

## ORIGINAL ARTICLE

# Robust Estimation and Inference for Time-Varying Unconditional Volatility

Adam Lee<sup>1</sup> | Rickard Sandberg<sup>2</sup> | Genaro Sucarrat<sup>3</sup>

<sup>1</sup>Department of Data Science and Analytics, BI Norwegian Business School, Oslo, Norway | <sup>2</sup>Centre for Data Analytics, Stockholm School of Economics, Stockholm, Sweden | <sup>3</sup>Department of Economics, BI Norwegian Business School, Oslo, Norway

**Correspondence:** Genaro Sucarrat ([genaro.sucarrat@bi.no](mailto:genaro.sucarrat@bi.no))

**Received:** 13 January 2025 | **Revised:** 29 September 2025 | **Accepted:** 4 November 2025

**Funding:** The authors received no specific funding for this work.

**Keywords:** ARCH | GARCH | MEM | robust estimation | time-varying unconditional volatility | volatility

## ABSTRACT

We derive a general and robust estimator of a large class of parametric specifications of time-varying unconditional volatility of financial returns, both univariate and multivariate, and establish the Consistency and Asymptotic Normality (CAN) of the estimator. A number of well-known and widely used parametric specifications, for many of which asymptotic results have not been specifically established, are contained in the class. The estimator is robust in the sense that the exact specification of the conditional volatility dynamics need not be known or estimated, and in the sense that the stochastic component need not be strictly stationary. The latter is especially important in light of recent findings, which document that financial returns are frequently characterised by a non-stationary zero-process. Our estimator is also robust to the well-known “curse of dimensionality” in multivariate models due to its equation-by-equation nature. While our estimator does not require the exact specification of the conditional volatility dynamics to be known or estimated, our results imply that the scaled GARCH(1,1) specification is well-defined under both correct and incorrect specifications. So we provide methods for its estimation in a second step. Also, due to the assumptions we rely upon, our results extend directly to the Multiplicative Error Model (MEM) interpretation of volatility models. This means our results can also be applied to other non-negative processes like volume, duration, realised volatility, dividends, unemployment, and so on. Three numerical applications illustrate the versatility of our results.

**JEL Classification:** C01, C13, C14, C22, C58

**MSC2020 Classification:** 60, 62

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

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

for all  $t$ . Henceforth, we refer to  $g_t$  as the unconditional volatility at  $t$ . The decomposition in Equation (1) implies  $E(\phi_t^2) = 1$  for all  $t$ . For some of the implications of this, see our discussion in relation to Assumption 4 further below. While a model of the conditional volatility dynamics is not needed for our main results, it is often of interest in empirical applications, since volatility prediction is commonly an important objective. A leading example is the scaled version of the GARCH(1,1) model of Bollerslev (1986):

$$\begin{aligned} \phi_t^2 &= h_t \eta_t^2, & E(\eta_t^2 | \mathcal{F}_{t-1}) &= 1, \\ h_t &= \omega + \alpha \phi_t^2 + \beta h_{t-1}, & t \in \mathbb{Z}, \end{aligned} \quad (2)$$

where  $\mathcal{F}_{t-1}$  is a suitable  $\sigma$ -algebra generated by past  $\phi_t^2$ 's. The conditional volatility or variance at  $t$  is thus  $\sigma_t^2 = g_t h_t$ , and the unconditional volatility or variance at  $t$  is  $E(\sigma_t^2) = g_t$ . In Section 4, we show that the scaled GARCH(1,1) is well-defined in our setup under both correct and incorrect specification. Other examples of  $h_t$  include scaled versions of Stochastic Volatility (SV) models (in which  $\sigma_t^2$  need not equal the conditional variance), and scaled versions of Dynamic Conditional Score (DCS) models. Henceforth, to simplify the exposition, we refer to both  $\sigma_t$  and  $\sigma_t^2$  as conditional volatility, since one is obtained from the other via a straightforward transformation.

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 governed by a finite-dimensional parameter  $\theta$ . 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 uses Gaussian Maximum Likelihood (ML) for estimation, whereas the latter employs 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 (2024),  $g_t$  is specified as a smooth transition function, and  $\phi_t^2$  is governed by a first-order GARCH model. In these papers, the Gaussian Quasi ML Estimator (QMLE) is used to estimate the parameter  $\theta$  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  $\phi_t^2$  is *iid*, and in the latter, the standardised error is assumed to be *iid*.<sup>1</sup> Next, Consistency and Asymptotic Normality (CAN) of the parameters of a GARCH model that governs  $\phi_t^2$  is established in the

infeasible case where  $\theta$  is known from the first step.<sup>2</sup> Theorem 7 in Silvennoinen and Teräsvirta (2024) 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.<sup>3</sup> 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 Online 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  $\theta$ . 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  $\phi_t \stackrel{iid}{\sim} N(0, 1)$  (see their assumption A5 in Section 5.2). This is generalised to the multivariate case in He et al. (2024), but  $\phi_t$  is still required to be  $\stackrel{iid}{\sim} N(0, 1)$  for all  $t$  conditional on the past (see assumption AV6 in the supplement to He et al. 2024).

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

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

so  $\{g_{t,T} : T \in \mathbb{N}, 1 \leq t \leq T\}$  forms a triangular array of functions from  $\Theta$  to  $(0, \infty)$ . We prove that the equation-by-equation Gaussian QMLE provides Consistent and Asymptotically Normal (CAN) estimators of  $\theta$  for a large number and widely used specifications in Equation (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 Equation (3) is  $g_{t,T}(\theta) = g(\theta, t/T)$ ,  $g : \Theta \times [0, 1] \rightarrow (0, \infty)$ , 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  $\phi_t^2$  in the estimation of  $\theta$ . Also, the  $\phi_t^2$ 's can be dependent (strongly mixing) over time. Our results thus hold for a large class of specifications of  $\phi_t^2$ , 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 specification assumptions on  $\phi_t^2$ .<sup>4</sup> Second, contrary to the parametric works cited above, we do not rely on the assumption that  $\{\phi_t^2\}$  or  $\{\eta_t^2\}$  is strictly stationary. Relaxing the strict stationarity assumption on  $\{\phi_t^2\}$  and  $\{\eta_t^2\}$  is important, since recent studies reveal that the zero-process of financial returns—both daily and intraday—is frequently non-stationary, see, e.g., Kolokolov et al. (2020), Sucarrat and Grønneberg (2022), Francq and Sucarrat (2023), Stauskas and Sucarrat (2025), Patilea and Raïssi (2024), and Kolokolov and Reno (2024). In particular, the standard assumption that  $\eta_t^2$  is *iid* is not compatible with a non-stationary zero-process. Sections 5.2 and 5.3 contain illustrations. A third attractive property of our estimator is its equation-by-equation nature (cf. Francq and Zakoian 2016). This reduces the numerical challenges (‘‘the curse

of dimensionality”) typically associated with multivariate models. Fourth, while our results do not require the estimation and explicit specification of a model of  $\phi_t^2$ , a model can nevertheless be estimated in a second step. In particular, an especially interesting implication of our results is that the scaled GARCH(1,1) prediction is well-defined under both correct and incorrect specification within our framework, even under certain types of non-stationarities of the stochastic component  $\phi_t^2$ . 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  $\phi_t^2/h_t$  is *iid*. A fifth 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 gives an empirical illustration. Sixth, 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  $\phi_t^2/h_t$ , see, e.g., He et al. (2019), He et al. (2024), and Silvennoinen and Teräsvirta (2024). 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  $\epsilon_t^2$  and  $\phi_t^2$ , not on  $\epsilon_t$  and  $\phi_t$ . 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  $\epsilon_t^2$  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 Equation (3), and derives 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, and how time-varying correlations can be estimated subsequently. Section 5 contains numerical illustrations of our results, whereas Section 6 concludes. The proofs of our results are contained in the Online Appendix.

