Econometric Theory, **35**, 2019, 943–977.
doi:10.1017/S0266466618000348

STATISTICAL INFERENCE FOR MEASUREMENT EQUATION SELECTION IN THE LOG-REALGARCH MODEL

YU-NING LI AND YI ZHANG
Zhejiang University

CAIYA ZHANG
Zhejiang University City College

This article investigates the statistical inference problem of whether a measurement equation is self-consistent in the logarithmic realized GARCH model (log-RealGARCH). First, we provide the sufficient and necessary conditions for the strict stationarity of both the log-RealGARCH model and the log-GARCH-X model. Under these conditions, strong consistency and asymptotic normality of the quasi-maximum likelihood estimators of these two models are obtained. Then, based on the asymptotic results, we propose a Hausman-type self-consistency test for diagnosing the suitability of the measurement equation in the log-RealGARCH model. Finally, the results of simulations and an empirical study are found to accord with the theoretical results.

1. INTRODUCTION

Forecasting of the volatility of stock return series, exchange rate return series or interest rate series is one of the critical issues in finance and is involved in capital asset pricing, portfolio selection, risk management and other financial decisions. To capture the unobserved volatility, the GARCH process and its extensions are widely used. These models have been thoroughly explored since the initial investigations of Engle (1982) and Bollerslev (1986).

The GARCH-X model is one type of extension of the GARCH model; see Han and Kristensen (2014), Han (2015), and reference therein. Covariates are

We especially thank two anonymous referees and Pentti Saikkonen, the co-editor, for helpful suggestions that greatly improved the article. We are also grateful for the comments from seminar and conference participants at the 2nd International Symposium on Interval Data Modelling: Theory and Applications (SIDM2016) and Contributed Sessions in Statistics (CSS09) of 2017 IMS-China International Conference on Statistics and Probability. This research is partly supported by the Zhejiang Provincial Natural Science Foundation (No. LY18A010005), the Research Project of Humanities and Social Science of Ministry of Education of China (No. 17YJA910003), the Fundamental Research Funds for the Central Universities and Major Project of the National Social Science Foundation of China (No.13&ZD163). Address correspondence to Yi Zhang, School of Mathematical Sciences, Zhejiang University, Hangzhou, China; e-mail: zhangyi63@zju.edu.cn and Caiya Zhang, Department of Statistics, Zhejiang University City College, Hangzhou, China; e-mail: zhangcy@zucc.edu.cn.

© Cambridge University Press 2018

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

added into the model, which are usually easily observed and carry information on the volatility. These covariates can be constructed from either low-frequency or high-frequency data, e.g., interest rate levels, bid-ask spreads, interest rate spreads, forward-spot spreads, futures open interest, information flow, and realized volatility. The inclusion of an additional covariate helps explain the volatilities, thereby leading to better performance in terms of both in-sample fit and out-of-sample forecasting. The covariate can also be considered in the log-GARCH model, which leads to the log-GARCH-X model. This model was recently studied by Sucarrat, Grønneberg, and Escribano (2016) and Francq and Sucarrat (2017) via the ARMA-X approach. However, as we do not know the future value of the covariate, only one-step forecasting is possible. Thus, we cannot perform multi-step forecasting using the (log-)GARCH-X model.

Within the (log-)GARCH-X models, 'X' is regarded as an exogenous variable. However, in practice, the covariates mentioned above may have values that shift dynamically as part of a functional relationship among volatilities or returns within the model. Therefore, it is reasonable to consider them as endogenous variables. Hansen, Huang, and Shek (2012) proposed the realized GARCH model (RealGARCH), in which a measurement equation is incorporated into the GARCH-X model, which unveils the hidden principle of the covariate and, in this sense, makes the model 'complete'. Owing to the existence of the measurement equation, one can easily forecast the covariate. Thus, multi-step forecasting can be performed iteratively. This model is named after the realized measure of volatility, which is used as the covariate. The realized volatility is constructed from intra-daily returns, which are used to consistently estimate the integrated volatility; see, for example, Barndorff-Nielsen and Shephard (2002).

The RealGARCH model has two main specifications: Hansen et al. (2012) documented that the RealGARCH model with log-linear specification (log-RealGARCH) outperforms the linear RealGARCH model, as the volatility of the linear RealGARCH model has a pronounced degree of heteroscedasticity. Therefore, the residuals of the log-RealGARCH model are more likely to be two independent normally distributed variables. An additional advantage of log-volatility is that it ensures a positive variance without positivity constraints, which is the same as in the EGARCH model of Nelson (1991) and the log-GARCH model of Geweke (1986) and Pantula (1986). The RealGARCH model also has various extensions, including the realized EGARCH model of Hansen and Huang (2016) and the realized HAR GARCH model of Huang, Liu, and Wang (2016).

This article mainly focuses on the problem of statistical inference of the measurement equation. As Hansen et al. (2012) discussed, the forms of the measurement equation are varied. Even if the variance equation is of the same form, different forms of the measurement equations induce different models, thereby leading to different estimators and, consequently, yielding different volatility forecasting performances. Thus, selecting a proper measurement equation is

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

important. The main contribution of this article is a test for the selection of the measurement equation in the RealGARCH model. We define the measurement equation (or the model) as self-consistent with respect to some data generation process (DGP) if the common parts of the estimators in the complete model (log-RealGARCH model) and the corresponding incomplete model (log-GARCH-X model) converge to the same limit almost surely. Based on this, we construct a Hausman-type test by comparing two estimates. Then, the efficiency of the estimators is discussed. We call this test a self-consistency test rather than the Hausman test since it tests the self-consistency of the measurement equation rather than the endogeneity of the covariate. As the test depends on consistency and asymptotic normality of the estimators of a specific model, our analysis hereafter only considers the log-RealGARCH model. However, it is easy to extend this test to other extensions of the RealGARCH model.

