JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **41**: 551–570 (2020)
Published online 04 February 2020 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12522

ORIGINAL ARTICLE

TWO-STEP ESTIMATION FOR TIME VARYING ARCH MODELS

YUANYUAN ZHANG,[a] RONG LIU,[b] QIN SHAO[b] AND LIJIAN YANG[a*]

[a] *Center for Statistical Science & Department of Industrial Engineering, Tsinghua University, Beijing, China*
[b] *Department of Mathematics and Statistics, University of Toledo, Toledo, OH, USA*

A time varying autoregressive conditional heteroskedasticity (ARCH) model is proposed to describe the changing volatility of a financial return series over long time horizon, along with two-step least squares and maximum likelihood estimation procedures. After preliminary estimation of the time varying trend in volatility scale, approximations to the latent stationary ARCH series are obtained, which are used to compute the least squares estimator (LSE) and maximum likelihood estimator (MLE) of the ARCH coefficients. Under elementary and mild assumptions, oracle efficiency of the two-step LSE for ARCH coefficients is established, that is, the two-step LSE is asymptotically as efficient as the infeasible LSE based on the unobserved ARCH series. As a matter of fact, the two-step LSE deviates from the infeasible LSE by $o_p\left(n^{-1/2}\right)$. The two-step MLE, however, does not enjoy such efficiency, but $n^{1/2}$ asymptotic normality is established for both the two-step MLE as well as its deviation from the infeasible MLE. Simulation studies corroborate the asymptotic theory, and application to the S&P 500 index daily returns from 1950 to 2018 indicates significant change in volatility scale over time.

Received 11 April 2019; Accepted 06 January 2020

Keywords: Asymptotic normality; B-spline; least squares; maximum likelihood; oracle efficiency.

JEL. MOS subject classification code: 62G08; 62M10.

1. INTRODUCTION

The purpose of this article is to describe the gradual change over time observed in volatility of financial returns that is unaccounted for by standard autoregressive conditional heteroskedasticity (ARCH) model. In managing risk about financial returns, much more information is in the volatility rather than the mean of the return series. The ARCH model proposed by Engle (1982) and the generalized ARCH (GARCH) model by Bollerslev (1986) are widely used tools to compute conditional volatility in financial time series, such as stock returns. Consider a sequence of n financial returns $\mathbf{X} = \left(X_1, \ldots, X_n\right)^{\mathrm{T}}$ with X_t denoting the return at time t, which is a realization of stochastic process $\left\{X_t\right\}_{-\infty}^{\infty}$ satisfying

$$X_t = \sigma_t \epsilon_t, \quad -\infty < t < \infty, \tag{1.1}$$

in which $\left\{\epsilon_t\right\}_{-\infty}^{\infty}$ are i.i.d. with $\mathsf{E}\left(\epsilon_t\right) = 0$, $\mathsf{var}\left(\epsilon_t\right) = 1$, called innovations, and with respect to σ-fields $\mathcal{F}_{t-1} = \sigma\left(X_{t-j}, j = 1, 2, \ldots\right)$, the following conditional variance σ_t^2 is called volatility

$$\sigma_t^2 \equiv \mathsf{E}\left(X_t^2 | \mathcal{F}_{t-1}\right), \quad -\infty < t < \infty.$$

* Correspondence to: Lijian Yang, Center for Statistical Science & Department of Industrial Engineering, Tsinghua University, Beijing 100084, China.
E-mail: yanglijian@mail.tsinghua.edu.cn

© 2020 John Wiley & Sons Ltd

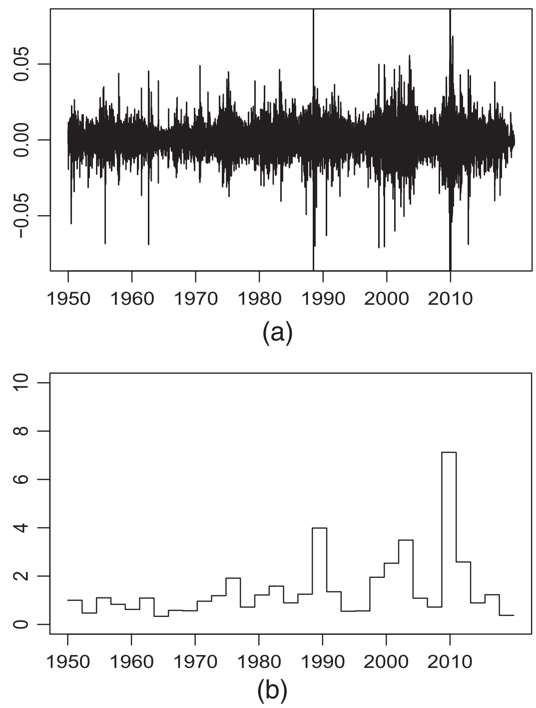

Figure 1. Plots of S&P 500 index daily returns, January 3, 1950–August 28, 2018: (a) the scatter plot of $\{y_{t,17276}\}_{t=1}^{17276}$; (b) the constant spline estimate $\hat{g}_1(u)$

An ARCH(p) model as in Engle (1982) specifies that σ_t^2 depends on some model parameters $\boldsymbol{\theta}_0 = (\omega_0, \alpha_{0,1}, \dots, \alpha_{0,p})^{\mathrm{T}}$,

$$\sigma_t^2 = \omega_0 + \sum_{k=1}^{p} \alpha_{0,k} X_{t-k}^2, \quad -\infty < t < \infty, \tag{1.2}$$

while the GARCH(p, q) model of Bollerslev (1986) stipulates that

$$\sigma_t^2 = \omega_0 + \sum_{k=1}^{p} \alpha_{0,k} X_{t-k}^2 + \sum_{k=1}^{q} \beta_{0,k} \sigma_{t-k}^2, \quad -\infty < t < \infty. \tag{1.3}$$

Both ARCH and GARCH models have been theoretically investigated and empirically applied under the assumption of stationarity, that is, the stochastic process $\{X_t\}_{-\infty}^{\infty}$ is strictly stationary. Such stationarity assumptions are also prevalent in various extensions of ARCH and GARCH models with different approaches, such as Engle and Ng (1993), Glosten *et al.* (1993), Hentschel (1995), Duan (1997), Hall and Yao (2003), Peng and Yao (2003), Hafner and Herwartz (2006), Yang (2006), Hafner (2008), Hafner and Preminger (2009a, 2009b), Hafner and Preminger (2017), Liu and Yang (2016).

This presumption of stationarity, however, has not gone unquestioned, especially over longer time period. Engle and Lee (1999), for instance, argued that the constant variance assumption for classic ARCH models is not realistic based on empirical findings. Take, for example, the classical Standard & Poors (S&P) 500-index daily returns series, with $n = 17{,}276$ observations from January 3, 1950 to August 28, 2018, depicted in Figure 1(a). The magnitude of variation has visibly increased from low in the first 30 years (1950–1980) to high in the last 18 years (2000–2018), the gradual change of which is accurately plotted in Figure 1(b).

© 2020 John Wiley & Sons Ltd

DOI: 10.1111/jtsa.12522

To remedy this deficiency of standard ARCH model, Dahlhaus and Subba Rao (2006) modified the constant coefficients into smooth functional coefficients to account for the time variation in conditional volatility. Specifically, equation (9) in Dahlhaus and Subba Rao (2006) specifies that

$$X_{t,n} = \sigma_{t,n}\epsilon_t, \sigma_{t,n}^2 = \omega_0\left(\frac{t}{n}\right) + \sum_{k=1}^{p}\alpha_{0,k}\left(\frac{t}{n}\right)X_{t-k,n}^2, \quad 1 \le t \le n, \tag{1.4}$$

in which $\omega_0(\cdot), \alpha_{0,k}(\cdot), 1 \le k \le p$ are unknown smooth functions defined on $[0,1]$. The above model is called time-varying ARCH, which is locally stationary, see also equation (2.1) of Truquet (2017) for a semi-parametric version of model (1.4), in which some of the $\omega_0(\cdot), \alpha_{0,k}(\cdot), 1 \le k \le p$ are constants. An unpleasant feature of the local stationarity in model (1.4) is the non-existence of a latent stationary series from which the observed series tractably deviates, unless all $\omega_0(\cdot), \alpha_{0,k}(\cdot), 1 \le k \le p$ are constants and the observed series itself is stationary. Another line of research on time-varying volatility is the smooth transition approach in Amado and Teräsvirta (2013, 2014a, 2014b), which is fully parametric, whereas model (1.4) is non-parametric.

In contrast, a more interpretable alternative to accommodate time-varyingness for volatility is the Spline-GARCH model of Engle and Rangel (2008). In this framework, the observed time series data is $\mathbf{Y}_n = \left(Y_{1,n}, \ldots, Y_{n,n}\right)^{\mathrm{T}}$, which satisfies

$$Y_{t,n} = \exp\left\{\vartheta\left(\frac{t}{n}\right)\right\}X_t, \quad 1 \le t \le n$$

where $\mathbf{X}$ is a stationary GARCH series according to (1.1) and (1.3), and $\vartheta(\cdot)$ is a quadratic spline function defined on [0, 1], a slowly changing trend corresponding to economic conditions. Compared to the non-parametric time-varying model (1.4), the Spline-GARCH model is one exponential spline function factor $\exp\{\vartheta(t/n)\}$ away from a latent stationary GARCH series with interpretable parameters $\omega_0, \alpha_{0,k}, \beta_{0,k}$. Compared to the fully parametric smooth transition model of Amado and Teräsvirta ((2013, 2014a, 2014b), the Spline-GARCH model is a great deal more flexible for adapting to changes in time. We therefore consider the Spline-GARCH approach a philosophically more preferable compromise.

Although the above Spline-GARCH model was applied to real data examples in Engle and Rangel (2008), it lacks both statistical theory and Monte Carlo evidence on its performance. To ameliorate this weakness, we propose the following time-varying volatility model in which the data $\mathbf{Y}_n$ satisfies:

$$Y_{t,n} = g\left(\frac{t}{n}\right)^{1/2}X_t, \quad 1 \le t \le n, \tag{1.5}$$

where $g(\cdot)$ is a positive unknown function on [0, 1], that is, there exist two positive numbers c_l and c_u such that $0 < c_l \le g(u) \le c_u < \infty$ for any $0 \le u \le 1$. This function $g(\cdot)$ does not need to be the exponent of spline as in the Spline GARCH model, only Hölder continuous of order greater $1/2$, see comments following Assumption (c2). The latent series $\mathbf{X}$ is a stationary GARCH series according to (1.3) or ARCH as in (1.2), only the latter one is treated in this article due to limit of space and the ARCH model being more intuitive and interpretable.

Maximum likelihood estimators (MLE) and least squares estimators (LSE) of $\left(\omega_0, \alpha_{0,1}, \ldots, \alpha_{0,p}\right)^{\mathrm{T}}$ in (1.2) can be computed from $\mathbf{X}$ as in Francq and Zakoian (2010) and Weiss (1986), which, under standard assumptions, enjoy asymptotic normality properties; however, they are 'infeasible' due to the fact that $\mathbf{X}$ is unobserved. The feasible analog of these estimators starts with estimating $g(\cdot)$ by a non-parametric estimator $\hat{g}(\cdot)$ in step one, and computing MLE and LSE in step two by using $\widehat{X}_t = \hat{g}(t/n)^{-1/2}Y_{t,n}, 1 \le t \le n$ in lieu of $X_t = g(t/n)^{-1/2}Y_{t,n}, 1 \le t \le n$. In response to one Reviewer, we point out that k-step ahead predictor ($k > 0$) $\mathbb{P}_n Y_{t+k,n}$ of $Y_{t+k,n}$ from $\left\{Y_{t,n}\right\}_{t=1}^n$ is easily obtained via the latent predictor $\mathbb{P}_n X_{t+k}$ of X_{t+k} from $\left\{\widehat{X}_t\right\}_{t=1}^n$: $\mathbb{P}_n Y_{t+k,n} = \hat{g}(n/n)\mathbb{P}_n X_{t+k} = \hat{g}(1)\mathbb{P}_n X_{t+k}$. This is due

 © 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522

to the scale factor $g(\cdot/n)$ changing from time t to $t+1$ at an infinitesimal magnitude $O(1/n)$, so the net variation in scale is negligible from time n to $n+k$ for a finite integer $k > 0$. On the other hand, the variation in scale over long span, for example, from time 1 to n (or $n+k$) is non-negligible according to model (1.5), unless the series is stationary (i.e., $g(\cdot) \equiv$ constant). For the S&P 500 daily returns, Figure 1(b) shows the minimum of $\hat{g}(\cdot) < 1$ and maximum > 7, hence the slow varyingness over 17,276 days has accumulated to pronounced scale variation. One reviewer has correctly pointed out that in practice, when sample size increases to $n+1$, the model estimation and forecasting should be re-calculated for finite sample optimality. The finite sample improvement due to such recalculation is, however, of the negligible magnitude $O(1/n)$.

