification is straightforward.

3.2 Piecewise constant models

Van Bellegem and Von Sachs (2004) specify $g_{m,t,T}$ as piecewise constant. This amounts to

$$g_{m,t,T}(\boldsymbol{\theta}_m) = \delta_{m,0} + \sum_{l=1}^{s_m} \delta_{m,l} I(t/T \geq c_{m,l}), \qquad \boldsymbol{\theta}_m = (\delta_{m,0}, \delta_{m,1}, \ldots, \delta_{m,s_m})', \tag{12}$$

where $I(A)$ is an indicator function equal to 1 if A holds and 0 otherwise. The values of the possible break-locations $c_{m,1}, \ldots, c_{m,s}$ are thus known and not estimated. To estimate $\boldsymbol{\theta}_m$, Van Bellegem and Von Sachs (2004) proposed the sample variance of each constant period. This does not allow for the joint estimation and inference of multiple break-sizes. Our results, by contrast, permit this.

In Escribano and Sucarrat (2018), $g_{t,T}$ 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 contained in their class of models:

$$\ln g_{m,t,T}(\boldsymbol{\theta}_m) = \delta_{m,0} + \sum_{l=1}^{s_m} \delta_{m,l} \cdot I(t/T \geq c_{m,l}), \qquad \boldsymbol{\theta}_m = (\delta_{m,0}, \delta_{m,1}, \ldots, \delta_{m,s_m})'. \tag{13}$$

Notice that (12) can always be re-written as (13). The advantage of this is that non-negativity constraints on $\boldsymbol{\theta}_m$ are not needed in (13). This simplifies estimation and inference under the null hypothesis that one or more of the coefficients are zero. The following result ensures that the high-level Assumptions 5 and 8 hold.

Proposition 4. *Suppose $g_{mt,T}(\boldsymbol{\theta}_m)$ is given by (13) with $c_{m,0} < c_{m,1} < \cdots < c_{m,s_m} < c_{m,s_m+1}$, where $c_{m,0} = 0$ and $c_{m,s_m+1} = 1$. Suppose further that Assumption 1 hold, that*

$\boldsymbol{\theta}_m^\star \in \boldsymbol{\Theta}_m$, that $\boldsymbol{\Theta}_m^*$ is an open, bounded and convex set that contains $\boldsymbol{\Theta}_m$, that Assumption 4 hold, and that $\mathcal{V}_m$ in Assumption 8 is contained in $\boldsymbol{\Theta}_m$. Then Assumptions 2 and 8 hold, and the limit $L_m(\boldsymbol{\theta}_m)$ in Assumption 5 exists and attains a unique minimum at $\boldsymbol{\theta}_m^\star \in \boldsymbol{\Theta}_m^*$.

Proof: See Section C.3 in the appendix.

3.3 Splines

Engle and Rangel (2008), and Brownlees and Gallo (2010), specify $g_{m,t,T}$ 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 Assumptions 2 and 8. 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

$$\ln g_{m,t,T}(\boldsymbol{\theta}_m) = \delta_{m,0} + \sum_{l=1}^{s_m} \delta_{m,l}(t/T - c_{m,l})^2 I(t/T \geq c_{m,l}), \quad \boldsymbol{\theta}_m = (\delta_{m,0}, \delta_{m,1}, \ldots, \delta_{m,s_m})', \tag{14}$$

where $I(A)$ is an indicator function equal to 1 if A holds and 0 otherwise, and the $c_{m,l}$'s are given knot-locations that are assumed known and therefore not estimated. The value s_m is the number of knots, and $\delta_{m,1}, \ldots, \delta_{m,s_m}$ are the knot-coefficients. Large values of s_m imply more frequent cycles, and the sharpness of each cycle is governed by the knot-coefficients. The following result ensures that the high-level Assumptions 5 and 8 hold.

Proposition 5. *Suppose $g_{t,T}(\boldsymbol{\theta})$ is given by (14) with $c_0 < c_1 < \cdots < c_s < c_{s+1}$, where*

$c_0 = 0$ and $c_{s+1} = 1$. Suppose further that Assumption 1 hold, that $\boldsymbol{\theta}_m^\star \in \boldsymbol{\Theta}_m$, that $\boldsymbol{\Theta}_m^*$ is an open, bounded and convex set that contains $\boldsymbol{\Theta}_m$, that Assumption 4 hold, and that $\mathcal{V}_m$ in Assumption 8 is contained in $\boldsymbol{\Theta}_m$. Then Assumptions 2 and 8 hold and the limit $L_m(\boldsymbol{\theta}_m)$ in Assumption 5 exists. Moreover, if the Hessian $\ddot{\boldsymbol{L}}_m(\boldsymbol{\theta}_m)$ is positive definite on $\boldsymbol{\Theta}_m^*$, $L_m(\boldsymbol{\theta}_m)$ attains a unique minimum at $\boldsymbol{\theta}_m^\star \in \boldsymbol{\Theta}_m^*$

Proof: See Section C.4 in the Appendix.

4 Estimation of conditional volatility

In empirical applications, it is often of interest to obtain estimates of the full conditional covariance matrix $E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'|\mathcal{F}_{t-1})$, where $\mathcal{F}_{t-1} = \sigma\{\boldsymbol{\epsilon}_u, u < t\}$. The conditional volatilities, the $\sigma_{m,t}^2$'s with $\sigma_{m,t}^2 = g_{m,t}h_{m,t}$, are on the diagonal of this matrix. In portfolio analysis, under the unpredictability of returns assumption $E(\boldsymbol{\epsilon}_t|\mathcal{F}_{t-1}) = \mathbf{0}$, the matrix must be positive definite to ensure the conditional variance (i.e. a measure of risk) of a weighted portfolio of asset returns is non-negative. Here, the conditional covariance matrix can be written as

$$E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'|\mathcal{F}_{t-1}) = \boldsymbol{G}_t^{1/2}\boldsymbol{H}_t^{1/2}\boldsymbol{R}_t\boldsymbol{H}_t^{1/2}\boldsymbol{G}_t^{1/2}, \tag{15}$$

where $\boldsymbol{G}_t^{1/2} := \mathrm{diag}\big(g_{1,t}^{1/2}, \ldots, g_{M,t}^{1/2}\big)$, $\boldsymbol{H}_t^{1/2} := \mathrm{diag}\big(h_{1,t}^{1/2}, \ldots, h_{M,t}^{1/2}\big)$ and $\boldsymbol{R}_t$ is a conditional correlation matrix that is either constant or time-varying. Both $\boldsymbol{G}_t^{1/2}$ and $\boldsymbol{H}_t^{1/2}$ are positive definite if their diagonal elements are strictly positive, and $E(\boldsymbol{\phi}_t\boldsymbol{\phi}_t'|\mathcal{F}_{t-1}) = \boldsymbol{H}_t^{1/2}\boldsymbol{R}_t\boldsymbol{H}_t^{1/2}$. As a consequence, if $\boldsymbol{R}_t$ is also positive definite, then $E(\boldsymbol{\phi}_t\boldsymbol{\phi}_t'|\mathcal{F}_{t-1})$ and (15) are also positive definite.[9]

Given first step estimates of the $g_{m,t}$'s, estimates of the $h_{m,t}$'s can be obtained in a second step. While, in principle, any model can be fitted in a second step under

[9] If two square matrices of the same size $\boldsymbol{A}$ and $\boldsymbol{B}$ are positive definite, then also $\boldsymbol{ABA}$ is positive definite.

suitable assumptions, here we study the second step estimation of the scaled GARCH(1,1) specfication under both correct and incorrect specification. We also outline how Dynamic Conditional Correlations (DCCs) can be estimated in a third step while ensuring positive definiteness of (15).

4.1 QML estimation of the scaled GARCH(1,1) model

To simplify notation we omit the subscript T in this subsection. If $\phi_{m,t}^2 := \epsilon_t^2/g_{m,t}(\boldsymbol{\theta}_m^\star)$ is governed by a scaled GARCH(1,1) model, then equation m is given by

$$\phi_{m,t} = \sqrt{h_{m,t}}\eta_{m,t}, \qquad \eta_{m,t} \sim iid(0,1), \tag{16}$$

$$h_{m,t} = \omega_m^\star + \alpha_m^\star \phi_{m,t-1}^2 + \beta_m^\star h_{m,t-1}, \qquad \omega_m^\star, \alpha_m^\star, \beta_m^\star > 0, \quad \omega_m^\star = 1 - \alpha_m^\star - \beta_m^\star. \tag{17}$$

It is the condition $\omega_m^\star = (1 - \alpha_m^\star - \beta_m^\star)$ in (17) which converts the standard GARCH(1,1) into a scaled version, i.e. $E(\phi_{m,t}^2) = 1$ for all t. Note that this condition is not restrictive (recall the discussion of Assumption 4(i) in Section 2.1). An implication of the condition is that only two parameters need to be estimated in the second step, namely $\boldsymbol{\vartheta}_m^\star := (\alpha_m^\star, \beta_m^\star)'$.

In the infeasible case, $\{\phi_{m,t}^2\}$ is observed and the consistency of the QMLE follows trivially under suitable assumptions, see Appendix D.1. In the feasible case, the second step QMLE estimator is similar – but *not* identical – to the target variance estimator of Francq et al. (2011). There, $g_{m,t}$ is constant (i.e. $g_{m,t} = g_m$ for all t), and the asymptotic variance can differ from that of the Ordinary QMLE. Here, the recursive parametrisation differs from that of the target variance estimator. The feasible second step QMLE of $\boldsymbol{\vartheta}_m^\star = (\alpha_m^\star, \beta_m^\star)'$ is

$$\widehat{\boldsymbol{\vartheta}}_{m,T} = \arg\min_{\boldsymbol{\vartheta}_m \in \boldsymbol{\Xi}_m} \frac{1}{T} \sum_{t=1}^{T} \ln \widehat{h}_{m,t} + \frac{\widehat{\phi}_{m,t}^2}{\widehat{h}_{m,t}},$$

where $\boldsymbol{\vartheta}_m = (\alpha_m, \beta_m)'$, $\widehat{\phi}^2_{m,t} = \frac{\epsilon^2_{m,t}}{\widehat{g}_{m,t}}$ and $\widehat{h}_{m,t} = (1-\alpha_m-\beta_m) + \alpha_m \widehat{\phi}^2_{m,t-1} + \beta_m \widehat{h}_{m,t-1}$. Appendix D.2 contains the simulation results of the Two-step QMLE. The results suggests it is consistent, and that its asymptotic variance is the same as that of the Ordinary QMLE, both when $g_{m,t}$ is constant and when it is time-varying, in the experiments investigated.

When the DGP of $\phi^2_{m,t}$ is not a GARCH(1,1), a scaled GARCH(1,1) specification provides mis-specified predictions of volatility. Still, even though the predictions are generated by a mis-specified model, they nevertheless possess several desirable properties that are typically associated with the predictions of a correctly specified conditional expectation. First, the prediction is unbiased for volatility in the unconditional sense, just *as if* it were the correct specification. To see this, suppose $\{\phi^2_{m,t}\}$ is not governed by a GARCH(1,1), and let

$$h_{m,t} = \omega_m + \alpha_m \phi^2_{m,t-1} + \beta_m h_{m,t-1}, \qquad \omega_m, \alpha_m, \beta_m > 0, \quad \omega_m = 1 - \alpha_m - \beta_m, \tag{18}$$

denote the scaled GARCH(1,1) prediction. It is straightforward to verify by backwards recursion that, for any pair (α_m, β_m) that satisfies the parameter constraints in (18),

$$E(h_{m,t}) \;=\; \frac{\omega_m}{1-\beta_m} + \alpha_m \sum_{i=1}^{\infty} \beta_m^{i-1} E(\phi^2_{m,t-i}) \;=\; 1 \qquad \text{for all } t.$$

Accordingly, $g_{m,t}h_{m,t}$ is unbiased for $E(\epsilon^2_{m,t})$ in the unconditional sense, since $E(g_{m,t}h_{m,t}) = g_{m,t} = E(\epsilon^2_{m,t})$ for all t. Note that $E(h_{m,t}) = 1$ also holds under certain types of non-stationarities of $\{\phi^2_{m,t}\}$, e.g. when the zero process is non-stationary (as in the illustrations in Sections 5.2 and 5.3). A second desirable property that characterises the scaled GARCH(1,1) prediction is that the S-steps-ahead prediction satisfies $\lim_{S\to\infty} E(h_{m,t+S}) = E(h_{m,t}) = 1$, just as if $h_{m,t}$ were the true DGP. Finally, a third desirable property the scaled GARCH(1,1) predictions possess under suitable regularity conditions when

(α_m, β_m) are estimated by QML, is weak identification in the sense of Sucarrat (2021b), i.e. $E(\phi^2_{m,t}/h_{m,t}) = 1$, see exercise 7.6 in Francq and Zakoïan (2019). In other words, under mis-specification, QML estimation under suitable assumptions ensures a necessary condition for weak identification holds.

In sum, the practical implication is that the predictions of a QML estimated scaled GARCH(1,1) specification are characterised by several desirable properties, both under correct and incorrect specification. Note that this also applies to the "cross-sectional" version of the scaled GARCH(1,1) prediction, $h_{m,t} = \omega + \alpha\phi_{m-1,t} + \beta h_{m-1,t}$, which we illustrate in Section 5.3.

4.2 Moment estimation of the scaled GARCH(1,1) model

It is well-known that the standard GARCH(1,1) admits a heteroscedastic ARMA(1,1) representation. If the DGP of $\phi^2_{m,t,T}$ is a scaled GARCH(1,1) as in (16)–(17), the representation is

$$\phi^2_{m,t,T} = \omega_m + (\alpha^\star_m + \beta^\star_m)\phi^2_{m,t-1,T} - \beta^\star_m u_{m,t-1,T} + u_{m,t,T}, \qquad \omega^\star_m = 1 - \alpha^\star_m - \beta^\star_m, \quad (19)$$

where $(\alpha^\star_m + \beta^\star_m)$ is the AR-parameter, $(-\beta^\star_m)$ is the MA-parameter and $u_{m,t,T} = \phi^2_{m,t,T} - h_{m,t,T}$ is a heteroscedastic error. Kristensen and Linton (2006) used this representation to derive a closed form estimator of the GARCH(1,1) parameters based on the autocovariance functions of the ARMA(1,1) model. Let $\gamma^\star_{m,t,T,j} = E\big((\phi^2_{m,t,T} - 1)(\phi^2_{m,t-j,T} - 1)\big) = E(\phi^2_{m,t,T}\phi^2_{m,t-j,T}) - 1$ denote the jth. autocovariance, $j = 0, 1, 2$, and let $\rho^\star_m(j) = \gamma^\star_{m,j}/\gamma^\star_{m,0}$ denote the jth. autocorrelation of $\phi^2_{m,t,T}$ under the assumption that the $\gamma^\star_{m,t,Tj}$'s are constant over t and T. For the scaled version of the standard GARCH(1,1), the expressions of $\alpha_m^\star$ and $\beta_m^\star$, respectively, are

$$\alpha_m^\star = \rho_m^\star(2)/\rho_m^\star(1) - \beta_m^\star, \tag{20}$$

$$\beta_m^\star = \frac{b_m^\star - \sqrt{(b_m^\star)^2 - 4}}{2}, \qquad b_m^\star = \frac{\rho_m^\star(2)^2/\rho_m^\star(1)^2 + 1 - 2\rho_m^\star(2)}{\rho_m^\star(2)/\rho_m^\star(1) - \rho_m^\star(1)}, \tag{21}$$

see Kristensen and Linton 2006, pp. 325-326. The sample counterparts of $\rho_m^\star(1)$ and $\rho_m^\star(2)$ can thus be used to obtain consistent estimates of $\alpha_m^\star$ and $\beta_m^\star$ under suitable assumptions. Note that, if $\phi_{m,t,T}^2$ is governed by (16)–(17), then the conditions in (17) imply that $\rho_m^\star(1) > 0$, $b_m^\star > 2$ and $b_m^\star < \infty$. If $\phi_{m,t,T}^2$ is not governed by (16)–(17), the expressions for $\alpha_m^\star$ and $\beta_m^\star$ may still be well-defined (and real valued). In other words, the expressions in (20) and (21) can be used to define a specific-valued scaled GARCH(1,1) prediction under mis-specification as discussed in Section 4.1. This is formalised in Proposition 6 of Appendix D.3.

Feasible estimators of $\gamma_{m,j}^\star$, $j = 0, 1, 2$, are given by

$$\widehat{\gamma}_{m,j}(\widehat{\boldsymbol{\theta}}_{m,T}) = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\phi}_{m,t,T}^2(\widehat{\boldsymbol{\theta}}_{m,T}) - 1\right)\left(\widehat{\phi}_{m,t-j,T}^2(\widehat{\boldsymbol{\theta}}_{m,T}) - 1\right), \tag{22}$$

where $\widehat{\phi}_{m,t,T}^2(\widehat{\boldsymbol{\theta}}_{m,T}) = \epsilon_{m,t,T}^2/g_{m,t,T}(\widehat{\boldsymbol{\theta}}_{m,T})$. Next, these can be used to obtain feasible estimators of $\alpha_m^\star$ and $\beta_m^\star$:

$$\widehat{\alpha}_m = \widehat{\rho}_m(2)/\widehat{\rho}_m(1) - \widehat{\beta}_m, \tag{23}$$

$$\widehat{\beta}_m = \frac{\widehat{b}_m - \sqrt{\widehat{b}_m^2 - 4}}{2}, \qquad \widehat{b}_m = \frac{\widehat{\rho}_m(2)^2/\widehat{\rho}_m(1)^2 + 1 - 2\widehat{\rho}_m(2)}{\widehat{\rho}_m(2)/\widehat{\rho}_m(1) - \widehat{\rho}_m(1)}, \tag{24}$$

where $\widehat{\rho}_m(1) = \widehat{\gamma}_{m,1}/\widehat{\gamma}_{m,0}$ and $\widehat{\rho}_m(2) = \widehat{\gamma}_{m,2}/\widehat{\gamma}_{m,0}$. Consistency of the estimators are established in Appendix D.4. It is worth underlining that the assumptions required for consistency are very mild, since consistency even holds under certain types of non-stationarities of

$\{\phi_t^2\}$ and mis-specification. A drawback of the estimators, however, is that weak identification in the sense of Sucarrat (2021b) may not hold. The proof of asymptotic normality of $(\widehat{\alpha}_m, \widehat{\beta}_m)'$, together with the additional assumptions required, are contained in Appendix D.5.

4.3 Estimation of conditional correlations

Let $\eta_{m,t} = \epsilon_{m,t}/\sqrt{g_{m,t}h_{m,t}}$, $m = 1, \ldots, M$, and let $\boldsymbol{\eta}_t = (\eta_{1,t}, \ldots, \eta_{M,t})'$. Accordingly, $\boldsymbol{\epsilon}_t = \boldsymbol{G}_t^{1/2}\boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t$ and $E(\boldsymbol{\eta}_t\boldsymbol{\eta}_t'|\mathcal{F}_{t-1}) = \boldsymbol{R}_t$. Note also that $Corr(\boldsymbol{\epsilon}_t|\mathcal{F}_{t-1}) = \boldsymbol{R}_t$ under the assumption that $E(\boldsymbol{\epsilon}_t|\mathcal{F}_{t-1}) = \mathbf{0}$ for all t. In applications, an estimator of $\boldsymbol{R}_t$ can be built with the standardised residuals $\widehat{\eta}_{m,t}$, where $\widehat{\eta}_{m,t} = \epsilon_{m,t}/\sqrt{\widehat{g}_{m,t}\widehat{h}_{m,t}}$. If $\boldsymbol{R}_t$ is constant over time, for example, the natural estimator is the sample estimator $\widehat{\boldsymbol{R}} = T^{-1}\sum_{t=1}^{T}\widehat{\boldsymbol{\eta}}_t\widehat{\boldsymbol{\eta}}_t'$, where $\widehat{\eta}_{m,t} = \epsilon_{m,t}/\sqrt{\widehat{g}_{m,t}\widehat{h}_{m,t}}$, $m = 1, \ldots, M$. If $\boldsymbol{R}_t$ is time-varying, then a natural candidate is the corrected Dynamic Conditional Correlation (cDCC) specification of Aielli (2013). Naturally, under mis-specification, $\boldsymbol{R}_t$ must be interpreted as a prediction that is not necessarily equal to $E(\boldsymbol{\eta}_t\boldsymbol{\eta}_t'|\mathcal{F}_{t-1})$.