The rest of this article is organized as follows: In Section 2, we discuss sufficient and necessary conditions for the strict stationarity of the log-RealGARCH and log-GARCH-X models. In Section 3, we establish the consistency and asymptotic normality of the quasi-maximum likelihood (QML) estimators under regularity conditions. In Section 4, the asymptotic efficiency of the estimators is discussed, and a self-consistency test is constructed for the measurement equation selection. Section 5 presents the small-sample performance of the QML estimators and the proposed test by simulation. Then, we apply the test to real data. Section 6 presents the conclusions. Appendix A contains mathematical proofs for the technical results in the article.

The following notations are used in this article: Denote the sets of all real numbers, complex numbers, and integers by $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Z}$, respectively. Let $C(K, \mathbb{R})$ denote all real-valued continuous functions on K. Let $||\cdot||$ signify the usual Euclidean norm for vectors and Frobenius norm for matrices. Recall that the Frobenuis norm of a matrix $\mathbf{A} = (a_{ij})$ is defined by $||\mathbf{A}|| = (\sum_{i,j} a_{ij}^2)^{1/2}$. We also define the norm of a continuous matrix-valued function $\mathbf{A}(s)$ on a compact set K by $||\mathbf{A}||_K = \sup_{s\in K} ||\mathbf{A}(s)||$. Standard terminologies and notations in probability are used throughout the article. Let $\mathcal{N}$ and $\mathcal{U}$ denote the Gaussian distribution and the uniform distribution, respectively. By n we denote the sample size, and all limits are taken as $n \to \infty$. Let $\ln^+(x)$ denote $\ln(x) \vee 0$, and let $a \vee b$ and $a \wedge b$ denote $\max\{a, b\}$ and $\min\{a, b\}$, respectively.

2. MODEL SPECIFICATION AND STRICT STATIONARITY

2.1. Log-RealGARCH and Log-GARCH-X Models

Denote the return by R_t and the observed covariate by X_t. Let $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$ be a filtration such that the time series $\{(R_t, X_t)\}$ are adapted to it, e.g., $\mathcal{F}_t = \sigma(R_s, X_s, s \le t)$. If $X_t > 0$ almost surely, then the realized GARCH(p, q) model with log-linear specification (log-RealGARCH(p, q)) is given by

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

$$\begin{cases} R_t = \sigma_t z_t, & \textbf{(1)} \\ \ln \sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \ln \sigma_{t-i}^2 + \sum_{j=1}^{q} \gamma_j \ln X_{t-j}, & \textbf{(2)} \\ \ln X_t = \mu + \varphi \ln \sigma_t^2 + \tau(z_t) + u_t. & \textbf{(3)} \end{cases}$$

The first equation is called a mean equation with innovation z_t and conditional variance σ_t^2. The conditional variance is $\mathcal{F}_{t-1}$-measurable, which is ruled by the second equation, i.e., the variance equation or GARCH equation. The last equation is called a measurement equation, and u_t is called the measurement error. The measurement equation allows us to use a realized measure that is computed from a shorter period, and the realized measure is no longer a consistent estimate of the volatility when $\mu \neq 0$ and $\varphi \neq 1$. If only the first two equations (1) and (2) are considered in a model without a measurement equation, then the covariate X_t is regarded as an exogenous variable, and this model is a log-GARCH-X(p, q) model.

We assume that z_t and u_t are i.i.d. mutually independent random variables, with mean zero and variances 1 and σ_u^2, respectively, as Hansen et al. (2012) did, to ensure the model's identifiability. The leverage function $\tau(\cdot)$ generates an asymmetric response in volatility to return shocks, i.e., the leverage effect. Generally, it can be constructed from Hermite polynomials, e.g., $\tau(z) = \tau_1 z + \tau_2(z^2 - 1) + \tau_3(z^3 - 3z)$. Since the form of the leverage function does not have a substantial impact on the asymptotic results, we only focus on the quadratic leverage function, namely, $\tau(z) = \tau_1 z + \tau_2(z^2 - 1)$, to be concise.

We use the following notations henceforth: Denote the unknown parameter by $\boldsymbol{\theta}$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \sigma_u^2)^T$, where $\boldsymbol{\theta}_1 = (\omega, \beta_1, \ldots, \beta_p, \gamma_1, \ldots, \gamma_q)^T$ and $\boldsymbol{\theta}_2 = (\mu, \varphi, \tau_1, \tau_2)^T$.

2.2. Strict Stationarity

The sufficient and necessary conditions for the existence of a strict stationary solution (stationarity, for short) of the GARCH(p, q) model were proved by Bougerol and Picard (1992). Straumann and Mikosch (2006) derived the stationarity conditions for other types of GARCH models with positivity constraints, e.g., the AGARCH and EGARCH models, using the top Lyapunov exponent technique.

We discuss the sufficient and necessary conditions for stationarity for the log-RealGARCH model and a sufficient condition for the log-GARCH-X model in this section. The sufficient and necessary conditions for the log-GARCH-X model are unobtainable since they rely on the unknown generation process of the exogenous variables $\ln X_t$.