One reviewer has pointed out that model (1.5) bears similarity to the model in Hafner and Linton (2010). We comment that Theorem 1, equation (25) of Hafner and Linton (2010) corresponds to our Proposition 2.2 for the infeasible MLE $\tilde{\boldsymbol{\alpha}}_{\text{MLE}}$, while Theorem 3, equation (31) of Hafner and Linton (2010) corresponds to our Theorem 2.2 on the data-based MLE $\hat{\boldsymbol{\alpha}}_{\text{MLE}}$. Such correspondence does not imply any logical consequence, however, as the time-varying ARCH model (1.5) is not nested as a special case of the GARCH model in Hafner and Linton (2010). In particular, our result on oracle efficiency of data-based LSE $\hat{\boldsymbol{\alpha}}_{\text{LSE}}$ has no counterpart in Hafner and Linton (2010). On the other hand, although the data-based MLE $\hat{\boldsymbol{\alpha}}_{\text{MLE}}$ is not oracally efficient due to its covariance matrix being more complicated than the infeasible MLE $\tilde{\boldsymbol{\alpha}}_{\text{MLE}}$, its covariance matrix is an exact replica of the semi-parametrically efficient MLE of Hafner and Linton (2010).

To make the trend scale function $g(\cdot)$ identifiable, one assumes that $\mathsf{E}X_t^2 \equiv 1$ in (1.1) and (1.2), and thus with the true model parameters $\boldsymbol{\alpha}_0 = \left(\alpha_{0,1}, \ldots, \alpha_{0,p}\right)^{\mathrm{T}}$,

$$1 \equiv \mathsf{E}\sigma_t^2 \equiv \omega_0 + \sum_{k=1}^{p} \alpha_{0,k} \mathsf{E}X_{t-k}^2, \omega_0 = 1 - \sum_{k=1}^{p} \alpha_{0,k},$$
$$\sigma_t^2(\boldsymbol{\alpha}_0) \equiv 1 - \sum_{k=1}^{p} \alpha_{0,k} + \sum_{k=1}^{p} \alpha_{0,k} X_{t-k}^2 = 1 + \sum_{k=1}^{p} \alpha_{0,k}\left(X_{t-k}^2 - 1\right). \tag{1.6}$$

The constant ω_0 is therefore not an independent parameter, as it relies on $\alpha_{0,1}, \ldots, \alpha_{0,p}$ from (1.6). The MLE and LSE of $\boldsymbol{\alpha}_0$ based on latent series $\mathbf{X}$ under the constraint (1.6) are similar in form as well as properties to unconstrained estimators of $\left(\omega_0, \alpha_{0,1}, \ldots, \alpha_{0,p}\right)^{\mathrm{T}}$ mentioned above. The main theoretical contribution of this article is establishing the asymptotic oracle efficiency of LSE and asymptotic normality of MLE for $\boldsymbol{\alpha}_0$ based on the observations $\left\{Y_{t,n}\right\}_{t=1}^{n}$ using $\left\{\hat{X}_t\right\}_{t=1}^{n}$ relative to those using $\left\{X_t\right\}_{t=1}^{n}$, given that the estimator $\hat{g}(\cdot)$ meets certain requirements. We then establish that B-spline estimator $\hat{g}(\cdot)$ satisfies those mild requirements. B-splines are renown for their simplicity and efficiency (see Xue and Yang, 2006; Wang and Yang, 2007 for details). Thus the two-step procedures to estimate $\boldsymbol{\alpha}_0$ provided in this article are both theoretically reliable and computationally efficient.

The article is organized as follows. Section 2 proposes two-step LSE and MLE procedures that estimate the parameter $\boldsymbol{\alpha}_0$ from the observations $\mathbf{Y}_n$ and elaborates on their asymptotic properties under a set of conditions on a pilot estimator of trend scale function $g(\cdot)$. Section 3 states the precise conditions on $g(\cdot)$ and formulates an appropriate B-spline estimator of $g(\cdot)$ from $\mathbf{Y}_n$ that fulfills the set of conditions in Section 2. Section 4 provides concrete steps to implement the LSE and MLE methods. Sections 5 and 6 illustrate the performance of the proposed procedures by simulations and a real data example. Section 7 concludes, while all technical proofs are in the Appendix and Supporting Information.

2. ESTIMATION FOR ARCH COEFFICIENTS

A two-step LSE and MLE based on observations $\mathbf{Y}_n$ are formulated for the parameter $\boldsymbol{\alpha}_0$, and the conditions under which the LSE is oracally as efficient as the corresponding LSE based on latent $\mathbf{X} = \left(X_1, \ldots, X_n\right)^{\mathrm{T}}$ are stated, while the MLE is $n^{1/2}$ asymptotically normal.

 © 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522

2.1. Least Squares Estimators

Since the ARCH process $\left\{X_t\right\}_{t=-\infty}^{\infty}$ is standardized (i.e., $\mathsf{E}X_t^2 \equiv 1$), one defines a related auxiliary process $\left\{Z_t\right\}_{t=-\infty}^{\infty}$ as

$$Z_t = X_t^2 - 1 = g\left(\frac{t}{n}\right)^{-1} Y_{t,n}^2 - 1,$$

which satisfies

$$Z_t = \sigma_t^2(\boldsymbol{\alpha}_0)\epsilon_t^2 - 1 = \sum_{k=1}^{p} \alpha_{0,k} Z_{t-k} + \xi_t, \quad \xi_t = \sigma_t^2(\boldsymbol{\alpha}_0)\left(\epsilon_t^2 - 1\right). \tag{2.1}$$

It is straightforward to show that $\mathsf{E}\left(\xi_t|\mathcal{F}_{t-1}\right) \equiv 0$ and that $\left\{\xi_t\right\}_{t=-\infty}^{\infty}$ is a strictly stationary sequence of martingale differences and hence a white noise sequence, so $\left\{Z_t\right\}_{t=-\infty}^{\infty}$ is a mean zero AR(p) sequence. The following assumptions are from Francq and Zakoian (2010) (see also Weiss, 1986).

(a) The ARCH process $\left\{X_t\right\}_{t=-\infty}^{\infty}$ is non-anticipative and strictly stationary and $P\left(\epsilon_t^2 = 1\right) < 1$. The true parameter $\boldsymbol{\alpha}_0$ is in the interior of parameter space $\Xi \subset \mathbb{R}^p$, a compact subset such that each vector $\boldsymbol{\alpha} = \left(\alpha_1, \dots, \alpha_p\right)^{\mathrm{T}} \in \Xi$ satisfies $\alpha_k \geq 0$ $(1 \leq k \leq p)$ and $\sum_{k=1}^{p} \alpha_k < 1$. Consequently, all roots of $1 - \sum_{k=1}^{p} \alpha_k z^k = 0$ lie outside the unit circle.

(b) $\mathsf{E}X_t^8 < \infty$.

If the ARCH process $\left\{X_t\right\}_{t=-\infty}^{\infty}$ were actually observed as $\mathbf{X}$, a simple would-be LSE $\widetilde{\boldsymbol{\alpha}}_{\mathrm{LSE}} = \left(\tilde{\alpha}_{1,\mathrm{LSE}}, \dots, \tilde{\alpha}_{p,\mathrm{LSE}}\right)^{\mathrm{T}}$ of $\boldsymbol{\alpha}_0$ follows

$$\widetilde{\boldsymbol{\alpha}}_{\mathrm{LSE}} = \underset{\boldsymbol{\alpha}\in\Xi}{\operatorname{argmin}} \sum_{t=p+1}^{n} \left(Z_t - \sum_{k=1}^{p} \alpha_k Z_{t-k}\right)^2, \tag{2.2}$$

which has a closed formula

$$\widetilde{\boldsymbol{\alpha}}_{\mathrm{LSE}} = \widetilde{\boldsymbol{\Gamma}}^{-1}\widetilde{\boldsymbol{\gamma}}, \tag{2.3}$$

where

$$\widetilde{\boldsymbol{\Gamma}} = \begin{pmatrix} \widetilde{\gamma}(1,1) & \cdots & \widetilde{\gamma}(p,1) \\ \vdots & \ddots & \vdots \\ \widetilde{\gamma}(p,1) & \cdots & \widetilde{\gamma}(p,p) \end{pmatrix}, \quad \widetilde{\boldsymbol{\gamma}} = \left(\widetilde{\gamma}(1,0), \dots, \widetilde{\gamma}(p,0)\right)^{\mathrm{T}},$$

$$\widetilde{\gamma}\left(k_1, k_2\right) = (n-p)^{-1} \sum_{t=p+1}^{n} Z_{t-k_1} Z_{t-k_2}, \quad 0 \leq k_1, k_2 \leq p.$$

The corresponding parameters are

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma(1,1) & \cdots & \gamma(p,1) \\ \vdots & \ddots & \vdots \\ \gamma(p,1) & \cdots & \gamma(p,p) \end{pmatrix}, \quad \boldsymbol{\gamma} = \left(\gamma(1,0), \dots, \gamma(p,0)\right)^{\mathrm{T}},$$

$$\gamma\left(k_1, k_2\right) = \mathsf{E}Z_{t-k_1} Z_{t-k_2}, \quad 0 \leq k_1, k_2 \leq p. \tag{2.4}$$

The above can be simplified as

$$\boldsymbol{\Gamma} = \mathsf{E}\left(\mathbf{M}\mathbf{M}^{\mathrm{T}}\right), \quad \mathbf{M} = (Z_p, \dots, Z_1)^{\mathrm{T}}. \tag{2.5}$$

Define next matrix

$$\boldsymbol{\Gamma}_\sigma = \mathsf{E}\left\{\sigma_t^4\left(\boldsymbol{\alpha}_0\right)\mathbf{M}\mathbf{M}^{\mathrm{T}}\right\}, \quad \mathbf{F} = \left(\mathsf{E}\epsilon^4 - 1\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_\sigma\boldsymbol{\Gamma}^{-1}. \tag{2.6}$$

The next result follows directly from Theorem 6.2 of Francq and Zakoian (2010) or Theorem 4.4 of Weiss (1986).

Proposition 2.1. Under Assumptions (a) and (b), as $n \to \infty$, $n^{1/2}\left(\widetilde{\boldsymbol{\alpha}}_{\mathrm{LSE}} - \boldsymbol{\alpha}_0\right) \xrightarrow{D} N(\mathbf{0}, \mathbf{F})$, with covariance matrix $\mathbf{F}$ defined in (2.6).

The would-be estimator $\widetilde{\boldsymbol{\alpha}}_{\mathrm{LSE}}$ depends on unobserved $\mathbf{X}$ and thus is infeasible, and the sample analog $\left\{\widehat{Z}_t\right\}_{t=1}^n$ of $\left\{Z_t\right\}_{t=1}^n$ are defined by

$$\widehat{Z}_t = \widehat{X}_t^2 - 1, \quad \widehat{X}_t = \hat{g}\left(\frac{t}{n}\right)^{-1/2} Y_{t,n} \tag{2.7}$$

with an appropriate non-parametric estimator $\hat{g}$ of g. A plug-in analog $\widehat{\boldsymbol{\alpha}}_{\mathrm{LSE}} = \left(\hat{\alpha}_{1,\mathrm{LSE}}, \ldots, \hat{\alpha}_{p,\mathrm{LSE}}\right)^{\mathrm{T}}$ to mimic $\widetilde{\boldsymbol{\alpha}}_{\mathrm{LSE}}$ in (2.2) is

$$\widehat{\boldsymbol{\alpha}}_{\mathrm{LSE}} = \underset{\alpha \in \Xi}{\operatorname{argmin}} \sum_{t=p+1}^{n}\left(\widehat{Z}_t - \sum_{k=1}^{p}\alpha_k \widehat{Z}_{t-k}\right)^2, \tag{2.8}$$

which has a closed formula similar to (2.3)

$$\widehat{\boldsymbol{\alpha}}_{\mathrm{LSE}} = \widehat{\boldsymbol{\Gamma}}^{-1}\widehat{\boldsymbol{\gamma}},$$

where

$$\widehat{\boldsymbol{\Gamma}} = \begin{pmatrix} \widehat{\gamma}(1,1) & \cdots & \widehat{\gamma}(p,1) \\ \vdots & \ddots & \vdots \\ \widehat{\gamma}(p,1) & \cdots & \widehat{\gamma}(p,p) \end{pmatrix}, \quad \widehat{\boldsymbol{\gamma}} = \left(\widehat{\gamma}(1,0), \ldots, \widehat{\gamma}(p,0)\right)^{\mathrm{T}},$$

$$\widehat{\gamma}\left(k_1, k_2\right) = (n-p)^{-1}\sum_{t=p+1}^{n}\widehat{Z}_{t-k_1}\widehat{Z}_{t-k_2}, \quad 0 \le k_1, k_2 \le p.$$

This two-step estimator $\widehat{\boldsymbol{\alpha}}_{\mathrm{LSE}}$ is feasible and called oracle under the following Assumption (c) on $\hat{g}(\cdot)$ that guarantees that the difference between $\widehat{\boldsymbol{\alpha}}_{\mathrm{LSE}}$ and the infeasible $\widetilde{\boldsymbol{\alpha}}_{\mathrm{LSE}}$ is negligible. Define the following measure of discrepancy between $\hat{g}(\cdot)$ and $g(\cdot)$

$$\Delta_{n,0} = \Delta_{n,0}(\hat{g}) = \max_{p+1 \le t \le n}\left|\hat{g}\left(\frac{t}{n}\right) - g\left(\frac{t}{n}\right)\right|. \tag{2.9}$$

Definition 1 is the same as the Definition 1 of Shao and Yang (2017).