## 2 | Consistency and Asymptotic Normality

### 2.1 | Consistency

Let  $\epsilon_{t,T} = (\epsilon_{1,t,T}, \dots, \epsilon_{M,t,T})'$  denote an  $M$ -dimensional multivariate return with  $M \in \mathbb{N}$ , and let

$$\begin{aligned} \epsilon_{m,t,T}^2 &= g_{m,t,T}(\theta_m^*) \phi_{m,t,T}^2, \\ m &= 1, \dots, M, \quad 1 \leq t \leq T, \quad T \in \mathbb{N}, \end{aligned} \quad (4)$$

where  $g_{m,t,T}$  is a deterministic function. Our estimator of  $\theta^* = (\theta_1^*, \dots, \theta_M^*)' \in \prod_{m=1}^M \Theta_m = \Theta$  is derived from the objective function

$$\begin{aligned} L_T(\theta) &= \sum_{m=1}^M L_{m,T}(\theta_m) \quad \text{with} \\ L_{m,T}(\theta_m) &= \frac{1}{T} \sum_{t=1}^T l_{m,t,T}(\theta_m, \epsilon_{m,t,T}^2), \end{aligned} \quad (5)$$

where  $\theta = (\theta_1', \dots, \theta_M')'$  and

$$\begin{aligned} l_{m,t,T}(\theta_m, \epsilon_{m,t,T}^2) &= \ln g_{m,t,T}(\theta_m) + \frac{\epsilon_{m,t,T}^2}{g_{m,t,T}(\theta_m)}, \quad m = 1, \dots, M. \end{aligned}$$

Minimisation of (5) leads to the Equation-by-Equation (EBE) Quasi Maximum Likelihood Estimator (QMLE):

$$\begin{aligned} \hat{\theta} &= \arg \min_{\theta \in \Theta} L_T(\theta) \\ &= (\hat{\theta}_1', \dots, \hat{\theta}_M')', \\ \hat{\theta}_m &= \arg \min_{\theta_m \in \Theta_m} L_{m,T}(\theta_m), \quad m = 1, \dots, M. \end{aligned} \quad (6)$$

This is an EBE estimator, since the parameters of equation  $m$ , i.e.,  $\theta_m$ , can be estimated separately from the parameters of the other equations. An attractive property of the EBE estimator is that it provides a solution to the “curse of dimensionality”. Note in that regard that an EBE estimator is not necessarily less efficient asymptotically than a system estimator (Francq and Zakoian 2016). Note also that a single-equation cannot be used to model unconditional volatility that is periodic (as in intraday data); this can be achieved, however, with a multiple-equation approach via the vector-of-seasons representation (this is illustrated in Section 5.3).

In establishing consistency of the EBE-QMLE, we rely on the following assumptions.

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

**Assumption 2.** For each  $m = 1, \dots, M$ , let  $\Theta_m^*$  be an open, convex set containing  $\Theta_m$ . For all  $1 \leq t \leq T$ ,  $T \in \mathbb{N}$ ,

i.  $g_{m,t,T}(\theta_m)$  is bounded away from zero and infinity, i.e.,

$$\begin{aligned} 0 &< \inf_{\theta_m \in \Theta_m^*, 1 \leq t \leq T, T \in \mathbb{N}} g_{m,t,T}(\theta_m) \\ &\leq \sup_{\theta_m \in \Theta_m^*, 1 \leq t \leq T, T \in \mathbb{N}} g_{m,t,T}(\theta_m) < \infty. \end{aligned}$$

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

$$\sup_{\theta_m \in \Theta_m^*, 1 \leq t \leq T, T \in \mathbb{N}} \|\dot{g}_{m,t,T}(\theta_m)\| < \infty.$$

**Assumption 3.** For each  $m = 1, \dots, M$ ,  $\{\epsilon_{m,t,T}^2 : t \in \mathbb{Z}, T \in \mathbb{N}\}$  forms a triangular array of a.s. non-negative random variables.

Let  $\alpha_{m,T}(k)$  be the  $\alpha$ -mixing coefficients corresponding to  $\{\epsilon_{m,t,T}^2 : t \in \mathbb{Z}\}$  and suppose that as  $k \rightarrow \infty$ ,

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

where  $\alpha_{m,T}(k) := \sup_{t \in \mathbb{Z}} \sup\{|P(F \cap G) - P(F)P(G)| : F \in \mathcal{F}_{m,-\infty,t}^t, G \in \mathcal{F}_{m,t+k,T}^\infty\}$  with  $\mathcal{F}_{m,-\infty,t}^t := \sigma(\epsilon_{m,s,T}^2 : s \leq t)$  and  $\mathcal{F}_{m,t+k,T}^\infty := \sigma(\epsilon_{m,s,T}^2 : s \geq t+k)$ .

**Assumption 4.** For each  $m = 1, \dots, M$ ,  $\phi_{m,t,T}^2 := \epsilon_{m,t,T}^2 / g_{m,t,T}(\theta_m^*)$  is a non-degenerate random variable such that:

- i.  $E(\phi_{m,t,T}^2) = 1$  for all  $1 \leq t \leq T$ ,  $T \in \mathbb{N}$ ;
- ii.  $\sup_{1 \leq t \leq T, T \in \mathbb{N}} E|\phi_{m,t,T}^2|^{1+\delta} < \infty$  for some  $\delta > 0$ .

**Assumption 5.** For each  $m = 1, \dots, M$ ,  $L_m(\theta_m) := \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(l_{m,t,T}(\theta_m, \epsilon_{m,t,T}^2))$  exists and attains a unique minimum at  $\theta_m^* \in \Theta_m$ .

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  $\{\phi_{m,t,T}^2\}$  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  $\phi_{m,t,T}^2$ , stationary or not, as long as 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 (2021). 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 Zakoian (2019, Ch. 3). Note that, in the definition of mixing size, the underlying mixing coefficients are defined across  $\sigma$ -fields generated by the  $\epsilon_{m,t,T}^2$ 's and *not* the  $\epsilon_{t,T}^2$ 's. Since the  $\sigma$ -fields generated by the former are contained in those 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 most conditional volatility models are stable by scaling (cf. Francq et al. 2018). For conditional volatility models that are not stable by scaling, the condition  $E(\phi_{m,t,T}^2) = 1$  may be restrictive. Note that Assumption 4(i) is compatible with  $\{\phi_{m,t,T}^2\}$  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  $E(\phi_{m,t,T}^2)$  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  $E(\epsilon_{m,t,T}^2) = g_{m,t,T}(\theta_m^*)$ . This facilitates interpretation. Assumption 4(ii) is a fairly mild moment

assumption. For example, it holds when  $\{\phi_{m,t,T}^2\}$  is governed by a stationary GARCH(1,1), as in Equation (2), with finite  $E(\phi_{m,t,T}^4)$ . Finally, Assumption 5 ensures the existence of a minimising  $\theta_m^*$  in the asymptotic analogue of the minimisation problem in Equation (6). These assumptions are sufficient for consistency of the  $\hat{\theta}_m$  estimators as defined in Equation (6).

**Theorem 1 (Consistency).** If Assumptions 1 to 5 hold, then  $\hat{\theta}_m \xrightarrow{P} \theta_m^*$  for each  $m = 1, \dots, M$ .