5 Numerical illustrations

5.1 An efficiency comparison

A question of practical interest is how the efficiency of the Two-step QMLE discussed in Section 4.1 compares with that of the Iterative QMLE proposed by Amado and Teräsvirta (2013), since the latter is substantially more demanding computationally. The latter is also more prone to the "curse of dimensionality" in multiple equation specifications. In single equation specifications, Step 1 of the Two-step QMLE coincides with the first-step of the first iteration of the Iterative QMLE. Table 1 contains the simulation results from a comparison, where the DGP is the same as the g-DGP 2 in Section 4.1. The upper part of Table 1 contains the results of the g_t parameters, whereas the lower part contains the results of the h_t parameters. For the g_t parameters, the Iterative QMLE is not always more efficient for $T \leq 10000$. As the sample size grows very large, however, the results suggest the Iterative QMLE is slightly more efficient. The discrepancies are so small, though, that simulation error cannot be ruled out entirely. For the h_t parameters, the numerical efficiency of the two estimators is similar across all sample sizes. Interestingly, the standard errors are very close to the asymptotic standard errors of the infeasible QMLE (see Appendix D.2), which suggests the prior estimation of the g_t parameters does not affect the efficiency of the h_t parameters in a second step.

5.2 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 (2022). Eros International plc. (EROS) was an Indian multinational 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 until November 2013. Thereafter, there are few zeros. The return series thus exhibits a clear break in the unconditional zero-probability, so the zero-process is non-stationary over the sample. The return process ϵ_t and the transformation $\phi_t^2 = \epsilon_t^2/E(\epsilon_t^2)$ are therefore also non-stationary. Again, to keep notation simple, we suppress the subscripts m (since $m = 1$) and T.

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 $E(\epsilon_t^2)$ 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 use it to investigate whether there are breaks at the aforementioned points of time. More precisely, 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 $(c_1, c_2, c_3)' = (0.427, 0.638, 0.855)$. Our 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 truncation lag is obtained as the integer part of $4(T/100)^{(2/9)}$. The t-ratios of the break-size estimates are 0.806, 5.181 and -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. Finally, the second step QMLE (see Section 4.1) returns an estimated scaled GARCH(1,1) specification equal to $\widehat{h}_t = 0.873 + 0.127\widehat{\phi}_{t-1}^2 + 0.000\widehat{h}_{t-1}$ with $T^{-1}\sum_{t=1}^T \widehat{\phi}_t^2/\widehat{h}_t = 1.044$. In other words, the optimal scaled GARCH(1,1) prediction – optimal in the sense that it is both unbiased unconditionally and satisfies the necessary condition for weak identification – is characterised by an ARCH parameter aqual to 0.127, and a GARCH parameter close to zero. Second step estimation with the moment method (see Section 4.2) gives estimates that violate the parameter conditions, and the value $T^{-1}\sum_{t=1}^T \widehat{\phi}_t^2/\widehat{h}_t$ is far from 1 (i.e. the necessary condition for weak identification fails). So we do not report these estimates.

5.3 Time-varying intraday periodic 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. More recently, He et al. (2019), and He et al. (in press), establish asymptotic results for two classes of periodic smooth transition models. But they do so under restrictive assumptions on ϕ_t (it is assumed *iid* normal in the former, conditionally *iid* normal in the latter). In our case, by contrast, it can be substantially dependent in unknown ways, both intradaily and interdaily. Also, it can be non-stationary. To illustrate our results, we use the vector of seasons representation to model the evolution of intraday unconditional volatility. In effect, our EBE estimator becomes a "period-by-period" estimator (see e.g. Escribano and Sucarrat 2018).

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 via the vector of seasons representation. Consider, for example, the intraday returns $\epsilon_{m,t}$, $m = 1, \ldots, M$, of day t. Often, the sample averages $T^{-1}\sum_{t=1}^{T}\epsilon_{m,t}^2$, $m = 1, \ldots, M$, are used to estimate the intraday unconditional volatilities $E(\epsilon_{1,t}^2), \ldots, E(\epsilon_{M,t}^2)$. The collection of sample averages is a special case of the period-by-period estimator. But it is only consistent in the special case where the unconditional intraday volatilities are constant across days, i.e. for each $m = 1, \ldots, M$ we have $E(\epsilon_{m,t_1}^2) = E(\epsilon_{m,t_2}^2)$ for all t_1, t_2. By contrast, period-by-period estimation as sketched here can also be used to estimate unconditional intraday volatilities that vary across days. Again, to simplify notation, we suppress the subscript T.

For illustration we use intraday hourly USD/EUR exchange rate returns. 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 (https://www.forexite.com), a currency trading platform, from 2 January 2017 to 31 December 2018. This corresponds to 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. 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 similar to that of Engle and Rangel (2008) with re-scaled time and four knots at equidistant locations, i.e.

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

for each period $m = 1, \ldots, M$. Table 2 contains the estimation results together with a Wald-test of $H_0 : \delta_{m,1} = \cdots = \delta_{m,4} = 0$. Under the null the unconditional volatility of period m is thus 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.

Since $E(\phi_{m,t}^2) = 1$ for all m and t, the intraday or "cross-sectional" scaled GARCH(1,1) prediction $h_{m,t} = \omega + \alpha\phi_{m-1,t} + \beta h_{m-1,t}$ is well-defined and characterised by the properties sketched in Section 4.1. In other words, it is straightforward to estimate a single, scaled GARCH(1,1) prediction of volatility for both within and across days, even when the $g_{m,t}$'s are time-varying and the ϕ_t^2's are non-stationary. The QML estimated specification is $\widehat{h}_{m,t} = 0.106 + 0.052\widehat{\phi}^2_{m-1,t} + 0.8418\widehat{h}_{m-1,t}$ with $(T\cdot M)^{-1}\sum_{t=1}^{T}\sum_{m=1}^{M}\widehat{\phi}^2_{m,t}/\widehat{h}_{m,t} = 1.0004$. Second step estimation with the moment method (see Section 4.2), by contrast, returns estimates that violate the parameter conditions, and the value $(T\cdot M)^{-1}\sum_{t=1}^{T}\sum_{m=1}^{M}\widehat{\phi}^2_{m,t}/\widehat{h}_{m,t}$ is not close to 1 (i.e. the necessary condition for weak identification fails). So we do not report these estimates.

6 Conclusions

We conclude by summarising our contributions. We derive a unified and general framework for consistent estimation and asymptotically normal inference in a large class of models of time-varying unconditional volatility, both univariate and multivariate. Our framework is based on the equation-by-equation version of the Gaussian QMLE, and it is characterised by several attractive properties. One is its ease of implementation, since the equation-by-equation nature of the estimator reduces the curse of dimensionality in multivariate models. Another attractive property is that the exact specification of the conditional volatility dynamics need not be known or estimated. However, in empirical applications, models of the conditional volatility dynamics can nevertheless be fitted in a second step, if desired. In particular, as we show, the scaled GARCH(1,1) is well-defined under both correct and incorrect specification, in both the univariate and multivariate cases. Our multivariate results can also be used to estimate non-stationary periodic volatility by framing the problem via the vector of seasons representation. This leads to a period-by-period estimator, whereby not only the variation in intraday unconditional volatility is modelled, but also the variation over days for each intraday period. Another novel property of our results is that they are valid when the zero-process of financial returns is non-stationary. This is important, since recent studies document that financial returns, both daily and intradaily, are widely characterised by a non-stationary zero-process. In the multivariate case, our results are also valid when the time-varying correlations are non-stationary, even when this is not due to a non-stationary zero-process. Next, due to the assumptions we rely upon, our results extend directly to the Multiplicative Error Model (MEM) interpretation of volatility models. Finally, we illustrated the usefulness of our results in three applications.

References

Aielli, G. P. (2013). Dynamic Conditional Correlations: On Properties and Estimation. *Journal of Business and Economic Statistics 31*, 282–299. http://dx.doi.org/10.1080/07350015.2013.771027.

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*. London: Routledge.

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.

Andrews, D. W. K. (1992). Generic uniform convergence. *Econometric Theory 8*(2), 241–257.

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

Bernstein, D. S. (2009). *Matrix Mathematics: Theory, Facts, and Formulas.* Princeton University Press.

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 (2024). Modeling nonstationary financial volatility with the R package tvgarch. *Journal of Statistical Software 108*(9), 1–38.

Carrasco, M. and X. Chen (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory 18*, pp. 17–39.

Davidson, J. (1994). *Stochastic limit theory.* Oxford University Press.

Davis, R. A. and T. Mikosch (2009). Probabilistic Properties of Stochastic Volatility Models. In T. Mikosch, T. Kreiss, J.-P. Davis, R. Andersen, and T. Gustav (Eds.), *Handbook of Financial Time Series*, pp. 255–268. Berlin: Springer.

de Jong, R. M. (1997). Central limit theorems for dependent heterogeneous random variables. *Econometric Theory 13*(3), 353–367.

de Jong, R. M. and J. Davidson (2000). Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica 68*(2), 407–423.

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.

Folland, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications.* New York, NY, USA: John Wiley & Sons, Inc.

Francq, C., L. Horvath, and J.-M. Zakoian (2011). Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics 9*, 619–656.

Francq, C., L. Horvath, and J.-M. Zakoian (2016). Variance Targeting of Estimation of Multivariate GARCH Models. *Journal of Financial Econometrics 14*, 353–382.

Francq, C. and G. Sucarrat (2023). Volatility Estimation when the Zero-Process is Nonstationary. *Journal of Business and Economic Statistics 41*, 53–66.

Francq, C. and L. Q. Thieu (2019). Qml inference for volatility models with covariates. *Econometric Theory 35*, 37–72. DOI: 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.

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

He, C., J. Kang, A. Silvennoinen, and T. Teräsvirta (2023a). Long monthly temperature series and the vector seasonal shifting mean and covariance autoregressive model. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2023.105494.

He, C., J. Kang, A. Silvennoinen, and T. Teräsvirta (2023b). Supplement to 'long monthly temperature series and the vector seasonal shifting mean and covariance autoregressive model'. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2023.105494.

He, C., J. Kang, T. Teräsvirta, and S. Zhang (2019). The shifting seasonal mean autoregressive model and seasonality in the central england monthly temperature, 1772-2016. *Econometrics and Statistics 12*, 1–24. https://doi.org/10.1016/j.ecosta.2019.05.005.

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

Horn, R. A. and C. R. Johnson (2013). *Matrix Analysis* (2 ed.). Cambridge University Press.

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.

Koo, B. and O. Linton (2015). Let's get LADE: Robust estimation of semiparametric multiplicative volatility models. *Econometric Theory 31*, 671–702.

Kristensen, D. and O. Linton (2006). A Closed-Form Estimator for the GARCH(1,1) Model. *Econometric Theory 22*, 323–337.

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.

Lindner, A. (2009). Stationarity, Mixing, Distributional Properties and Moments of GARCH(p, q)–Processes. In T. Mikosch, T. Kreiss, J.-P. Davis, R. Andersen, and T. Gustav (Eds.), *Handbook of Financial Time Series*, pp. 43–84. Berlin: Springer.

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.

Newey, W. K. and D. McFadden (1994). Chapter 36 large sample estimation and hypothesis testing. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume 4*, Volume 4 of *Handbook of Econometrics*, pp. 2111–2245. Elsevier.

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.

Rudin, W. (1976). *Principles of Mathematical Analysis* (3 ed.). McGraw Hill, Inc.

Silvennoinen, A. and T. Teräsvirta (2021). Consistency and asymptotic normality of maximum likelihood estimators of a multiplicative time-varying smooth transition correlation garch model. *Econometrics and Statistics*. https://doi.org/10.1016/j.ecosta.2021.07.008.

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.

Stauskas, O. and G. Sucarrat (2023). Testing the Zero-Process of Intraday Financial Returns for Non-Stationary Periodicity. Working Paper.

Sucarrat, G. (2021a). garchx: Flexible and Robust GARCH-X Modelling. *The R Journal 13*, 276–291. https://journal.r-project.org/archive/2021/RJ-2021-057/.

Sucarrat, G. (2021b). 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 (2022). Risk Estimation with a Time Varying Probability of Zero Returns. *Journal of Financial Econometrics 20*, 278–309.

Sundaram, R. K. (1996). *A First Course in Optimization Theory*. Cambridge University Press.

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

van der Vaart, A. W. (1998). *Asymptotic Statistics*. New York, NY, USA: Cambridge University Press.

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.

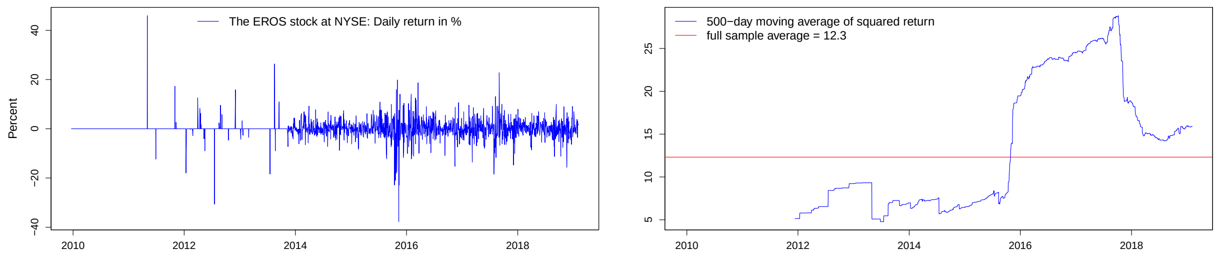

Figure 1: Daily log-returns in % of the EROS stock at NYSE (left) and 500-day moving average of squared returns (right), 21 December 2009 – 4 February 2021 (see Section 5.2). Datasource: Bloomberg

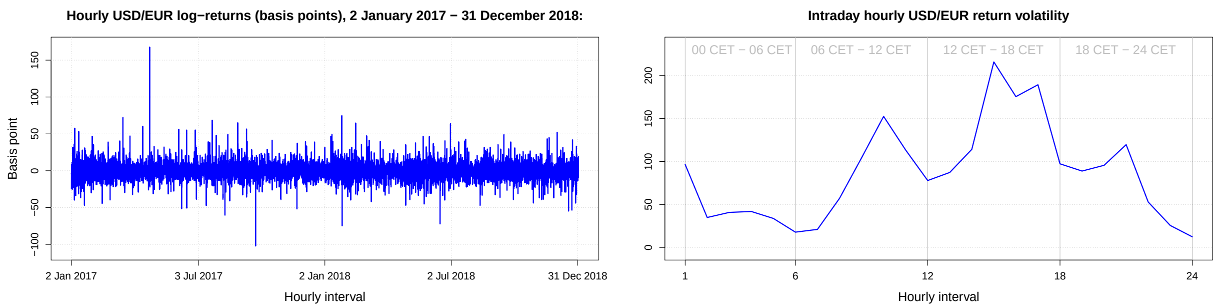

Figure 2: Hourly log-returns in basis points of the USD/EUR exchange rate (left) and estimates (assuming constancy over t) of its intraday hourly volatility (right), 2 January 2017 – 31 December 2018 (see Section 5.3). Datasource: Forexite

Table 1: Comparison of the Two-step QMLE and the Iterative QMLE of Amado and Teräsvirta (2013), see Section 5.1

	T	$m(\widehat{\delta}_0)$	$se(\widehat{\delta}_0)$	$m(\widehat{\delta}_1)$	$se(\widehat{\delta}_1)$	$m(\widehat{\gamma})$	$se(\widehat{\gamma})$	$m(\widehat{c})$	$se(\widehat{c})$
g_t:	Two-step QMLE (step 1):								
	2000	−0.0306	0.1421	0.1928	0.7559	6.8617	19.326	0.0197	0.1180
	5000	−0.0146	0.0933	0.0848	0.4203	1.6698	6.8159	0.0099	0.0686
	10000	−0.0050	0.0569	0.0269	0.1961	0.6592	2.9829	0.0031	0.0339
	20000	−0.0039	0.0386	0.0170	0.1260	0.2612	1.8592	0.0011	0.0237
	40000	−0.0032	0.0262	0.0156	0.0898	0.0720	1.3271	0.0013	0.0167
	80000	−0.0018	0.0171	0.0062	0.0572	0.0337	0.8697	0.0002	0.0111
	Iterative QMLE:								
	2000	−0.0380	0.1548	7.9542	145.1647	10.4640	40.8952	0.0240	0.1189
	5000	−0.0206	0.1086	0.1939	2.4315	1.3270	10.5000	0.0089	0.0570
	10000	−0.0088	0.0718	0.0248	0.2045	0.4501	2.6239	0.0030	0.0295
	20000	−0.0055	0.0490	0.0127	0.1459	0.1787	1.6676	0.0015	0.0223
	40000	−0.0037	0.0306	0.0106	0.0863	0.0476	1.1624	0.0010	0.0144
	80000	−0.0018	0.0171	0.0044	0.0510	0.0177	0.7819	0.0003	0.0097

	T	$m(\widehat{\omega})$	$se(\widehat{\omega})$	$m(\widehat{\alpha})$	$se(\widehat{\alpha})$	$m(\widehat{\beta})$	$se(\widehat{\beta})$
h_t:	Two-step QMLE (step 2):						
	2000	–	–	−0.0006	0.0197	−0.0180	0.0525
	5000	–	–	−0.0003	0.0121	−0.0057	0.0271
	10000	–	–	0.0000	0.0087	−0.0031	0.0192
	20000	–	–	0.0000	0.0059	−0.0017	0.0135
	40000	–	–	0.0001	0.0042	−0.0009	0.0092
	80000	–	–	0.0001	0.0032	−0.0005	0.0068
	Iterative QMLE:						
	2000	0.0189	0.0587	0.0000	0.0200	−0.0195	0.0524
	5000	0.0069	0.0231	0.0000	0.0122	−0.0064	0.0271
	10000	0.0037	0.0144	0.0001	0.0087	−0.0034	0.0192
	20000	0.0022	0.0103	0.0001	0.0059	−0.0019	0.0135
	40000	0.0011	0.0069	0.0001	0.0042	−0.0010	0.0092
	80000	0.0006	0.0049	0.0001	0.0032	−0.0006	0.0068

DGP: $\epsilon_t = \sqrt{g_t}\phi_t$, $\phi_t = \sqrt{h_t}\eta_t$, $\eta_t \overset{iid}{\sim} N(0,1)$, $h_t = 0.1 + 0.1\phi_{t-1}^2 + 0.8h_{t-1}$, $g_t = 0.5 + 1.5\big(1 + \exp(-10(t/T - 0.5)\big)^{-1}$. 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). The Two-step QMLE is implemented with own code. The Iterative QMLE is implemented with the `tvgarch()` function of the CRAN package **tvgarch** (Campos-Martins and Sucarrat, 2024).