Definition 1. For some constants $C > 0$ and $0 < \rho < 1$, a deterministic vector $\mathbf{a} = \left(a_0, \ldots, a_n\right)^{\mathrm{T}}$ is called (C, ρ)-exponentially bounded if $|a_j| \le C\rho^j$ for any $0 \le j \le n-1$; a random vector $\zeta = \left(\zeta_1, \ldots, \zeta_n\right)^{\mathrm{T}}$ is called (C, ρ)-exponentially correlated if $\mathsf{E}(\zeta) = 0$ and $|\mathsf{E}\left(\zeta_k\zeta_l\right)| \le C\rho^{|k-l|}, 1 \le k, l \le n$.

 © 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522

For any integer $m' \geq 0$ and $\delta \in [0,1]$, denote by $C^{(m',\delta)}[0,1]$ the space of functions whose m'-th derivatives satisfy Hölder conditions of order δ, that is

$$C^{(m',\delta)}[0,1] = \left\{ \phi : [0,1] \to \mathbb{R} \mid \|\phi\|_{m',\delta} = \sup_{0 \leq x < y \leq 1} \frac{\left|\phi^{(m')}(x) - \phi^{(m')}(y)\right|}{|x-y|^{\delta}} < \infty \right\}.$$

(c) The trend estimator $\hat{g}(\cdot)$ satisfies

$$\Delta_{n,0} = o_p\left(n^{-1/4}\right), \tag{2.10}$$

$$\|\hat{g}\|_{0,r} = \mathcal{O}_p(1), \|g\|_{0,r} < \infty, \text{for some } r > \frac{1}{2}, \tag{2.11}$$

$$S_{n,\zeta} = (n-p)^{-1} \sum_{t=p+1}^{n} \zeta_t \frac{g(t/n) - \hat{g}(t/n)}{g(t/n)} = o_p\left(n^{-1/2}\right), \tag{2.12}$$

where

$$\zeta_t = Z_{t-k}, \xi_t \quad \text{or} \quad Z_{t-k}\xi_t,$$

so $\left(\zeta_{p+1}, \ldots, \zeta_n\right)$ is a $\left(C_\zeta, \rho_\zeta\right)$-exponentially correlated sequence of random variables according to Definition 1.

The next theorem concerning $\hat{\boldsymbol{\alpha}}_{\text{LSE}}$ is the first main theoretical result.

Theorem 2.1. Under Assumptions (a)–(c), as $n \to \infty, n^{1/2}\left(\hat{\boldsymbol{\alpha}}_{\text{LSE}} - \tilde{\boldsymbol{\alpha}}_{\text{LSE}}\right) \xrightarrow{P} 0$, and consequently $n^{1/2}\left(\hat{\boldsymbol{\alpha}}_{\text{LSE}} - \boldsymbol{\alpha}_0\right) \xrightarrow{D} N(\mathbf{0}, \mathbf{F})$, with covariance matrix $\mathbf{F}$ defined in (2.6).

2.2. Maximum Likelihood Estimators

Similar to $\tilde{\boldsymbol{\alpha}}_{\text{LSE}}$, an infeasible MLE $\tilde{\boldsymbol{\alpha}}_{\text{MLE}} = \left(\tilde{\alpha}_{1,\text{MLE}}, \ldots, \tilde{\alpha}_{p,\text{MLE}}\right)^{\text{T}}$ of $\boldsymbol{\alpha}_0$ is defined using unobserved $\mathbf{X}$:

$$\begin{aligned} \tilde{\boldsymbol{\alpha}}_{\text{MLE}} &= \operatorname*{argmin}_{\boldsymbol{\alpha} \in \Xi} Q_n(\boldsymbol{\alpha}), \\ Q_n(\boldsymbol{\alpha}) &= (n-p)^{-1} \sum_{t=p+1}^{n} \left\{ \log \sigma_t^2(\boldsymbol{\alpha}) + \frac{X_t^2}{\sigma_t^2(\boldsymbol{\alpha})} \right\}, \\ \sigma_t^2(\boldsymbol{\alpha}) &= 1 - \sum_{k=1}^{p} \alpha_k + \sum_{k=1}^{p} \alpha_k X_{t-k}^2, \quad t = p+1, \ldots, n, \end{aligned} \tag{2.13}$$

where $\boldsymbol{\alpha} = \left(\alpha_1, \ldots, \alpha_p\right)^{\text{T}}$ denotes a candidate value of $\boldsymbol{\alpha}_0 = \left(\alpha_{0,1}, \ldots, \alpha_{0,p}\right)^{\text{T}}$. The function $Q_n(\boldsymbol{\alpha})$ is negatively proportional to the log likelihood of $\mathbf{X}$ if the distribution of ϵ_t were normal, see Francq and Zakoian (2010). Similar to $\hat{\boldsymbol{\alpha}}_{\text{LSE}}$, a two-step estimator $\hat{\boldsymbol{\alpha}}_{\text{MLE}} = \left(\hat{\alpha}_{1,\text{MLE}}, \ldots, \hat{\alpha}_{p,\text{MLE}}\right)^{\text{T}}$ is defined as:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_{\text{MLE}} &= \operatorname*{argmin}_{\boldsymbol{\alpha} \in \Xi} \hat{Q}_n(\boldsymbol{\alpha}), \\ \hat{Q}_n(\boldsymbol{\alpha}) &= (n-p)^{-1} \sum_{t=p+1}^{n} \left\{ \log \hat{\sigma}_t^2(\boldsymbol{\alpha}) + \frac{\hat{X}_t^2}{\hat{\sigma}_t^2(\boldsymbol{\alpha})} \right\}, \end{aligned} \tag{2.14}$$

© 2020 John Wiley & Sons Ltd

$$\widehat{\sigma}_t^2(\boldsymbol{\alpha}) = 1 - \sum_{k=1}^{p} \alpha_k + \sum_{k=1}^{p} \alpha_k \widehat{X}_{t-k}^2, \quad t = p+1, \ldots, n,$$

where $\widehat{X}_t^2$ is defined in (2.7).

Applying Theorem 7.2 of Francq and Zakoian (2010) and some algebra, we obtain the asymptotic distribution of the infeasible estimator $\widetilde{\boldsymbol{\alpha}}_{\text{MLE}}$ as follows:

Proposition 2.2. Under Assumptions (a)–(c), $\widetilde{\boldsymbol{\alpha}}_{\text{MLE}}$ is strongly consistent and asymptotically normally distributed:

$$\widetilde{\boldsymbol{\alpha}}_{\text{MLE}} \xrightarrow{a.s.} \boldsymbol{\alpha}_0, n^{1/2}\left(\widetilde{\boldsymbol{\alpha}}_{\text{MLE}} - \boldsymbol{\alpha}_0\right) \xrightarrow{D} N\left(\mathbf{0}, \left(\mathsf{E}\epsilon^4 - 1\right)\mathbf{J}^{-1}\right) \text{ as } n \to \infty$$

where

$$\mathbf{J} = \mathsf{E}\left\{\sigma_{p+1}^{-4}\left(\boldsymbol{\alpha}_0\right)\mathbf{M}\mathbf{M}^{\mathrm{T}}\right\} \tag{2.15}$$

is a positive definite matrix, with vector $\mathbf{M}$ defined in (2.5).

It turns out that unlike $\widehat{\boldsymbol{\alpha}}_{\text{LSE}}$, the two-step estimator $\widehat{\boldsymbol{\alpha}}_{\text{MLE}}$ does not enjoy oracle efficiency, but $\widehat{\boldsymbol{\alpha}}_{\text{MLE}} - \boldsymbol{\alpha}_0$ is $n^{1/2}$ asymptotically normal under the additional constraint as follows:

(d) The trend estimator $\hat{g}(\cdot)$ satisfies

$$\begin{aligned} S_n &= (n-p)^{-1} \sum_{t=p+1}^{n} \frac{g(t/n) - \hat{g}(t/n)}{g(t/n)} \\ &= (n-p)^{-1} \sum_{t=p+1}^{n} Z_t + o_p\left(n^{-1/2}\right), \end{aligned} \tag{2.16}$$

while (2.12) holds, with $1 \le k, l \le p$, for $S_{n,\zeta}$ with the following (C, ρ)-exponentially correlated sequences

$$\zeta_t = \frac{Z_{t-k}\xi_t}{\sigma_t^4(\boldsymbol{\alpha}_0)}, \frac{\xi_t}{\sigma_t^4(\boldsymbol{\alpha}_0)}, \frac{Z_{t-k}Z_{t-l}\xi_t}{\sigma_t^6(\boldsymbol{\alpha}_0)}, \frac{Z_{t-k}\xi_t}{\sigma_t^6(\boldsymbol{\alpha}_0)}, \text{ or } \frac{Z_{t-k}}{\sigma_t^4(\boldsymbol{\alpha}_0)} - \mathsf{E}\left(\frac{Z_{p+1-k}}{\sigma_{p+1}^4(\boldsymbol{\alpha}_0)}\right).$$

Theorem 2.2. Under Assumptions (a), (b) and (d), as $n \to \infty$,

$$n^{1/2}\begin{pmatrix} \widetilde{\boldsymbol{\alpha}}_{\text{MLE}} - \boldsymbol{\alpha}_0 \\ \widehat{\boldsymbol{\alpha}}_{\text{MLE}} - \widetilde{\boldsymbol{\alpha}}_{\text{MLE}} \end{pmatrix} \xrightarrow{D} N\left(\mathbf{0}_{2p\times1}, \begin{pmatrix} \left(\mathsf{E}\epsilon^4 - 1\right)\mathbf{J}^{-1} & \boldsymbol{\Sigma}_{\text{cov}}^{\mathrm{T}} \\ \boldsymbol{\Sigma}_{\text{cov}} & \mathbf{J}^{-1}\boldsymbol{\Sigma}_{\text{diff}}\mathbf{J}^{-1} \end{pmatrix}\right).$$

in which $\mathbf{J}$ is as defined in (2.15) and

$$\begin{aligned} \boldsymbol{\Sigma}_{\text{cov}} &= \omega_0 \mathbf{J}^{-1} \sum_{l=1}^{+\infty} \left\{\mathsf{E}\sigma_{t-l}^{-4}(\boldsymbol{\alpha}_0) Z_t Z_{t-l-k} \xi_{t-l}\right\}_{k=1}^{p} \mathsf{E}\left\{\sigma_{p+1}^{-4}(\boldsymbol{\alpha}_0)\mathbf{M}\right\}^{\mathrm{T}} \mathbf{J}^{-1}, \\ \boldsymbol{\Sigma}_{\text{diff}} &= \omega_0^2 \left\{\sum_{k=-\infty}^{+\infty} \gamma(0,k)\right\} \mathsf{E}\left\{\sigma_{p+1}^{-4}(\boldsymbol{\alpha}_0)\mathbf{M}\right\} \mathsf{E}\left\{\sigma_{p+1}^{-4}(\boldsymbol{\alpha}_0)\mathbf{M}\right\}^{\mathrm{T}} \end{aligned}$$

with vector $\mathbf{M}$ defined in (2.5). Consequently, for
$\boldsymbol{\Sigma} = \mathbf{J}^{-1}\boldsymbol{\Sigma}_{\text{diff}}\mathbf{J}^{-1} + \left(\mathsf{E}\epsilon^4 - 1\right)\mathbf{J}^{-1} + \boldsymbol{\Sigma}_{\text{cov}} + \boldsymbol{\Sigma}_{\text{cov}}^{\mathrm{T}}, n^{1/2}\left(\widehat{\boldsymbol{\alpha}}_{\text{MLE}} - \boldsymbol{\alpha}_0\right) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma})$.

 © 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522

While the infeasible MLE $\widetilde{\boldsymbol{\alpha}}_{\mathrm{MLE}}$ is more efficient than the infeasible LSE $\widetilde{\boldsymbol{\alpha}}_{\mathrm{LSE}}$ by standard theory in Francq and Zakoian (2010), that is, $\mathbf{F} \geq \left(\mathsf{E}\epsilon^4 - 1\right) \mathbf{J}^{-1}$, the same is not necessarily true for the two step estimators. In fact, it is unclear to us if the two-step MLE $\widehat{\boldsymbol{\alpha}}_{\mathrm{MLE}}$ is more efficient than the two-step LSE $\widehat{\boldsymbol{\alpha}}_{\mathrm{MLE}}$ or less, due to the complicated extra terms in $\boldsymbol{\Sigma}$ in Theorem 2.2. This is the price one pays for having to remove first the nuance time scale $g(\cdot)$.