## 2.2 | Asymptotic Normality Equation-by-Equation

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

**Assumption 6.**  $\theta^* \in \text{int}(\Theta)$ .

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

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

- i.  $\theta_m \mapsto g_{m,t,T}(\theta_m)$  is twice continuously differentiable on a neighbourhood  $\mathcal{V}_m$  of  $\theta_m^*$  in  $\Theta_m$ .
- ii. On  $\mathcal{V}_m$ , define

$$S_{m,T}(\theta_m) := \frac{1}{T} \sum_{t=1}^T l_{m,t,T}(\theta_m, \epsilon_{m,t,T}^2),$$

and

$$\hat{A}_{m,T}(\theta_m) := \frac{1}{T} \sum_{t=1}^T \ddot{l}_{m,t,T}(\theta_m, \epsilon_{m,t,T}^2),$$

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

- iii. There are deterministic functions  $\varphi_{m,t,T} : \mathcal{V}_m \rightarrow \mathbb{R}$  and random variables  $v_{m,t,T}$  such that,

$$\|\ddot{l}_{m,t,T}(\theta_m, \epsilon_{m,t,T}^2)\| \leq \varphi_{m,t,T}(\theta_m) v_{m,t,T}, \quad \theta_m \in \mathcal{V}_m,$$

where

$$\sup_{\theta_m \in \mathcal{V}_m} \sup_{1 \leq t \leq T, T \in \mathbb{N}} \varphi_{m,t,T}(\theta_m) < \infty, \quad \sup_{1 \leq t \leq T, T \in \mathbb{N}} E v_{m,t,T}^2 < \infty.$$

- iv. There exist random variables  $\psi_{m,t,T}$  such that for  $\theta_m, \theta'_m \in \mathcal{V}_m$ ,

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

where

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

**Assumption 9.** For each  $m = 1, \dots, M$ , 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  $m = 1, \dots, M$ , as  $T \rightarrow \infty$

$$\mathbf{B}_{m,T} := \text{Var} \left( T^{-1/2} \sum_{t=1}^T \mathbf{l}_{m,t,T}(\boldsymbol{\theta}_m^*, \epsilon_{m,t,T}^2) \right) \rightarrow \mathbf{B}_m^*,$$

with  $\mathbf{B}_m^*$  positive definite.

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

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

where  $\mathcal{V}_m$  is as in Assumption 8.  $\mathbf{A}_m^* := \mathbf{A}_m(\boldsymbol{\theta}_m^*)$  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 in the Online Appendix). A sufficient condition for Assumption 8 (given Assumptions 2 and 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 5 in the Online Appendix. 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  $\mathbf{B}_{m,T} = O(1)$  (cf. Lemma 2); Assumption 10 further ensures that  $\mathbf{B}_{m,T}$  converges to a positive definite limit. Similarly Assumptions 2, 4, and 8 suffice that each  $\mathbf{A}_{m,T}(\boldsymbol{\theta}_m) = O(1)$  (cf. Lemma 3); existence of the limit is assumed in Assumption 11. Note that the limits in Assumptions 10 and 11 will not exist if  $g_{m,t,T}$  is non-stationary periodic, as is common in intraday data. However, as we illustrate in Section 5.3, in that case our multivariate results can be used in combination with the vector-of-seasons representation. Note also that even if the limits in Assumptions 10 and 11 exist, they may not be positive definite. So, an important role played by the assumptions is to ensure that positive definiteness holds.

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

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

**Corollary 1.** Suppose Assumptions 1 to 11 hold. Then  $\hat{\mathbf{A}}_{m,T}(\hat{\boldsymbol{\theta}}_m) \xrightarrow{P} \mathbf{A}_m^*$  for  $m = 1, \dots, M$ .

Proposition 6 in the Online Appendix demonstrates how to verify the Assumptions required by Theorem 2 in a specific class of

models (a piecewise constant specification with  $\{\phi_{m,t,T}^2\}$  covariance stationary). In particular, this Proposition applies when the DGP of  $\{\phi_{m,t,T}^2\}$  is a strictly stationary scaled GARCH process.

It is worth emphasising that the autocovariance structure of the  $\phi_{m,t,T}^2$ 's only affects the asymptotic variance in Theorem 2 via  $\mathbf{B}_m^*$ , not via  $\mathbf{A}_m^*$ . To see this, note that the term  $E(\mathbf{l}_{m,t,T}(\boldsymbol{\theta}, \epsilon_{m,t,T}^2))$  in Assumption 11 depends only on  $g_{m,t,T}$  and its first derivatives when evaluated at  $\boldsymbol{\theta}_m^*$ . Accordingly, the limit  $\mathbf{A}_m^*$  does not depend on the specific model that governs  $\phi_{m,t,T}^2$ , or on any other of the statistical properties of  $\{\phi_{m,t,T}^2\}$ . By contrast, it can be shown that each entry in  $\mathbf{B}_{m,T}$  in Assumption 10 is a sum of the autocovariances of  $\phi_{m,t,T}^2$ . If the autocovariances are all non-negative, as in the standard GARCH (cf. Section 2.5.1 in Francq and Zakoian 2019), the diagonal entries of the limit  $\mathbf{B}_m^*$  will be higher the stronger and more persistent the autocovariances of  $\phi_{m,t,T}^2$  are. Still, since the off-diagonals of  $\mathbf{B}_m^*$  and the entries in  $\mathbf{A}_m^{*-1}$  can be negative, it is not clear what the overall effect is on the asymptotic variance due to stronger and more persistent autocovariances. Under more specific assumptions, the asymptotic variance simplifies; see Proposition 6 in Appendix B for the case where  $g_{m,t,T}$  is piecewise constant and  $\{\phi_{m,t,T}^2\}$  is covariance stationary.

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

$$\begin{aligned} \hat{\mathbf{B}}_{m,T} &:= \sum_{j=-T}^T k_m(j/\kappa_{m,T}) \hat{\Gamma}_{m,T}(j), \\ \hat{\Gamma}_{m,T}(j) &:= \frac{1}{T} \sum_{t=1}^{T-j} \mathbf{l}_{m,t,T}(\hat{\boldsymbol{\theta}}_m, \epsilon_{m,t,T}^2) \mathbf{l}_{m,t,T}(\hat{\boldsymbol{\theta}}_m, \epsilon_{m,t+j,T}^2)' \quad (j \geq 0), \\ \hat{\Gamma}_{m,T}(j) &:= \hat{\Gamma}_{m,T}(-j)' \quad (j < 0). \end{aligned} \quad (7)$$

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  $\mathcal{K}$  of De Jong and Davidson (2000, p. 409), defined as:

$$\begin{aligned} \mathcal{K} &:= \left\{ k : \mathbb{R} \rightarrow [-1, 1] : k(0) = 1, k(x) = k(-x), \right. \\ &\quad \times \int |k(x)| dx < \infty, \int |\phi(\xi)| d\xi < \infty, \\ &\quad \left. k \text{ is continuous at 0 and at all but a finite number of points} \right\}, \end{aligned}$$

where  $\phi(\xi) := \frac{1}{2\pi} \int k(x) e^{i\xi x} dx$ .

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

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

Most kernels considered in the literature satisfy Assumption 12. This includes, amongst others, 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), demonstrating that—under our Assumptions— $\hat{\mathbf{B}}_{m,T}$  is consistent for  $\mathbf{B}_m^*$ .

**Proposition 1.** Suppose Assumptions 1 to 13 hold. Then  $\hat{\mathbf{B}}_{m,T} \xrightarrow{P} \mathbf{B}_m^*$  for  $m = 1, \dots, M$ .