The log-RealGARCH model has an ARMA representation, which is expressed as follows and was presented by Hansen et al. (2012):

$$\ln \sigma_t^2 = \mu_1 + \sum_{i=1}^{p} \beta_i \ln \sigma_{t-i}^2 + \sum_{j=1}^{q} \varphi \gamma_j \ln \sigma_{t-j}^2 + \sum_{j=1}^{q} \gamma_j w_{t-j}, \, t \in \mathbb{Z}, \quad \textbf{(4)}$$

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

where $\mu_1 = \omega + \gamma_\bullet \mu$, $\gamma_\bullet = \sum_{j=1}^{q} \gamma_j$, and $w_t = \tau(z_t) + u_t$.

Let $k = p \wedge q$, $\mathcal{A}_{\boldsymbol{\theta}_1}(z) = \sum_{j=1}^{q} \gamma_j z^j$ and $\mathcal{C}_{\boldsymbol{\theta}}(z) = 1 - \sum_{i=1}^{k} (\beta_i + \varphi \gamma_j) z^i$, with the convention $\beta_i = 0\,(\gamma_j = 0)$ if $i > q\,(j > p)$. We can rewrite equation (4) in backshift operator notation:

$$\mathcal{C}_{\boldsymbol{\theta}}(B) \ln \sigma_t^2 = \mu_1 + \mathcal{A}_{\boldsymbol{\theta}_1}(B) w_t, \tag{5}$$

where B is the backshift operator.

The sufficient and necessary stationarity conditions of $\{\ln \sigma_t^2\}$ are given in the following proposition, which follows directly from the causality of the classical ARMA time series (cf. Brockwell and Davis, 2009, Thms. 3.1.1 and 3.1.3).

PROPOSITION 2.1. *Assume that the polynomials $\mathcal{A}_{\boldsymbol{\theta}_1}(z)$ and $\mathcal{C}_{\boldsymbol{\theta}}(z)$ have no common roots. Then, with respect to the log-RealGARCH model, there exists a stationary solution $\{\ln \sigma_t^2\}$ if and only if $\mathcal{C}_{\boldsymbol{\theta}}(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$.*

Next, we turn to the log-GARCH-X model. Let $\mathcal{B}_{\boldsymbol{\theta}_1}(z) = 1 - \sum_{i=1}^{p} \beta_i z^i$. The variance equation (2) can be rewritten as

$$\mathcal{B}_{\boldsymbol{\theta}_1}(B) \ln \sigma_t^2 = \omega_1 + \mathcal{A}_{\boldsymbol{\theta}_1}(B) \ln X_t. \tag{6}$$

The stationarity of the log-GARCH-X model is given in the following proposition. Similar results were also discussed by Francq and Sucarrat (2017).

PROPOSITION 2.2. *Suppose that $\{\ln X_t\}$ is a stationary and ergodic process with $E[\ln^+(|\ln X_t|)] < \infty$. Assume that the polynomials $\mathcal{A}_{\boldsymbol{\theta}_1}(z)$ and $\mathcal{B}_{\boldsymbol{\theta}_1}(z)$ have no common roots. If $\mathcal{B}_{\boldsymbol{\theta}_1}(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$, then with respect to the log-GARCH-X model, there exists a stationary solution $\{\ln \sigma_t^2\}$.*

Remark 2.1. The condition $\mathsf{E}[\ln^+(|\ln X_t|)] < \infty$ is a very weak moment condition. Note that $\mathsf{E}|\ln X_t|^\eta < \infty$, for some $\eta > 0$, implies immediately the condition $\mathsf{E}[\ln^+(|\ln X_t|)] < \infty$; see Lemma 2.5.3 in Straumann (2005).

Remark 2.2. For a log-RealGARCH model, the linear representation (6) is only a sufficient condition for the existence of the stationary solution $\ln \sigma_t^2$. Combining Propositions 2.1 and 2.2, we obtain that the stationarity of $\ln \sigma_t^2$ in the log-RealGARCH model does not necessarily ensure that $\ln \sigma_t^2$ can be written as a linear combination of $\{\ln X_t\}$.

Proposition 2.2 holds for the log-RealGARCH model. Moreover, we have the following proposition:

PROPOSITION 2.3. *Assume that the polynomials $\mathcal{A}_{\boldsymbol{\theta}_1}(z)$ and $\mathcal{B}_{\boldsymbol{\theta}_1}(z)$ have no common roots. If $\mathcal{B}_{\boldsymbol{\theta}_1}(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$, then the solution $\ln \sigma_t^2$ of the log-RealGARCH model has a linear representation in terms of $\{\ln X_t\}$.*

Remark 2.3. Similarly to the discussion in Remark 2.1, the moment condition of $\ln X_t$ is fulfilled by $\mathsf{E}[w_t] = 0$ thus $\mathsf{E}|w_t| < \infty$ (see Proposition 2.1 and Lemma

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

A.2). $\{\ln X_t\}$ is a stationary and ergodic process because $\{\ln \sigma_t^2\}$ is stationary and ergodic.

The stationarity property and linear representation will play an essential role in the estimation for both the log-GARCH-X model and the log-RealGARCH model. If the stationary solution $\{\ln \sigma_t^2\}$ exists, then because of the ARMA representation (4), it must be ergodic (cf. Billingsley, 1995, pp. 284 and 495). For the log-RealGARCH model, since $\{w_t\}$ are i.i.d. random variables, $\{(R_t, \ln X_t)\}$ are stationary and ergodic.