Table 2: Spline estimates of intraday hourly volatility (see Section 5.3)

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 (13.4496)	21.374 (35.3855)	−6.012 (42.0653)	53.597 (52.1279)	515	21.853 [0.0004]
2	3.775 (0.2522)	1.133 (5.5681)	−11.494 (17.2062)	9.289 (27.5929)	53.769 (44.4827)	516	13.659 [0.0085]
3	3.643 (0.1799)	2.215 (5.6773)	−8.954 (18.0857)	18.192 (29.0796)	−22.852 (51.8098)	516	0.518 [0.9717]
4	3.881 (0.2042)	−1.513 (5.8472)	−0.526 (17.8325)	10.394 (24.7791)	−31.197 (36.2248)	516	3.739 [0.4425]
5	3.374 (0.2480)	−3.385 (5.6155)	17.916 (25.1050)	−22.709 (50.8935)	−58.255 (65.4081)	516	11.252 [0.0239]
6	2.735 (0.1639)	4.456 (4.8265)	−13.627 (14.8996)	17.628 (21.9421)	−6.619 (35.6283)	516	2.678 [0.6131]
7	2.828 (0.1880)	2.119 (4.8305)	2.422 (14.8866)	−18.812 (23.0809)	24.475 (41.8878)	516	4.935 [0.2940]
8	3.857 (0.1761)	0.895 (4.7521)	−5.108 (15.6562)	27.629 (27.3992)	−85.692 (50.4148)	515	4.604 [0.3303]
9	4.521 (0.1046)	1.519 (3.3635)	1.150 (11.0441)	−10.082 (17.3547)	−6.229 (26.6335)	516	9.899 [0.0422]
10	5.017 (0.1083)	−4.345 (4.1078)	20.107 (14.1018)	−44.509 (23.5979)	86.918 (39.2136)	517	5.068 [0.2804]
11	4.639 (0.1883)	−6.359 (4.2830)	31.118 (13.0861)	−61.819 (19.7582)	82.749 (32.9971)	517	12.851 [0.0120]
12	4.270 (0.1340)	3.207 (6.0620)	−9.070 (19.7664)	7.739 (27.8529)	8.046 (31.5354)	517	1.285 [0.8638]
13	4.557 (0.2465)	−1.464 (5.8333)	4.095 (16.2241)	−6.101 (20.1854)	−3.786 (27.0236)	517	6.345 [0.1748]
14	4.543 (0.1308)	−2.035 (5.9661)	19.024 (20.3574)	−47.267 (30.2696)	52.408 (33.8309)	517	6.326 [0.1761]
15	5.459 (0.1745)	−2.838 (6.1909)	10.719 (19.1024)	−19.411 (26.1782)	11.935 (39.5786)	517	4.168 [0.3838]
16	5.119 (0.1370)	2.594 (3.5312)	−9.415 (11.2713)	11.557 (18.2850)	10.993 (29.4824)	517	4.806 [0.3078]
17	5.409 (0.2083)	−4.002 (6.1470)	10.676 (18.9849)	−19.079 (26.0639)	51.860 (33.0372)	517	6.545 [0.1620]
18	4.525 (0.1408)	2.129 (4.2691)	−5.967 (14.9249)	8.191 (25.3200)	−22.307 (35.8577)	517	1.757 [0.7804]
19	4.538 (0.2072)	−8.461 (4.6763)	28.788 (14.3697)	−45.468 (22.1037)	89.723 (37.5709)	517	11.391 [0.0225]
20	4.812 (0.2824)	−12.499 (6.1907)	42.377 (18.0686)	−66.247 (25.3002)	78.278 (37.1904)	515	7.317 [0.1201]
21	4.762 (0.2543)	12.607 (8.3223)	−51.998 (27.7935)	68.146 (41.7346)	13.428 (60.2769)	516	13.748 [0.0081]
22	4.247 (0.2998)	−9.138 (6.8886)	26.477 (22.1289)	−37.022 (36.0346)	50.927 (50.1777)	516	3.522 [0.4746]
23	3.192 (0.1530)	−0.677 (6.5806)	−1.918 (21.7796)	17.689 (34.8812)	−34.275 (58.9946)	412	3.196 [0.5257]
24	2.203 (0.1849)	20.118 (4.9578)	−68.402 (15.6431)	80.915 (24.5484)	−18.442 (37.1556)	413	34.318 [0.0000]

The estimated model is $\ln g_{m,t} = \delta_{m,0} + \sum_{l=1}^{4} \delta_{m,l}(t/T - c_l)^2 I(t/T \geq c_l)$ 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).

A Proofs of main results

A.1 Proof of Theorem 1

By Theorem 5.7 in van der Vaart (1998) it suffices to show (i) uniform convergence in probability of $L_{m,T}$ to L_m over $\mathbf{\Theta}_m$ and (ii) that the maximiser of L_m (over $\mathbf{\Theta}_m$) is well-separated.

(i) Uniform convergence in probability

Let $\bar{L}_{m,T}(\boldsymbol{\theta}_m) := \frac{1}{T}\sum_{t=1}^{T} E\left[l_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})\right]$. We first show that $L_{m,T}(\boldsymbol{\theta}_m) - \bar{L}_{m,T}(\boldsymbol{\theta}_m)$ converges to zero in probability, pointwise in $\boldsymbol{\theta}_m \in \mathbf{\Theta}_m$. Let $X_{t,T} := l_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T}) - E\left[l_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})\right]$ and let $\mathcal{F}_{t,T} := \sigma(\{X_{i,T} : 1 \le i \le t\})$. Combining Assumptions 2 and 4 yields that $\sup_{1\le t\le T, T\in\mathbb{N}} E|X_{t,T}|^p < \infty$ for some $p > 1$. Hence $\{X_{t,T} : 1 \le t \le T,\ T \in \mathbb{N}\}$ is uniformly integrable and, given Assumption 3, $\{X_{t,T}/T, \mathcal{F}_{t,T} : 1 \le t \le T,\ T \in \mathbb{N}\}$ forms a L_1-mixingale array with respect to the constants $c_{t,T} = 1/T$ by Theorems 14.1 & 14.2 in Davidson (1994). Therefore by Theorem 19.11 in Davidson (1994), $E\left|L_{m,T}(\boldsymbol{\theta}_m) - \bar{L}_{m,T}(\boldsymbol{\theta}_m)\right| \to 0$ (for any fixed $\boldsymbol{\theta}_m \in \mathbf{\Theta}_m$), implying the required convergence in probability.

Next observe that for $\boldsymbol{\theta}_m, \boldsymbol{\theta}'_m \in \mathbf{\Theta}^*_m$, by the mean value theorem there is a $c \in [0, 1]$ such that for $\boldsymbol{\theta}^\dagger_m := \boldsymbol{\theta}_m(1 - c) + c\,\boldsymbol{\theta}'_m \in \Theta^*_m$,

$$\begin{aligned}
&\left|l_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T}) - l_{m,t,T}(\boldsymbol{\theta}'_m, \epsilon^2_{m,t,T})\right| \\
&\quad \le \|\theta_m - \theta'_m\| \left[\frac{\left\|\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}^\dagger_m)\right\|}{|g_{m,t,T}(\boldsymbol{\theta}^\dagger_m)|} + \frac{\epsilon^2_{m,t,T}\left\|\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}^\dagger_m)\right\|}{g^2_{m,t,T}(\boldsymbol{\theta}^\dagger_m)}\right] \\
&\quad \le C\|\boldsymbol{\theta}_m - \boldsymbol{\theta}'_m\| \times (1 + \epsilon^2_{m,t,T}),
\end{aligned} \tag{25}$$

where C is an absolute constant and the last line follows from Assumption 2. By (25) and Jensen's inequality

$$\begin{aligned}
|\bar{L}_{m,T}(\boldsymbol{\theta}_m) - \bar{L}_{m,T}(\boldsymbol{\theta}'_m)| &\le \frac{1}{T}\sum_{t=1}^{T}\left|E\left[l_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T}) - l_{m,t,T}(\boldsymbol{\theta}'_m, \epsilon^2_{m,t,T})\right]\right| \\
&\le C\|\boldsymbol{\theta}_m - \boldsymbol{\theta}'_m\|\frac{1}{T}\sum_{t=1}^{T} E\left(1 + \epsilon^2_{m,t,T}\right) \\
&\le \mathsf{L}\|\boldsymbol{\theta}_m - \boldsymbol{\theta}'_m\|,
\end{aligned}$$

where the last line follows from $\sup_{1\le t\le T, T\in\mathbb{N}} E\epsilon^2_{m,t,T} < \infty$ which is ensured by Assumptions 2 and 4. Hence $\{\bar{L}_{m,T} : T \in \mathbb{N}\}$ is uniformly equicontinuous (on Θ^*_m).

We next establish a Lipschitz–type bound on $L_{m,T}$. By (25)

$$\begin{aligned}
|L_{m,T}(\boldsymbol{\theta}_m) - L_{m,T}(\boldsymbol{\theta}'_m)| &\le \frac{1}{T}\sum_{t=1}^{T}\left|l_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T}) - l_{m,t,}(\boldsymbol{\theta}'_m, \epsilon^2_{m,t,T})\right| \\
&\le C\|\boldsymbol{\theta}_m - \boldsymbol{\theta}'_m\| \times \frac{1}{T}\sum_{t=1}^{T} 1 + \epsilon^2_{m,t,T}.
\end{aligned}$$

Assumptions 2 and 4 together ensure that $\sup_{1\le t\le T, T\in\mathbb{N}} E\epsilon^2_{m,t,T} < \infty$ and so $\frac{1}{T}\sum_{t=1}^{T} 1 + \epsilon^2_{m,t,T} = O_P(1)$ by Markov's inequality. This, along with the preceding paragraph, verifies

Assumption SE-1 in Andrews (1992) and hence $\{L_{m,T} - \bar{L}_{m,T} : T \in \mathbb{N}\}$ is stochastically equicontinuous by Lemma 1 of Andrews (1992). In view of Assumption 1 and the pointwise convergence in probability previously established, Theorem 1 in Andrews (1992) therefore yields that

$$\sup_{\boldsymbol{\theta}_m \in \boldsymbol{\Theta}_m} |L_{m,T}(\boldsymbol{\theta}_m) - \bar{L}_{m,T}(\boldsymbol{\theta}_m)| \xrightarrow{P} 0.$$

By Assumption 5, $\bar{L}_{m,T} \to L_m$ pointwise on $\boldsymbol{\Theta}_m$. Since $\{\bar{L}_{m,T} : T \in \mathbb{N}\}$ is equicontinuous on $\boldsymbol{\Theta}_m$, this convergence is in fact uniform on $\boldsymbol{\Theta}_m$:

$$\sup_{\boldsymbol{\theta}_m \in \boldsymbol{\Theta}_m} |\bar{L}_{m,T}(\boldsymbol{\theta}_m) - L_m(\boldsymbol{\theta}_m)| \to 0.$$

Combination of the last two displays yields (i).

(ii) Well-separated point of maximum

By Assumption 5, L_m has a unique maximiser. Since $\boldsymbol{\Theta}_m$ is compact by Assumption 1, the same is true of $\{\boldsymbol{\theta} \in \boldsymbol{\Theta}_m : \|\boldsymbol{\theta} - \boldsymbol{\theta}_m^\star\| \geq \varepsilon\}$ for any $\varepsilon > 0$. It is therefore sufficient to note that L_m is continuous on $\boldsymbol{\Theta}_m$, which follows by the uniform limit theorem (e.g. Theorem 7.12 in Rudin (1976)) since each $\bar{L}_{m,T}$ is continuous and $\bar{L}_{m,T} \to L_m$ uniformly on $\boldsymbol{\Theta}_m$.

A.2 Proof of Theorem 2

We verify the conditions in Theorem 3.1 of Newey and McFadden (1994).[10] $\hat{\boldsymbol{\theta}}_m$ minimises $L_{m,T}(\boldsymbol{\theta}_m)$ over $\boldsymbol{\Theta}_m$ by definition and is consistent for $\boldsymbol{\theta}_m^\star$ by Theorem 1. Condition (i) holds by Assumption 6. Condition (ii) holds by Assumption 8, the chain rule and the definition of $l_{m,t,T}$. For condition (iii) we show that

$$\boldsymbol{S}_{m,T}(\boldsymbol{\theta}_m^\star) = T^{-1/2} \sum_{t=1}^{T} \dot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m^\star, \epsilon_{m,t,T}^2) \rightsquigarrow \mathcal{N}(0, \boldsymbol{B}_m^\star). \tag{26}$$

For this, firstly note that

$$\dot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m^\star, \epsilon_{m,t,T}^2) = \frac{\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m^\star)}{g_{m,t,T}(\boldsymbol{\theta}_m^\star)} - \frac{\epsilon_{m,t,T}^2 \dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m^\star)}{g_{m,t,T}^2(\boldsymbol{\theta}_m^\star)} = \frac{\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m^\star)}{g_{m,t,T}(\boldsymbol{\theta}_m^\star)} \left(1 - \phi_{m,t,T}^2\right), \tag{27}$$

and hence by Assumption 4,

$$E\left[\dot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m^\star, \epsilon_{m,t,T}^2)\right] = \frac{\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m^\star)}{g_{m,t,T}(\boldsymbol{\theta}_m^\star)} E\left[1 - \phi_{m,t,T}^2\right] = 0.$$

By Assumption 2

$$\|\dot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m^\star, \epsilon_{m,t,T}^2)\| = \left\| \frac{\|\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m^\star)\|}{|g_{m,t,T}(\boldsymbol{\theta}_m^\star)|} + \frac{\epsilon_{m,t,T}^2 \|\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m^\star)\|}{g_{m,t,T}^2(\boldsymbol{\theta}_m^\star)} \right\| \leq C(1 + \epsilon_{m,t,T}^2). \tag{28}$$

[10] Theorem 3.1 in Newey and McFadden (1994) applies to maximisers; our estimator is a minimiser. Multipliying by -1 in the appropriate places allows the application of this result under the conditions shown.

For notational convenience we drop the arguments of $\dot{\boldsymbol{l}}_{m,t,T}$, $g_{m,t,T}$ and $\dot{\boldsymbol{g}}_{m,t,T}$ for the remainder of this part of the proof. Let $Z_{t,T} := T^{-1/2}\lambda'\dot{\boldsymbol{l}}_{m,t,T}$ for $\|\lambda\|_2 = 1$. $\sigma_T := \|\sum_{t=1}^{T} Z_{t,T}\|_{L_2}$ and $X_{t,T} = Z_{t,T}/\sigma_T$.[11] Let $\mathcal{F}_{m,t,T} := \sigma(\epsilon^2_{m,1,T}, \ldots, \epsilon^2_{m,t,T})$. We will verify the conditions of Corollary 2 in de Jong (1997). (a) follows as $X_{t,T}$ is a mean-zero random variable with $\|\sum_{t,T} X_{t,T}\|_{L_2} = 1$ and each $X_{t,T}$ is $\mathcal{F}_{m,t,T}$–measurable. For (b) set $c_{t,T} := \max\{\|Z_{t,T}\|_{L_2}, 1\}/\sigma_T$. By the moment bounds in Assumption 7 and (28), one concludes that

$$\sup_{1\le t\le T, T\in\mathbb{N}} \|X_{t,T}/c_{t,T}\|_{L_{r_m}} \le \sup_{1\le t\le T, T\in\mathbb{N}} \sigma_T\|X_{t,T}\|_{L_{r_m}} = \sup_{1\le t\le T, T\in\mathbb{N}} \|Z_{t,T}\|_{L_{r_m}} < \infty, \tag{29}$$

For (c), since each $X_{t,T}$ is $\mathcal{F}_{m,t,T}$–measurable (and in L_2), it is trivially L_2 – NED (of any size) on $(\epsilon^2_{m,t,T})_{1\le t\le T, T\in\mathbb{N}}$ and by Assumption 9 this latter array is α-mixing of size $-\rho_m$, with $\rho_m := r_m/(r_m - 2)$.[12] Finally, we note that by the moment bounds in Assumption 4 and (28)

$$Tc^2_{t,T} \le \frac{1}{\sigma_T^2}\max\{\|\dot{\boldsymbol{l}}_{m,t,T}\|^2_{L_2}, 1\} \lesssim \frac{1}{\sigma_T^2}.$$

By Assumption 10,

$$\sigma_T^2 = \left\|\sum_{t=1}^{T} Z_{t,T}\right\|^2_{L_2} = \lambda'\boldsymbol{B}_{m,T}\lambda \to \lambda'\boldsymbol{B}^\star_m\lambda > 0. \tag{30}$$

Combination of the preceding displays permits the conclusion that $c^2_{t,T} = O(T^{-1})$, establishing the final condition of Corollary 2 of de Jong (1997) with $\beta = \gamma = 0$. Therefore, $\sum_{t=1}^{T} X_{t,T} \rightsquigarrow \mathcal{N}(0,1)$. In conjunction with (30) and Slutsky's Theorem this implies $\sum_{t=1}^{T} Z_{t,T} \rightsquigarrow \mathcal{N}(0, \lambda'\boldsymbol{B}^\star_m\lambda)$. Hence (26) holds by the Cramér – Wold Theorem.

For condition (iv) we will first show the pointwise convergence of $\hat{\boldsymbol{A}}_{m,T} - \boldsymbol{A}_{m,T}$ to zero on $\mathcal{V}_m$. By Jensen's inequality and part iii of Assumption 8,

$$E\left\|\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})\right\| \le \varphi_{m,t,T}(\boldsymbol{\theta}_m)\mathbb{E}|\upsilon_{m,t,T}| \le C_1\mathbb{E}|\upsilon_{m,t,T}|$$

for some $C_1 \in (0,\infty)$ and where $\sup_{1\le t\le T, T\in\mathbb{N}} E\upsilon^2_{m,t,T} < \infty$. Hence for $\delta \in (0,1]$,

$$E\left\|\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})\right\|^{1+\delta} \le C, \tag{31}$$

for some $C \in (0,\infty)$, implying that $(\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T}))_{1\le t\le T, T\in\mathbb{N}}$ is uniformly integrable. The same is therefore true of $(X_{t,T})_{1\le t\le T, T\in\mathbb{N}}$ for $X_{t,T} := \ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T}) - E\left[\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})\right]$. Equation (31) along with Assumption 3 and Theorems 14.1 and 14.2 in Davidson (1994) imply that $(X_{t,T}/T)_{1\le t\le T, T\in\mathbb{N}}$ forms a L_1 – mixingale array with respect to the constants

[11] That σ_T is finite follows by Assumption 7; that it is (at least eventually) non-zero follows from (30) below.

[12] The constants $d_{t,T}$ in the definition of NED (cf. e.g. Definition 17.2 in Davidson (1994) or Definition 2 in de Jong (1997)) can be set to $c_{t,T}$ to ensure the boundedness condition holds, as the ν_m sequence can be taken to equal zero for each m.

$c_{t,T} = 1/T$.[13] Therefore

$$\hat{\boldsymbol{A}}_{m,T}(\boldsymbol{\theta}_m) - \boldsymbol{A}_{m,T}(\boldsymbol{\theta}_m) = \frac{1}{T}\sum_{t=1}^{T} \ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T}) - E\left[\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})\right] \xrightarrow{P} 0 \tag{32}$$

by Theorem 19.11 in Davidson (1994). By Assumption 8 iv, for each pair of indices (k, j) and $\boldsymbol{\theta}_m, \boldsymbol{\theta}'_m \in \mathcal{V}_m$,

$$\begin{aligned}
|\boldsymbol{A}_{m,T,k,j}(\boldsymbol{\theta}_m) - \boldsymbol{A}_{m,T,k,j}(\boldsymbol{\theta}'_m)| &= \left| \frac{1}{T}\sum_{t=1}^{T} e'_k E\left[\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T}) - \ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}'_m, \epsilon^2_{m,t,T})\right] e_j \right| \\
&\le \frac{1}{T}\sum_{t=1}^{T} E\left\|\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T}) - \ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}'_m, \epsilon^2_{m,t,T})\right\| \\
&\le \frac{1}{T}\sum_{t=1}^{T} E|\psi_{m,t,T}| \|\boldsymbol{\theta}_m - \boldsymbol{\theta}'_m\| \\
&\le C\|\boldsymbol{\theta}_m - \boldsymbol{\theta}'_m\|,
\end{aligned}$$

for some constant $C \in (0, \infty)$. This implies that for each (k, j) pair, $\{\boldsymbol{A}_{m,T,k,j} : T \in \mathbb{N}\}$ is uniformly equicontinuous on $\mathcal{V}_m$. Similarly,

$$|\hat{\boldsymbol{A}}_{m,T,k,j}(\boldsymbol{\theta}_m) - \hat{\boldsymbol{A}}_{m,T,k,j}(\boldsymbol{\theta}'_m)| \le \left[\frac{1}{T}\sum_{t=1}^{T} \psi_{m,t,T}\right] \|\boldsymbol{\theta}_m - \boldsymbol{\theta}'_m\|.$$

with $\frac{1}{T}\sum_{t=1}^{T} \psi_{m,t,T} = O_P(1)$. In combination with the uniform equicontinuity of $\{\boldsymbol{A}_{m,T,k,j} : T \in \mathbb{N}\}$, this verifies Assumption SE-1 in Andrews (1992). Hence, for each pair of indices (k, j), $\{\hat{\boldsymbol{A}}_{m,T,k,j} - \boldsymbol{A}_{m,T,k,j} : T \in \mathbb{N}\}$ is stochastically equicontinuous by Lemma 1 of Andrews (1992). In view of (32) and since $\mathcal{V}_m$ is totally bounded as a subset of a compact metric space, Theorem 1 in Andrews (1992) applied to each pair (k, j) therefore yields that

$$\sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \|\hat{\boldsymbol{A}}_{m,T}(\boldsymbol{\theta}_m) - \boldsymbol{A}_{m,T}(\boldsymbol{\theta}_m)\| \xrightarrow{P} 0. \tag{33}$$

By Assumption 11, $\|\boldsymbol{A}_{m,T}(\boldsymbol{\theta}_m) - \boldsymbol{A}_m(\boldsymbol{\theta}_m)\| \to 0$. Since $\{\boldsymbol{A}_{m,T} : T \in \mathbb{N}\}$ is uniformly equicontinuous on $\mathcal{V}_m$ (as noted above), the convergence is uniform:

$$\sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \|\boldsymbol{A}_{m,T}(\boldsymbol{\theta}_m) - \boldsymbol{A}_m(\boldsymbol{\theta}_m)\| \to 0 \tag{34}$$

Combining this with equation (33) demonstrates that the convergence in condition (iv) holds. That $\boldsymbol{A}_m$ is continuous at $\boldsymbol{\theta}^\star_m$ follows from the uniform limit theorem.