### 2.3 | Joint Asymptotic Normality

We next establish the joint asymptotic normality of  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1', \dots, \hat{\boldsymbol{\theta}}_M')'$ . This is required for hypothesis tests on coefficients across equations. Let  $\boldsymbol{\epsilon}_{i,T}^2 := (\epsilon_{1,i,T}^2, \dots, \epsilon_{M,i,T}^2)'$ . We can re-write the objective function in Equation (5) as

$$L_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{i=1}^T l_{i,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}_{i,T}^2),$$

$$l_{i,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}_{i,T}^2) := \sum_{m=1}^M l_{m,i,T}(\boldsymbol{\theta}_m, \epsilon_{m,i,T}^2). \quad (8)$$

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

$$\dot{l}_{i,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}_{i,T}^2) = (\dot{l}_{1,i,T}(\boldsymbol{\theta}_1, \epsilon_{1,i,T}^2)', \dots, \dot{l}_{M,i,T}(\boldsymbol{\theta}_M, \epsilon_{M,i,T}^2))',$$

$$\ddot{l}_{i,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}_{i,T}^2) = \text{diag}(\ddot{l}_{1,i,T}(\boldsymbol{\theta}_1, \epsilon_{1,i,T}^2), \dots, \ddot{l}_{M,i,T}(\boldsymbol{\theta}_M, \epsilon_{M,i,T}^2)).$$

Note that under Assumption 11,

$$\mathbf{A}_T(\boldsymbol{\theta}) := \frac{1}{T} \sum_{i=1}^T E[\ddot{l}_{i,T}(\boldsymbol{\theta}, \boldsymbol{\epsilon}_{i,T}^2)] \rightarrow \mathbf{A}(\boldsymbol{\theta})$$

with  $\mathbf{A}(\boldsymbol{\theta}) := \text{diag}(\mathbf{A}_1(\boldsymbol{\theta}_1), \dots, \mathbf{A}_M(\boldsymbol{\theta}_M))$ ,  $\boldsymbol{\theta} \in \mathcal{V}$ , (9)

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

$$\hat{\mathbf{A}}_T(\boldsymbol{\theta}) := \text{diag}(\hat{\mathbf{A}}_{1,T}(\boldsymbol{\theta}_1), \dots, \hat{\mathbf{A}}_{M,T}(\boldsymbol{\theta}_M)). \quad (10)$$

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 of  $\{\epsilon_{i,T}^2 : i \in \mathbb{Z}, T \in \mathbb{N}\}$ , i.e.,  $\alpha_T(k) := \sup_{i \in \mathbb{Z}} \sup\{|P(F \cap G) - P(F)P(G)| : F \in \mathcal{F}_{-\infty,T}^i, G \in \mathcal{F}_{i+k,T}^\infty\}$  with  $\mathcal{F}_{-\infty,T}^i := \sigma(\epsilon_{s,T}^2 : s \leq i)$  and  $\mathcal{F}_{i+k,T}^\infty := \sigma(\epsilon_{s,T}^2 : s \geq i+k)$ , then

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

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

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

$$\mathbf{B}_T := \text{Var}\left(T^{-1/2} \sum_{i=1}^T \dot{l}_{i,T}(\boldsymbol{\theta}^*, \boldsymbol{\epsilon}_{i,T}^2)\right) \rightarrow \mathbf{B}^*,$$

with  $\mathbf{B}^*$  positive definite.

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

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

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

$$\hat{\mathbf{B}}_T := \sum_{j=-T}^T k(j/\kappa_T) \hat{\boldsymbol{\Gamma}}_T(j),$$

$$\hat{\boldsymbol{\Gamma}}_T(j) := \frac{1}{T} \sum_{i=1}^{T-j} \dot{l}_{i,T}(\hat{\boldsymbol{\theta}}, \boldsymbol{\epsilon}_{i,T}^2) \dot{l}_{i+j,T}(\hat{\boldsymbol{\theta}}, \boldsymbol{\epsilon}_{i+j,T}^2)' \quad (j \geq 0),$$

$$\hat{\boldsymbol{\Gamma}}_T(j) := \hat{\boldsymbol{\Gamma}}_T(-j)' \quad (j < 0). \quad (11)$$

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

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

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

**Proposition 2.** Suppose Assumptions 1 to 8, 11, and 14 to 17 hold. Then  $\hat{\mathbf{B}}_T \xrightarrow{P} \mathbf{B}^*$ .

### 3 | Examples of $g_{i,T}$

Here we provide examples of  $g_{m,i,T}(\boldsymbol{\theta}_m)$  and derive verifiable conditions that ensure the high-level Assumptions 5 and 8 hold. Our choice of examples comprises the most common classes of parametric versions of  $g_{m,i,T}$ . In Section 3.1, we establish conditions for the piecewise constant class of models. Specifications within this class have proved particularly useful in testing whether a break occurred at a known time-point due to, say, a policy decision, and in quantifying that break. The piecewise constant model is especially popular among practitioners due to its ease of estimation and interpretability. In Section 3.2, we establish conditions for the most common variant of smooth transition models. In contrast to the piecewise constant model, smooth transition models allow breaks to be gradual or “smooth”. Moreover, in addition to an estimate of the (total) break-size, the smooth transition model also provides estimates of the locations of the breakpoints, and estimates of the speed of transitions. Sums of smooth transition terms are thus capable of providing a detailed and interpretable characterisation of virtually any dynamics of  $g_{m,i,T}$ . The smooth transition model can also be viewed as a generalisation of the piecewise constant model, since the latter is obtained as a special case (in the limit) when the speed of transition parameters tends to infinity. Finally, in Section 3.3, we establish conditions for a class of splines that have proved particularly useful and flexible in the econometric literature.

### 3.1 | Piecewise Constant Models

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

$$g_{m,t,T}(\theta_m) = \delta_{m,0} + \sum_{l=1}^{s_m} \delta_{m,l} \cdot I(t/T \geq c_{m,l}),$$

$$\theta_m = (\delta_{m,0}, \delta_{m,1}, \dots, \delta_{m,s_m})', \quad (12)$$

where  $I(A)$  is an indicator function equal to 1 if  $A$  holds and 0 otherwise. The values of the possible break-locations  $c_{m,1}, \dots, c_{m,s_m}$  are thus known and not estimated. To estimate  $\theta_m$ , 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.

The log-linear version of a piecewise constant specification (cf. Escribano and Sucarrat 2018) is given by

$$\ln g_{m,t,T}(\theta_m) = \delta_{m,0} + \sum_{l=1}^{s_m} \delta_{m,l} \cdot I(t/T \geq c_{m,l}),$$

$$\theta_m = (\delta_{m,0}, \delta_{m,1}, \dots, \delta_{m,s_m})'. \quad (13)$$

Notice that (12) can always be re-written as (13). The advantage of this is that non-negativity constraints on  $\theta_m$  are not needed in Equation (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 3.** Suppose  $g_{m,t,T}(\theta_m)$  is given by (13) with  $c_{m,0} < c_{m,1} < \dots < c_{m,s_m} < c_{m,s_m+1}$ , where  $c_{m,0} = 0$  and  $c_{m,s_m+1} = 1$ . Suppose further that Assumption 1 holds, that  $\theta_m^* \in \Theta_m$ , that  $\Theta_m^*$  is an open, bounded and convex set that contains  $\Theta_m$ , that Assumption 4 holds, and that  $\mathcal{V}_m$  in Assumption 8 is contained in  $\Theta_m$ . Then Assumptions 2 and 8 hold, and the limit  $L_m(\theta_m)$  in Assumption 5 exists and attains a unique minimum at  $\theta_m^* \in \Theta_m^*$ .