3. B-SPLINE ESTIMATE FOR TREND FUNCTION

Both Theorems 2.1 and 2.2 are highly desirable results which guarantee that estimate $\widehat{\boldsymbol{\alpha}}_{\mathrm{LSE}}$ computed from observations $\mathbf{Y}_n = \left(Y_{1,n}, \ldots, Y_{n,n}\right)^{\mathrm{T}}$ is asymptotically as efficient as $\widetilde{\boldsymbol{\alpha}}_{\mathrm{LSE}}$ computed from the latent $\mathbf{X} = \left(X_1, \ldots, X_n\right)^{\mathrm{T}}$, and $\widehat{\boldsymbol{\alpha}}_{\mathrm{MLE}}$ is $n^{1/2}$ asymptotically normal. Both depend crucially on estimator $\hat{g}(\cdot)$ of $g(\cdot)$ satisfying (2.10) and (2.11) in Assumption (c) and (2.16) in Assumption (d). Naturally one wonders if any such estimator $\hat{g}(\cdot)$ exists, and in this section, we present some elementary conditions which ensure that a B-spline estimator of $g(\cdot)$ satisfies Assumptions (c) and (d).

Denote a sequence of interior knots $\tau_1, \ldots, \tau_N, 0 < \tau_1 < \cdots < \tau_N < 1, \tau_j = jh, 1 \leq j \leq N$, which divide the interval $[0,1]$ into subintervals of equal length $h = (N+1)^{-1}, J_j = \left[jh, (j+1)h\right), j = 0, \ldots, N-1$ and $J_N = [Nh, 1]$. For an integer $m > 0$, let $G_N^{(m-2)} = G_N^{(m-2)}[0,1]$ be the space of functions that are polynomial of degree $m-1$ on each J_j and have a continuous $(m-2)$-th derivative on $[0,1]$, and denote its B-spline basis as $\mathbf{b}_m(u) = (b_{-m+1,m}(u), \ldots, b_{N,m}(u))^{\mathrm{T}}$, see Chapter IX of de Boor (2001).

Denote for any function $\varphi(\cdot)$ in $L^2[0,1]$ the norm as $\|\varphi\|_2 = \left\{\int_0^1 \varphi^2(x)\mathrm{d}x\right\}^{1/2}$. For any $u \in [0,1]$, the standardized B-spline basis $\mathbf{B}_m(u)$ is defined as

$$
\begin{aligned}
\mathbf{B}_m(u) &= (B_{-m+1,m}(u), \ldots, B_{N,m}(u)), \\
B_{j,m}(u) &= \frac{b_{j,m}(u)}{\|b_{j,m}\|_2} = \frac{b_{j,m}(u)}{\left\{\int_0^1 b_{j,m}^2(u)du\right\}^{1/2}}, \quad -m+1 \leq j \leq N.
\end{aligned}
$$

The B-spline estimator $\hat{g}_m(\cdot)$ of $g(\cdot)$ is defined as

$$
\hat{g}_m(\cdot) = \underset{\varphi(\cdot)\in G_N^{(m-2)}[0,1]}{\operatorname{argmin}} \sum_{t=p+1}^{n} \left\{Y_{t,n}^2 - \varphi\left(\frac{t}{n}\right)\right\}^2, \tag{3.1}
$$

and clearly

$$
\hat{g}_m(u) = \mathbf{B}_m(u)\,\hat{\lambda}, \hat{\lambda} = \underset{\lambda\in\mathbb{R}^{N+m}}{\operatorname{argmin}} \sum_{t=p+1}^{n} \left\{Y_{t,n}^2 - \mathbf{B}_m\left(\frac{t}{n}\right)\lambda\right\}^2.
$$

B spline smoothing is preferred over kernel smoothing as an intermediate step for estimating the parameters in semi-parametric models because it achieves the same theoretical efficiency but is immensely faster when sample size is large, see Xue and Yang (2006), Wang and Yang (2007, 2009), Liu and Yang (2016) for related computational consideration. One reviewer was concerned that the B spline estimator $\hat{g}_m(u)$ could take negative values. Our view is that (1) theoretically, the B spline estimator $\hat{g}_m(u)$ is guaranteed non-negative if one takes the default constant B spline ($m = 1$), under minimum requirement on the trend smoothness ($m' + \delta > 1/2$), see the comments following Assumption (c2); (2) computationally, the negativity anomaly has never occurred to real or simulated data examples even with cubic spline $m = 4$, due to the sample size $n \geq 10{,}000$ so asymptotics kicks in very nicely. Such large sample size is entirely realistic as the slowly varying phenomenon is observed only over long horizon, see comments in the beginning of Section 5.

 © 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522

For sequences of real numbers c_n and d_n, one writes $c_n \gg d_n$ to mean $d_n/c_n \to 0$, as $n \to \infty$. The following Assumptions on $g(\cdot)$ and N are adapted from Shao and Yang (2017):

(c1) The trend function $g(\cdot) \in C^{(m',\delta)}[0,1]$ with $m' + \delta \le m$, $\min_{u\in[0,1]} g(u) = c_l > 0$, $\max_{u\in[0,1]} g(u) = c_u$.
(c2) The number of interior knots $N = N_n$ satisfies

$$n^{1/2} \gg N \gg \begin{cases} n^{1/4(m'+\delta)} & \text{for LSE} \\ n^{1/2(m'+\delta+1/2)} & \text{for MLE} \end{cases}. \tag{3.2}$$

Note that Assumption (c2) is satisfiable as long as the smoothness order of g, $m' + \delta > 1/2$, which ensures that the Hölder continuity order of g to be $r = \min(m' + \delta, 1) > 1/2$, as required in (2.11), and this is close to minimal. Notice also that the conditions for MLE automatically imply the conditions for LSE: if $n^{1/2} \gg N \gg n^{1/2(m'+\delta+1/2)}$, then $m' + \delta > 1/2$ hence $n^{1/2(m'+\delta+1/2)} \gg n^{1/4(m'+\delta)}$. These facts allow one in particular to use constant spline ($m = 1$), which is fastest to compute and guarantees non-negativity of the estimator $\hat{g}_m(\cdot)$, see Section 4 for details.

Theorem 3.1. Under Assumptions (a), (b), (c1) and (c2), the B-spline estimator $\hat{g}_m(\cdot)$ defined in (3.1) satisfies Assumptions (c) and (d).

The proofs of all Theorems 2.1–3.1 are in the Appendix.

4. IMPLEMENTATION

In simulation studies and real data analysis, the number of interior knots $N = N_n$ is computed according to the formula $N = \min\left(\left[c_1 n^{1/2(m'+\delta+1/2)} \log n + c_2\right], b\right)$, where c_1, c_2, b are non-negative tuning parameters. The simulation experiments indicate that the simple choice of $c_1 = 0.1$, $c_2 = 3$ and $b = 37$ works quite well, so these are set as the default values. The default values for m' and δ are 0 and 1 respectively, whereas the default value for m is 1.

A plug-in estimator for the LSE along with its asymptotic covariance matrix in Theorem 2.1 is calculated as follows: $\hat{\gamma}(k_1, k_2) = (n-p)^{-1} \sum_{t=p+1}^{n} \hat{Z}_{t-k_1} \hat{Z}_{t-k_2}$, where $\hat{Z}_t = \hat{X}_t^2 - 1 = \hat{g}_m(t/n)^{-1} Y_{t,n}^2 - 1$; $\hat{\sigma}_t^2(\hat{\boldsymbol{\alpha}}_{\text{LSE}}) = 1 - \sum_{k=1}^{p} \hat{\alpha}_{k,\text{LSE}} + \sum_{k=1}^{p} \hat{\alpha}_{k,\text{LSE}} \hat{X}_{t-k}^2$, where $\hat{\boldsymbol{\alpha}}_{\text{LSE}} = (\hat{\alpha}_{1,\text{LSE}}, \ldots, \hat{\alpha}_{p,\text{LSE}})^{\text{T}}$ is the LSE in (2.8); $\hat{\epsilon}^2 = \hat{X}_t^2 \hat{\sigma}_t^{-2}(\hat{\boldsymbol{\alpha}}_{\text{LSE}})$; finally

$$\begin{aligned} \hat{\boldsymbol{\Gamma}}_\sigma &= (n-p)^{-1} \sum_{t=p+1}^{n} \left\{ \hat{\sigma}_t^4(\hat{\boldsymbol{\alpha}}_{\text{LSE}}) \hat{\mathbf{M}}_t \hat{\mathbf{M}}_t^{\text{T}} \right\}, \\ \hat{\mathbf{F}} &= \left\{ (n-p)^{-1} \sum_{t=p+1}^{n} \hat{X}_t^4 \hat{\sigma}_t^{-4}(\hat{\boldsymbol{\alpha}}_{\text{LSE}}) - 1 \right\} \hat{\boldsymbol{\Gamma}}^{-1} \hat{\boldsymbol{\Gamma}}_\sigma^{-1} \hat{\boldsymbol{\Gamma}}, \end{aligned} \tag{4.1}$$

where $\hat{\mathbf{M}}_t = (\hat{Z}_{t-1}, \ldots, \hat{Z}_{t-p})^{\text{T}}$.

Similar to the calculation of $\hat{\boldsymbol{\alpha}}_{\text{LSE}}$ above, a plug-in estimator $\hat{\boldsymbol{\alpha}}_{\text{MLE}} = (\hat{\alpha}_{1,\text{MLE}}, \ldots, \hat{\alpha}_{p,\text{MLE}})^{\text{T}}$ is derived according to (2.14), then $\hat{\sigma}_t^2(\hat{\boldsymbol{\alpha}}_{\text{MLE}}) = 1 - \sum_{k=1}^{p} \hat{\alpha}_{k,\text{MLE}} + \sum_{k=1}^{p} \hat{\alpha}_{k,\text{MLE}} \hat{X}_{t-k}^2$, $\hat{\epsilon}^2 = \hat{X}_t^2 \hat{\sigma}_t^{-2}(\hat{\boldsymbol{\alpha}}_{\text{MLE}})$ and $\sum_{k_2=0}^{n-1} \hat{\gamma}(0, k_2) = n^{-1} \sum_{k_2=0}^{n-1} \sum_{t=1}^{n-k_2} \hat{Z}_t \hat{Z}_{t+k_2}$. Then one again computes $\hat{\mathbf{M}}_t = (\hat{Z}_{t-1}, \ldots, \hat{Z}_{t-p})^{\text{T}}$, and

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{\text{diff}} = &\left(1 - \sum_{k=1}^{p} \hat{\alpha}_{k,\text{MLE}}\right)^2 \left\{ \sum_{k_2=0}^{n-1} \hat{\gamma}(0, k_2) \right\} (n-p)^{-1} \sum_{t=p+1}^{n} \left\{ \hat{\sigma}_t^{-4}(\hat{\boldsymbol{\alpha}}_{\text{MLE}}) \hat{\mathbf{M}}_t \right\} \\ &\times (n-p)^{-1} \sum_{t=p+1}^{n} \left\{ \hat{\sigma}_t^{-4}(\hat{\boldsymbol{\alpha}}_{\text{MLE}}) \hat{\mathbf{M}}_t \right\}^{\text{T}}, \end{aligned}$$