3. QMLEs AND THEIR ASYMPTOTIC PROPERTIES

Quasi-maximum likelihood estimation (QMLE) is the most popular method for estimating GARCH-type models; see Lee and Hansen (1994). Francq et al. (2013) established the strong consistency and the asymptotic normality of the QML estimator of the log-GARCH model under mild assumptions, which provides inspiration for our proof. An extra term of the measurement error in the log-likelihood function makes the proof of the asymptotic normality of the QML estimator of the log-RealGARCH model nontrivial.

Let Θ denotes the parameter space of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \sigma_u^2)^T$. For simplicity, we assume $\Theta = \Theta_1 \times \Theta_2 \times \Theta_3$. The true value of parameter $\boldsymbol{\theta}$ is denoted by $\boldsymbol{\theta}^\circ = (\boldsymbol{\theta}_1^{\circ T}, \boldsymbol{\theta}_2^{\circ T}, \sigma_u^{2\circ})^T$. To construct QMLEs for both models, we must rewrite the variance equation. Replacing $\ln \sigma_t^2$ with $\ln h_t(\boldsymbol{\theta}_1)$, we have

$$\ln h_t(\boldsymbol{\theta}_1) = \omega + \sum_{i=1}^{p} \beta_i \ln h_{t-i}(\boldsymbol{\theta}_1) + \sum_{j=1}^{q} \gamma_j \ln X_{t-j}.$$

We have now established a function $\ln h_t(\boldsymbol{\theta}_1)$, which equals $\ln \sigma_t^2$ when $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^\circ$. Note that $\ln h_t(\boldsymbol{\theta}_1)$ may not be well-defined in this recursive way. Therefore, to give a more specific definition of $\ln h_t$, we assume the following:

C1: $\mathcal{A}_{\boldsymbol{\theta}_1}(z)$ and $\mathcal{B}_{\boldsymbol{\theta}_1}(z)$ do not have any common roots, and $|\mathcal{B}_{\boldsymbol{\theta}_1}(z)| = 0 \Rightarrow |z| > 1, \forall \boldsymbol{\theta}_1 \in \Theta_1$.

If $\ln X_t$ is stationary and ergodic with $\mathrm{E}[\ln^+(|\ln X_t|)] < \infty$ and Assumption C1 holds, we may define

$$\ln(h_t(\boldsymbol{\theta}_1)) := \mathcal{B}_{\boldsymbol{\theta}_1}^{-1}(B)\left(\omega + \mathcal{A}_{\boldsymbol{\theta}_1}(B) \ln X_t\right), \boldsymbol{\theta}_1 \in \Theta_1. \tag{7}$$

The function $\ln h_t$ is a stationary and ergodic random element on $C(\Theta_1, \mathbb{R})$ with $\ln\ h_t(\boldsymbol{\theta}_1^\circ) = \ln \sigma_t^2$. Then h_t is defined by $\exp(\ln h_t)$, which is positive for all $\boldsymbol{\theta}_1 \in \Theta_1$ almost surely.

For the log-RealGARCH model, because of the existence of the measurement equation, we require g_t to be defined as follows:

$$g_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \ln X_t - \mu - \varphi \ln h_t(\boldsymbol{\theta}_1) - \tau\left(\frac{R_t}{h_t^{1/2}(\boldsymbol{\theta}_1)}\right).$$

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

Function g_t is a stationary and ergodic random element on $C(\Theta_1 \times \Theta_2, \mathbb{R})$ with $g_t(\theta_1^\circ, \theta_2^\circ) = u_t$.

Note that $h_t(\boldsymbol{\theta}_1)$ and $g_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ will be used in the likelihood function. In the QMLE, we must estimate $h_t(\boldsymbol{\theta}_1)$ and $g_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. With any fixed initial values $\{\varsigma_t, t = 1-p, \ldots, 0\}$, the estimator of $\ln h_t(\boldsymbol{\theta}_1)$ is recursively defined by

$$\ln \widehat{h}_t(\boldsymbol{\theta}_1) = \begin{cases} \varsigma_t, & t = 1-p, \ldots, 0, \\ \omega + \sum_{i=1}^{p} \beta_i \ln \widehat{h}_{t-i}(\boldsymbol{\theta}_1) + \sum_{j=1}^{\min\{q, t-1\}} \gamma_j \ln X_{t-j}, & t > 0, \end{cases} \tag{8}$$

and

$$\widehat{g}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \ln X_t - \mu - \varphi \ln \widehat{h}_t(\boldsymbol{\theta}_1) - \tau \left(\frac{R_t}{\widehat{h}_t^{1/2}(\boldsymbol{\theta}_1)} \right).$$

3.1. Estimation of the Log-RealGARCH Model

The infeasible quasi-log-likelihood function of the log-RealGARCH model is given by

$$l_t(\boldsymbol{\theta}) = \ln h_t(\boldsymbol{\theta}_1) + \frac{R_t^2}{h_t(\boldsymbol{\theta}_1)} + \ln(\sigma_u^2) + \frac{g_t^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\sigma_u^2},$$

which depends on mutually independent normally distributed errors. This is why the estimation method is called the (Gaussian) QMLE. However, we do not assume that z_t and u_t are normally distributed. Let $\widehat{l}_t$ denote the estimator of the log-likelihood function, which is given by

$$\widehat{l}_t(\boldsymbol{\theta}) = \ln(\widehat{h}_t(\boldsymbol{\theta}_1)) + \frac{R_t^2}{\widehat{h}_t(\boldsymbol{\theta}_1)} + \ln(\sigma_u^2) + \frac{\widehat{g}_t^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\sigma_u^2}.$$

Then, a QML estimator of $\boldsymbol{\theta}^\circ$ is defined as any measurable solution $\widehat{\boldsymbol{\theta}}_n$ of

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta} \in \Theta} \widehat{L}_n(\boldsymbol{\theta}),$$

where $\widehat{L}_n(\boldsymbol{\theta}) = n^{-1} \sum_{t=1}^{n} \widehat{l}_t(\boldsymbol{\theta})$.