Condition (v) holds by Assumption 11. The claimed result follows by Theorem 3.1 in Newey and McFadden (1994).

[13] This choice of $c_{t,T}$ evidently satisfies conditions (b), (c) of Theorem 19.11 in Davidson (1994).

A.3 Proof of Corollary 1

By Theorem 1, with probability approaching one, $\hat{\boldsymbol{\theta}}_m \in \mathcal{V}_m$. Therefore,

$$\|\hat{\boldsymbol{A}}_{m,T}(\hat{\boldsymbol{\theta}}_m) - \boldsymbol{A}_m^\star\| \leq \sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \|\hat{\boldsymbol{A}}_{m,T}(\boldsymbol{\theta}_m) - \boldsymbol{A}_m(\boldsymbol{\theta}_m)\| + \|\boldsymbol{A}_m(\hat{\boldsymbol{\theta}}_m) - \boldsymbol{A}_m(\boldsymbol{\theta}_m^\star)\|,$$

with probability approaching one. By (33) and (34) in the Proof of Theorem 2,

$$\sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \|\hat{\boldsymbol{A}}_{m,T}(\boldsymbol{\theta}_m) - \boldsymbol{A}_m(\boldsymbol{\theta}_m)\| \xrightarrow{P} 0.$$

The Proof of Theorem 2 also demonstrated that $\boldsymbol{A}_m$ is continuous at $\boldsymbol{\theta}_m^\star$. Combination of this, $\hat{\boldsymbol{\theta}}_m \xrightarrow{P} \boldsymbol{\theta}_m^\star$ and the preceding two displays proves the claim.

A.4 Proof of Proposition 1

We verify Assumptions 1 – 4 of de Jong and Davidson (2000), in order to apply their Theorem 2.2 with

$$X_{t,T}(\boldsymbol{\theta}_m) := T^{-1/2}\dot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon_{m,t,T}^2), \quad X_{t,T} := X_{t,T}(\boldsymbol{\theta}_m^\star).$$

Their Assumption 1 holds by our Assumption 12. For their Assumption 2, let $\mathcal{F}_{m,t,T} := \sigma(\epsilon_{m,1,T}^2, \ldots, \epsilon_{m,t,T}^2)$, $d_{t,T} := c_{t,T} = T^{-1/2}$.[14] Since each $X_{t,T}$ is $\mathcal{F}_{m,t,T}$–measurable (and in L_2), it is trivially L_2 – NED (of any size) on $(\epsilon_{m,t,T}^2)_{1\leq t\leq T, T\in\mathbb{N}}$ and by Assumption 9 this latter array is α-mixing of size $-\rho_m$, with $\rho_m := r_m/(r_m-2)$. Their condition (2.7) is evidently satisfied by this choice of $c_{t,T}$, whilst (2.6) holds by Assumptions 2 and 7 since (cf. (29))

$$\left(\|X_{t,T}\|_{L_{r_m}} + d_{t,T}\right) c_{t,T}^{-1} = \|\dot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon_{m,t,T}^2)\|_{L_{r_m}} + 1.$$

Their Assumption 3 holds by Assumption 13. Part (a) of their Assumption 4 is implied by Theorem 2; part (b) of their Assumption 4 is implied by the (uniform) equicontinuity on $\mathcal{V}_m$ of $\{\boldsymbol{A}_{m,T} : T \in \mathbb{N}\}$, which was noted to hold in the Proof of Theorem 2. For part (c) of their Assumption 4, we note that the rate condition in (ii) is satisfied under Assumption 13, whilst (2.10) holds by Assumption 7 and Assumption 8 iii since for some $C \in (0, \infty)$,

$$\begin{aligned}\sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \left\| \sum_{t=1}^T [\nabla_{\boldsymbol{\theta}_m} X_{t,T}(\boldsymbol{\theta}_m)'] [\nabla_{\boldsymbol{\theta}_m} X_{t,T}(\boldsymbol{\theta}_m)] \right\| &\leq \sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \frac{1}{T} \sum_{t=1}^T \left\| \ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon_{m,t,T}^2) \right\|^2 \\ &\leq \frac{1}{T} \sum_{t=1}^T C^2 \upsilon_{m,t,T}^2 \\ &= O_P(1).\end{aligned}$$

Finally, their (2.8) holds as – also by Assumption 7 and Assumption 8 iii –

$$E \sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \|\nabla_{\boldsymbol{\theta}_m} X_{t,T}(\boldsymbol{\theta}_m)\|^2 = E \sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \frac{1}{T} \left\| \ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon_{m,t,T}^2) \right\|^2 \leq \frac{C^2}{T} E\upsilon_{m,t,T}^2 = O(T^{-1}).$$

Hence the claim follows by Theorem 2.2 in de Jong and Davidson (2000).

[14] Cf. footnote 12.

A.5 Proof of Theorem 3

The proof is essentially analogous to that of Theorem 2; we give it for completeness.

We verify the conditions in Theorem 3.1 of Newey and McFadden (1994).[15] $\hat{\boldsymbol{\theta}}$ minimises $L_T(\boldsymbol{\theta})$ over Θ by definition and is consistent for $\boldsymbol{\theta}^\star$ by Theorem 1. Condition (i) holds by Assumption 6. Condition (ii) holds by Assumption 8, the chain rule and the definition of $l_{m,t,T}$. For condition (iii) we show that

$$S_T(\boldsymbol{\theta}^\star) = T^{-1/2} \sum_{t=1}^{T} \dot{\boldsymbol{l}}_{t,T}(\boldsymbol{\theta}^\star, \boldsymbol{\epsilon}_{t,T}^2) \rightsquigarrow \mathcal{N}(0, \boldsymbol{B}^\star). \tag{35}$$

For this, firstly note that by equation (27) and Assumption 4, $E\left[\dot{\boldsymbol{l}}_{t,T}(\boldsymbol{\theta}^\star, \boldsymbol{\epsilon}_{t,T}^2)\right] = 0$. For notational convenience we drop the arguments of $\dot{\boldsymbol{l}}_{t,T}$ for the remainder of this part of the proof. Let $Z_{t,T} := T^{-1/2}\lambda'\dot{\boldsymbol{l}}_{t,T}$ for $\|\lambda\|_2 = 1$. $\sigma_T := \|\sum_{t=1}^T Z_{t,T}\|_{L_2}$ and $X_{t,T} = Z_{t,T}/\sigma_T$.[16] Let $\mathcal{F}_{t,T} := \sigma(\boldsymbol{\epsilon}_{1,T}^2, \ldots, \boldsymbol{\epsilon}_{t,T}^2)$. We will verify the conditions of Corollary 2 in de Jong (1997). (a) follows as $X_{t,T}$ is a mean-zero random variable with $\|\sum_{t,T} X_{t,T}\|_{L_2} = 1$ and each $X_{t,T}$ is $\mathcal{F}_{t,T}$–measurable. For (b) set $c_{t,T} := \max\{\|Z_{t,T}\|_{L_2}, 1\}/\sigma_T$. By the moment bounds in Assumption 7 and (28), one concludes that

$$\sup_{1\le t\le T, T\in\mathbb{N}} \|X_{t,T}/c_{t,T}\|_{L_r} \le \sup_{1\le t\le T, T\in\mathbb{N}} \sigma_T\|X_{t,T}\|_{L_r} = \sup_{1\le t\le T, T\in\mathbb{N}} \|Z_{t,T}\|_{L_r} < \infty. \tag{36}$$

For (c), since each $X_{t,T}$ is $\mathcal{F}_{t,T}$–measurable (and in L_2), it is trivially L_2 – NED (of any size) on $(\boldsymbol{\epsilon}_{t,T}^2)_{1\le t\le T, T\in\mathbb{N}}$ and by Assumption 14 this latter array is α-mixing of size $-\rho$, with $\rho := r/(r-2)$.[17] Finally, we note that by the moment bounds in Assumption 4 and (28)

$$Tc_{t,T}^2 \le \frac{1}{\sigma_T^2}\max\{\|\dot{\boldsymbol{l}}_{t,T}\|_{L_2}^2, 1\} \lesssim \frac{1}{\sigma_T^2}.$$

By Assumption 15,

$$\sigma_T^2 = \left\|\sum_{t=1}^T Z_{t,T}\right\|_{L_2}^2 = \lambda'\boldsymbol{B}_T\lambda \to \lambda'\boldsymbol{B}^\star\lambda > 0. \tag{37}$$

Combination of the preceding displays permits the conclusion that $c_{t,T}^2 = O(T^{-1})$, establishing the final condition of Corollary 2 of de Jong (1997) with $\beta = \gamma = 0$. Therefore, $\sum_{t=1}^T X_{t,T} \rightsquigarrow \mathcal{N}(0,1)$. In conjunction with (37) and Slutsky's Theorem this implies $\sum_{t=1}^T Z_{t,T} \rightsquigarrow \mathcal{N}(0, \lambda'\boldsymbol{B}^\star\lambda)$. Hence (35) holds by the Cramér – Wold Theorem.

For condition (iv) we note that by (33) and (34) (which can be established in the present setting in exactly the same way as in the proof of Theorem 2) we have

$$\sup_{\boldsymbol{\theta}_m\in\mathcal{V}_m} \|\hat{\boldsymbol{A}}_{m,T}(\boldsymbol{\theta}_m) - \boldsymbol{A}_m(\boldsymbol{\theta}_m)\| \xrightarrow{P} 0,$$

[15] Cf. footnote 10.

[16] That σ_T is finite follows by Assumption 7; that it is (at least eventually) non-zero follows from (37) below.

[17] Cf. footnote 12.

hence (cf. (9))

$$\sup_{\theta\in\mathcal{V}} \|\hat{\boldsymbol{A}}_T(\boldsymbol{\theta}) - \boldsymbol{A}(\boldsymbol{\theta})\| \xrightarrow{P} 0.$$

Condition (v) holds by Assumption 11. The claimed result follows by Theorem 3.1 in Newey and McFadden (1994).

A.6 Proof of Proposition 2

We verify Assumptions 1 – 4 of de Jong and Davidson (2000), in order to apply their Theorem 2.2 with

$$X_{t,T}(\boldsymbol{\theta}) \coloneqq T^{-1/2}\dot{\boldsymbol{l}}_{t,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}^2_{t,T}), \quad X_{t,T} \coloneqq X_{t,T}(\boldsymbol{\theta}^\star).$$

Their Assumption 1 holds by our Assumption 16. For their Assumption 2, let $\mathcal{F}_{t,T} \coloneqq \sigma(\boldsymbol{\epsilon}^2_{1,T}, \dots, \boldsymbol{\epsilon}^2_{t,T})$, $d_{t,T} \coloneqq c_{t,T} = T^{-1/2}$.[18] Since each $X_{t,T}$ is $\mathcal{F}_{t,T}$–measurable (and in L_2), it is trivially L_2 – NED (of any size) on $(\boldsymbol{\epsilon}^2_{t,T})_{1\le t\le T, T\in\mathbb{N}}$ and by Assumption 14 this latter array is α-mixing of size $-\rho$, with $\rho \coloneqq r/(r-2)$ with r as in Assumption 14. Their condition (2.7) is evidently satisfied by this choice of $c_{t,T}$, whilst (2.6) holds by Assumptions 2 and 7 since

$$(\|X_{t,T}\|_{L_r} + d_{t,T})\, c^{-1}_{t,T} = \|\dot{\boldsymbol{l}}_{t,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}^2_{t,T})\|_{L_{rm}} + 1.$$

Their Assumption 3 holds by Assumption 17. Part (a) of their Assumption 4 is implied by Theorem 3; part (b) of their Assumption 4 is implied by the (uniform) equicontinuity on $\mathcal{V}$ of $\{\boldsymbol{A}_T : T \in \mathbb{N}\}$, which follows from the uniform equicontinuity on $\mathcal{V}_m$ of $\{\boldsymbol{A}_{m,T} : T \in \mathbb{N}\}$ for $m = 1, \dots, M$ (as was noted to hold in the Proof of Theorem 2 and can be established in exactly the same way in the present setting). For part (c) of their Assumption 4, we note that the rate condition in (ii) is satisfied under Assumption 17, whilst (2.10) holds by Assumption 7 and Assumption 8 iii since for some $C \in (0, \infty)$,

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\mathcal{V}} \left\| \sum_{t=1}^T [\nabla_{\boldsymbol{\theta}} X_{t,T}(\boldsymbol{\theta})']\,[\nabla_{\boldsymbol{\theta}} X_{t,T}(\boldsymbol{\theta})] \right\| &\le \sup_{\boldsymbol{\theta}\in\mathcal{V}} \frac{1}{T}\sum_{t=1}^T \left\| \ddot{\boldsymbol{l}}_{t,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}^2_{t,T}) \right\|^2 \\
&\le \frac{1}{T}\sum_{t=1}^T C^2 \sum_{m=1}^M \upsilon^2_{m,t,T} \\
&= O_P(1).
\end{aligned}$$

Finally, their (2.8) holds as – also by Assumption 7 and Assumption 8 iii –

$$E\sup_{\boldsymbol{\theta}\in\mathcal{V}} \|\nabla_{\boldsymbol{\theta}} X_{t,T}(\boldsymbol{\theta})\|^2 = E\sup_{\boldsymbol{\theta}\in\mathcal{V}} \frac{1}{T}\left\| \ddot{\boldsymbol{l}}_{t,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}^2_{t,T}) \right\|^2 \le \frac{C^2}{T}\sum_{m=1}^M E\upsilon^2_{m,t,T} = O(T^{-1}).$$

Hence the claim follows by Theorem 2.2 in de Jong and Davidson (2000).

[18] Cf. footnote 12.

B Auxiliary results

Lemma 1. *Suppose that Assumptions 2 and 8(i) hold and that*

$$\sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \sup_{1 \le t \le T, T \in \mathbb{N}} \|\ddot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)\| < \infty. \tag{38}$$

Then, Assumption 8(iii) holds with $\upsilon_{m,t,T} \coloneqq 1 + \epsilon^2_{m,t,T}$.

Proof. By direct calculation the (i,j)-th element of $\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})$ is

$$\begin{aligned}
&[\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})]_{i,j} \\
&\quad = \epsilon^2_{m,t,T} \left(\frac{2[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_i [\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_j}{g_{m,t,T}(\boldsymbol{\theta}_m)^3} - \frac{[\ddot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_{i,j}}{g_{m,t,T}(\boldsymbol{\theta}_m)^2} \right) \\
&\qquad + \left(\frac{[\ddot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_{i,j}}{g_{m,t,T}(\boldsymbol{\theta}_m)} - \frac{[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_i [\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_j}{g_{m,t,T}(\boldsymbol{\theta}_m)^2} \right).
\end{aligned}$$

By Assumption 2, 8(i), and equation (38)

$$0 < c_\rho \le |g_{m,t,T}|^\rho \le C_{0,\rho} < \infty, \qquad |[\dot{g}_{m,t,T}]_k| \le C_1 < \infty, \qquad \text{and} \qquad |[\ddot{g}_{m,t,T}]_{k,l}| \le C_2 < \infty$$

uniformly in t, T and $\boldsymbol{\theta}_m \in \mathcal{V}_m$, $\rho = 1, 2, 3$ and $k, l = 1, \ldots, K$, where K is the dimension of $\boldsymbol{\theta}_m$. Therefore, for $C_0 \coloneqq \max\{2C_1^2 c_3^{-1} + C_2 c_2^{-1}, C_2 c_1^{-1} + C_1^2 c_2^{-1}\} < \infty$,

$$\begin{aligned}
\|\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})\|_\infty &= \max_{1 \le i,j \le K} |[\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})]_{i,j}| \\
&\le \epsilon^2_{m,t,T}(2C_1^2 c_3^{-1} + C_2 c_2^{-1}) + C_2 c_1^{-1} + C_1^2 c_2^{-1} \\
&\le C_0(1 + \epsilon^2_{m,t,T}).
\end{aligned}$$

Since $\|\cdot\|_\infty$ is a norm on the space of $K \times K$ matrices (e.g. Horn and Johnson, 2013, p. 342) and all norms on the same finite dimensional vector space are equivalent, this implies that

$$\|\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m, \epsilon^2_{m,t,T})\| \le C(1 + \epsilon^2_{m,t,T}),$$

for some constant $C > 0$ and hence the claim follows with $\varphi_{m,t,T}(\boldsymbol{\theta}_m) \coloneqq C$. $\square$

Lemma 2. *Under Assumptions 2, 4 and 9, $\boldsymbol{B}_{m,T} = O(1)$ for $\boldsymbol{B}_{m,T}$ as in Assumption 10.*

Proof. By Davydov's inequality (e.g. Davidson, 1994, Corollary 14.3), (28) and Assumption 4 for $1 \le t, s \le T$, any indices l, k, and $r_m = 2 + \delta_m/2$,

$$|\mathrm{Cov}(\dot{\boldsymbol{l}}_{m,t,T,l}, \dot{\boldsymbol{l}}_{m,s,T,k})| \le 6 \|\dot{\boldsymbol{l}}_{m,t,T,l}\|_{L_{r_m}} \|\dot{\boldsymbol{l}}_{m,s,T,k}\|_{L_{r_m}} \alpha_{m,T}(|t-s|)^{1-2/r_m} \lesssim \sup_{T \in \mathbb{N}} \alpha_{m,T}(|t-s|)^{1-2/r_m}.$$

Hence

$$\begin{aligned}e_T := \left|E\left[T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\dot{\boldsymbol{l}}_{m,t,T,l}\dot{\boldsymbol{l}}_{m,s,T,k}\right]\right| &\lesssim T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\sup_{T\in\mathbb{N}}\alpha_{m,T}(|t-s|)^{1-2/r_m}\\ &= \sum_{k=-T}^{T}\left(1-\frac{|k|}{T}\right)\sup_{T\in\mathbb{N}}\alpha_{m,T}(|k|)^{1-2/r_m}\\ &\leq 2\sum_{k=0}^{T}\left(1-\frac{k}{T}\right)\sup_{T\in\mathbb{N}}\alpha_{m,T}(k)^{1-2/r_m}\end{aligned}$$