### 3.2 | 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:

$$g_{m,t,T}(\theta_m) = \delta_{m,0} + \sum_{l=1}^{s_m} \frac{\delta_{m,l}}{1 + \exp(-\gamma_{m,l}(t/T - c_{m,l}))}, \quad (14)$$

where  $\theta_m = (\delta_m', \gamma_m', c_m')'$  with  $\delta_m = (\delta_{m,0}, \delta_{m,1}, \dots, \delta_{m,s_m})'$ ,  $\gamma_m = (\gamma_{m,1}, \dots, \gamma_{m,s_m})'$  and  $c_m = (c_{m,1}, \dots, c_{m,s_m})'$ . For  $l = 1, \dots, s_m$ , 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  $\delta_{m,1} = \dots = \delta_{m,s_m} = 0$ . 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 4.** Suppose  $g_{m,t,T}(\theta_m)$  is given by (14) with  $\delta_{m,1} \neq 0, \dots, \delta_{m,s_m} \neq 0$ , with  $\gamma_{m,1} \neq 0, \dots, \gamma_{m,s_m} \neq 0$ , and with

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

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

### 3.3 | Splines

Engle and Rangel (2008), and Brownlees and Gallo (2010), specify  $g_{m,t,T}$  as a deterministic spline. The former uses 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

$$\ln g_{m,t,T}(\theta_m) = \delta_{m,0} + \sum_{l=1}^{s_m} \delta_{m,l} (t/T - c_{m,l})^2 I(t/T \geq c_{m,l}),$$

$$\theta_m = (\delta_{m,0}, \delta_{m,1}, \dots, \delta_{m,s_m})', \quad (15)$$

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  $\delta_{m,1}, \dots, \delta_{m,s_m}$  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  $g_{m,t,T}(\theta)$  is given by (15) with  $c_0 < c_1 < \dots < c_s < c_{s+1}$ , where  $c_0 = 0$  and  $c_{s+1} = 1$ . Suppose further that Assumption 1 holds, that  $\theta_m^* \in \Theta_m$ , that  $\Theta_m^*$  is an open, bounded and convex set that contains  $\Theta_m$ , that Assumption 4 holds, and that  $\mathcal{V}_m$  in Assumption 8 is contained in  $\Theta_m$ . Then Assumptions 2 and 8 hold, and the limit  $L_m(\theta_m)$  in Assumption 5 exists. Moreover, if the Hessian  $\ddot{L}_m(\theta_m)$  is positive definite on  $\Theta_m^*$ ,  $L_m(\theta_m)$  attains a unique minimum at  $\theta_m^* \in \Theta_m^*$ .

## 4 | Estimation of Conditional Volatility

In empirical applications, it is often of interest to obtain estimates of the full conditional covariance matrix  $E(\epsilon_t \epsilon_t' | \mathcal{F}_{t-1})$ , where  $\mathcal{F}_{t-1} = \sigma\{\epsilon_u, u < t\}$ . 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  $E(\epsilon_t | \mathcal{F}_{t-1}) = \mathbf{0}$ , 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. In this case, the conditional

covariance matrix can be written as

$$E(\epsilon_t \epsilon_t' | \mathcal{F}_{t-1}) = \mathbf{G}_t^{1/2} \mathbf{H}_t^{1/2} \mathbf{R}_t \mathbf{H}_t^{1/2} \mathbf{G}_t^{1/2}, \quad (16)$$

where  $\mathbf{G}_t^{1/2} := \text{diag}(g_{1,t}^{1/2}, \dots, g_{M,t}^{1/2})$ ,  $\mathbf{H}_t^{1/2} := \text{diag}(h_{1,t}^{1/2}, \dots, h_{M,t}^{1/2})$  and  $\mathbf{R}_t$  is a conditional correlation matrix that is either constant or time-varying. Both  $\mathbf{G}_t^{1/2}$  and  $\mathbf{H}_t^{1/2}$  are positive definite if their diagonal elements are strictly positive, and  $E(\phi_t \phi_t' | \mathcal{F}_{t-1}) = \mathbf{H}_t^{1/2} \mathbf{R}_t \mathbf{H}_t^{1/2}$ . As a consequence, if  $\mathbf{R}_t$  is also positive definite, then  $E(\phi_t \phi_t' | \mathcal{F}_{t-1})$  and (16) are also positive definite.<sup>5</sup>

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 suitable assumptions, here we study the second step estimation of the scaled GARCH(1,1) specification under both correct and incorrect specifications. In other words,  $g_{m,t} h_{m,t}$  constitutes a prediction of  $\sigma_{m,t}^2 = g_{m,t} E(\phi_{m,t}^2 | \mathcal{F}_{t-1})$ , where  $h_{m,t}$  is either correctly or incorrectly specified for  $E(\phi_{m,t}^2 | \mathcal{F}_{t-1})$ . We also outline how Dynamic Conditional Correlations (DCCs) can be estimated in a third step while ensuring positive definiteness of (16).

#### 4.1 | Estimation of a Scaled GARCH(1,1) Model

In the first part of this subsection, the properties of  $\{\phi_{m,t,T}^2\}$  do not vary with  $T$ . To emphasise this, we omit the subscript  $T$ . Suppose  $\phi_{m,t}^2 := \epsilon_t^2 / g_{m,t}(\theta_m^*)$  is governed by a strictly stationary scaled GARCH(1,1), where  $\theta_m^*$  is the true parameter. This means

$$\phi_{m,t}^2 = h_{m,t} \eta_{m,t}^2, \quad E(\eta_{m,t}^2 | \mathcal{F}_{m,t-1}) = 1, \quad (17)$$

$$h_{m,t} = \omega_m^* + \alpha_m^* \phi_{m,t-1}^2 + \beta_m^* h_{m,t-1}, \quad \omega_m^*, \alpha_m^*, \beta_m^* > 0, \quad \omega_m^* = 1 - \alpha_m^* - \beta_m^*, \quad (18)$$

for all  $T$ , where  $\mathcal{F}_{m,t-1} = \sigma\{\phi_{m,u}^2, u < t\}$ . It is the condition  $\omega_m^* = (1 - \alpha_m^* - \beta_m^*)$  in Equation (18) which makes the GARCH(1,1) a scaled version, i.e.,  $E(\phi_{m,t}^2) = 1$  for all  $t$ . An implication of the condition is that only two parameters need to be estimated in the second step, namely  $\theta_m^* := (\alpha_m^*, \beta_m^*)'$ . In the infeasible case,  $\{\phi_{m,t}^2\}$  is observed and the consistency of the standard GARCH 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  $\theta_m^* = (\alpha_m^*, \beta_m^*)'$  is

$$\hat{\theta}_{m,T} = \arg \min_{\theta_m \in \Xi_m} \frac{1}{T} \sum_{t=1}^T \ln \hat{h}_{m,t} + \frac{\hat{\phi}_{m,t}^2}{\hat{h}_{m,t}},$$

where  $\theta_m = (\alpha_m, \beta_m)'$ ,  $\hat{\phi}_{m,t}^2 = \frac{\epsilon_{m,t}^2}{\hat{g}_{m,t}}$  and  $\hat{h}_{m,t} = (1 - \alpha_m - \beta_m) + \alpha_m \hat{\phi}_{m,t-1}^2 + \beta_m \hat{h}_{m,t-1}$ . Appendix D.2 contains the simulation results of the Two-step QMLE. Based on the results, a reasonable conjecture is that the asymptotic properties of the Two-step

QMLE are the same as those of the ordinary QMLE, both when  $g_{m,t}$  is constant and when it is time-varying, in the experiments investigated.