To discuss the asymptotic properties, we require the following assumptions:

- **C2**: Let $\tau(z) = \tau_1 z + \tau_2(z^2 - 1)$. For any real number $(\widetilde{\tau}_1, \widetilde{\tau}_2) \neq (0, 0)$, $\widetilde{\tau}_1 z_t + \widetilde{\tau}_2 z_t^2$ does not degenerate, $\forall t \in \mathbb{Z}$.
- **C3**: $\mathcal{A}_{\boldsymbol{\theta}_1}(z)$ and $\mathcal{C}_{\boldsymbol{\theta}}(z)$ do not have any common zeros, and $|\mathcal{C}_{\boldsymbol{\theta}}(z)| = 0 \Rightarrow |z| > 1$, $\forall \boldsymbol{\theta} \in \Theta$.
- **C4**: There exists $s > 0$ such that $\mathsf{E}[\exp(s|\ln X_t|)] < \infty$.

Assumption C2 is necessary for the identifiability of the log-RealGARCH models, which means the distribution of z_t is not concentrated at two points. Assumption C3 ensures that the model is future-independent, i.e., $\ln \sigma_t^2$ is $\mathcal{F}_{t-1}$-measurable.

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

Assumption C4 on $|\ln X_t|$ suggests that the density function of X_t around 0 does not explode too fast; see Proposition 3.2 of Francq et al. (2013). By the measurement equation (3), and the representation (5), Assumption C4 can be rewritten in the expression of w_t°, as follows,

$$\mathsf{E}[\exp(s|\mu^\circ + \varphi^\circ \mathcal{C}_{\boldsymbol{\theta}^\circ}^{-1}(B)(\mu_1^\circ + \mathcal{A}_{\boldsymbol{\theta}_1^\circ}(B)w_t^\circ) + w_t^\circ|)] < \infty,$$

where $w_t^\circ = \tau^\circ(z_t) + u_t = \tau_1^\circ z_t + \tau_2^\circ z_t^2 + u_t$. However, we still use the formulation in Assumption C4 for simplicity and clarity.

THEOREM 3.1. *Suppose that* $\{\ln\sigma_t^2\}$ *is a stationary log-RealGARCH*(p,q) *process defined by equations (1)–(3) with parameters* $\boldsymbol{\theta}^\circ$. *If* $\boldsymbol{\theta}^\circ \in \Theta$, *where* Θ *is compact, and assumptions C1–C3 hold, then* $\widehat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}^\circ$ *as* $n \to \infty$.

Now, we establish the asymptotic normality of QMLE. For convenience, we use the following notation throughout the article: Let $\mathsf{E}^\circ[\cdot] = \mathsf{E}[\cdot|\boldsymbol{\theta} = \boldsymbol{\theta}^\circ]$, $\nabla_1 = \partial/\partial\boldsymbol{\theta}_1$, $\nabla_2 = \partial/\partial\boldsymbol{\theta}_2$, and $\nabla_{12} = \partial/\partial(\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$. Define $\mathbf{A} = \mathsf{E}^\circ[\nabla_1(\ln h_t)\nabla_1(\ln h_t)^T]$, $\mathbf{F} = \mathsf{E}[z_t^4 - 3]\mathbf{A}$, $\mathbf{B}_{11} = \frac{2}{\sigma_u^{2\circ}}\mathsf{E}^\circ[\nabla_1 g_t \nabla_1 g_t^T]$, $\mathbf{B}_{21} = \frac{2}{\sigma_u^{2\circ}}\mathsf{E}^\circ[\nabla_2 g_t \nabla_1 g_t^T]$, and $\mathbf{B}_{22} = \frac{2}{\sigma_u^{2\circ}}\mathsf{E}^\circ[\nabla_2 g_t \nabla_2 g_t^T]$. We have the following theorem:

THEOREM 3.2. *Suppose that the assumptions in Theorem 3.1 and Assumption C4 are satisfied. If* $\boldsymbol{\theta}^\circ$ *is in the interior of* Θ, $\mathsf{E}|z_t|^4 < \infty$ *and* $\mathsf{E}|u_t|^4 < \infty$, *then*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^\circ) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}),$$

where $\mathbf{V} = \mathbf{H}^{-1}\mathbf{J}\mathbf{H}^{-1}$,

$$\mathbf{J} = \begin{bmatrix} \mathbf{F} + 2\mathbf{A} + 2\mathbf{B}_{11} & \bullet & \bullet \\ 2\mathbf{B}_{21} & 2\mathbf{B}_{22} & \bullet \\ -2\frac{\mathsf{E}[u_t^3]}{(\sigma_u^{2\circ})^3}\mathsf{E}^\circ[\nabla_1 g_t^T] & -2\frac{\mathsf{E}[u_t^3]}{(\sigma_u^{2\circ})^3}\mathsf{E}^\circ[\nabla_2 g_t^T] & \frac{\mathsf{E}[u_t^4]-(\sigma_u^{2\circ})^2}{(\sigma_u^{2\circ})^4} \end{bmatrix}, \text{ and}$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{A} + \mathbf{B}_{11} & \mathbf{B}_{21}^T & \mathbf{0} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0}^T & \frac{1}{(\sigma_u^{2\circ})^2} \end{bmatrix}.$$

In addition, the asymptotic covariance matrix $\mathbf{V}$ *is singular if and only if* $u_t^2 \equiv \sigma_u^{2\circ}$ *for all* t.