 © 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522

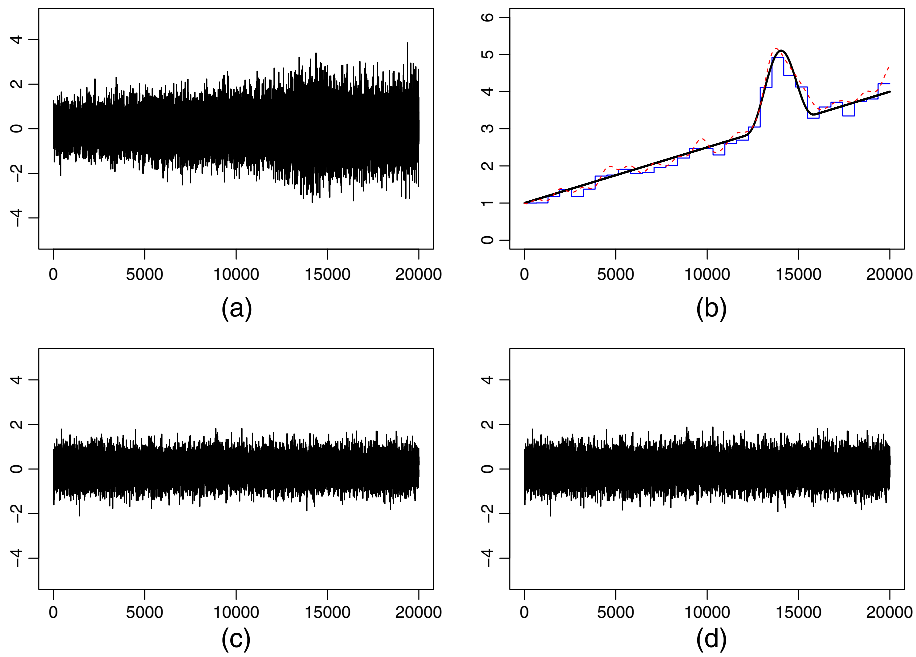

Figure 2. Plots of ARCH(9) with $\boldsymbol{\alpha}_0 = (0.133, 0.096, 0.080, 0.079, 0.081, 0.061, 0.056, 0.085, 0.094)^{\mathrm{T}}$ and $n = 20{,}000$: (a) the scatter plot of $\{y_{t,n}\}_{t=1}^{20{,}000}$; (b) the true function $g(u)$ (thick), the constant spline estimate $\hat{g}_1(u)$ (solid), and the cubic spline estimate $\hat{g}_4(u)$ (dashed) ; (c) the scatter plot of $\{x_t\}_{t=1}^{20{,}000}$; (d) the scatter plot of $\{\hat{x}_t\}_{t=1}^{20{,}000}$ [Color figure can be viewed at wileyonlinelibrary.com]

$$\hat{\mathbf{J}} = (n-p)^{-1} \sum_{t=p+1}^{n} \left\{ \hat{\sigma}_t^{-4}(\hat{\boldsymbol{\alpha}}_{\mathrm{MLE}}) \hat{\mathbf{M}}_t \hat{\mathbf{M}}_t^{\mathrm{T}} \right\}, \hat{\epsilon}^2 = \hat{X}_t^2 \hat{\sigma}_t^{-2} \left(\hat{\boldsymbol{\alpha}}_{\mathrm{MLE}} \right).$$

Finally, the plug-in asymptotic covariance matrix $\hat{\boldsymbol{\Sigma}}_{\mathrm{cov}}$ of MLE method in Theorem 2.2 is derived, and

$$\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{J}}^{-1} \hat{\boldsymbol{\Sigma}}_{\mathrm{diff}}^{-1} \hat{\mathbf{J}}^{-1} + \left\{ (n-p)^{-1} \sum_{t=p+1}^{n} \hat{X}_t^4 \hat{\sigma}_t^{-4} \left(\hat{\boldsymbol{\alpha}}_{\mathrm{MLE}} \right) - 1 \right\} \hat{\mathbf{J}}^{-1} + \hat{\boldsymbol{\Sigma}}_{\mathrm{cov}} + \hat{\boldsymbol{\Sigma}}_{\mathrm{cov}}^{\mathrm{T}}.$$

5. SIMULATION

Sample paths are generated 1000 times for an ARCH(9) model with parameter $\boldsymbol{\alpha}_0 = (0.133, 0.096, 0.080, 0.079, 0.081, 0.061, 0.056, 0.085, 0.094)^{\mathrm{T}}$ that satisfies Assumption (a). Sample sizes used are 10,000, 15,000, 20,000, all quite large as correctly pointed out by one reviewer. These large sample sizes are in fact very realistic since the time varying phenomenon manifests itself only over long time horizons. For instance, the S&P 500 daily returns data examined in the next section is of size 17,276.

The dashed line in Figure 2(b) is a smooth bounded positive function $g(u)$ which is defined as

$$g(u) = 1 + 3u + 2\left\{1 - 100(u-0.7)^2\right\}^3 I_{\{|u-0.7| \le 0.1\}}(u),$$

where $I_A(u)$ is the indicator function for the set A, namely, $I_A(u) = 1$ for $u \in A$ and 0 otherwise. It is a slowly increasing linear function for $\{u : |u-0.7| > 0.1\}$ and 6th order polynomial function for $\{u : |u-0.7| \le 0.1\}$

© 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522

Table I. Sample means and standard errors of estimates based on $\{Y_{t,n}\}_{t=1}^{n}$

	$n = 10{,}000$		$n = 15{,}000$		$n = 20{,}000$		
	Mean	SE	Mean	SE	Mean	SE	True value
$\hat{\alpha}_{1,\mathrm{LSE},Y_{t,n}}$	0.067	0.010	0.070	0.009	0.085	0.007	$\alpha_{0,1} = 0.133$
$\hat{\alpha}_{1,\mathrm{MLE},Y_{t,n}}$	0.050	0.009	0.060	0.008	0.098	0.007	
$\hat{\alpha}_{2,\mathrm{LSE},Y_{t,n}}$	0.035	0.010	0.045	0.009	0.076	0.008	$\alpha_{0,2} = 0.096$
$\hat{\alpha}_{2,\mathrm{MLE},Y_{t,n}}$	0.048	0.010	0.067	0.008	0.090	0.007	
$\hat{\alpha}_{3,\mathrm{LSE},Y_{t,n}}$	0.015	0.010	0.032	0.009	0.070	0.007	$\alpha_{0,3} = 0.080$
$\hat{\alpha}_{3,\mathrm{MLE},Y_{t,n}}$	0.035	0.009	0.050	0.008	0.082	0.007	
$\hat{\alpha}_{4,\mathrm{LSE},Y_{t,n}}$	0.037	0.009	0.025	0.008	0.066	0.007	$\alpha_{0,4} = 0.079$
$\hat{\alpha}_{4,\mathrm{MLE},Y_{t,n}}$	0.056	0.009	0.038	0.008	0.078	0.007	
$\hat{\alpha}_{5,\mathrm{LSE},Y_{t,n}}$	0.030	0.010	0.040	0.009	0.042	0.009	$\alpha_{0,5} = 0.081$
$\hat{\alpha}_{5,\mathrm{MLE},Y_{t,n}}$	0.026	0.009	0.036	0.008	0.039	0.007	
$\hat{\alpha}_{6,\mathrm{LSE},Y_{t,n}}$	0.030	0.009	0.045	0.008	0.048	0.007	$\alpha_{0,6} = 0.061$
$\hat{\alpha}_{6,\mathrm{MLE},Y_{t,n}}$	0.050	0.009	0.068	0.008	0.069	0.007	
$\hat{\alpha}_{7,\mathrm{LSE},Y_{t,n}}$	0.052	0.010	0.067	0.009	0.068	0.008	$\alpha_{0,7} = 0.056$
$\hat{\alpha}_{7,\mathrm{MLE},Y_{t,n}}$	0.045	0.009	0.057	0.008	0.059	0.007	
$\hat{\alpha}_{8,\mathrm{LSE},Y_{t,n}}$	0.098	0.009	0.089	0.008	0.090	0.007	$\alpha_{0,8} = 0.085$
$\hat{\alpha}_{8,\mathrm{MLE},Y_{t,n}}$	0.069	0.009	0.078	0.008	0.079	0.007	
$\hat{\alpha}_{9,\mathrm{LSE},Y_{t,n}}$	0.015	0.010	0.030	0.009	0.032	0.007	$\alpha_{0,9} = 0.094$
$\hat{\alpha}_{9,\mathrm{MLE},Y_{t,n}}$	0.035	0.010	0.042	0.008	0.043	0.007	

Table II. Sample means and standard errors of estimates based on $\{\hat{X}_t\}_{t=1}^{n}$

	$n = 10{,}000$		$n = 15{,}000$		$n = 20{,}000$		
	Mean	SE	Mean	SE	Mean	SE	True value
$\hat{\alpha}_{1,\mathrm{LSE}}$	0.120	0.016	0.130	0.009	0.132	0.008	$\alpha_{0,1} = 0.133$
$\hat{\alpha}_{1,\mathrm{MLE}}$	0.128	0.010	0.132	0.008	0.134	0.007	
$\hat{\alpha}_{2,\mathrm{LSE}}$	0.088	0.010	0.093	0.010	0.095	0.008	$\alpha_{0,2} = 0.096$
$\hat{\alpha}_{2,\mathrm{MLE}}$	0.089	0.010	0.090	0.009	0.097	0.007	
$\hat{\alpha}_{3,\mathrm{LSE}}$	0.083	0.008	0.076	0.010	0.078	0.008	$\alpha_{0,3} = 0.080$
$\hat{\alpha}_{3,\mathrm{MLE}}$	0.071	0.012	0.078	0.008	0.082	0.007	
$\hat{\alpha}_{4,\mathrm{LSE}}$	0.075	0.011	0.076	0.009	0.076	0.009	$\alpha_{0,4} = 0.079$
$\hat{\alpha}_{4,\mathrm{MLE}}$	0.072	0.010	0.075	0.009	0.080	0.007	
$\hat{\alpha}_{5,\mathrm{LSE}}$	0.073	0.010	0.079	0.007	0.078	0.006	$\alpha_{0,5} = 0.081$
$\hat{\alpha}_{5,\mathrm{MLE}}$	0.075	0.009	0.076	0.006	0.080	0.005	
$\hat{\alpha}_{6,\mathrm{LSE}}$	0.054	0.010	0.057	0.007	0.063	0.007	$\alpha_{0,6} = 0.061$
$\hat{\alpha}_{6,\mathrm{MLE}}$	0.067	0.008	0.058	0.005	0.060	0.006	
$\hat{\alpha}_{7,\mathrm{LSE}}$	0.050	0.010	0.052	0.006	0.054	0.007	$\alpha_{0,7} = 0.056$
$\hat{\alpha}_{7,\mathrm{MLE}}$	0.060	0.008	0.053	0.004	0.057	0.006	
$\hat{\alpha}_{8,\mathrm{LSE}}$	0.072	0.011	0.080	0.008	0.084	0.009	$\alpha_{0,8} = 0.085$
$\hat{\alpha}_{8,\mathrm{MLE}}$	0.079	0.008	0.082	0.007	0.087	0.007	
$\hat{\alpha}_{9,\mathrm{LSE}}$	0.091	0.010	0.090	0.008	0.093	0.006	$\alpha_{0,9} = 0.094$
$\hat{\alpha}_{9,\mathrm{MLE}}$	0.089	0.009	0.100	0.007	0.096	0.006	

with a small ridge. The scatter plot of a sample path $\{y_{t,20{,}000}\}_{t=1}^{20{,}000}$ in Figure 2(a) shows that the variance looks relatively stable at the beginning and becomes unsteady in the middle. The thin and dashed curves in Figure 2(b) are respectively the constant B-spline estimate $\hat{g}_1(u)$ and cubic B-spline estimate $\hat{g}_4(u)$ of the sample path in Figure 2(a). Both of the estimates $\hat{g}_1(u)$ and $\hat{g}_4(u)$ approximate the trend function $g(u)$ (thick) well. After removing $\hat{g}_1(u)$ from $\{Y_{t,n}\}_{t=1}^{n}$, $\{\hat{X}_t\}_{t=1}^{20{,}000}$ in Figure 2(d) looks quite similar to $\{X_t\}_{t=1}^{20{,}000}$ in Figure 2(c).