We now turn to the case where  $\{\phi_{m,t,T}^2\}$  is not governed by a GARCH, where it is not necessarily strictly stationary, and where the properties of  $\{\phi_{m,t,T}^2\}$  can vary with  $T$ . In this case, a scaled GARCH(1,1) specification provides mis-specified predictions of volatility. However, 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  $\{\phi_{m,t,T}^2\}$  is not governed by a GARCH(1,1), and let

$$h_{m,t,T} = \omega_m + \alpha_m \phi_{m,t-1,T}^2 + \beta_m h_{m,t-1,T}, \quad \omega_m, \alpha_m, \beta_m > 0, \quad \omega_m = 1 - \alpha_m - \beta_m, \quad (19)$$

denote the scaled GARCH(1,1) prediction, where  $\alpha_m$  and  $\beta_m$  are real-valued parameters that satisfy the constraints in Equation (19). It is straightforward to verify by backwards recursion that, for any pair  $(\alpha_m, \beta_m)$  that satisfies the parameter constraints in Equation (19),

$$E(h_{m,t,T}) = \frac{\omega_m}{1 - \beta_m} + \alpha_m \sum_{i=1}^{\infty} \beta_m^{i-1} E(\phi_{m,t-i,T}^2) = 1 \quad \text{for all } t \text{ and } T.$$

Accordingly,  $g_{m,t,T} h_{m,t,T}$  is unbiased for  $E(\epsilon_{m,t,T}^2)$  in the unconditional sense, since  $E(g_{m,t,T} h_{m,t,T}) = g_{m,t,T} E(h_{m,t,T}) = E(\epsilon_{m,t,T}^2)$  for all  $t$  and  $T$ . We emphasise that  $E(h_{m,t,T}) = 1$  holds under certain types of non-stationarities of  $\{\phi_{m,t,T}^2\}$ , 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  $\lim_{S \rightarrow \infty} E(h_{m,t+S,T}) = E(h_{m,t,T}) = 1$  for all  $t$  and  $T$ , just as if  $h_{m,t,T}$  were the true DGP. Finally, a third desirable property the scaled GARCH(1,1) predictions possess under suitable regularity conditions when  $(\alpha_m, \beta_m)$  are estimated by QML, is weak identification in the sense of Sucarrat (2021), i.e.,  $E(\phi_{m,t,T}^2 / h_{m,t,T}) = 1$ , see exercise 7.6 in Francq and Zakoian (2019). In other words, under mis-specification, QML estimation under suitable assumptions ensures that a necessary condition for weak identification holds.

Finally, it is worth mentioning that the moment estimators of Kristensen and Linton (2006) can also be used to estimate the parameters of a scaled GARCH(1,1) specification, both under correct and incorrect specifications. The details of this are contained in Section D.3 of the Online Appendix.

#### 4.2 | Estimation of Conditional Correlations

Let  $\eta_{m,t} = \epsilon_{m,t} / \sqrt{g_{m,t} h_{m,t}}$ ,  $m = 1, \dots, M$ , and let  $\eta_t = (\eta_{1,t}, \dots, \eta_{M,t})'$ . Accordingly,  $\epsilon_t = \mathbf{G}_t^{1/2} \mathbf{H}_t^{1/2} \eta_t$  and  $E(\eta_t \eta_t' | \mathcal{F}_{t-1}) = \mathbf{R}_t$ . Note also that  $\text{Corr}(\epsilon_t | \mathcal{F}_{t-1}) = \mathbf{R}_t$  under the assumption that  $E(\epsilon_t | \mathcal{F}_{t-1}) = \mathbf{0}$  for all  $t$ . In applications, an estimator of

$R_t$  can be built with the standardised residuals  $\hat{\eta}_{m,t}$ , where  $\hat{\eta}_{m,t} = \epsilon_{m,t} / \sqrt{\hat{g}_{m,t} \hat{h}_{m,t}}$ . If  $R_t$  is constant over time, e.g., the natural estimator is the sample estimator  $\hat{R} = T^{-1} \sum_{t=1}^T \hat{\eta}_t \hat{\eta}_t'$ , where  $\hat{\eta}_{m,t} = \epsilon_{m,t} / \sqrt{\hat{g}_{m,t} \hat{h}_{m,t}}$ ,  $m = 1, \dots, M$ . If  $R_t$  is time-varying, then a natural candidate is the Dynamic Conditional Correlations (DCCs) specification of Engle (2002) or Aielli (2013). Under mis-specification,  $R_t$  must be interpreted as a prediction that is not necessarily equal to  $E(\eta_t \eta_t' | \mathcal{F}_{t-1})$ . Finally, note that our framework is compatible with time-varying unconditional correlations  $E(\eta_t \eta_t')$ .

## 5 | Numerical Illustrations

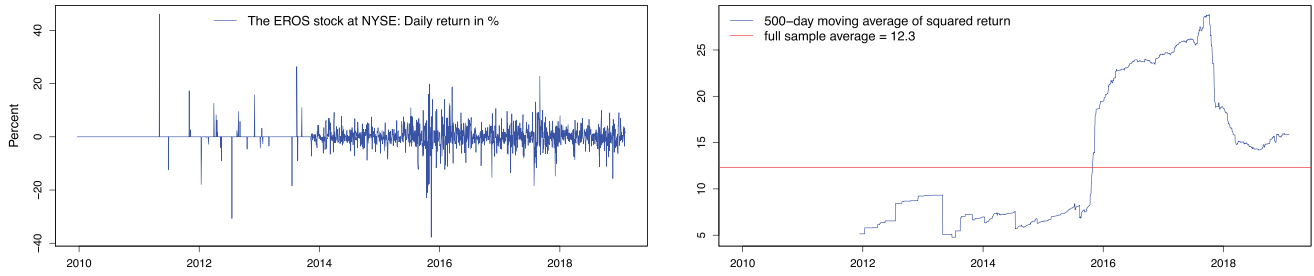
### 5.1 | An Efficiency Comparison

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

**TABLE 1** | Comparison of the Two-step QMLE and the Iterative QMLE of Amado and Teräsvirta (2013), see Section 5.1.

	$T$	$m(\hat{\delta}_0)$	$se(\hat{\delta}_0)$	$m(\hat{\delta}_1)$	$se(\hat{\delta}_1)$	$m(\hat{\gamma})$	$se(\hat{\gamma})$	$m(\hat{c})$	$se(\hat{c})$
$g_t$ :	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
	10,000	−0.0050	0.0569	0.0269	0.1961	0.6592	2.9829	0.0031	0.0339
	20,000	−0.0039	0.0386	0.0170	0.1260	0.2612	1.8592	0.0011	0.0237
	40,000	−0.0032	0.0262	0.0156	0.0898	0.0720	1.3271	0.0013	0.0167
	80,000	−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
	10,000	−0.0088	0.0718	0.0248	0.2045	0.4501	2.6239	0.0030	0.0295
	20,000	−0.0055	0.0490	0.0127	0.1459	0.1787	1.6676	0.0015	0.0223
	40,000	−0.0037	0.0306	0.0106	0.0863	0.0476	1.1624	0.0010	0.0144
	80,000	−0.0018	0.0171	0.0044	0.0510	0.0177	0.7819	0.0003	0.0097
	$T$	$m(\hat{\omega})$	$se(\hat{\omega})$	$m(\hat{\alpha})$	$se(\hat{\alpha})$	$m(\hat{\beta})$	$se(\hat{\beta})$		
$h_t$ :	Two-step QMLE (step 2)								
	2000	—	—	−0.0006	0.0197	−0.0180	0.0525		
	5000	—	—	−0.0003	0.0121	−0.0057	0.0271		
	10,000	—	—	0.0000	0.0087	−0.0031	0.0192		
	20,000	—	—	0.0000	0.0059	−0.0017	0.0135		
	40,000	—	—	0.0001	0.0042	−0.0009	0.0092		
	80,000	—	—	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		
	10,000	0.0037	0.0144	0.0001	0.0087	−0.0034	0.0192		
	20,000	0.0022	0.0103	0.0001	0.0059	−0.0019	0.0135		
	40,000	0.0011	0.0069	0.0001	0.0042	−0.0010	0.0092		
	80,000	0.0006	0.0049	0.0001	0.0032	−0.0006	0.0068		