Remark 3.1. If a higher-order Hermite polynomial leverage function $\tau(z)$ is used, we require a higher-order moment condition of z_t. In general, $\mathsf{E}[\tau^2(z_t)] < \infty$ is a necessary condition for asymptotic normality.

Remark 3.2. If a different measurement equation than the log-linear measurement equation (3) is used, e.g., a linear measurement equation or a log-linear

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

measurement equation with extra lag terms, we can still establish the asymptotic normality with proper stationary assumption and moment assumption. The asymptotic variance matrix is expressed in the same form as in Theorem 3.2.

3.2. Estimation of the Log-GARCH-X Model

The log-GARCH-X model does not have the measurement equation. Thus, we only need to estimate the true parameters $\boldsymbol{\theta}_1^{\circ(GX)}$ of the model. GX is the abbreviation for the log-GARCH-X model for distinguishing the true value $\boldsymbol{\theta}^{\circ}$ of the log-RealGARCH model.

The infeasible quasi-log-likelihood function of the log-GARCH-X model can be viewed as the partial log-likelihood function of the log-RealGARCH model, which is given by

$$l_t^{(GX)}(\boldsymbol{\theta}_1) = \ln(h_t(\boldsymbol{\theta}_1)) + \frac{R_t^2}{h_t(\boldsymbol{\theta}_1)}.$$

The estimator of this function is given by $\widetilde{l}_t^{(GX)}(\boldsymbol{\theta}_1) = \ln(\widehat{h}_t(\boldsymbol{\theta}_1)) + \frac{R_t^2}{\widehat{h}_t(\boldsymbol{\theta}_1)}$.

Then, the QML estimator of $\boldsymbol{\theta}_1^{\circ(GX)}$ is defined as any measurable solution $\widetilde{\boldsymbol{\theta}}_{1,n}$ of

$$\widetilde{\boldsymbol{\theta}}_{1,n} = \underset{\boldsymbol{\theta}_1 \in \Theta_1}{\arg\min}\, \widetilde{L}_n^{(GX)}(\boldsymbol{\theta}_1),$$

where $\widetilde{L}_n^{(GX)}(\boldsymbol{\theta}_1) = n^{-1}\sum_{t=1}^{n} \widetilde{l}_t^{(GX)}(\boldsymbol{\theta}_1)$.

Since $\mathsf{E}[l_t^{(GX)}(\boldsymbol{\theta}_1)]$ attains its minimum value at $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{\circ(GX)}$, we have the following theorem:

THEOREM 3.3. *Suppose that* $\{\ln \sigma_t^2\}$ *is a stationary log-GARCH-X*(p, q) *process, which is defined by equations (1)–(2) with parameters* $\boldsymbol{\theta}_1^{\circ(GX)}$. *Assume that* $\{\ln X_t\}$ *is a nonconstant stationary and ergodic process with* $E[\ln X_t] < \infty$. *In addition, if assumption C1 holds and* $\boldsymbol{\theta}_1^{\circ(GX)} \in \Theta_1$, *where* Θ_1 *is compact, then*

$$\widetilde{\boldsymbol{\theta}}_{1,n} \xrightarrow{a.s.} \boldsymbol{\theta}_1^{\circ(GX)}, \text{ as } n \to \infty.$$

THEOREM 3.4. *Suppose that the assumptions in Theorem 3.3 and Assumption C4 are satisfied. If* $\boldsymbol{\theta}_1^{\circ(GX)}$ *is in the interior of* Θ_1, $E[z_t^4] < \infty$ *and* $E[\ln X_t]^2 < \infty$, *then*

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}}_{1,n} - \boldsymbol{\theta}_1^{\circ(GX)}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{W}),$$

where the asymptotic covariance matrix $\mathbf{W}$ *is given by*

$$\mathbf{W} = \mathbf{A}^{-1}(\mathbf{F} + 2\mathbf{A})\mathbf{A}^{-1} = E[z_t^4 - 1](E^{\circ}[\nabla_1(\ln h_t)\nabla_1(\ln h_t)^T])^{-1}.$$

The proofs of Theorems 3.3 and 3.4 are similar to the proofs of Theorems 3.1 and 3.2. Thus, we omit the proofs.

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

3.3. Relation Between Log-RealGARCH and Log-GARCH-X

The log-RealGARCH model is a complete form of the log-GARCH-X model. Thus, they share various similar attributes. For example, there is no minimum value for the volatility, and there are no positivity constraints on the parameters. Theoretically, if $\{\ln\sigma_t^2\}$ is a stationary log-RealGARCH(p,q) process, then it also follows a stationary log-GARCH-X(p,q) process. Since the log-RealGARCH(p,q) process and the log-GARCH-X(p,q) process follow the same variance equation, the values of the common parameters in the variance equations must be the same (by definition, $\boldsymbol{\theta}_1^{\circ}$ is one of the available parameters for the log-GARCH-X model, and the uniqueness of $\boldsymbol{\theta}_1^{\circ(GX)}$ is given by Lemma A.5). Therefore, we have

$$\mathsf{H}_0 : \boldsymbol{\theta}_1^{\circ(GX)} = \boldsymbol{\theta}_1^{\circ}, \tag{9}$$

if the data follows a DGP of log-RealGARCH(p,q). Using this equality, we will solve the measurement equation selection problem in the next section.