The estimates $\hat{\boldsymbol{\alpha}}_{\mathrm{MLE},Y_{t,n}} = \left(\hat{\alpha}_{1,\mathrm{MLE},Y_{t,n}}, \dots, \hat{\alpha}_{p,\mathrm{MLE},Y_{t,n}}\right)^{\mathrm{T}}$ in Table I are computed using the model (1.2) from the observations $\{Y_{t,n}\}_{t=1}^{n}$. It is not surprising that the estimates $\hat{\boldsymbol{\alpha}}_{\mathrm{MLE},Y_{t,n}}$ are very far from the true values $\boldsymbol{\alpha}_0$, as

 © 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522

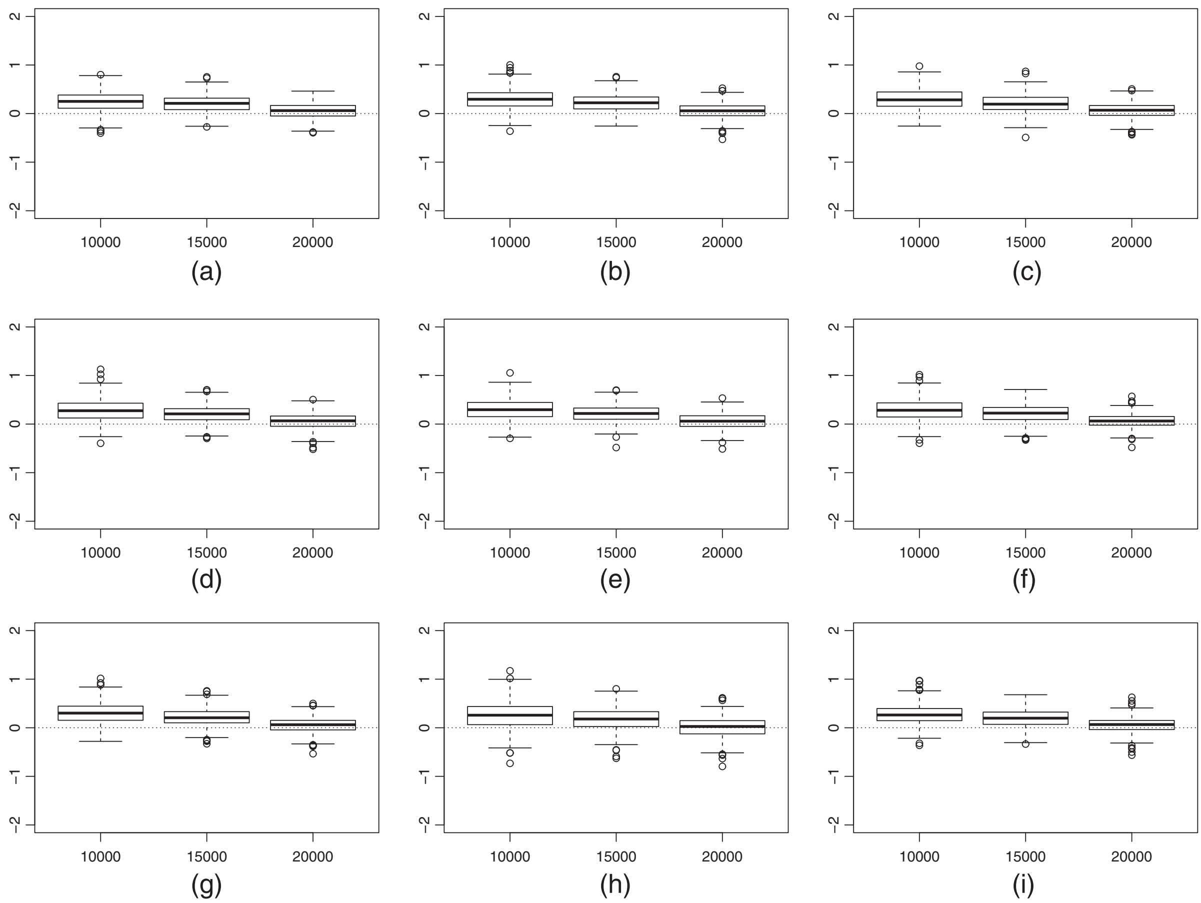

Figure 3. Boxplots of $n^{1/2}\left(\widehat{\boldsymbol{\alpha}}_{\mathrm{LSE}}-\widetilde{\boldsymbol{\alpha}}_{\mathrm{LSE}}\right)$ with $\boldsymbol{\alpha}_0 = (0.133, 0.096, 0.080, 0.079, 0.081, 0.061, 0.056, 0.085, 0.094)^{\mathrm{T}}$ and $n = 10{,}000, 15{,}000, 20{,}000$: (a) $n^{1/2}\left(\hat{\alpha}_{1,\mathrm{LSE}}-\tilde{\alpha}_{1,\mathrm{LSE}}\right)$; (b) $n^{1/2}\left(\hat{\alpha}_{2,\mathrm{LSE}}-\tilde{\alpha}_{2,\mathrm{LSE}}\right)$; (c) $n^{1/2}\left(\hat{\alpha}_{3,\mathrm{LSE}}-\tilde{\alpha}_{3,\mathrm{LSE}}\right)$; (d) $n^{1/2}\left(\hat{\alpha}_{4,\mathrm{LSE}}-\tilde{\alpha}_{4,\mathrm{LSE}}\right)$; (e) $n^{1/2}\left(\hat{\alpha}_{5,\mathrm{LSE}}-\tilde{\alpha}_{5,\mathrm{LSE}}\right)$; (f) $n^{1/2}\left(\hat{\alpha}_{6,\mathrm{LSE}}-\tilde{\alpha}_{6,\mathrm{LSE}}\right)$; (g) $n^{1/2}\left(\hat{\alpha}_{7,\mathrm{LSE}}-\tilde{\alpha}_{7,\mathrm{LSE}}\right)$; (h) $n^{1/2}\left(\hat{\alpha}_{8,\mathrm{LSE}}-\tilde{\alpha}_{8,\mathrm{LSE}}\right)$; (i) $n^{1/2}\left(\hat{\alpha}_{9,\mathrm{LSE}}-\tilde{\alpha}_{9,\mathrm{LSE}}\right)$

$\left\{Y_{t,n}\right\}_{t=1}^{n}$ is not a realization of the ARCH(9). The MLE and LSE estimates in Table II are obtained from the cubic B-spline estimates $\left\{\hat{g}_1(t/n)\right\}_{t=1}^{n}$ and the residual sequence $\left\{\hat{X}_t\right\}_{t=1}^{n}$. They are much closer to the true parameters $\boldsymbol{\alpha}_0$ than $\widehat{\boldsymbol{\alpha}}_{\mathrm{MLE},Y_{t,n}}$, which coincides with the theoretical results in Sections 2 and 3. Moreover, it is worth mentioning that according to the sample standard deviation, the MLE estimates perform better than the LSE estimates.

It is insightful to investigate the difference between the estimates from the data set $\left\{\widehat{X}_t\right\}_{t=1}^{n}$ and those from $\left\{X_t\right\}_{t=1}^{n}$. In addition to $\widehat{\boldsymbol{\alpha}}_{\mathrm{MLE}}$ and $\widehat{\boldsymbol{\alpha}}_{\mathrm{LSE}}$, $\widetilde{\boldsymbol{\alpha}}_{\mathrm{MLE}}$ and $\widetilde{\boldsymbol{\alpha}}_{\mathrm{LSE}}$ are also computed from the simulated ARCH series $\left\{X_t\right\}_{t=1}^{n}$. The boxplots in Figure 3 show that when the sample size increases, the difference between $\widehat{\boldsymbol{\alpha}}_{\mathrm{LSE}}$ and $\widetilde{\boldsymbol{\alpha}}_{\mathrm{LSE}}$ decreases to zero faster than the rate $n^{-1/2}$, which corroborates with the theoretical results in Section 2. In contrast, the difference between $\widehat{\boldsymbol{\alpha}}_{\mathrm{MLE}}$ and $\widetilde{\boldsymbol{\alpha}}_{\mathrm{MLE}}$ is asymptotically Gaussian at the rate $n^{-1/2}$ according to the boxplots in Figure 4 in the Supporting Information.

The boxplots in Figure 5 in the Supporting Information show that when the sample size increases, the difference between $\widehat{\boldsymbol{\alpha}}_{\mathrm{LSE}}$ and $\boldsymbol{\alpha}_0$ is asymptotically Gaussian at the rate $n^{-1/2}$, which validates the theoretical results in Section 2. A similar phenomenon for the difference between $\widetilde{\boldsymbol{\alpha}}_{\mathrm{LSE}}$ and $\boldsymbol{\alpha}_0$ is depicted in the boxplots of Figure 6

 © 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522

Table III. Coverage frequencies of LSE and MLE confidence intervals from 1000 replications

		Nominal significance level			
		0.950		0.990	
n	Coefficient	LSE	MLE	LSE	MLE
10,000	$\alpha_{0,1}$	0.939	0.930	0.985	0.980
	$\alpha_{0,2}$	0.929	0.920	0.980	0.972
	$\alpha_{0,3}$	0.938	0.932	0.980	0.970
	$\alpha_{0,4}$	0.940	0.928	0.981	0.980
	$\alpha_{0,5}$	0.930	0.930	0.978	0.979
	$\alpha_{0,6}$	0.939	0.932	0.980	0.981
	$\alpha_{0,7}$	0.942	0.940	0.982	0.981
	$\alpha_{0,8}$	0.930	0.920	0.980	0.980
	$\alpha_{0,9}$	0.940	0.936	0.979	0.980
15,000	$\alpha_{0,1}$	0.946	0.940	0.986	0.979
	$\alpha_{0,2}$	0.948	0.942	0.981	0.978
	$\alpha_{0,3}$	0.947	0.946	0.984	0.980
	$\alpha_{0,4}$	0.950	0.948	0.984	0.980
	$\alpha_{0,5}$	0.948	0.946	0.983	0.980
	$\alpha_{0,6}$	0.946	0.947	0.981	0.982
	$\alpha_{0,7}$	0.949	0.948	0.983	0.981
	$\alpha_{0,8}$	0.945	0.945	0.981	0.982
	$\alpha_{0,9}$	0.948	0.945	0.981	0.981
20,000	$\alpha_{0,1}$	0.950	0.946	0.989	0.986
	$\alpha_{0,2}$	0.948	0.947	0.989	0.989
	$\alpha_{0,3}$	0.948	0.948	0.990	0.987
	$\alpha_{0,4}$	0.952	0.950	0.988	0.989
	$\alpha_{0,5}$	0.948	0.951	0.988	0.986
	$\alpha_{0,6}$	0.949	0.950	0.989	0.987
	$\alpha_{0,7}$	0.949	0.949	0.990	0.989
	$\alpha_{0,8}$	0.950	0.948	0.987	0.988
	$\alpha_{0,9}$	0.949	0.950	0.989	0.990

in the Supporting Information. The same is observed for $\widehat{\boldsymbol{\alpha}}_{\text{MLE}}$ and $\boldsymbol{\alpha}_0$, $\widetilde{\boldsymbol{\alpha}}_{\text{MLE}}$ and $\boldsymbol{\alpha}_0$ according to the boxplots in Figures 7 and 8 in the Supporting Information.

In addition, to illustrate how well the distribution of a finite sample is approximated by the normal distributions, the empirical confidence levels based on $\left\{\widehat{X}_t\right\}_{t=1}^{n}$ are calculated using the asymptotic distributions in Theorems 2.1 and 2.2. These empirical confidence levels are essentially the relative frequencies that the true value $\boldsymbol{\alpha}_0$ is in the confidence intervals constructed using either the LSE or MLE and their respective asymptotic standard errors in Theorems 2.1 and 2.2. The true significance levels or nominal levels used are 0.95 and 0.99. The empirical confidence levels from the LSE and MLE are respectively summarized in Table III. The relative frequency approaches the nominal level with increasing sample size, which confirms Theorems 2.1 and 2.2. Moreover, in all scenarios, the true parameter is more likely to fall within the LSE confidence intervals than the corresponding MLE confidence intervals.

6. APPLICATION

The application of the proposed method is illustrated by the analysis of the S&P500 index series $\left\{Y_{t,n}\right\}_{t=1}^{n}$. The data set in our analysis includes the observations as early as January 3, 1950 and as late as August 28, 2018. There are a total of $n = 17{,}277$ closing prices, which were downloaded from https://finance.yahoo.com. More precisely, this series contains the differences of logarithms of daily opening and closing prices for about 68 years. The scatter plot of $\left\{Y_{t,n}\right\}_{t=1}^{n}$ in Figure 1(a) shows conspicuous non-constant pattern.

There are more frequent and more pronounced spikes since the end of 1990s. The pattern of clustering in the past 10 years is conspicuously changed compared with that before, which suggests our proposed model is more appropriate than the classic ARCH for the entire data set.

 © 2020 John Wiley & Sons Ltd *J. Time Ser. Anal.* **41**: 551–570 (2020)
DOI: 10.1111/jtsa.12522

In the implementation of the two-step estimation procedure, the constant B-spline ($m = 1$) and the number of knots $N = 30$ are used. A pronounced non-constant trend is exhibited by the estimate $\hat{g}_1(t/n)$ in Figure 1(b). The scatter plot of the residual sequence $\hat{X}_t = \{\hat{g}_1(t/n)\}^{-1/2} Y_{t,n}$ in Figure 9, which is showed in the Supporting Information, resembles the clustering pattern of an autoregressive conditional heteroskedasticity series.

Before proceeding to the second step of estimation, an ARCH(9) model is selected for $\{\hat{X}_t\}$ according to the Bayesian information criterion. The LSE with the error margins of a 95% confidence intervals are $\hat{\alpha}_{1,\mathrm{LSE}} = 0.133 \pm 0.050$, $\hat{\alpha}_{2,\mathrm{LSE}} = 0.096 \pm 0.052$, $\hat{\alpha}_{3,\mathrm{LSE}} = 0.080 \pm 0.048$, $\hat{\alpha}_{4,\mathrm{LSE}} = 0.079 \pm 0.050$, $\hat{\alpha}_{5,\mathrm{LSE}} = 0.81 \pm 0.046$, $\hat{\alpha}_{6,\mathrm{LSE}} = 0.061 \pm 0.045$, $\hat{\alpha}_{7,\mathrm{LSE}} = 0.056 \pm 0.048$, $\hat{\alpha}_{8,\mathrm{LSE}} = 0.085 \pm 0.050$, $\hat{\alpha}_{9,\mathrm{LSE}} = 0.094 \pm 0.056$, and the MLE with the error margin of a 95% confidence intervals are $\hat{\alpha}_{1,\mathrm{MLE}} = 0.129 \pm 0.035$, $\hat{\alpha}_{2,\mathrm{MLE}} = 0.098 \pm 0.030$, $\hat{\alpha}_{3,\mathrm{MLE}} = 0.082 \pm 0.036$, $\hat{\alpha}_{4,\mathrm{MLE}} = 0.080 \pm 0.030$, $\hat{\alpha}_{5,\mathrm{MLE}} = 0.075 \pm 0.028$, $\hat{\alpha}_{6,\mathrm{MLE}} = 0.056 \pm 0.030$, $\hat{\alpha}_{7,\mathrm{MLE}} = 0.058 \pm 0.032$, $\hat{\alpha}_{8,\mathrm{MLE}} = 0.082 \pm 0.029$, $\hat{\alpha}_{9,\mathrm{MLE}} = 0.096 \pm 0.032$. Neither the estimates nor the error margins of the two methods are identical. The differences are sometimes noticeable, and the error margins of the MLEs are smaller.