Note: DGP:  $\epsilon_t = \sqrt{g_t} \phi_t$ ,  $\phi_t = \sqrt{h_t} \eta_t$ ,  $\eta_t \stackrel{iid}{\sim} N(0, 1)$ ,  $h_t = 0.1 + 0.1 \phi_{t-1}^2 + 0.8 h_{t-1}$ ,  $g_t = 0.5 + 1.5(1 + \exp(-10(t/T - 0.5)))^{-1}$ .  $T$ , sample size.  $m(\hat{x})$ , average bias of estimate  $\hat{x}$  across replications (no. of replications = 1000).  $se(\hat{x})$ , sample standard deviation of estimate  $\hat{x}$  across replications. All computations in R (R Core Team 2021). The Two-step QMLE is implemented with our own code. The Iterative QMLE is implemented with the `tvqarch()` function of the CRAN package `tvqarch` (Campos-Martins and Sucarrat 2024).



**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.

the simulation results from a comparison, where the parameter values of the DGP correspond to what is commonly found empirically. 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  $T \leq 10,000$ . As the sample size grows very large, however, the results suggest the Iterative QMLE is slightly more efficient. 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  $\phi_t^2$  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. At 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 a 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  $\epsilon_t$  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

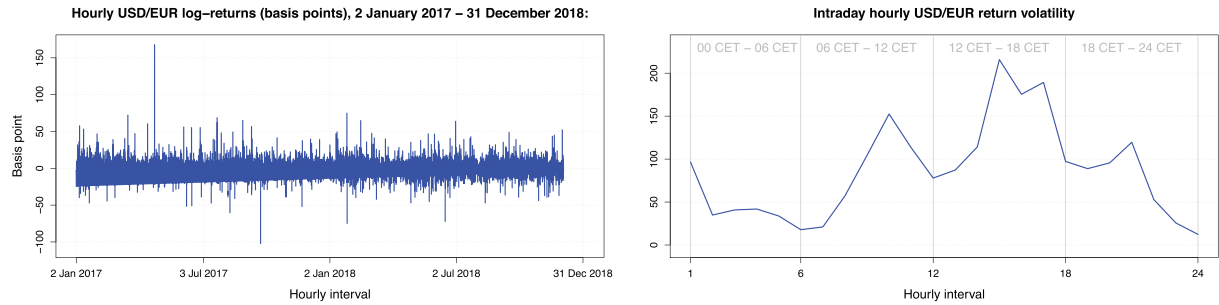
these correspond to  $(c_1, c_2, c_3)' = (0.427, 0.638, 0.855)$ . Our estimated model is

$$\begin{aligned} \ln \hat{g}_t = & 1.795 + \frac{0.351}{(0.4174)} I(t/T \geq c_1) + \frac{1.215}{(0.2342)} I(t/T \geq c_2) \\ & - \frac{0.912}{(0.2545)} I(t/T \geq c_3). \end{aligned}$$

The numbers in parentheses are the standard errors of the estimates. These are computed as the square root of the diagonal of  $\hat{\Sigma}/T$ , where  $\hat{\Sigma} = \hat{A}^{-1} \hat{B} \hat{A}^{-1}$  is the estimate of the asymptotic coefficient covariance. A Bartlett kernel is used in the computation of  $\hat{B}$ , 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  $\hat{h}_t = 0.873 + 0.127\hat{\phi}_{t-1}^2 + 0.000\hat{h}_{t-1}$  with  $T^{-1} \sum_{t=1}^T \hat{\phi}_t^2 / \hat{h}_t = 1.044$ . 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 equal to 0.127, and a GARCH parameter close to zero. Second step estimation with the moment method (see the Online Appendix) gives estimates that violate the parameter conditions, and the value  $T^{-1} \sum_{t=1}^T \hat{\phi}_t^2 / \hat{h}_t$  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  $g_t(\theta)$  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. (2024), establish asymptotic results for two classes of periodic smooth transition models. But they do so under restrictive assumptions on  $\phi_t$  (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).



**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 2** | Spline estimates of intraday hourly volatility (see Section 5.3).

$m$	$\hat{\delta}_{m,0}$ (s.e.)	$\hat{\delta}_{m,1}$ (s.e.)	$\hat{\delta}_{m,2}$ (s.e.)	$\hat{\delta}_{m,3}$ (s.e.)	$\hat{\delta}_{m,4}$ (s.e.)	$T_m$	$\chi^2(4)$ [p]
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]

Note: The estimated model is  $\ln g_{m,t} = \delta_{m,0} + \sum_{j=1}^4 \delta_{m,j} (t/T - c_j)^2 I(t/T \geq c_j)$  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  $H_0 : \delta_{m,1} = \dots = \delta_{m,4} = 0$  ( $p$ -value in square brackets).

The common practice of estimating the intraday unconditional volatilities with cross-day averages of squared returns is a special case of period-by-period estimation via the vector of seasons representation. Consider, e.g., the intraday returns  $\epsilon_{m,t}$ ,  $m = 1, \dots, M$ , of day  $t$ . Often, the sample averages  $T^{-1} \sum_{t=1}^T \epsilon_{m,t}^2$ ,  $m = 1, \dots, M$ , are used to estimate the intraday unconditional volatilities  $E(\epsilon_{1,t}^2), \dots, E(\epsilon_{M,t}^2)$ . 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  $m = 1, \dots, M$  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 volatilities 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.,

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

$$(c_1, c_2, c_3, c_4) = (0.2, 0.4, 0.6, 0.8),$$

for each period  $m = 1, \dots, M$ . Table 2 contains the estimation results together with a Wald-test of  $H_0 : \delta_{m,1} = \dots = \delta_{m,4} = 0$ . 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 volatilities 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  $\phi_t^2$ ’s are non-stationary. The QML estimated

specification is  $\hat{h}_{m,t} = 0.106 + 0.052 \hat{\phi}_{m-1,t}^2 + 0.8418 \hat{h}_{m-1,t}$  with  $(T \cdot M)^{-1} \sum_{t=1}^T \sum_{m=1}^M \hat{\phi}_{m,t}^2 / \hat{h}_{m,t} = 1.0004$ . Second step estimation with the moment method (see the Online Appendix), by contrast, returns estimates that violate the parameter conditions, and the value  $(T \cdot M)^{-1} \sum_{t=1}^T \sum_{m=1}^M \hat{\phi}_{m,t}^2 / \hat{h}_{m,t}$  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 general and robust estimator of a large class of parametric models of time-varying unconditional volatility, both univariate and multivariate, and establish its consistency and asymptotic normality. Our estimator 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, within our framework, 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.