Another relation between the log-RealGARCH model and the log-GARCH-X model can be identified from the log-likelihood function. Let $\boldsymbol{\theta}_2 = \mathbf{0}$ in $\widehat{l}_t(\boldsymbol{\theta})$. We have

$$\begin{aligned}\widehat{l}_t(\boldsymbol{\theta}_1, \mathbf{0}, \sigma_u^2) &= \ln(\widehat{h}_t(\boldsymbol{\theta}_1)) + \frac{R_t^2}{\widehat{h}_t(\boldsymbol{\theta}_1)} + \ln(\sigma_u^2) + \frac{(\ln X_t)^2}{\sigma_u^2} \\ &= \widetilde{l}_t^{(GX)}(\boldsymbol{\theta}_1) + \ln(\sigma_u^2) + \frac{(\ln X_t)^2}{\sigma_u^2}.\end{aligned}$$

Based on this equation, we can use the restricted log-RealGARCH model to estimate the log-GARCH-X model since $\sum_{t=1}^n \widehat{l}_t(\boldsymbol{\theta}_1, \mathbf{0}, \sigma_u^2)$ attains its minimum when $\boldsymbol{\theta}_1 = \widetilde{\boldsymbol{\theta}}_{1,n}$ and $\sigma_u^2 = n^{-1}\sum_{t=1}^n (\ln X_t)^2$.

4. INFERENCE OF MEASUREMENT EQUATION

The measurement equation describes the dynamic properties of $\ln X_t$, which makes multi-period-ahead forecasting feasible. As Hansen et al. (2012) noted, the measurement equation can be of any form, e.g., $\ln X_t = f(\sigma_t, z_t, u_t)$. It is essential to select the form of the measurement equation. A more suitable choice of measurement equation will lead to better performance of the model, whereas a misspecified measurement equation may cause self-inconsistency and model inadequacy.

First, we briefly discuss the test of model adequacy. One classical way to test it is to perform portmanteau tests on the residuals $\{\widehat{z}_t\}$ and $\{\widehat{u}_t\}$. We can use the Ljung-Box Q-test (LBQ test) to assesses the null hypothesis that the series of residuals exhibits no autocorrelation for a fixed number M of lags. The test statistic is defined by

$$Q(M; \{a_t\}) = n(n+2)\sum_{m=1}^{M} \frac{\widehat{\rho}_{\{a_t\},m}^2}{n-m} \tag{10}$$

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

for any time series $\{a_t\}$, where $\widehat{\rho}_{\{a_t\},m}$ is the sample autocorrelation of $\{a_t\}$ at lag m. The test statistic follows a $\chi^2(M)$ distribution asymptotically if $\{a_t\}$ is a white noise. We may need to adjust the degrees of freedom due to the use of residual series; however, we leave this for a future study and use $\chi^2(M)$ as the null distribution. The main weakness is that we may make fewer Type I errors but more Type II errors, thereby underestimating the model inadequacy. For the log-RealGARCH model, we must test the autocorrelation of both errors $\{z_t\}$ and $\{u_t\}$. Meanwhile, to determine whether there is an ARCH effect, we must repeat the test over the squared residuals (or absolute residuals); see Mcleod and Li (1983). In this article, we apply the LBQ test with $M = 20$ to the residual series $\{\widehat{z}_t\}$ and $\{\widehat{u}_t\}$ as well as $\{|\widehat{z}_t|\}$ and $\{|\widehat{u}_t|\}$.

To assess the self-consistency of the model, we construct a self-consistency test based on equation (9). Under the null hypothesis, $\boldsymbol{\theta}_1^\circ$ is the limit of both $\widehat{\boldsymbol{\theta}}_{1,n}$ and $\widetilde{\boldsymbol{\theta}}_{1,n}$. Thus only when the estimators $\widehat{\boldsymbol{\theta}}_{1,n}$ and $\widetilde{\boldsymbol{\theta}}_{1,n}$ are close to each other can a log-RealGARCH(p,q) model be applicable. In general, when we face a choice among different measurement equations, the corresponding estimates $\widehat{\boldsymbol{\theta}}_{1,n}$ should be different. Then the estimate $\widetilde{\boldsymbol{\theta}}_{1,n}$ can serve as a benchmark to determine the proper measurement equation. The estimate $\widehat{\boldsymbol{\theta}}_{1,n}$ should be close to the benchmark when an appropriate measurement equation is selected.

In the following theorem, we show the joint asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$ and $\widetilde{\boldsymbol{\theta}}_{1,n}$.