7. CONCLUSIONS

A time varying ARCH model is proposed to take into account the slow deterministic change in volatility, which includes the classic autoregressive conditional heteroskedasticity model as a special case. Two-step LSE and MLE procedures are provided for parameters of the unobserved ARCH process, based on residuals computed from the observations $\mathbf{Y}_n$ with a time varying scale. Oracle efficiency for the LSE and $n^{1/2}$ asymptotic normality for both the MLE and its deviation from the infeasible MLE are established under some conditions on the pilot estimator of a time varying trend. A B-spline estimator is shown to be one such trend estimator, with easy implementation by built-in functions of the statistical computing environment R (2018). The two-step LSE and MLE are not only theoretically optimal and computationally simple, but intuitively easy to interpret as well. Thus they are highly recommended for analyzing time varying financial volatility.

ACKNOWLEDGMENTS

This work is part of the first author's doctoral dissertation under the supervision of the last author, and has been supported in part by National Natural Science Foundation of China awards 11371272 and 11771240. Helpful comments from two reviewers led to improved and more focused exposition as well as sharper theoretical result.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study is openly available in Yahoo Finance at https://finance.yahoo.com, under the heading S&P 500. It is also uploaded as a supplementary file S&P500.XLSX.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Amado C, Teräsvirta T. 2013. Modelling volatility by variance decomposition. *Journal of Econometrics* **175**: 142–153.
Amado C, Teräsvirta T. 2014a. Modelling changes in the unconditional variance of long stock return series. *Journal of Empirical Finance* **25**: 15–35.
Amado C, Teräsvirta T. 2014b. Conditional correlation models of autoregressive conditional heteroskedasticity with nonstationary GARCH equations. *Journal of Business and Economic Statistics* **32**: 69–87.
Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307–327.
Brockwell PJ, Davis RA. 1991. *Time Series: Theory and Methods*, 2nd ed. New York: Springer.
Duan JC. 1997. Augmented GARCH(p,q) process and its diffusion limit. *Journal of Econometrics* **79**: 97–127.

 © 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522

Dahlhaus R, Subba Rao S. 2006. Statistical inference for time-varying ARCH processes. *The Annals of Statistics* **34**: 1075–1114.
de Boor C. 2001. *A Practical Guide to Splines*. New York: Springer.
Engle RF. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of inflation in the United Kingdom. *Econometrica* **50**: 987–1007.
Engle RF, Ng V. 1993. Measuring and testing the impact of news on volatility. *Journal of Finance* **48**: 1749–1778.
Engle RF, Rangel JG. 2008. The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies* **21**: 1187–1122.
Engle RF, Lee G. 1999. *A long-run and short-run component model of stock return volatility*. A Festschrift in Honour of Clive WJ Granger: Cointegration, Causality, and Forecasting.
Francq C, Zakoian JM. 2010. *GARCH Models: Structure, Statistical Inference and Financial Applications*. Chichester: Wiley.
Glosten LR, Jaganathan R, Runkle DE. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**: 1779–1801.
Hafner CM, Herwartz H. 2006. Volatility impulse responses for multivariate GARCH models: an exchange rate illustration. *Journal of International Money and Finance* **25**: 719–740.
Hafner CM. 2008. Temporal aggregation of multivariate GARCH processes. *Journal of Econometrics* **142**: 467–483.
Hafner CM, Preminger A. 2009a. On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis* **100**: 2044–2054.
Hafner CM, Preminger A. 2009b. On asymptotic theory for a factor GARCH model. *Econometric Theory* **25**: 336–363.
Hafner CM, Linton O. 2010. Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics* **159**: 55–73.
Hafner CM, Preminger A. 2017. On asymptotic theory for ARCH(∞) models. *Journal of Time Series Analysis* **38**: 865–879.
Hall P, Yao Q. 2003. Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* **71**: 285–317.
Hentschel L. 1995. All in the family: nesting symmetric and asymmetric GARCH models. *Journal of Financial Econometrics* **39**: 71–104.
Liu R, Yang L. 2016. Spline estimation of a semiparametric GARCH model. *Econometric Theory* **32**: 1023–1054.
Peng L, Yao Q. 2003. Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**: 967–975.
R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Shao Q, Yang L. 2017. Oracally efficient estimation and consistent model selection for auto-regressive moving average time series with trend. *Journal of the Royal Statistical Society: Series B* **79**: 507–524.
Truquet L. 2017. Parameter stability and semiparametric inference in time varying auto-regressive conditional heteroscedasticity models. *Journal of the Royal Statistical Society: Series B* **79**: 1391–1414.
Wang L, Yang L. 2007. Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *The Annals of Statistics* **35**: 2474–2503.
Wang L, Yang L. 2009. Spline estimation of single index model. *Statistica Sinica* **19**: 765–783.
Weiss AA. 1986. Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory* **2**: 107–131.
Xue L, Yang L. 2006. Additive coefficient modelling via polynomial spline. *Statistica Sinica* **16**: 1423–1446.
Yang L. 2006. A semiparametric GARCH model for foreign exchange volatility. *Journal of Econometrics* **130**: 365–384.

APPENDIX A.

We denote bounded by $\mathcal{O}_p$ and uniformly bounded by $\mathcal{U}_p$, a constant by C, the value of which depends on different contexts.

In addition to $\Delta_{n,0}$ defined in (2.9), we also need the following measures of discrepancy

$$\Delta_{n,1} = \Delta_{n,1}(\hat{g}) = \max_{p+1\le t\le n} \left| \frac{\widehat{Z}_t - Z_t}{Y_{t,n}^2} - \frac{g(t/n) - \hat{g}(t/n)}{g^2(t/n)} \right|, \tag{A.1}$$

$$\Delta_{n,2} = \Delta_{n,2}(\hat{g}) = \max_{p+1\le t\le n} \left| \frac{\widehat{Z}_t - Z_t}{Y_{t,n}^2} \right|, \tag{A.2}$$

$$\Delta_{n,3} = \Delta_{n,3}(\hat{g}) = \max_{1\le k\le p}\max_{p+1\le t\le n} \left| \frac{\hat{g}((t-k)/n)}{g((t-k)/n)} - \frac{\hat{g}(t/n)}{g(t/n)} \right|. \tag{A.3}$$

Lemma A.1. Under Assumption (c), as $n \to \infty$, if $\Delta_{n,0} = o_p(1)$

$$\Delta_{n,1} \le \left(c_l - \Delta_{n,0}\right)^{-1} c_l^{-2} \Delta_{n,0}^2 = \mathcal{O}_p\left(\Delta_{n,0}^2\right) = o_p\left(n^{-1/2}\right), \tag{A.4}$$

$$\Delta_{n,2} \le \Delta_{n,1} + c_l^{-2} \Delta_{n,0} = \mathcal{O}_p\left(\Delta_{n,0}\right) = o_p\left(n^{-1/4}\right), \tag{A.5}$$

$$\Delta_{n,3} \le c_l^{-2}\left\{\|\hat{g}\|_{0,r}\, c_l + \|g\|_{0,r}\left(c_u + \Delta_{n,0}\right)\right\}\left(\frac{p}{n}\right)^r = \mathcal{O}_p\left(n^{-r}\right) = o_p\left(n^{-1/2}\right). \tag{A.6}$$

Proof. See the Supporting Information. □

Proof of Theorem 2.1. See the Supporting Information. □

Proof of Theorem 2.2. We apply the Taylor expansions to the first derivative $\partial Q_n(\boldsymbol{\alpha})/\partial \boldsymbol{\alpha}$ and obtain the following

$$\left.\frac{\partial Q_n(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = \left(\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial \alpha_k \partial \alpha_l}\right|_{\boldsymbol{\alpha}=r_k\boldsymbol{\alpha}_0+\left(1-r_k\right)\widetilde{\boldsymbol{\alpha}}_{\mathrm{MLE}}}\right)_{k,l=1}^p \left(\boldsymbol{\alpha}_0 - \widetilde{\boldsymbol{\alpha}}_{\mathrm{MLE}}\right),$$

for some $r_k \in [0,1]$, $1 \le k \le p$.

Since $\left\{\sigma_t^{-4}(\boldsymbol{\alpha}_0) Z_{t-k}\xi_t\right\}_{t=-\infty}^{+\infty}$ is strictly stationary with mean zero and finite variance according to Assumption (b), $(n-p)^{-1}\sum_{t=p+1}^n \sigma_t^{-4}(\boldsymbol{\alpha}_0) Z_{t-k}\xi_t = \mathcal{O}_p\left(n^{-1/2}\right)$ for each $1 \le k \le p$. According to (A.10) in the Supporting Information

$$\begin{aligned}\left.\frac{\partial Q_n(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} &= -(n-p)^{-1}\left\{\sum_{t=p+1}^n \frac{Z_{t-k}\left(\sum_{l=1}^p \alpha_{0,l} Z_{t-l} - Z_t\right)}{\sigma_t^4(\boldsymbol{\alpha}_0)}\right\}_{k=1}^p \\ &= -\left\{(n-p)^{-1}\sum_{t=p+1}^n \frac{Z_{t-k}\xi_t}{\sigma_t^4(\boldsymbol{\alpha}_0)}\right\}_{k=1}^p = \mathcal{O}_p\left(n^{-1/2}\right).\end{aligned} \tag{A.7}$$

Combining (A.12) in the Supporting Information and (A.7)

$$\left.\frac{\partial \widehat{Q}_n(\boldsymbol{\alpha})}{\partial \alpha_k}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = \left.\frac{\partial Q_n(\boldsymbol{\alpha})}{\partial \alpha_k}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} + \mathcal{O}_p\left(n^{-1/2}\right) = \mathcal{O}_p\left(n^{-1/2}\right).$$

Note that due to the consistency of $\widetilde{\boldsymbol{\alpha}}_{\mathrm{MLE}}$, the boundedness in probability of $\left\{\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^{\mathrm{T}}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\right\}^{-1}$ and Lemma A.5 in the Supporting Information, denote $\mathbf{r}_k = r_k\boldsymbol{\alpha}_0 + \left(1-r_k\right)\widetilde{\boldsymbol{\alpha}}_{\mathrm{MLE}}$, $\mathbf{s}_k = s_k\boldsymbol{\alpha}_0 + \left(1-s_k\right)\widehat{\boldsymbol{\alpha}}_{\mathrm{MLE}}$,

$$\left(\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial \alpha_k \partial \alpha_l}\right|_{\boldsymbol{\alpha}=\mathbf{r}_k}\right)_{k,l=1}^p - \mathbf{J} = \left(\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial \alpha_k \partial \alpha_l}\right|_{\boldsymbol{\alpha}=\mathbf{r}_k}\right)_{k,l=1}^p - \left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^{\mathrm{T}}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0},$$

 © 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522

$$\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^{\mathrm{T}}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} - \mathbf{J} = o_p(1),$$

$$\widetilde{\boldsymbol{\alpha}}_{\mathrm{MLE}} - \boldsymbol{\alpha}_0 = \left\{\left(\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial\alpha_k\partial\alpha_l}\right|_{\boldsymbol{\alpha}=\mathbf{r}_k}\right)_{k,l=1}^{p}\right\}^{-1}\left\{(n-p)^{-1}\sum_{t=p+1}^{n}\frac{Z_{t-k}\xi_t}{\sigma_t^4(\boldsymbol{\alpha}_0)}\right\}_{k=1}^{p}$$

$$\widetilde{\boldsymbol{\alpha}}_{\mathrm{MLE}} - \boldsymbol{\alpha}_0 = \mathbf{J}^{-1}\left\{(n-p)^{-1}\sum_{t=p+1}^{n}\frac{Z_{t-k}\xi_t}{\sigma_t^4(\boldsymbol{\alpha}_0)}\right\}_{k=1}^{p} + o_p\left(n^{-1/2}\right). \tag{A.8}$$