By Assumption 9, for all sufficiently large $k \geq K \in \mathbb{N}$, $\sup_{T\in\mathbb{N}}\alpha_{m,T}(k)^{1-2/r_m} \lesssim k^{-r_m/(r_m-2)-\varepsilon}$. Hence for some constant $C > 0$

$$e_T \lesssim C + \sum_{k=K}^{T} k^{-r_m/(r_m-2)-\varepsilon} \leq C + \sum_{k=0}^{\infty} k^{-r_m/(r_m-2)-\varepsilon} < \infty$$

since $r_m/(r_m-2)+\varepsilon > 1$. Hence $\boldsymbol{B}_{m,T} = O(1)$. □

Lemma 3. *If Assumptions 2, 4 and parts i, iii of Assumption 8 hold, then $\boldsymbol{A}_{m,T}(\boldsymbol{\theta}_m) = O(1)$ for each $\boldsymbol{\theta}_m \in \mathcal{V}_m$.*

Proof. By Jensen's inequality and part iii of Assumption 8,

$$\left\|E\left[\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m,\epsilon^2_{m,t,T})\right]\right\| \leq E\left\|\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m,\epsilon^2_{m,t,T})\right\| \leq \varphi_{m,t,T}(\boldsymbol{\theta}_m)E|\upsilon_{m,t,T}|,$$

with $E|\upsilon_{m,t,T}|^2 \leq C_1 < \infty$ and $\varphi_{m,t,T}(\boldsymbol{\theta}_m) \leq C_2 < \infty$. Putting $C = \sqrt{C_1}C_2$, one then has that

$$\left\|E\left[\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m,\epsilon^2_{m,t,T})\right]\right\| \leq C.$$

Hence $T^{-1}\sum_{t=1}^{T}\left\|E\left[\ddot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}_m,\epsilon^2_{m,t,T})\right]\right\|$ is bounded above by C, which implies the result. □

Lemma 4. *Under Assumptions 2, 4 and 14, $B_T = O(1)$ for $\boldsymbol{B}_T$ as in Assumption 15.*

Proof. The proof follows analogously to that of Lemma 2 on replacing r_m with r, δ_m with $\min\{\delta_1,\ldots,\delta_M\}$ and $\alpha_{m,T}$ with α_T. □

Lemma 5. *In the setting of Theorem 6, for any $\overline{\boldsymbol{\theta}}_T \xrightarrow{P} \boldsymbol{\theta}^\star$,*

$$\frac{\partial\gamma_m(\overline{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}'} \xrightarrow{P} \overline{\boldsymbol{F}}^\star_m = \left(\overline{\boldsymbol{F}}_{m,0}(\boldsymbol{\theta}^\star),\overline{\boldsymbol{F}}_{m,1}(\boldsymbol{\theta}^\star),\overline{\boldsymbol{F}}_{m,2}(\boldsymbol{\theta}^\star)\right)'.$$

Proof. Dropping the m in the notation, let $\boldsymbol{F}_T(\boldsymbol{\theta}) := \frac{1}{T}\sum_{t=1}^{T} b_{t,T}(\boldsymbol{\theta})$ on $\mathcal{V}$. It is clearly sufficient to show that $\boldsymbol{F}_T$ converges to $\boldsymbol{F}^\star$ uniformly in probability. Initially, we show pointwise convergence in probability of $\boldsymbol{F}_T - \overline{\boldsymbol{F}}_T$ to zero on $\mathcal{V}$. Let $X_{t,T} := \boldsymbol{b}_{t,T}(\boldsymbol{\theta}) - E[\boldsymbol{b}_{t,T}(\boldsymbol{\theta})]$ and $\mathcal{F}_{t,T} := \sigma(X_{t,T}, X_{t-1,T},\ldots)$. Assumptions 2,20 allow the application of Theorems 14.1 & 14.2 in Davidson (1994) to permit the conclusion that $(X_{t,T}/T,\mathcal{F}_{t,T})_{1\leq t\leq T, T\in\mathbb{N}}$ is an L_1 – mixingale arrray with respect to the constants $c_{t,T} = 1/T$. Therefore Theorem 19.11 in

Davidson (1994) yields that

$$\boldsymbol{F}_T(\boldsymbol{\theta}) - \overline{\boldsymbol{F}}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{b}_{t,T}(\boldsymbol{\theta}) - E[\boldsymbol{b}_{t,T}(\boldsymbol{\theta})] \xrightarrow{P} 0. \tag{39}$$

By Assumption 21, one has

$$\begin{aligned} \|\overline{\boldsymbol{F}}_T(\boldsymbol{\theta}) - \overline{\boldsymbol{F}}_T(\boldsymbol{\theta}')\| &\leq \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\,\|b_{t,T}(\boldsymbol{\theta}) - b_{t,T}(\boldsymbol{\theta}')\| \\ &\leq \frac{1}{T}\sum_{t=1}^{T} E\tilde{\psi}_{t,T}\|\boldsymbol{\theta} - \boldsymbol{\theta}'\| \\ &\leq C\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|, \end{aligned}$$

for some constant $C \in (0, \infty)$. Hence each $\overline{\boldsymbol{F}}_T$ is Lipschitz on $\mathcal{V}$ with a common Lipschitz constant and so uniformly equicontinuous on $\mathcal{V}$. Similarly,

$$\|\boldsymbol{F}_T(\boldsymbol{\theta}) - \boldsymbol{F}_T(\boldsymbol{\theta}')\| \leq \left[\frac{1}{T}\sum_{t=1}^{T} \tilde{\psi}_{t,T}\right] \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|,$$

with $\frac{1}{T}\sum_{t=1}^{T} \tilde{\psi}_{t,T} = O_P(1)$. In combination with the uniform equiconitinuity of $\overline{\boldsymbol{F}}_T$, this verifies Assumption SE-1 in Andrews (1992) (elementwise). Hence Lemma 1 of Andrews (1992) yields that for each co-ordinate k, $\{\boldsymbol{F}_{T,k} - \overline{\boldsymbol{F}}_{T,k} : T \in \mathbb{N}\}$ is stochastically equicontinuous on $\mathcal{V}$. Given (39) and that $\mathcal{V}$ is totally bounded as a subset of a compact metric space, Theorem 1 in Andrews (1992) applied to each co-ordinate k yields that

$$\sup_{\boldsymbol{\theta} \in \mathcal{V}} \|\boldsymbol{F}_T(\boldsymbol{\theta}) - \overline{\boldsymbol{F}}_T(\boldsymbol{\theta})\| \xrightarrow{P} 0.$$

By Assumption 21, $\overline{\boldsymbol{F}}_T(\boldsymbol{\theta}) - \overline{\boldsymbol{F}}^{\star} \to 0$ as $T \to \infty$. Since, as noted above, $\{\overline{\boldsymbol{F}}_T : T \in \mathbb{N}\}$ is uniformly equicontinuous on the totally bounded set $\mathcal{V}$, this convergence is uniform. Combining this with the preceding display yields

$$\sup_{\boldsymbol{\theta} \in \mathcal{V}} \|\boldsymbol{F}_T(\boldsymbol{\theta}) - \overline{\boldsymbol{F}}^{\star}(\boldsymbol{\theta})\| \xrightarrow{P} 0.$$

Since $\bar{\theta}_T \xrightarrow{P} \boldsymbol{\theta}$, and $\mathcal{V}$ is a neighbourhood of $\boldsymbol{\theta}$, the claim follows. □

Lemma 6. *Let $\mathcal{J} \coloneqq \{x \in \mathbb{R}^s : x_1 > x_2 > \cdots > x_s > 0\}$ and $X : \mathbb{R}^s \to \mathbb{R}^{s \times s}$ be the function:*

$$X(x) \coloneqq \begin{bmatrix} x_1 & x_2 & \dots & x_{s-1} & x_s \\ x_2 & x_2 & \dots & x_{s-1} & x_s \\ x_3 & x_3 & \dots & x_{s-1} & x_s \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{s-1} & x_{s-1} & \dots & x_{s-1} & x_s \\ x_s & x_s & \dots & x_s & x_s \end{bmatrix}.$$

Then the image of $\mathcal{J}$ under X is a subset of the convex cone of positive definite $s \times s$ matrices.

Proof. It suffices to show that each $X(x)$ is positive definite. Any such $X(x)$ is clearly symmetric. If $s = 1$ then $x \in \mathcal{J}$ iff $x_1 > 0$ and $X(x) = [x_1] \succ 0$. Now suppose that the conclusion holds for matrices of size $k-1 \times k-1$. We will show that the image of $\mathcal{J}_k := \{x \in \mathbb{R}^k : x_1 > \cdots > x_k > 0\}$ under X_k is a subset of convex cone of positive definite $k \times k$ matrices where

$$X_k(x) := \begin{bmatrix} x_1 & x_2 & \dots & x_{k-1} & x_k \\ x_2 & x_2 & \dots & x_{k-1} & x_k \\ x_3 & x_3 & \dots & x_{k-1} & x_k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{k-1} & x_{k-1} & \dots & x_{k-1} & x_k \\ x_k & x_k & \dots & x_k & x_k \end{bmatrix}.$$

In particular, by the induction hypothesis

$$A = \begin{bmatrix} x_2 & \dots & x_{k-1} & x_k \\ x_3 & \dots & x_{k-1} & x_k \\ \vdots & \ddots & \vdots & \vdots \\ x_{k-1} & \dots & x_{k-1} & x_k \\ x_k & \dots & x_k & x_k \end{bmatrix},$$

(the matrix obtained by removing the first column and row of $X_k(x)$) is positive definite since $x_2 > x_3 > \cdots > x_k > 0$. Define $a := [x_2, \dots, x_k]'$. By Proposition 8.2.4 in Bernstein (2009) it suffices to show that $x_1 - a'A^{-1}a > 0$. As $x_1 > x_2$ it is enough to note that $a'A^{-1}a = x_2$ since

$$A^{-1}a = d \qquad \iff \qquad a = Ad,$$

and choosing $d = e_1$ satisfies the right hand side. □

C Proofs of the results in Section 3

C.1 A useful lemma for verification of Assumption 8

Verifying Assumption 8 requires the choice of the functions $\varphi_{m,t,T}$ and $\psi_{m,t,T}$. For the class of functions that is the main target of our theory, third order differentiability with respect to $\boldsymbol{\theta}_m$ along with certain domination conditions simplifies the choices of $\varphi_{m,t,T}$ and $\psi_{m,t,T}$. This is recorded formally in Lemma 7 below. Next, the Lemma is used to show that Assumption 8 holds for the class of functions that is the main target of our theory in Sections C.2 – C.4.

Lemma 7. *Suppose that Assumptions 2 and 4 hold, and that each $g_{m,t,T}$ is three-times differentiable on a neighbourhood $\mathcal{V}_m$ of $\boldsymbol{\theta}_m^\star$. Let $\dddot{g}_{m,t,T,(j,l,k)} := \frac{\partial[\ddot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_{j,l}}{\partial[\boldsymbol{\theta}_m]_k}$ denote the (j,l,k) entry of the array of third order derivatives and suppose also that for some functions $\overline{g_m}(\boldsymbol{\theta}_m)$, $\overline{\overline{g_m}}(\boldsymbol{\theta}_m)$,*

$$\|\ddot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)\| \le \overline{g_m}(\boldsymbol{\theta}_m) \le \sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \overline{g_m}(\boldsymbol{\theta}_m) < \infty.$$

and

$$|\dddot{g}_{m,t,T,(j,l,k)}(\boldsymbol{\theta}_m)| \leq \overline{\overline{g_m}}(\boldsymbol{\theta}_m) \leq \sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \overline{\overline{g_m}}(\boldsymbol{\theta}_m) < \infty.$$

Then Assumption 8 holds.

Note that the boundedness of the dominating functions $\overline{g_m}$ and $\overline{\overline{g_m}}$ is automatic if $\mathcal{V}_m$ is taken to be compact and $\overline{g_m}$ and $\overline{\overline{g_m}}$ are continuous.

Proof. That Assumption 8 part i holds follows immediately from the assumption of three – times differentiability. Given this, part ii is simply a definition and requires no proof.

For part iii it suffices to note that equation (38) holds as

$$\sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \|\ddot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)\| \leq \sup_{\boldsymbol{\theta}_m \in \mathcal{V}_m} \overline{g_m}(\boldsymbol{\theta}_m) < \infty,$$

and hence we may apply Lemma 1.

Finally, for part iv, note that the derivative of the (i, j) element of $\ddot{\boldsymbol{l}}_{m,t,T}$ with respect to $[\boldsymbol{\theta}_m]_k$ has the form

$$\begin{aligned}
&\epsilon^2_{m,t,T}\left(\frac{2[\ddot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_{i,k}[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_j}{g_{m,t,T}(\boldsymbol{\theta}_m)^3} + \frac{2[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_i[\ddot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_{j,k}}{g_{m,t,T}(\boldsymbol{\theta}_m)^3} + \frac{2[\ddot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_{i,j}[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_k}{g_{m,t,T}(\boldsymbol{\theta}_m)^3}\right) \\
&\quad - \epsilon^2_{m,t,T}\left(\frac{6[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_i[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_j[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_k}{g_{m,t,T}(\boldsymbol{\theta}_m)^4} + \frac{\dddot{g}_{m,t,T,(i,j,k)}(\boldsymbol{\theta}_m)}{g_{m,t,T}(\boldsymbol{\theta}_m)^2}\right) \\
&\quad + \left(\frac{-[\ddot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_{i,j}[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_k}{g_{m,t,T}(\boldsymbol{\theta}_m)^2} + \frac{\dddot{g}_{m,t,T,(i,j,k)}(\boldsymbol{\theta}_m)}{g_{m,t,T}(\boldsymbol{\theta}_m)} - \frac{[\ddot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_{i,k}[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_j}{g_{m,t,T}(\boldsymbol{\theta}_m)^2}\right) \\
&\quad + \left(\frac{-[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_i[\ddot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_{j,k}}{g_{m,t,T}(\boldsymbol{\theta}_m)^2} + \frac{2[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_i[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_j[\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)]_k}{g_{m,t,T}(\boldsymbol{\theta}_m)^3}\right)
\end{aligned}$$

which, given the boundedness of each 0th, 1st, 2nd and 3rd derivative uniformly over $\boldsymbol{\theta}_m$ and t, T already established, this is bounded by $C(\epsilon^2_{m,t,T} + 1)$ for some constant $C \in (0, \infty)$.[19] Moreover, this bound is uniform over all i, j, k indices. Hence by the mean value theorem, the required inequality holds with $\psi_{m,t,T} := C(\epsilon^2_{m,t,T} + 1)$. That $\sup_{1 \leq t \leq T, T \in \mathbb{N}} E|\psi_{m,t,T}| < \infty$ then follows directly from Assumptions 2 and 4. □

C.2 Proof of Proposition 3

To simplify notation we omit the subscript m. Let $G(r, \gamma, c) := 1 + \exp(-\gamma(r - c))$. The gradient and Hessian of this function are

$$\dot{\boldsymbol{G}}(r, \gamma, c) = ((c - r), \gamma)' \exp(-\gamma(r - c)),$$

$$\ddot{\boldsymbol{G}}r, \gamma, c) = \begin{bmatrix} (c - r)^2 & \gamma(c - r) + 1 \\ \gamma(c - r) + 1 & \gamma^2 \end{bmatrix} \exp(-\gamma(r - c)).$$

[19]Specifically $g_{m,t,T}(\boldsymbol{\theta}_m)$ is bounded above and below by Assumption 2; the first derivatives $\dot{\boldsymbol{g}}_{m,t,T}$ have their norm bounded above by Assumption 2. The boundedness of $\|\ddot{\boldsymbol{g}}_{m,t,T}\|$ holds by the $\overline{g_m}$ – domination hypothesis of this Lemma, whilst the boundedness of each $|\dddot{\boldsymbol{g}}_{m,t,T,(i,j,k)}(\boldsymbol{\theta}_m,)\|$ holds by the $\overline{\overline{g_m}}$ – domination hypothesis of this Lemma.

Moreover, define $\dddot{\boldsymbol{G}}^{(1)}(r,\gamma,c)$ as

$$\frac{\partial\ddot{\boldsymbol{G}}(r,\gamma,c)}{\partial\gamma}=\begin{bmatrix}(c-r)^3 & 2(c-r)+\gamma(c-r)^2\\ 2(c-r)+\gamma(c-r)^2+1 & 2\gamma+(c-r)\gamma^2\end{bmatrix}\exp(-\gamma(r-c)),$$

and $\dddot{\boldsymbol{G}}^{(2)}(r,\gamma,c)$ as

$$\frac{\partial\ddot{\boldsymbol{G}}(r,\gamma,c)}{\partial c}=\begin{bmatrix}2(c-r)+\gamma(c-r)^2 & 2\gamma+(c-r)\gamma^2\\ 2\gamma+(c-r)\gamma^2 & \gamma^3\end{bmatrix}\exp(-\gamma(r-c)).$$

Then with $r=t/T$, $\dot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta}):=\dot{\boldsymbol{g}}(\boldsymbol{\theta},r):=(\dot{\boldsymbol{g}}_{t,T}^{(\delta)}(\boldsymbol{\theta})',\dot{\boldsymbol{g}}_{t,T}^{(\gamma)}(\boldsymbol{\theta})',\dot{\boldsymbol{g}}_{t,T}^{(c)}(\boldsymbol{\theta})')'$ where,

$$\begin{aligned}\dot{\boldsymbol{g}}_{t,T}^{(\delta)}(\boldsymbol{\theta})&=\left(1,G(r,\gamma_1,c_1)^{-1},\ldots,G(r,\gamma_s,c_s)^{-1}\right)\\ \dot{\boldsymbol{g}}_{t,T}^{(\gamma)}(\boldsymbol{\theta})&=-\left(\delta_1 G(r,\gamma_1,c_1)^{-2}\dot{\boldsymbol{G}}_1(r,\gamma_1,c_1),\ldots,\delta_s G(r,\gamma_s,c_s)^{-2}\dot{\boldsymbol{G}}_1(r,\gamma_s,c_s)\right)\\ \dot{\boldsymbol{g}}_{t,T}^{(c)}(\boldsymbol{\theta})&=-\left(\delta_1 G(r,\gamma_1,c_1)^{-2}\dot{\boldsymbol{G}}_2(r,\gamma_1,c_1),\ldots,\delta_s G(r,\gamma_s,c_s)^{-2}\dot{\boldsymbol{G}}_2(r,\gamma_s,c_s)\right).\end{aligned}$$

The Hessian $\ddot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta})$ is

$$\ddot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta})=\begin{bmatrix}\mathbf{0}'_{1+s,1+s} & \mathbf{0}'_{1+s,s} & \mathbf{0}'_{1+s,s}\\ \mathbf{0}_{s,1+s} & \ddot{\boldsymbol{g}}_{t,T}^{(\gamma\gamma)}(\boldsymbol{\theta}) & \ddot{\boldsymbol{g}}_{t,T}^{(\gamma c)}(\boldsymbol{\theta})\\ \mathbf{0}_{s,1+s} & \ddot{\boldsymbol{g}}_{t,T}^{(\gamma c)}(\boldsymbol{\theta}) & \ddot{\boldsymbol{g}}_{t,T}^{(cc)}(\boldsymbol{\theta})\end{bmatrix},$$

where $\ddot{\boldsymbol{g}}_{t,T}^{(\gamma\gamma)}(\boldsymbol{\theta})$, $\ddot{\boldsymbol{g}}_{t,T}^{(\gamma c)}(\boldsymbol{\theta})$ and $\ddot{\boldsymbol{g}}_{t,T}^{(cc)}(\boldsymbol{\theta})$ are diagonal with (respectively) entries

$$\begin{aligned}[\ddot{\boldsymbol{g}}_{t,T}^{(\gamma\gamma)}(\boldsymbol{\theta})]_{ll}&=\delta_l\left[2G(r,\gamma_l,c_l)^{-3}\dot{\boldsymbol{G}}_1(r,\gamma_l,c_l)^2-G(r,\gamma_l,c_l)^{-2}\ddot{\boldsymbol{G}}_{11}(r,\gamma_l,c_l)\right]\\ [\ddot{\boldsymbol{g}}_{t,T}^{(\gamma c)}(\boldsymbol{\theta})]_{ll}&=\delta_l\left[2G(r,\gamma_l,c_l)^{-3}\dot{\boldsymbol{G}}_1(r,\gamma_l,c_l)\dot{\boldsymbol{G}}_2(r,\gamma_l,c_l)-G(r,\gamma_l,c_l)^{-2}\ddot{\boldsymbol{G}}_{12}(r,\gamma_l,c_l)\right]\\ [\ddot{\boldsymbol{g}}_{t,T}^{(cc)}(\boldsymbol{\theta})]_{ll}&=\delta_l\left[2G(r,\gamma_l,c_l)^{-3}\dot{\boldsymbol{G}}_2(r,\gamma_l,c_l)^2-G(r,\gamma_l,c_l)^{-2}\ddot{\boldsymbol{G}}_{22}(r,\gamma_l,c_l)\right].\end{aligned}$$

Each third order derivative is thus either zero or has the form

$$\begin{aligned}\delta_l\Big[&-6G(r,\gamma_l,c_l)^{-4}\dot{\boldsymbol{G}}_i(r,\gamma_l,c_l)\dot{\boldsymbol{G}}_j(r,\gamma_l,c_l)\dot{\boldsymbol{G}}_k(r,\gamma_l,c_l)\\ &+2G(r,\gamma_l,c_l)^{-3}\dot{\boldsymbol{G}}_i(r,\gamma_l,c_l)\ddot{\boldsymbol{G}}_{jk}(r,\gamma_l,c_l)+2G(r,\gamma_l,c_l)^{-3}\ddot{\boldsymbol{G}}_{ik}(r,\gamma_l,c_l)\dot{\boldsymbol{G}}_j(r,\gamma_l,c_l)\\ &+2G(r,\gamma_l,c_l)^{-3}\ddot{\boldsymbol{G}}_{ij}(r,\gamma_l,c_l)\dot{\boldsymbol{G}}_k(r,\gamma_l,c_l)-G(r,\gamma_l,c_l)^{-2}\dddot{\boldsymbol{G}}_{ij}^{(k)}(r,\gamma_l,c_l)\Big],\end{aligned}$$

for $(i,j,k)\in\{1,2\}^3$. It follows straightforwardly that Assumption 2 holds and that on a suitable compact subset of $\boldsymbol{\Theta}^*$, both $\|\ddot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta})\|$ and the $|\dddot{g}_{t,T,(j,l,k)}(\boldsymbol{\theta})|$ in Lemma 7 are uniformly bounded by sufficiently large constants that serve as dominating functions. In consequence, (13) satisfies Assumption 8.