## Acknowledgments

We are grateful to the Associate Editor, two anonymous referees, Steffen Grønneberg, Fabian Harang, Dennis Kristensen, Mika Meitz, Ovidijus Stauskas 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. Rickard Sandberg acknowledges financial support from Jan Walland’s and Tom Hedelius’ Foundation, grant no. P22-0264. The data used in this paper is available on request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Data Availability Statement

The data of this study are available on request.

## Endnotes

<sup>1</sup> See the assumption that  $h_t = 1$  for all  $t$  in Theorem 1 of (Amado and Teräsvirta 2013, 145), and Theorem 3 in Silvennoinen and Teräsvirta (2024).

<sup>2</sup> See Amado and Teräsvirta (2013, Theorem 2), and Silvennoinen and Teräsvirta (2024, Theorem 6).

<sup>3</sup> 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  $g_{i,T}$ 's and how this may affect invertibility (i.e., the asymptotic irrelevance of the initial values of the GARCH recursion at the true parameter value), and due to the estimation error in the second step estimation of the parameters of  $\phi_t^2$  (cf. Francq and Zakoian 2019, 190, Francq et al. 2011, and Francq et al. 2016).

<sup>4</sup> Our dependence assumption could be relaxed further at the cost of increased conceptual complexity. Overall, we believe that the conditions we impose strike a good balance in this regard: They are relatively simple whilst being sufficiently weak for the theory to apply to many cases of practical interest.

<sup>5</sup> If two square matrices of the same size  $A$  and  $B$  are positive definite, then also  $ABA$  is positive definite.

## References

- Aielli, G. P. 2013. "Dynamic Conditional Correlations: On Properties and Estimation." *Journal of Business & Economic Statistics* 31: 282–299.
- Amado, C., and T. Teräsvirta. 2013. "Modelling Volatility by Variance Decomposition." *Journal of Econometrics* 175: 142–153.
- Amado, C., and T. Teräsvirta. 2014. "Modelling Changes in the Unconditional Variance of Long Stock Return Series." *Journal of Empirical Finance* 25: 15–35.
- Amado, C., and T. Teräsvirta. 2017. "Specification and Testing of Multiplicative Time-Varying GARCH Models With Applications." *Econometric Reviews* 36: 421–446.
- Andersen, T. G., and T. Bollerslev. 1997. "Intraday Periodicity and Volatility Persistence in Financial Markets." *Journal of Empirical Finance* 4: 115–158.
- Andreou, E., and E. Ghysels. 2008. "Quality Control for Structural Credit Risk Models." *Journal of Econometrics* 146: 364–375.
- Bates, D. S. 2000. "Post-'87 Crash Fears in the S& P 500 Futures Option Market." *Journal of Econometrics* 94: 181–238.
- Bollerslev, T. 1986. "Generalized Autoregressive Conditional Heteroscedasticity." *Journal of Econometrics* 31: 307–327.
- Brownlees, C., and G. Gallo. 2010. "Comparison of Volatility Measures: A Risk Management Perspective." *Journal of Financial Econometrics* 8: 29–56.
- Campos-Martins, S., and G. Sucarrat. 2024. "Modeling Nonstationary Financial Volatility With the R Package Tvargarch." *Journal of Statistical Software* 108: 1–38.
- Carrasco, M., and X. Chen. 2002. "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models." *Econometric Theory* 18: 17–39.
- Davis, R. A., and T. Mikosch. 2009. "Probabilistic Properties of Stochastic Volatility Models." In *Handbook of Financial Time Series*, edited by T. Mikosch, T. Kreiss, J.-P. Davis, R. Andersen, and T. Gustav, 255–268. Springer.
- De Jong, R. M., and J. Davidson. 2000. "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices." *Econometrica* 68, no. 2: 407–423.
- Engle, R. 2002. "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models." *Journal of Business & Economic Statistics* 20: 339–350.
- 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 Modelling* 25: 850–867.
- 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.
- Francq, C., O. Wintenberger, and J.-M. Zakoian. 2018. "Goodness-Of-Fit Tests for Log and Exponential GARCH Models." *Test* 27: 27–51.
- Francq, C., and J.-M. Zakoian. 2016. "Estimating Multivariate Volatility Models Equation by Equation." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 78: 613–635.
- Francq, C., and J.-M. Zakoian. 2019. *GARCH Models*. 2nd ed. Wiley.
- Hafner, C., and O. Linton. 2010. "Efficient Estimation of a Multivariate Multiplicative Volatility Model." *Journal of Econometrics* 159: 55–73.
- He, C., J. Kang, A. Silvennoinen, and T. Teräsvirta. 2019. "The Shifting Seasonal Mean Autoregressive Model and Seasonality in the Central England Monthly Temperature, 1772-2016." *Econometrics and Statistics* 12: 1–24.
- He, C., J. Kang, A. Silvennoinen, and T. Teräsvirta. 2024. "Long Monthly Temperature Series and the Vector Seasonal Shifting Mean and Covariance Autoregressive Model." *Journal of Econometrics* 239, no. 1: 105494.
- Hillebrand, E. 2005. "Neglecting Parameter Changes in GARCH Models." *Journal of Econometrics* 129: 121–138.
- Jiang, F., D. Li, and K. Zhu. 2021. "Adaptive Inference for a Semiparametric Generalized Autoregressive Conditional Heteroskedasticity Model." *Journal of Econometrics* 224: 306–329.
- Kim, K. H., and T. Kim. 2016. "Capital Asset Pricing Model: A Time-Varying Volatility Approach." *Journal of Empirical Finance* 37: 268–281.

- Kolokolov, A., G. Livieri, and D. Pirino. 2020. "Statistical Inferences for Price Staleness." *Journal of Econometrics* 218: 32–81.
- Kolokolov, A., and R. Reno. 2024. "Jumps or Staleness?" *Journal of Business and Economic Statistics* 42: 516–532.
- 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, Structural Breaks, and the GARCH Model." *Journal of Business & Economic Statistics* 8: 225–234.
- Lindner, A. 2009. "Stationarity, Mixing, Distributional Properties and Moments of  $GARCH(p < /i >, q < /i >)$ -processes." In *Handbook of Financial Time Series*, edited by T. Mikosch, T. Kreiss, J.-P. Davis, R. Andersen, and T. Gustav, 43–84. 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." *Review of Economics and Statistics* 86: 378–390.
- Pastor, L., and R. F. Stambaugh. 2001. "The Equity Premium and Structural Breaks." *Journal of Finance* 56: 1207–1239.
- Patilea, V., and H. Raïssi. 2024. "Powers Correlation Analysis of Returns With a Non-Stationary Zero-Process." *Journal of Financial Econometrics* 22: 1345–1371.
- 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*. R Foundation for Statistical Computing.
- Silvennoinen, A., and T. Teräsvirta. 2024. "Consistency and Asymptotic Normality of Maximum Likelihood Estimators of a Multiplicative Time-Varying Smooth Transition Correlation GARCH Model." *Econometrics and Statistics* 32: 57–72.
- 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. 2025. "Testing the Zero-Process of Intraday Financial Returns for Non-Stationary Periodicity." *Journal of Financial Econometrics* 23: nbaf013.
- Sucarrat, G. 2021. "Identification of Volatility Proxies as Expectation of Squared Financial Return." *International Journal of Forecasting* 37: 1677–1690.
- Sucarrat, G., and S. Grønneberg. 2022. "Risk Estimation With a Time Varying Probability of Zero Returns." *Journal of Financial Econometrics* 20: 278–309.
- Van Bellegem, S., and R. Von Sachs. 2004. "Forecasting Economic Time-Series With Unconditional Time-Varying Variance." *International Journal of Forecasting* 20: 611–627.
- 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.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** jtsa70034-sup-0001-Supinfo.pdf.