Similarly, one obtains

$$\left.\frac{\partial\widehat{Q}_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = \left(\left(\left.\frac{\partial^2\widehat{Q}_n(\boldsymbol{\alpha})}{\partial\alpha_k\partial\alpha_l}\right|_{\boldsymbol{\alpha}=\mathbf{s}_k}\right)_{k,l=1}^{p}\right)\left(\boldsymbol{\alpha}_0 - \widehat{\boldsymbol{\alpha}}_{\mathrm{MLE}}\right),$$

for some $s_k \in [0,1]$, $1 \le k \le p$. Thus one can write $\widetilde{\boldsymbol{\alpha}}_{\mathrm{MLE}} - \widehat{\boldsymbol{\alpha}}_{\mathrm{MLE}}$ as

$$\begin{aligned}&\left\{\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^{\mathrm{T}}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\right\}^{-1}\left\{\left.\frac{\partial\widehat{Q}_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} - \left.\frac{\partial Q_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\right\}\\
&+\left[\left\{\left(\left.\frac{\partial^2\widehat{Q}_n(\boldsymbol{\alpha})}{\partial\alpha_k\partial\alpha_l}\right|_{\boldsymbol{\alpha}=\mathbf{s}_k}\right)_{k,l=1}^{p}\right\}^{-1} - \left\{\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^{\mathrm{T}}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\right\}^{-1}\right]\left.\frac{\partial\widehat{Q}_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\\
&+\left[\left\{\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^{\mathrm{T}}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\right\}^{-1} - \left\{\left(\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial\alpha_k\partial\alpha_l}\right|_{\boldsymbol{\alpha}=\mathbf{r}_k}\right)_{k,l=1}^{p}\right\}^{-1}\right]\left.\frac{\partial Q_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}.\end{aligned}$$

Again using the consistency of $\widetilde{\boldsymbol{\alpha}}_{\mathrm{MLE}}$ and $\widehat{\boldsymbol{\alpha}}_{\mathrm{MLE}}$, the boundedness in probability of $\left\{\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^{\mathrm{T}}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\right\}^{-1}$ and Lemma A.5 in the Supporting Information, one has

$$\left|\left\{\left(\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial\alpha_k\partial\alpha_l}\right|_{\boldsymbol{\alpha}=\mathbf{r}_k}\right)_{k,l=1}^{p}\right\}^{-1}\right| + \left|\left\{\left(\left.\frac{\partial^2\widehat{Q}_n(\boldsymbol{\alpha})}{\partial\alpha_k\partial\alpha_l}\right|_{\boldsymbol{\alpha}=s\mathbf{s}_k}\right)_{k,l=1}^{p}\right\}^{-1}\right| = O_p(1),$$

$$\left\{\left(\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial\alpha_k\partial\alpha_l}\right|_{\boldsymbol{\alpha}=\mathbf{r}_k}\right)_{k,l=1}^{p}\right\}^{-1} = \left\{\left.\frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^{\mathrm{T}}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\right\}^{-1} + o_p(1),$$

 © 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522

$$\left\{ \left(\left. \frac{\partial^2 \hat{Q}_n(\boldsymbol{\alpha})}{\partial \alpha_k \partial \alpha_l} \right|_{\boldsymbol{\alpha}=\mathbf{s}_k} \right)_{k,l=1}^{p} \right\}^{-1} = \left\{ \left. \frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^{\mathrm{T}}} \right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} \right\}^{-1} + o_p(1),$$

and hence

$$\begin{aligned} &\tilde{\boldsymbol{\alpha}}_{\mathrm{MLE}} - \hat{\boldsymbol{\alpha}}_{\mathrm{MLE}} \\ &\quad = \left\{ \left. \frac{\partial^2 Q_n(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^{\mathrm{T}}} \right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} \right\}^{-1} \left\{ \left. \frac{\partial \hat{Q}_n(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} - \left. \frac{\partial Q_n(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} \right\} \\ &\qquad + o_p\left(n^{-1/2}\right), \\ &\hat{\boldsymbol{\alpha}}_{\mathrm{MLE}} - \tilde{\boldsymbol{\alpha}}_{\mathrm{MLE}} \\ &\quad = \omega_0 \mathbf{J}^{-1} \mathsf{E}\left\{ \sigma_{p+1}^{-4}(\boldsymbol{\alpha}_0)\mathbf{M} \right\} (n-p)^{-1} \sum_{t=p+1}^{n} Z_t + o_p\left(n^{-1/2}\right) \end{aligned} \tag{A.9}$$

with vector $\mathbf{M}$ defined in (2.5).

For vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p$, it follows from (A.8) and (A.9) that

$$\begin{aligned} &\mathbf{a}^{\mathrm{T}}\left(\tilde{\boldsymbol{\alpha}}_{\mathrm{MLE}} - \boldsymbol{\alpha}_0\right) + \mathbf{b}^{\mathrm{T}}\left(\hat{\boldsymbol{\alpha}}_{\mathrm{MLE}} - \tilde{\boldsymbol{\alpha}}_{\mathrm{MLE}}\right) \\ &\quad = \mathbf{a}^{\mathrm{T}}\mathbf{J}^{-1} \left\{ (n-p)^{-1} \sum_{t=p+1}^{n} \sigma_t^{-4}(\boldsymbol{\alpha}_0) Z_{t-k} \xi_t \right\}_{k=1}^{p} \\ &\qquad + \mathbf{b}^{\mathrm{T}} \left\{ \omega_0 \mathbf{J}^{-1} \left[\mathsf{E}\left\{ \sigma_{p+1}^{-4}(\boldsymbol{\alpha}_0)\mathbf{M} \right\} \right] (n-p)^{-1} \sum_{t=p+1}^{n} Z_t \right\} + o_p\left(n^{-1/2}\right). \end{aligned}$$

Thus

$$\mathbf{a}^{\mathrm{T}}\left(\tilde{\boldsymbol{\alpha}}_{\mathrm{MLE}} - \boldsymbol{\alpha}_0\right) + \mathbf{b}^{\mathrm{T}}\left(\hat{\boldsymbol{\alpha}}_{\mathrm{MLE}} - \tilde{\boldsymbol{\alpha}}_{\mathrm{MLE}}\right) = (n-p)^{-1} \sum_{t=p+1}^{n} \eta_t + o_p\left(n^{-1/2}\right),$$

in which

$$\eta_t = \mathbf{a}^{\mathrm{T}}\mathbf{J}^{-1} \left\{ \sigma_t^{-4}(\boldsymbol{\alpha}_0) Z_{t-k} \right\}_{k=1}^{p} \xi_t + \omega_0 \left\{ \mathsf{E}\sigma_{p+1}^{-4}(\boldsymbol{\alpha}_0)\mathbf{M} \right\}^{\mathrm{T}} \mathbf{J}^{-1}\mathbf{b} Z_t.$$

Since according to (2.1), $\xi_t = \sigma_t^2(\boldsymbol{\alpha}_0)\left(\epsilon_t^2 - 1\right)$ with $\epsilon_t^2 - 1$ being mean zero and independent of $\mathcal{F}_{t-1}$, hence $\mathsf{E}\left(\xi_t | \mathcal{F}_{t-1}\right) \equiv 0$. So $\left\{\eta_t\right\}_{-\infty}^{+\infty}$ is strictly stationary with mean zero and finite variance according to Assumption (b). Furthermore, for $l = 0, 1, \ldots$

$$\begin{aligned} \mathsf{E}\eta_t \eta_{t-l} = {} & \delta_{0l}\left(\mathsf{E}\epsilon^4 - 1\right) \mathbf{a}^{\mathrm{T}}\mathbf{J}^{-1}\mathbf{a} \\ & + \omega_0^2 \gamma(0, l)\, \mathbf{b}^{\mathrm{T}} \mathsf{E}\left\{ \sigma_{p+1}^{-4}(\boldsymbol{\alpha}_0)\mathbf{M} \right\} \mathsf{E}\left\{ \sigma_{p+1}^{-4}(\boldsymbol{\alpha}_0)\mathbf{M} \right\}^{\mathrm{T}} \mathbf{b} \\ & + \omega_0 \mathbf{a}^{\mathrm{T}}\mathbf{J}^{-1} \left\{ \mathsf{E}\sigma_{t-l}^{-4}(\boldsymbol{\alpha}_0) Z_t Z_{t-l-k} \xi_{t-l} \right\}_{k=1}^{p} \mathsf{E}\left\{ \sigma_{p+1}^{-4}(\boldsymbol{\alpha}_0)\mathbf{M} \right\}^{\mathrm{T}} \mathbf{J}^{-1}\mathbf{b}, \end{aligned}$$

 © 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522

in which δ_{0l} is the Kronecker symbol. According to Theorem 7.1.2 of Brockwell and Davis (1991), one has

$$n^{1/2} \times (n-p)^{-1} \sum_{t=p+1}^{n} \eta_t \xrightarrow{D} N(0, \nu),$$

where

$$\begin{aligned}
\nu &= \left(\mathsf{E}\epsilon^4 - 1\right) \mathbf{a}^{\mathrm{T}} \mathbf{J}^{-1} \mathbf{a} + \omega_0^2 \sum_{l=-\infty}^{+\infty} \gamma(0,l)\, \mathbf{b}^{\mathrm{T}} \mathsf{E}\left\{\sigma_{p+1}^{-4}(\boldsymbol{\alpha}_0)\mathbf{M}\right\} \mathsf{E}\left\{\sigma_{p+1}^{-4}(\boldsymbol{\alpha}_0)\mathbf{M}\right\}^{\mathrm{T}} \mathbf{b} \\
&\quad + 2\omega_0 \mathbf{a}^{\mathrm{T}} \mathbf{J}^{-1} \sum_{l=1}^{+\infty} \left\{\mathsf{E}\sigma_{t-l}^{-4}(\boldsymbol{\alpha}_0) Z_t Z_{t-l-k} \xi_{t-l}\right\}_{k=1}^{p} \mathsf{E}\left\{\sigma_{p+1}^{-4}(\boldsymbol{\alpha}_0)\mathbf{M}\right\}^{\mathrm{T}} \mathbf{J}^{-1}\mathbf{b} \\
&= \left(\mathsf{E}\epsilon^4 - 1\right) \mathbf{a}^{\mathrm{T}} \mathbf{J}^{-1} \mathbf{a} + \mathbf{b}^{\mathrm{T}} \mathbf{J}^{-1} \boldsymbol{\Sigma}_{\mathrm{diff}} \mathbf{J}^{-1} \mathbf{b} \\
&\quad + 2\omega_0 \mathbf{a}^{\mathrm{T}} \mathbf{J}^{-1} \sum_{l=1}^{+\infty} \left\{\mathsf{E}\sigma_{t-l}^{-4}(\boldsymbol{\alpha}_0) Z_t Z_{t-l-k} \xi_{t-l}\right\}_{k=1}^{p} \mathsf{E}\left\{\sigma_{p+1}^{-4}(\boldsymbol{\alpha}_0)\mathbf{M}\right\}^{\mathrm{T}} \mathbf{J}^{-1}\mathbf{b} \\
&= \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \left(\mathsf{E}\epsilon^4 - 1\right)\mathbf{J}^{-1} & \boldsymbol{\Sigma}_{\mathrm{cov}}^{\mathrm{T}} \\ \boldsymbol{\Sigma}_{\mathrm{cov}} & \mathbf{J}^{-1}\boldsymbol{\Sigma}_{\mathrm{diff}}\mathbf{J}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}.
\end{aligned}$$

Applying the Cramér-Wold device,

$$n^{1/2} \begin{pmatrix} \widetilde{\boldsymbol{\alpha}}_{\mathrm{MLE}} - \boldsymbol{\alpha}_0 \\ \widehat{\boldsymbol{\alpha}}_{\mathrm{MLE}} - \widetilde{\boldsymbol{\alpha}}_{\mathrm{MLE}} \end{pmatrix} \xrightarrow{D} N\left(\mathbf{0}_{2p\times 1}, \begin{pmatrix} \left(\mathsf{E}\epsilon^4 - 1\right)\mathbf{J}^{-1} & \boldsymbol{\Sigma}_{\mathrm{cov}}^{\mathrm{T}} \\ \boldsymbol{\Sigma}_{\mathrm{cov}} & \mathbf{J}^{-1}\boldsymbol{\Sigma}_{\mathrm{diff}}\mathbf{J}^{-1} \end{pmatrix}\right).$$

Hence, for $\boldsymbol{\Sigma} = \mathbf{J}^{-1}\boldsymbol{\Sigma}_{\mathrm{diff}}\mathbf{J}^{-1} + \left(\mathsf{E}\epsilon^4 - 1\right)\mathbf{J}^{-1} + \boldsymbol{\Sigma}_{\mathrm{cov}} + \boldsymbol{\Sigma}_{\mathrm{cov}}^{\mathrm{T}}$,

$$n^{1/2}\left(\widehat{\boldsymbol{\alpha}}_{\mathrm{MLE}} - \boldsymbol{\alpha}_0\right) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

The proof is complete. □

Proof of Theorem 3.1. See the Supporting Information. □

 © 2020 John Wiley & Sons Ltd
DOI: 10.1111/jtsa.12522