To establish the existence of the limit $L(\boldsymbol{\theta})$, decompose $E(l_{t,T}(\boldsymbol{\theta},\epsilon_{t,T}^2))=\ln g_{t,T}(\boldsymbol{\theta})+g_{t,T}(\boldsymbol{\theta}^\star)/g_{t,T}(\boldsymbol{\theta})$. Note by the form in (11),

$$g_{t,T}(\boldsymbol{\theta}):=g(\boldsymbol{\theta},r):=\delta_0+\sum_{l=1}^{s}\delta_l G(r,\gamma_l,c_l)^{-1},\qquad r=t/T.$$

As $|g(\boldsymbol{\theta}, r)|$ is bounded away from zero and infinity the same is true of $|\ln g(\boldsymbol{\theta}, r) + g(\boldsymbol{\theta}^{\star}, r)/g(\boldsymbol{\theta}, r)|$. As $\ln g(\boldsymbol{\theta}, r) + g(\boldsymbol{\theta}^{\star}, r)/g_{t,T}(\boldsymbol{\theta}, r)$ is also continuous in r it is Riemann integrable and hence

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \left(\ln g_{t,T}(\boldsymbol{\theta}) + g_{t,T}(\boldsymbol{\theta}^{\star})/g_{t,T}(\boldsymbol{\theta})\right) \\
&= \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \left(\ln g(\boldsymbol{\theta}, t/T) + g(\boldsymbol{\theta}^{\star}, t/T)/g(\boldsymbol{\theta}, t/T)\right) \\
&= \int_0^1 \left(\ln g(\boldsymbol{\theta}, r) + g(\boldsymbol{\theta}^{\star}, r)/g(\boldsymbol{\theta}, r)\right) \mathrm{d}r.
\end{aligned}
$$

Given the continuity and boundedness of the partial derivatives noted above, we may calculate $\dot{\boldsymbol{L}}(\boldsymbol{\theta})$ by differentiation inside the integral (e.g. Folland, 1999, Theorem 2.27)

$$
\dot{\boldsymbol{L}}(\boldsymbol{\theta}) = \int_0^1 \dot{\boldsymbol{g}}(\boldsymbol{\theta}, r)\frac{1}{g(\boldsymbol{\theta}, r)} - \dot{\boldsymbol{g}}(\boldsymbol{\theta}, r)\frac{g(\boldsymbol{\theta}^{\star}, r)}{g(\boldsymbol{\theta}, r)^2}\, \mathrm{d}r.
$$

It is clear that $\dot{\boldsymbol{L}}(\boldsymbol{\theta}^{\star}) = \mathbf{0}$. Similarly, given continuity and boundedness of the second partial derivatives noted above, we may calculate $\ddot{\boldsymbol{L}}(\boldsymbol{\theta})$ by differentiating under the integral, hence $\ddot{\boldsymbol{L}}(\boldsymbol{\theta})$ exists.

C.3 Proof of Proposition 4

To simplify notation we omit the subscript m. The first, second and third partial derivatives of $g_{t,T}(\boldsymbol{\theta})$ are

$$
\begin{aligned}
\dot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta}) &= g_{t,T}(\boldsymbol{\theta})\cdot\big(1, I(t/T \geq c_1), \ldots, I(t/T \geq c_s)\big)', \\
\ddot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta}) &= g_{t,T}(\boldsymbol{\theta})\cdot \\
&\begin{pmatrix}
1 & I(t/T \geq c_1) & \cdots & I(t/T \geq c_s) \\
I(t/T \geq c_1) & I(t/T \geq c_1)\cdot I(t/T \geq c_1) & \cdots & I(t/T \geq c_1)\cdot I(t/T \geq c_s) \\
\vdots & \vdots & \ddots & \vdots \\
I(t/T \geq c_s) & I(t/T \geq c_s)\cdot I(t/T \geq c_1) & \cdots & I(t/T \geq c_s)\cdot I(t/T \geq c_s)
\end{pmatrix}, \\
\frac{\partial^3 g_{t,T}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\partial\theta_k} &= \ddot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta})\cdot\frac{\partial g_{t,T}(\boldsymbol{\theta})}{\partial\theta_k} \qquad \text{for} \qquad k = 1, \ldots, s+1.
\end{aligned}
$$

It follows straightforwardly that $\dot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta})$ satisfies Assumption 2 and that on a suitable compact subset of $\boldsymbol{\Theta}^{*}$, both $\|\ddot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta})\|$ and the $|\dddot{g}_{t,T,(j,l,k)}(\boldsymbol{\theta})|$ in Lemma 7 are uniformly bounded by sufficiently large constants that serve as dominating functions. In consequence, (13) satisfies Assumption 8.

The limit in Assumption 5 can be written as $L(\boldsymbol{\theta}) = L_1(\boldsymbol{\theta}) + L_2(\boldsymbol{\theta})$, where

$$
L_1(\boldsymbol{\theta}) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \ln g_{t,T}(\boldsymbol{\theta}) \qquad \text{and} \qquad L_2(\boldsymbol{\theta}) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\big(\epsilon_t^2/g_{t,T}(\boldsymbol{\theta})\big).
$$

The first term is

$$L_1(\boldsymbol{\theta}) = \delta_0 + \int_{c_1}^{1} \delta_1 dx + \cdots + \int_{c_s}^{1} \delta_s dx \;=\; \delta_0 + \sum_{l=1}^{s} \delta_l (1-c_l) \;=\; \sum_{l=0}^{s} \delta_l (1-c_l),$$

where $c_0 = 0$. The second term is, using that $E(\epsilon_{t,T}^2) = g_{t,T}(\boldsymbol{\theta}^\star)$ for all t, T,

$$L_2(\boldsymbol{\theta}) \;=\; \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{E(\epsilon_t^2)}{g_{t,T}(\boldsymbol{\theta})} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{g_{t,T}(\boldsymbol{\theta}^\star)}{g_{t,T}(\boldsymbol{\theta})} = \int_0^1 \frac{g(\boldsymbol{\theta}^\star, x)}{g(\boldsymbol{\theta}, x)} dx.$$

Let $g(\boldsymbol{\theta}^\star, \boldsymbol{\theta}, x) = g(\boldsymbol{\theta}^\star, x)/g(\boldsymbol{\theta}, x)$, so that

$$g(\boldsymbol{\theta}^\star, \boldsymbol{\theta}, x) \;=\; \exp\big(\delta_0^\star - \delta_0\big) \cdot \exp\big((\delta_1^\star - \delta_1) I(x \geq c_1)\big) \;\cdots\; \exp\big((\delta_s^\star - \delta_s) I(x \geq c_s)\big)$$

and

$$\int_0^1 g(\boldsymbol{\theta}^\star, \boldsymbol{\theta}, x) dx \;=\; \int_0^{c_1} g(\boldsymbol{\theta}^\star, \boldsymbol{\theta}, x) dx + \int_{c_1}^{c_2} g(\boldsymbol{\theta}^\star, \boldsymbol{\theta}, x) dx + \cdots + \int_{c_s}^{1} g(\boldsymbol{\theta}^\star, \boldsymbol{\theta}, x) dx,$$

where

$$\begin{aligned}
\int_0^{c_1} g(\boldsymbol{\theta}^\star, \boldsymbol{\theta}, x) dx \;&=\; \exp\big(\delta_0^\star - \delta_0\big) \cdot (c_1 - c_0) \\
\int_{c_1}^{c_2} g(\boldsymbol{\theta}^\star, \boldsymbol{\theta}, x) dx \;&=\; \exp\big(\delta_0^\star - \delta_0\big) \cdot \exp\big(\delta_1^\star - \delta_1\big) \cdot (c_2 - c_1) \\
&\;\vdots \\
\int_{c_s}^{1} g(\boldsymbol{\theta}^\star, \boldsymbol{\theta}, x) dx \;&=\; \exp\big(\delta_0^\star - \delta_0\big) \cdot \exp\big(\delta_1^\star - \delta_1\big) \cdots \exp\big(\delta_s^\star - \delta_s\big) \cdot (1 - c_s).
\end{aligned}$$

The sum of these terms is

$$\int_0^1 g(\boldsymbol{\theta}^\star, \boldsymbol{\theta}, x) dx = \sum_{l=0}^{s} (c_{l+1} - c_l) \prod_{k=0}^{l} \exp(\delta_k^\star - \delta_k) \qquad \text{with} \qquad c_0 = 0 \;\text{ and }\; c_{s+1} = 1,$$

In conclusion,

$$L(\boldsymbol{\theta}) = \sum_{l=0}^{s} \delta_l (1 - c_l) + \sum_{l=0}^{s} (c_{l+1} - c_l) \prod_{k=0}^{l} \exp(\delta_k^\star - \delta_k),$$

and from Assumption 2(i) it follows that $\big|L(\boldsymbol{\theta})\big| < \infty$ on $\boldsymbol{\Theta}^*$.

For $i = 0, 1, 2, \ldots, s$

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \delta_i} = (1 - c_i) - \sum_{l=i}^{s} (c_{l+1} - c_l) \prod_{k=0}^{l} \exp(\delta_k^\star - \delta_k) \qquad \text{with} \qquad c_0 = 0 \;\text{ and }\; c_{s+1} = 1.$$

Hence, at $\boldsymbol{\theta}_m = \boldsymbol{\theta}_m^\star$ the derivatives are

$$
\begin{aligned}
\frac{\partial L(\boldsymbol{\theta}^\star)}{\partial \delta_0} &= (1-c_0) - \big((c_1-c_0)+(c_2-c_1)+\cdots+(1-c_s)\big) \;=\; (1-c_0)-1 \;=\; 0\\
\frac{\partial L(\boldsymbol{\theta}^\star)}{\partial \delta_1} &= (1-c_1) - \big((c_2-c_1)+\cdots+(1-c_s)\big) \;=\; (1-c_1)-(1-c_1) \;=\; 0\\
&\;\vdots\\
\frac{\partial L(\boldsymbol{\theta}^\star)}{\partial \delta_{s-1}} &= (1-c_{s-1}) - \big((c_s-c_{s-1})+(1-c_s)\big) \;=\; (1-c_{s-1})-(1-c_{s-1}) \;=\; 0\\
\frac{\partial L(\boldsymbol{\theta}^\star)}{\partial \delta_s} &= (1-c_s)-(1-c_s) \;=\; 0.
\end{aligned}
$$

Accordingly, $\boldsymbol{\theta}^\star$ is a stationary point of $L(\boldsymbol{\theta})$.

By differentiating $\partial L(\boldsymbol{\theta})/\partial \delta_i$ with respect to δ_j, it follows that the generic $(i+1, j+1)$ entry of $\ddot{\boldsymbol{L}}(\boldsymbol{\theta})$, with $i, j \in \{0, 1, 2, \ldots, s\}$, is

$$
\ddot{L}_{ij}(\boldsymbol{\theta}) = \sum_{l=i}^{s} I(l \geq j) \cdot a_l(\boldsymbol{\theta}) \;=\; \sum_{l=\max\{i,j\}}^{s} a_l(\boldsymbol{\theta}),
$$

where $a_l(\boldsymbol{\theta}) \coloneqq (c_{l+1}-c_l)\prod_{k=0}^{l}\exp(\delta_k^\star-\delta_k) > 0$. Fix any $\boldsymbol{\theta}$ and set $a_l \coloneqq a_l(\boldsymbol{\theta})$. Then

$$
\ddot{\boldsymbol{L}}(\boldsymbol{\theta}) = \begin{bmatrix}
\sum_{l=0}^{s} a_l & \sum_{l=1}^{s} a_l & \cdots & \sum_{l=s-1}^{s} a_l & \sum_{l=s}^{s} a_l\\
\sum_{l=1}^{s} a_l & \sum_{l=1}^{s} a_l & \cdots & \sum_{l=s-1}^{s} a_l & \sum_{l=s}^{s} a_l\\
\sum_{l=2}^{s} a_l & \sum_{l=2}^{s} a_l & \cdots & \sum_{l=s-1}^{s} a_l & \sum_{l=s}^{s} a_l\\
\vdots & \vdots & \ddots & \vdots & \vdots\\
\sum_{l=s-1}^{s} a_l & \sum_{l=s-1}^{s} a_l & \cdots & \sum_{l=s-1}^{s} a_l & \sum_{l=s}^{s} a_l\\
\sum_{l=s}^{s} a_l & \sum_{l=s}^{s} a_l & \cdots & \sum_{l=s}^{s} a_l & \sum_{l=s}^{s} a_l
\end{bmatrix}.
$$

As $a_l > 0$, we have $\sum_{l=0}^{s} a_l > \sum_{l=1}^{s} a_l > \cdots > \sum_{l=s}^{s} a_l = a_s > 0$. Let $x_i \coloneqq \sum_{l=i}^{s} a_l$. Then

$$
\ddot{\boldsymbol{L}}(\boldsymbol{\theta}) = \begin{bmatrix}
x_0 & x_1 & \cdots & x_{s-1} & x_s\\
x_1 & x_1 & \cdots & x_{s-1} & x_s\\
x_2 & x_2 & \cdots & x_{s-1} & x_s\\
\vdots & \vdots & \ddots & \vdots & \vdots\\
x_{s-1} & x_{s-1} & \cdots & x_{s-1} & x_s\\
x_s & x_s & \cdots & x_s & x_s
\end{bmatrix}.
$$

This is positive definite for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}^*$ by Lemma 6 and hence $L(\boldsymbol{\theta})$ is strictly convex since the set $\boldsymbol{\Theta}^*$ is open and convex by Assumption 2. Next, part 1 of Theorem 7.13 in Sundaram (1996) implies $\boldsymbol{\theta}^\star$ is a global minimum, and Theorem 7.14 (Sundaram, 1996) implies the set of minimisers of L over $\boldsymbol{\Theta}^*$ is either empty or a singleton. In conclusion, $\boldsymbol{\theta}^\star$ is the unique minimiser of $L(\boldsymbol{\theta})$ on $\boldsymbol{\Theta}^*$.

C.4 Proof of Proposition 5

To simplify notation we omit the subscript m. Let $G(r,\delta,c) := \exp\left(\delta_0 + \sum_{l=1}^{s} \delta_l (r-c_l)^2 I(r \geq c_l)\right)$, where $r = t/T$. The gradient is

$$\dot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta}) = G(r,\delta,c)\left(1, (r-c_1)^2 I(r \geq c_1), \ldots, (r-c_s)^2 I(r \geq c_s), \right)',$$

and the Hessian has $(i+1, j+1)$-th entry

$$[\ddot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta})]_{i+1,j+1} = \begin{cases} G(r,\delta,c)^2 & \text{if } i=j=0 \\ G(r,\delta,c)^2 (r-c_j)^2 I(r \geq c_j) & \text{if } i=0, j>0 \\ G(r,\delta,c)^2 (r-c_i)^2 I(r \geq c_i) & \text{if } i>0, j=0 \\ G(r,\delta,c)^2 (r-c_i)^2 I(r \geq c_i)(r-c_j)^2 I(r \geq c_j) & \text{otherwise} \end{cases}.$$

The third partial derivatives are the product of $G(r,\delta,c)^3$ and three terms from $\{1\} \cup \{(r-c_l)^2 I(r \geq c_l) : l = 1, \ldots, s\}$. It follows straightforwardly that $\dot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta})$ satisfies Assumption 2 and that on a suitable compact subset of $\boldsymbol{\Theta}^*$, both $\|\ddot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta})\|$ and the $|\dddot{g}_{t,T,(j,l,k)}(\boldsymbol{\theta})|$ in Lemma 7 are uniformly bounded by sufficiently large constants that serve as dominating functions. In consequence, (13) satisfies Assumption 8.

The limit $L(\boldsymbol{\theta})$ in Assumption 5 is made up of two terms:

$$L(\boldsymbol{\theta}) = L_1(\boldsymbol{\theta}) + L_2(\boldsymbol{\theta}) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \ln g_{t,T}(\boldsymbol{\theta}) + \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left(\epsilon_t^2 / g_{t,T}(\boldsymbol{\theta})\right).$$

The first term is

$$\begin{aligned} L_1(\boldsymbol{\theta}) &= \delta_0 + \int_{c_1}^{1} \delta_1 (x-c_1)^2 dx + \cdots + \int_{c_s}^{1} \delta_s (x-c_s)^2 dx \\ &= \delta_0 + \sum_{l=1}^{s} \delta_l \int_{c_l}^{1} (x-c_l)^2 dx \\ &= \delta_0 - \frac{1}{3}\sum_{l=1}^{s} \delta_l (c_l - 1)^3. \end{aligned}$$

The second term in the limit $L(\boldsymbol{\theta})$ is, using that $E(\epsilon_{t,T}^2) = g_{t,T}(\boldsymbol{\theta}^\star)$ for all t, T,

$$L_2(\boldsymbol{\theta}) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \frac{g_{t,T}(\boldsymbol{\theta}^\star)}{g_{t,T}(\boldsymbol{\theta})}.$$

Noting that we can write

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\frac{g_{t,T}(\boldsymbol{\theta}^\star)}{g_{t,T}(\boldsymbol{\theta})} &= \frac{1}{T}\sum_{t=1}^{T}\exp\bigl(\delta_0^\star-\delta_0\bigr)\cdot\prod_{l=1}^{s}\exp\bigl((\delta_l^\star-\delta_l)(t/T-c_l)^2 I(t/T\geq c_l)\bigr)\\ &= \frac{1}{T}\sum_{t=1}^{T} g(\boldsymbol{\theta}^\star,\boldsymbol{\theta},t/T)\\ &= \frac{1}{T}\sum_{c_0\leq x<c_1} g(\boldsymbol{\theta}^\star,\boldsymbol{\theta},x)+\cdots+\frac{1}{T}\sum_{c_s\leq x<c_{s+1}} g(\boldsymbol{\theta}^\star,\boldsymbol{\theta},x)\end{aligned}$$

where $x=t/T$, letting $T\to\infty$ gives

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} g(\boldsymbol{\theta}^\star,\boldsymbol{\theta},t/T)=\int_0^{c_1} g(\boldsymbol{\theta}^\star,\boldsymbol{\theta},x)dx+\int_{c_1}^{c_2} g(\boldsymbol{\theta}^\star,\boldsymbol{\theta},x)dx+\cdots+\int_{c_s}^{1} g(\boldsymbol{\theta}^\star,\boldsymbol{\theta},x)dx,$$

where

$$\begin{aligned}\int_0^{c_1} g(\boldsymbol{\theta}^\star,\boldsymbol{\theta},x)dx &= \exp\bigl(\delta_0^\star-\delta_0\bigr)\cdot c_1\\ \int_{c_1}^{c_2} g(\boldsymbol{\theta}^\star,\boldsymbol{\theta},x)dx &= \exp\bigl(\delta_0^\star-\delta_0\bigr)\cdot\int_{c_1}^{c_2}\exp\bigl((\delta_1^\star-\delta_1)(x-c_1)^2\bigr)dx\\ &\;\;\vdots\\ \int_{c_s}^{1} g(\boldsymbol{\theta}^\star,\boldsymbol{\theta},x)dx &= \exp\bigl(\delta_0^\star-\delta_0\bigr)\cdot\int_{c_s}^{1}\prod_{l=1}^{s}\exp\bigl((\delta_l^\star-\delta_l)(x-c_l)^2\bigr)dx.\end{aligned}$$

Their sum is

$$\int_0^1 g(\boldsymbol{\theta}^\star,\boldsymbol{\theta},x)dx = \exp\bigl(\delta_0^\star-\delta_0\bigr)\cdot\left(c_1+\sum_{l=1}^{s}\int_{c_l}^{c_{l+1}}\prod_{k=1}^{l}\exp\bigl((\delta_k^\star-\delta_k)(x-c_k)^2\bigr)dx\right),$$

where $c_{s+1}=1$. Hence

$$L(\boldsymbol{\theta})=\delta_0-\sum_{l=1}^{s}\frac{\delta_l}{3}(c_l-1)^3+\exp\bigl(\delta_0^\star-\delta_0\bigr)\left(c_1+\sum_{l=1}^{s}\int_{c_l}^{c_{l+1}}\prod_{k=1}^{l}\exp\bigl((\delta_k^\star-\delta_k)(x-c_k)^2\bigr)dx\right).$$

We now show that $\dot{\boldsymbol{L}}(\boldsymbol{\theta}^\star)=\mathbf{0}$. For $i=0$, $\frac{\partial L(\boldsymbol{\theta})}{\partial\delta_0}=1-L_2(\boldsymbol{\theta})$ and it is straightforward to verify that $L_2(\boldsymbol{\theta}^\star)=1$ and therefore that $\frac{\partial L(\boldsymbol{\theta})}{\partial\delta_0}=0$. Next, for $i=1,\ldots,s$,

$$\frac{\partial L(\boldsymbol{\theta})}{\partial\delta_i}=-\frac{1}{3}(c_i-1)^3-\exp(\delta_0^\star-\delta_0)\left(\sum_{l=i}^{s}\int_{c_l}^{c_{l+1}}(x-c_i)^2\prod_{k=1}^{l}\exp\left((\delta_k^\star-\delta_k)(x-c_k)^2\right)\,\mathrm{d}x\right).$$

so, at $\boldsymbol{\theta}=\boldsymbol{\theta}^\star$,

$$\frac{\partial L(\boldsymbol{\theta})}{\partial\delta_i}=-\frac{1}{3}(c_i-1)^3-\sum_{l=i}^{s}\int_{c_l}^{c_{l+1}}(x-c_i)^2dx=\int_{c_i}^{1}(x-c_i)^2dx-\int_{c_i}^{1}(x-c_i)^2dx=0.$$

The second derivatives are ($j \geq 1$)

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \delta_0^2} = L_2(\boldsymbol{\theta})$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \delta_0 \partial \delta_j} = \exp(\delta_0^\star - \delta_0) \left(\sum_{l=j}^{s} \int_{c_l}^{c_{l+1}} (x - c_j)^2 \prod_{k=1}^{l} \exp\left((\delta_k^\star - \delta_k)(x - c_k)^2 \right) \mathrm{d}x \right)$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \delta_i \partial \delta_j} = \exp(\delta_0^\star - \delta_0) \left(\sum_{l=i}^{s} I(j \leq l) \int_{c_l}^{c_{l+1}} (x - c_i)^2 (x - c_j)^2 \prod_{k=1}^{l} \exp\left((\delta_k^\star - \delta_k)(x - c_k)^2 \right) \mathrm{d}x \right).$$

If $\ddot{\boldsymbol{L}}(\boldsymbol{\theta})$ is positive definite on $\boldsymbol{\Theta}^*$ then $L(\boldsymbol{\theta})$ has a unique minimum over this set by the same argument which concludes the proof of Proposition 4.

D Proofs of results and simulations in Section 4

D.1 The infeasible case

In the infeasible case, $\{\phi_{m,t}^2\}$ is observed and the QMLE is

$$\widehat{\boldsymbol{\vartheta}}_m^\star = \arg\min_{\boldsymbol{\vartheta}_m \in \boldsymbol{\Xi}} \frac{1}{T} \sum_{t=1}^{T} l_t(\boldsymbol{\vartheta}_m, \phi_{m,t}^2), \qquad l_t(\boldsymbol{\vartheta}_m, \phi_{m,t}^2) = \ln h_t(\boldsymbol{\vartheta}_m) + \frac{\phi_{m,t}^2}{h_t(\boldsymbol{\vartheta}_m)}. \tag{40}$$

To emphasise that this estimator is infeasible, we add $\star$ as superscript. For this estimator to be strongly consistent (see Theorem 4 below), a compactness assumption is needed:

Assumption 18. *For each $m = 1, \ldots, M$, $\boldsymbol{\vartheta}_m^\star \in \boldsymbol{\Xi}_m$ and $\boldsymbol{\Xi}_m$ is compact.*

Next, the following additional assumption ensures the process $\{\phi_{m,t}^2\}$ is β-mixing with exponential decay, which implies that also the α-mixing assumption in Assumption 3 holds.

Assumption 19 (condition η in Carrasco and Chen 2002)**.** *For each $m = 1, \ldots, M$, the probability distribution of $\eta_{m,t}$ has a continuous density (with respect to the Lebesgue measure on the real line), and its density is positive on $(-\infty, \infty)$.*

Theorem 4 (Consistency, infeasible case)**.** *Suppose $\{\phi_{m,t}^2\}$ is governed by (16)–(17), and that Assumption 18 holds. Then $\widehat{\boldsymbol{\vartheta}}_m^\star \xrightarrow{P} \boldsymbol{\vartheta}_m^\star$, $m = 1, \ldots, M$. If, in addition, Assumption 19 holds, then $\{\phi_{m,t}^2\}$ is β-mixing with exponential decay, $m = 1, \ldots, M$.*

Proof of Theorem 4. To simplify notation, we omit the subscript m. Strong consistency of $\widehat{\boldsymbol{\vartheta}}^\star$ follows if the four assumptions of Theorem 7.1 in Francq and Zakoïan (2019) hold. In the GARCH(1,1) case, the four assumptions are:

- A1 $\boldsymbol{\vartheta}^\star \in \boldsymbol{\Xi}$ and $\boldsymbol{\Xi}$ is compact
- A2 The top Lyapunov exponent is strictly negative
- A3 η_t^2 has non-degenerate distribution and $E(\eta_t^2) = 1$
- A4 If the GARCH order $p > 0$, then $\mathcal{A}_{\boldsymbol{\vartheta}^\star}(z) = \alpha^\star z$ and $\mathcal{B}_{\boldsymbol{\vartheta}^\star}(z) = 1 - \beta^\star z$ have no common roots, $\mathcal{A}_{\boldsymbol{\vartheta}^\star}(1) \neq 0$ and $(\alpha^\star + \beta^\star) \neq 0$

A1 holds by Assumption 18. A2 holds since $-\infty \leq E\big(\ln(\alpha^\star\eta_t^2 + \beta^\star)\big) < 0$, see Theorem 2.1 in Francq and Zakoïan (2019, p. 22). Note that $(\alpha^\star + \beta^\star) < 1$ implies $-\infty \leq E\big(\ln(\alpha^\star\eta_t^2 + \beta^\star)\big) < 0$, since $E\big(\ln(\alpha^\star\eta_t^2 + \beta^\star)\big) \leq \ln E(\alpha^\star\eta_t^2 + \beta^\star) = \ln(\alpha^\star + \beta^\star) < 0$ by Jensen's inequality. A3 holds due to $\eta_t \sim iid(0,1)$ in (16). For $p > 0$ to occur in A4, we must have $\beta^\star > 0$. In this case, $\mathcal{A}_{\boldsymbol{\vartheta}^\star}(z) = \alpha^\star z$ and $\mathcal{B}_{\boldsymbol{\vartheta}^\star}(z) = 1 - \beta^\star z$ have no common roots, since the only root of the former is $z = 0$ and the latter has no roots. Also, $\mathcal{A}_{\boldsymbol{\vartheta}^\star}(1) \neq 0$ and $(\alpha^\star + \beta^\star) \neq 0$, since $\alpha^\star > 0$ in (17). So $\widehat{\boldsymbol{\vartheta}}^\star \xrightarrow{P} \boldsymbol{\vartheta}^\star$. That $\{\phi_t^2\}$ is β-mixing with exponential decay if Assumption 19 also holds, follows from Corollary 6 in Carrasco and Chen (2002). $\square$

D.2 Simulations

Table 3: Comparison of the Ordinary QMLE and the Two-step QMLE (see Section 4.1)

g-DGP	T	$m(\widehat{\omega})$	$se(\widehat{\omega})$	$ase(\widehat{\omega})$	$m(\widehat{\alpha})$	$se(\widehat{\alpha})$	$ase(\widehat{\alpha})$	$m(\widehat{\beta})$	$se(\widehat{\beta})$	$ase(\widehat{\beta})$
1:	Ordinary QMLE:									
	10000	0.0017	0.0135	0.0132	0.0002	0.0087	0.0088	−0.0019	0.0190	0.0190
	20000	0.0009	0.0094	0.0094	0.0002	0.0061	0.0062	−0.0011	0.0133	0.0134
	40000	0.0003	0.0068	0.0066	0.0000	0.0045	0.0044	−0.0002	0.0097	0.0095
	80000	0.0002	0.0047	0.0047	−0.0001	0.0031	0.0031	−0.0001	0.0068	0.0067
	Two-step QMLE:									
	10000	0.0014	0.0136	–	−0.0001	0.0087	–	−0.0016	0.0190	–
	20000	0.0087	0.0095	–	−0.0003	0.0063	–	−0.0005	0.0136	–
	40000	0.0004	0.0068	–	0.0000	0.0046	–	−0.0005	0.0098	–
	80000	0.0002	0.0048	–	−0.0001	0.0032	–	−0.0001	0.0069	–

Relative efficiency (two-step *vs.* ordinary):

T	$\frac{se_2(\widehat{\omega})}{se_1(\widehat{\omega})}$	$\frac{se_2(\widehat{\alpha})}{se_1(\widehat{\alpha})}$	$\frac{se_2(\widehat{\beta})}{se_1(\widehat{\beta})}$
10000	1.0044	1.0021	0.9997
20000	1.0103	1.0377	1.0196
40000	1.0007	1.0127	1.0129
80000	1.0107	1.0377	1.0206

2: Two-step QMLE:

T	$m(\widehat{\delta}_0)$	$m(\widehat{\delta}_1)$	$m(\widehat{\gamma})$	$m(\widehat{c})$	$m(\widehat{\alpha})$	$se(\widehat{\alpha})$	$m(\widehat{\beta})$	$se(\widehat{\beta})$
10000	−0.0050	0.0269	0.6592	0.0031	0.0000	0.0087	−0.0031	0.0192
20000	−0.0039	0.0170	0.2612	0.0011	0.0000	0.0059	−0.0017	0.0135
40000	−0.0032	0.0156	0.0720	0.0013	0.0001	0.0042	−0.0009	0.0092
80000	−0.0018	0.0062	0.0337	0.0002	0.0001	0.0032	−0.0005	0.0068

DGP: $\epsilon_t = g_t\phi_t$, $\phi_t = \sqrt{h_t}\eta_t$, $\eta_t \overset{iid}{\sim} N(0,1)$, $h_t = 0.1 + 0.1\phi_{t-1}^2 + 0.8h_{t-1}$. g-DGP 1: $g_t = 1$ for all t. g-DGP 2: $g_t = 0.5 + 1.5\big(1 + \exp(-10(t/T - 0.5)\big)^{-1}$. 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. $ase(\widehat{x})$, asymptotic standard deviation of Ordinary estimate $\widehat{x}$. $se_1(\cdot)$ and $se_2(\cdot)$, the standard errors of the Ordinary and Two-step QMLEs, respectively. All computations in R (R Core Team, 2021). The Ordinary estimator is estimated with the `garchx()` function of the CRAN package **garchx** (Sucarrat, 2021a). The asymptotic standard errors are obtained with the `garchxAvar()` function of the same package. The Two-step QMLE is implemented with own code.

D.3 Proof of Proposition 6

Proposition 6 (the moment-based scaled GARCH(1,1) prediction)**.** *Suppose Assumptions 1–4 hold and that* $\gamma^\star_{m,t,T,j} = \gamma^\star_{m,j}$ *for all* t, T, $|\gamma^\star_{m,j}| < \infty$, $j = 0, 1, 2$, $m = 1, \dots, M$. *If, in addition,*

$$\gamma^\star_{m,1}, \gamma^\star_{m,2} > 0, \qquad \frac{\gamma^\star_{m,2}}{\gamma^\star_{m,1}} \neq \frac{\gamma^\star_{m,1}}{\gamma^\star_{m,0}}, \qquad \gamma^\star_{m,1} > \gamma^\star_{m,2}, \qquad \frac{\gamma^\star_{m,2}}{\gamma^\star_{m,1}} > 0, \qquad b_m > 2, \tag{41}$$

then the moment based scaled GARCH(1,1) prediction

$$h_{m,t} = (1 - \alpha^\star_m - \beta^\star_m) + \alpha^\star_m \phi^2_{m,t-1} + \beta^\star_m h_{m,t-1}, \qquad \alpha^\star_m, \beta^\star_m > 0, \qquad \alpha^\star_m + \beta^\star_m < 1, \tag{42}$$

exists for $m = 1, \dots, M$ *with* $\alpha^\star_m$ *and* $\beta^\star_m$ *given by* (21) *and* (20)*, respectively.*

For notational convenience we omit the subscript m. $\alpha^\star$ and $\beta^\star$ are well defined and finite provided $\gamma^\star_1 \neq 0$ and $(\gamma^\star_2/\gamma^\star_1) \neq (\gamma^\star_1/\gamma^\star_0)$ which ensure that $\rho^\star(1) \neq 0$ and $\rho^\star(2)/\rho^\star(1) - \rho^\star(1) \neq 0$. From (20) it follows that $\alpha^\star + \beta^\star = \gamma^\star_2/\gamma^\star_1$. Therefore $\gamma^\star_2 < \gamma^\star_1$ and $\gamma^\star_2/\gamma^\star_1 > 0$ suffice for $0 < \alpha^\star + \beta^\star < 1$. Next, setting the expression for $\beta^\star$ in (20) to 0 yields a contradiction, so $\beta^\star$ cannot be 0. As $b^\star > 2$, $(b^\star)^2 - 4 > 0$ and hence $\beta^\star$ is real, strictly positive and less than 1. The existence of the prediction in (42) follows from these observations.

D.4 Proof of Theorem 5

Theorem 5 (consistency of $\widehat{\alpha}_m$ and $\widehat{\beta}_m$)**.** *Suppose the assumptions of Theorem 1, the assumptions of Proposition 6 and Assumption 7 hold. Then* $(\widehat{\alpha}_m, \widehat{\beta}_m)' \xrightarrow{P} (\alpha^\star_m, \beta^\star_m)'$ *for* $m = 1, \dots, M$.

For notational convenience we omit the subscript m. By a mean-value expansion,

$$\widehat{\gamma}_j(\widehat{\boldsymbol{\theta}}_T) = \widehat{\gamma}_j(\boldsymbol{\theta}^\star) + \frac{\partial \widehat{\gamma}_j(\overline{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star),$$

where $\overline{\boldsymbol{\theta}}_T$ is a mean-value between $\widehat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}^\star$. Given Theorem 1 it follows that $\widehat{\gamma}_j(\widehat{\boldsymbol{\theta}}_T) \xrightarrow{P} \gamma^\star_j$ if $\widehat{\gamma}_j(\boldsymbol{\theta}^\star) \xrightarrow{P} \gamma^\star_j$ and $\frac{\partial \widehat{\gamma}_j(\overline{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}'} = O_P(1)$. We first show that $\widehat{\gamma}_j(\boldsymbol{\theta}^\star)$ is consistent. At $\boldsymbol{\theta}^\star$,

$$\begin{aligned}
\widehat{\gamma}_j(\boldsymbol{\theta}^\star) &= \frac{1}{T}\sum_{t=1}^{T}\Big(\phi^2_{t,T} - 1\Big)\Big(\phi^2_{t-j,T} - 1\Big) \\
&= \frac{1}{T}\sum_{t=1}^{T}\phi^2_{t,T}\phi^2_{t-j,T} - \phi^2_{t,T} - \phi^2_{t-j,T} + 1, \\
&= \frac{1}{T}\sum_{t=1}^{T} a(\phi^2_{t,T}, \phi^2_{t-j,T}),
\end{aligned}$$

where $a(\phi^2_{t,T}, \phi^2_{t-j,T})$ is defined to be the summand to make the notation more compact. The assumption in Proposition 6 implies that $\gamma^\star_{t,j,T}$ is constant across t, T for $j = 0, 1, 2$ and so $Ea\big(\phi^2_{t,T}\phi^2_{t-j,T}\big) = \gamma^\star_j$ is constant over t, T for $j = 0, 1, 2$. Let $X_{t,T} := a(\phi^2_{t,T}, \phi^2_{t-j,T}) - E\big(a(\phi^2_{t,T}, \phi^2_{t-j,T})\big)$ and let $\mathcal{F}_{t,T} := \sigma(\{X_{i,T} : 1 \leq i \leq t\})$. Assumption 7 yields that

$\sup_{1\le t\le T, T\in\mathbb{N}} E|X_{t,T}|^p < \infty$ for some $p>1$. Hence $\{X_{t,T} : 1\le t\le T,\ T\in\mathbb{N}\}$ is uniformly integrable and, given Assumption 3, $\{X_{t,T}/T, \mathcal{F}_{t,T} : 1\le t\le T,\ T\in\mathbb{N}\}$ forms a L_1-mixingale array with respect to the constants $c_{t,T} = 1/T$ by Theorems 14.1 and 14.2 in Davidson (1994). Therefore, by Theorem 19.11 in Davidson (1994),

$$E\left|\frac{1}{T}\sum_{t=1}^{T} a(\phi_{t,T}^2, \phi_{t-j,T}^2) - \frac{1}{T}\sum_{t=1}^{T} E\big(a(\phi_{t,T}^2, \phi_{t-j,T}^2)\big)\right| \to 0,$$

which implies $\widehat{\gamma}_j(\boldsymbol{\theta}^\star) \xrightarrow{P} \gamma_j^\star$, $j=0,1,2$. We now show that $\frac{\partial \widehat{\gamma}_j(\overline{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}'} = O_P(1)$. We have

$$\frac{\partial \widehat{\gamma}_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \widehat{\phi}_{t,T}^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \widehat{\phi}_{t-j,T}^2(\boldsymbol{\theta}) + \widehat{\phi}_{t,T}^2(\boldsymbol{\theta}) \frac{\partial \widehat{\phi}_{t-j,T}^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \widehat{\phi}_{t,T}^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \widehat{\phi}_{t-j,T}^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

with

$$\widehat{\phi}_{t-j,T}^2(\boldsymbol{\theta}) = \frac{g_{t-j,T}(\boldsymbol{\theta}^\star)\phi_{t-j,T}^2}{g_{t-j,T}(\boldsymbol{\theta})}, \qquad \frac{\partial \widehat{\phi}_{t-j,T}^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{g_{t-j,T}(\boldsymbol{\theta}^\star)\phi_{t-j,T}^2}{\big(g_{t-j,T}(\boldsymbol{\theta})\big)^2} \dot{\boldsymbol{g}}_{t-j,T}(\boldsymbol{\theta}), \qquad j=0,1,2.$$

For more compact notation, let $\frac{\partial \widehat{\gamma}_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = F_{j,T}(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T} b_{t,T}(\phi_{t,T}^2, \phi_{t-j,T}^2, \boldsymbol{\theta})$, i.e.

$$\begin{aligned} b_{t,T}(\phi_{t,T}^2, \phi_{t-j,T}^2, \boldsymbol{\theta}) := {} & \phi_{t,T}^2 \left(\frac{g_{t,T}(\boldsymbol{\theta}^\star)\dot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta})}{g_{t,T}(\boldsymbol{\theta})^2}\right) + \phi_{t-j,T}^2 \left(\frac{g_{t-j,T}(\boldsymbol{\theta}^\star)\dot{\boldsymbol{g}}_{t-j,T}(\boldsymbol{\theta})}{g_{t-j,T}(\boldsymbol{\theta})^2}\right) \\ & - \phi_{t,T}^2\phi_{t-j,T}^2 \left(\frac{g_{t,T}(\boldsymbol{\theta}^\star)\dot{\boldsymbol{g}}_{t,T}(\boldsymbol{\theta})}{g_{t,T}(\boldsymbol{\theta})^2}\frac{g_{t-j,T}(\boldsymbol{\theta}^\star)}{g_{t-j,T}(\boldsymbol{\theta})} + \frac{g_{t-j,T}(\boldsymbol{\theta}^\star)\dot{\boldsymbol{g}}_{t-j,T}(\boldsymbol{\theta})}{g_{t-j,T}(\boldsymbol{\theta})^2}\frac{g_{t,T}(\boldsymbol{\theta}^\star)}{g_{t,T}(\boldsymbol{\theta})}\right). \end{aligned}$$

By Assumption 2 and the fact that $\overline{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}^\star$ given Theorem 1 as it is a mean-value between $\hat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}^\star$, the absolute value of each of the terms in parenthesis is bounded above by some constant C on sets E_n with $P(E_n)\to 1$. On these sets

$$F_{j,T}(\overline{\boldsymbol{\theta}}) \le C\frac{1}{T}\sum_{t=1}^{T} \phi_{t,T}^2 + \phi_{t-j,T}^2 + \phi_{t,T}^2\phi_{t-j,T}^2.$$

By Markov's inequality and Assumption 7

$$P\left(\frac{1}{T}\sum_{t=1}^{T} \phi_{t,T}^2 + \phi_{t-j,T}^2 + \phi_{t,T}^2\phi_{t-j,T}^2 > M\right) \le M^{-1}\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[\phi_{t,T}^2 + \phi_{t-j,T}^2 + \phi_{t,T}^2\phi_{t-j,T}^2\right] \lesssim M^{-1}.$$

Combine these observations to conclude that $F_{j,T}(\overline{\boldsymbol{\theta}}) = O_P(1)$. Thus, $\widehat{\gamma}_{m,j}(\widehat{\boldsymbol{\theta}}_T) \xrightarrow{P} \gamma_{m,j}^\star$ and by the continuous mapping theorem $(\widehat{\alpha}_m, \widehat{\beta}_m)' \xrightarrow{P} (\alpha_m, \beta_m)'$.

D.5 Proof of Theorem 6

To prove asymptotic normality of the least squares estimators, we require the following additional conditions.

Assumption 20. *For each $m=1,\dots,M$, $\sup_{1\le t\le T, T\in\mathbb{N}} E|\phi_{m,t,T}^4|^{2+\tilde{\delta}_m}$ for some $\tilde{\delta}_m > 0$.*

Additionally, the strong mixing coefficients $\alpha_{m,T}(k)$ *satisfy*

$$\sup_{T\in\mathbb{N}} \alpha_{m,T}(k) = O(k^{-\tilde{\rho}_m-\varepsilon}),$$

for some $\varepsilon > 0$ *where* $\tilde{\rho}_m := \tilde{r}_m/(\tilde{r}_m - 2)$ *with* $\tilde{r}_m := 2 + \tilde{\delta}_m$.

For the following condition, let $\overline{\boldsymbol{F}}_{m,T}$ be defined (on $\mathcal{V}_m$) as

$$\overline{\boldsymbol{F}}_{m,T}(\boldsymbol{\theta}_m) := \frac{1}{T}\sum_{t=1}^{T} E[\boldsymbol{b}_{m,t,T}(\boldsymbol{\theta}_m)], \qquad \boldsymbol{b}_{m,t,T} := \begin{pmatrix}\boldsymbol{b}_{m,t,T,0} & \boldsymbol{b}_{m,t,T,1} & \boldsymbol{b}_{m,t,T,2}\end{pmatrix}',$$

where for $j = 0, 1, 2$,

$$\begin{aligned}\boldsymbol{b}_{m,t,T,j}(\boldsymbol{\theta}_m) := {} & \phi^2_{m,t,T}\left(\frac{g_{m,t,T}(\boldsymbol{\theta}^\star_m)\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)}{g_{m,t,T}(\boldsymbol{\theta}_m)^2}\right) + \phi^2_{m,t-j,T}\left(\frac{g_{m,t-j,T}(\boldsymbol{\theta}^\star_m)\dot{\boldsymbol{g}}_{m,t-j,T}(\boldsymbol{\theta}_m)}{g_{m,t-j,T}(\boldsymbol{\theta}_m)^2}\right)\\ & - \phi^2_{m,t,T}\phi^2_{m,t-j,T}\left(\frac{g_{m,t,T}(\boldsymbol{\theta}^\star_m)\dot{\boldsymbol{g}}_{m,t,T}(\boldsymbol{\theta}_m)}{g_{m,t,T}(\boldsymbol{\theta}_m)^2}\frac{g_{m,t-j,T}(\boldsymbol{\theta}^\star_m)}{g_{m,t-j,T}(\boldsymbol{\theta}_m)}\right)\\ & - \phi^2_{m,t,T}\phi^2_{m,t-j,T}\left(\frac{g_{m,t-j,T}(\boldsymbol{\theta}^\star_m)\dot{\boldsymbol{g}}_{m,t-j,T}(\boldsymbol{\theta}_m)}{g_{m,t-j,T}(\boldsymbol{\theta}_m)^2}\frac{g_{m,t,T}(\boldsymbol{\theta}^\star_m)}{g_{m,t,T}(\boldsymbol{\theta}_m)}\right).\end{aligned}$$

Assumption 21. *For each* $m = 1, \ldots, M$*, the limit*

$$\overline{\boldsymbol{F}}^\star_{m,j}(\boldsymbol{\theta}_m) := \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E[\boldsymbol{b}_{m,t,T,j}(\boldsymbol{\theta}_m)], \qquad \boldsymbol{\theta}_m \in \mathcal{V}_m, \tag{43}$$

exists where $\mathcal{V}_m$ *is as in Assumption 8. Additionally, there are random variables functions* $\tilde{\psi}_{m,t,T}$ *such that for* $\boldsymbol{\theta}_m, \boldsymbol{\theta}'_m \in \mathcal{V}_m$

$$\|\boldsymbol{b}_{m,t,T}(\boldsymbol{\theta}_m) - \boldsymbol{b}_{m,t,T}(\boldsymbol{\theta}'_m)\| \le \tilde{\psi}_{m,t,T}\|\boldsymbol{\theta}_m - \boldsymbol{\theta}'_m\|,$$

with $\sup_{1\le t\le T, T\in\mathbb{N}} E|\tilde{\psi}_{m,t,T}| < \infty$.

The following Assumption strengthens Assumption 10.

Assumption 22. *For each* $m = 1, \ldots, M$*, as* $T \to \infty$,

$$\boldsymbol{D}_{m,T} := \mathrm{Var}\left(T^{-1/2}\sum_{t=1}^{T}\boldsymbol{y}_{m,t,T}\right) \to \boldsymbol{D}^\star_m,$$

for

$$\boldsymbol{y}_{m,t,T} := \begin{bmatrix} \dot{\boldsymbol{l}}_{m,t,T}(\boldsymbol{\theta}^\star_m, \epsilon^2_{m,t,T}) \\ \phi^4_{m,t,T} - E(\phi^4_{m,t,T}) \\ \phi^2_{m,t,T}\phi^2_{m,t-1,T} - E(\phi^2_{m,t,T}\phi^2_{m,t-1,T}) \\ \phi^2_{m,t,T}\phi^2_{m,t-2,T} - E(\phi^2_{m,t,T}\phi^2_{m,t-2,T}) \end{bmatrix},$$

with $\boldsymbol{D}^\star_m$ *positive definite.*

Theorem 6 (AN of $(\widehat{\alpha}_m, \widehat{\beta}_m)$)**.** *Suppose the assumptions of Theorems 1, 2, Proposition 6 and Assumptions 20 – 22 hold. Then* $\sqrt{T}(\widehat{\boldsymbol{\gamma}}_m - \boldsymbol{\gamma}^\star_m) \xrightarrow{D} N(\boldsymbol{0}, \boldsymbol{\Sigma}^\star_{m,\gamma})$ *with* $\boldsymbol{\Sigma}^\star_{m,\gamma} =$

$\boldsymbol{C}_m^\star \boldsymbol{D}_m^\star [\boldsymbol{C}_m^\star]'$, *$m = 1, \dots, M$, and*

$$\sqrt{T}\begin{pmatrix} \widehat{\alpha}_m - \alpha_m^\star \\ \widehat{\beta}_m - \beta_m^\star \end{pmatrix} \xrightarrow{D} N\big(\mathbf{0}, \boldsymbol{\Upsilon}_m(\boldsymbol{\gamma}_m^\star)\boldsymbol{\Sigma}_{m,\gamma}\boldsymbol{\Upsilon}_m(\boldsymbol{\gamma}_m^\star)'\big), \qquad m = 1, \dots, M, \tag{44}$$

where $\boldsymbol{\Upsilon}_m(\boldsymbol{\gamma}_m^\star) = \Big(\frac{\partial \alpha_m(\boldsymbol{\gamma}_m^\star)}{\partial \boldsymbol{\gamma}_m'}, \frac{\partial \beta_m(\boldsymbol{\gamma}_m^\star)}{\partial \boldsymbol{\gamma}_m'}\Big)'$.

The exact expressions for $\boldsymbol{C}_m^\star$ and $\boldsymbol{D}_m^\star$ are contained in the proof. A consistent estimator of $\boldsymbol{C}_m^\star$ can be derived by using the estimates of $g_{t,T}(\boldsymbol{\theta}_m^\star)$, $\gamma_{m,0}^\star$, $\gamma_{m,1}^\star$ and $\gamma_{m,2}^\star$ as ingredients, whereas a consistent estimator of $\boldsymbol{D}_m^\star$ of the HAC type can be derived along the same lines as in Proposition 1.

Proof of Theorem 6. For notational convenience we omit the subscript m. Let $\widehat{\boldsymbol{\gamma}} = \big(\widehat{\gamma}_0(\widehat{\boldsymbol{\theta}}), \widehat{\gamma}_1(\widehat{\boldsymbol{\theta}}), \widehat{\gamma}_2(\widehat{\boldsymbol{\theta}})\big)'$, where the expression for $\widehat{\gamma}_j(\widehat{\boldsymbol{\theta}})$ is contained in (22). By a mean-value expansion,

$$\widehat{\boldsymbol{\gamma}}(\widehat{\boldsymbol{\theta}}_T) = \widehat{\boldsymbol{\gamma}}(\boldsymbol{\theta}^\star) + \frac{\partial \boldsymbol{\gamma}(\overline{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star),$$

where $\overline{\boldsymbol{\theta}}_T$ is a mean value between $\widehat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}^\star$. Let $\boldsymbol{\gamma}^\star = (\gamma_0^\star, \gamma_1^\star, \gamma_2^\star)'$, so that we can write

$$\begin{aligned} \sqrt{T}\big(\widehat{\boldsymbol{\gamma}}(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{\gamma}^\star\big) &= \sqrt{T}\big(\widehat{\boldsymbol{\gamma}}(\boldsymbol{\theta}^\star) - \boldsymbol{\gamma}^\star\big) + \frac{\partial \boldsymbol{\gamma}(\overline{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}'}\sqrt{T}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) \\ &= \sqrt{T}\big(\widehat{\boldsymbol{\gamma}}(\boldsymbol{\theta}^\star) - \boldsymbol{\gamma}^\star\big) + \overline{\boldsymbol{F}}^\star\sqrt{T}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) + o_P(1), \end{aligned}$$

since $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) = O_P(1)$ by Theorem 2 and where $\overline{\boldsymbol{F}}^\star = \big(\overline{F}_0(\boldsymbol{\theta}^\star), \overline{F}_1(\boldsymbol{\theta}^\star), \overline{F}_2(\boldsymbol{\theta}^\star)\big)'$ is the probability limit of $\partial \boldsymbol{\gamma}(\overline{\boldsymbol{\theta}}_T)/\partial \boldsymbol{\theta}'$ as $T \to \infty$ which is established in Lemma 5.

Let $\boldsymbol{x}_{t,T} = \big(\phi_{t,T}^4 - E(\phi_{t,T}^4), \phi_{t,T}^2\phi_{t-1,T}^2 - E(\phi_{t,T}^2\phi_{t-1,T}^2), \phi_{t,T}^2\phi_{t-2,T}^2 - E(\phi_{t,T}^2\phi_{t-2,T}^2)\big)'$. Then

$$\sqrt{T}\big(\widehat{\boldsymbol{\gamma}}(\boldsymbol{\theta}^\star) - \boldsymbol{\gamma}^\star\big) = \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t,T}\right).$$

(The proof of) Theorem 2 (implicitly) uses the result that (cf. Theorem 3.1 in Newey and McFadden (1994))

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) = [\boldsymbol{A}^\star]^{-1}\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{l}_{t,T}(\boldsymbol{\theta}^\star, \epsilon_{t,T}^2)\right) + o_P(1).$$

Combine these to obtain

$$\begin{aligned} \sqrt{T}\big(\widehat{\boldsymbol{\gamma}}(\widehat{\boldsymbol{\theta}}_T) - \boldsymbol{\gamma}^\star\big) &= \overline{\boldsymbol{F}}^\star\sqrt{T}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) + \sqrt{T}\big(\widehat{\boldsymbol{\gamma}}(\boldsymbol{\theta}^\star) - \boldsymbol{\gamma}^\star\big) + o_P(1), \\ &= \boldsymbol{C}^\star\, T^{-1/2}\sum_{t=1}^{T}\boldsymbol{y}_{t,T} + o_P(1), \end{aligned}$$

where

$$\boldsymbol{C}^\star = \begin{pmatrix} \overline{\boldsymbol{F}}^\star[\boldsymbol{A}^\star]^{-1} & \underset{(3\times 3)}{\boldsymbol{I}} \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{y}_{t,T} = \Big(\boldsymbol{l}_{t,T}(\boldsymbol{\theta}^\star, \epsilon_{t,T}^2)', \boldsymbol{x}_{t,T}'\Big)'.$$

with $E(\boldsymbol{y}_{t,T}) = \mathbf{0}$ for all t, T by Assumption 4. We now show

$$T^{-1/2} \sum_{t=1}^{T} \boldsymbol{y}_{t,T} \xrightarrow{D} N(\mathbf{0}, \boldsymbol{D}^\star), \qquad \boldsymbol{D}^\star := \lim_{T\to\infty} \mathrm{Var}\left(T^{-1/2} \sum_{t=1}^{T} \boldsymbol{y}_{t,T}\right).$$

Notice that the upper-left $(K \times K)$ block of $\boldsymbol{D}^\star$, where K is the dimension of $\boldsymbol{\theta}^\star$, is equal to $\boldsymbol{B}^\star$ in Assumption 10. Let $Z_{t,T} := T^{-1/2}\lambda' \boldsymbol{y}_{t,T}$ for $\|\lambda\|_2 = 1$, and let $\sigma_T := \|\sum_{t=1}^{T} Z_{t,T}\|_{L_2}$ and $X_{t,T} = Z_{t,T}/\sigma_T$. That σ_T is finite follows from Assumption 20; that it is (at least eventually) non-zero follows from Assumption 22. Let $\mathcal{F}_{t,T} := \sigma(\epsilon_{1,T}^2, \ldots, \epsilon_{t,T}^2)$. We will verify the conditions of Corollary 1 in de Jong (1997). (a) follows as $X_{t,T}$ is a mean-zero random variable with $\|\sum_{t,T} X_{t,T}\|_{L_2} = 1$. For (b) set $c_{t,T} := \max\{\|Z_{t,T}\|_{L_2}, 1\}/\sigma_T$. By the moment bounds in Assumption 20

$$\sup_{1\le t\le T, T\in\mathbb{N}} \|X_{t,T}/c_{t,T}\|_{L_{r_m}} \le \sup_{1\le t\le T, T\in\mathbb{N}} \sigma_T \|X_{t,T}\|_{L_{r_m}} = \sup_{1\le t\le T, T\in\mathbb{N}} \|Z_{t,T}\|_{L_{r_m}} < \infty, \tag{45}$$

For (c), since each $X_{t,T}$ is $\mathcal{F}_{t,T}$–measurable (and in L_2), it is trivially L_2 – NED (of any size) on $(\epsilon_{t,T}^2)_{1\le t\le T, T\in\mathbb{N}}$ and by Assumption 20 this latter array is α-mixing of size $-\tilde{\rho}_m$. Finally, we note that by the moment bounds in Assumption 20 and

$$T c_{t,T}^2 \le \frac{1}{\sigma_T^2} \max\{\|\boldsymbol{y}_{t,T}\|_{L_2}^2, 1\} \lesssim \frac{1}{\sigma_T^2}.$$

By Assumption 22

$$\sigma_T^2 = \left\| \sum_{t=1}^{T} Z_{t,T} \right\|_{L_2}^2 = \lambda' \boldsymbol{D}_T \lambda \to \lambda' \boldsymbol{D}^\star \lambda > 0. \tag{46}$$

Combination of the preceding displays permits the conclusion that $c_{t,T}^2 = O(T^{-1})$, establishing the final condition of Corollary 1 of de Jong (1997) with $\beta = \gamma = 0$. Therefore, $\sum_{t=1}^{T} X_{t,T} \xrightarrow{D} N(0, 1)$. In conjunction with (46) and Slutsky's Theorem this implies $\sum_{t=1}^{T} Z_{t,T} \xrightarrow{D} N(0, \lambda' \boldsymbol{D}^\star \lambda)$. Hence $T^{-1/2} \sum_{t=1}^{T} \boldsymbol{y}_{t,T} \xrightarrow{D} N(\mathbf{0}, \boldsymbol{D}^\star)$ holds by the Cramér – Wold Theorem. Next, applying Slutsky's theorem again gives $\sqrt{T}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\star) \xrightarrow{D} N\big(\mathbf{0}, \boldsymbol{C}^\star \boldsymbol{D}^\star [\boldsymbol{C}^\star]'\big)$. Apply the delta method to obtain (44). $\square$