THEOREM 4.1. *Suppose that the assumptions in Theorem 3.2 are satisfied. We have*

$$\sqrt{n}((\widehat{\boldsymbol{\theta}}_n^T, \widetilde{\boldsymbol{\theta}}_{1,n}^T)^T - (\boldsymbol{\theta}^{\circ T}, \boldsymbol{\theta}_1^{\circ T})^T) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{U}),$$

and the asymptotic covariance matrix $\mathbf{U}$ *is given by*

$$\mathbf{U} = \begin{bmatrix} \mathbf{V} & \mathbf{U}_{21}^T \\ \mathbf{U}_{21} & \mathbf{W} \end{bmatrix},$$

where

$$\mathbf{U}_{21} = \left[E[z_t^4 - 1]\mathbf{G}^{-1}, \quad E[z_t^4 - 1]\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\mathbf{G}^{-1}, \quad \mathbf{0} \right],$$

and

$$\mathbf{G} = \mathbf{A} + \mathbf{B}_{11} - \mathbf{B}_{21}^T\mathbf{B}_{22}^{-1}\mathbf{B}_{21}.$$

Let $\mathbf{V}_{11}$ be the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{1,n} - \boldsymbol{\theta}_1^\circ)$. Then, it is a submatrix of $\mathbf{V}$. Note that $\mathbf{W}$ is the asymptotic covariance matrix of $\sqrt{n}(\widetilde{\boldsymbol{\theta}}_{1,n} - \boldsymbol{\theta}_1^\circ)$. We have the following corollary:

COROLLARY 4.1. *Under the assumptions of Theorem 4.1, we have*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{1,n} - \widetilde{\boldsymbol{\theta}}_{1,n}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma),$$

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

where

$$\Sigma = \mathbf{V}_{11} + \mathbf{W} - 2E[z_t^4 - 1]\mathbf{G}^{-1}.$$

In addition, the asymptotic covariance matrix Σ *is singular with rank* $p+q$.

With Corollary 4.1, we can estimate the standard error of $\widehat{\boldsymbol{\theta}}_{1,n} - \widetilde{\boldsymbol{\theta}}_{1,n}$ by plug-in method and can construct an element-wise t-test for determining whether $\widehat{\boldsymbol{\theta}}_{1,n}$ and $\widetilde{\boldsymbol{\theta}}_{1,n}$ are close to each other.

Finally, to test $\mathrm{H}_0 : \boldsymbol{\theta}_1^{\circ} = \boldsymbol{\theta}_1^{\circ(GX)}$, we construct the self-consistency statistic

$$S_n = n(\widehat{\boldsymbol{\theta}}_{1,n} - \widetilde{\boldsymbol{\theta}}_{1,n})^T \widehat{\Sigma}^{+}(\widehat{\boldsymbol{\theta}}_{1,n} - \widetilde{\boldsymbol{\theta}}_{1,n}),$$

where $\widehat{\Sigma}^{+}$ is the plug-in estimator of Σ^{+} and Σ^{+} is the Moore-Penrose generalized inverse of Σ. The consistency of the estimator can be similarly established by the argument in Remark 5.6.2 of Straumann (2005). Then, under the assumptions of Theorem 4.1, S_n asymptotically follows a χ^2-distribution with $p+q$ degree of freedom. Thus, testing of H_0 at the asymptotic level $\alpha \in (0,1)$ can be achieved by using the critical region $\{S_n > \chi^2_{1-\alpha}(p+q)\}$.

Remark 4.1. In general, a rejection of the null hypothesis could be evidence that either the GARCH equation or the measurement equation (even the mean equation) is misspecified. However, in our article, we investigate the measurement equation selection problem assuming that the GARCH equation is properly specified. Thus, we do not analyze the misspecification problem of the GARCH equation. There is no common method for dealing with this problem for the RealGARCH model. However, for other types of GARCH models, we refer the interested reader to the sign and size-bias test by Engle and Ng (1993), the Lagrange multiplier test by Lundbergh and Teräsvirta (2002) and Halunga and Orme (2009), and the rank-based test by Andreou and Werker (2015) as valuable references.

Remark 4.2. This test is a Hausman-like test; see Hausman (1978). The Hausman test evaluates the consistency of an estimator compared to a less-efficient estimator that is also consistent. The Hausman test is mainly used to determine the endogeneity of a variable. However, in our case, the realized volatility is already known as an endogenous variable, and the self-consistency test is used to evaluate the specification of the measurement equation.

As we have discussed in Section 3, $\widehat{\boldsymbol{\theta}}_{1,n}$ and $\widetilde{\boldsymbol{\theta}}_{1,n}$ are both consistent and asymptotically normally distributed under the null hypothesis. It is natural to ask which estimator performs better. We have the following theorem:

THEOREM 4.2. *Suppose that the assumptions in Theorem 3.2 are satisfied. We have*

$$\mathbf{V}_{11} \leq \mathbf{W},$$

i.e., $\mathbf{W} - \mathbf{V}_{11}$ *is a positive-semidefinite matrix.*

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

Theorem 4.2 states that if the measurement equation is correctly specified, the asymptotic variance of each element of $\widehat{\boldsymbol{\theta}}_{1,n}$ is not larger than that of the corresponding element of $\widetilde{\boldsymbol{\theta}}_{1,n}$. Therefore, the estimator $\widehat{\boldsymbol{\theta}}_{1,n}$ is more efficient.

In summary, according to the previous discussion, we select the measurement equation according to the following procedures: First, we fit a log-GARCH-X model and a log-RealGARCH model with a specific measurement equation. Then, we test the hypothesis H_0. If it is not rejected, the measurement equation is selected. Otherwise, the measurement equation leads to self-inconsistency, and we must try another measurement equation and test it again.

5. NUMERICAL EXPERIMENTS AND EMPIRICAL APPLICATIONS

We evaluate the in-sample and out-of-sample performances of the log-RealGARCH model and the log-GARCH-X model via the partial log-likelihood (PLLH) and the root-mean-squared error (RMSE). RMSE is defined by

$$\mathrm{RMSE}(\widehat{\sigma}; i) = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\sigma_t^{2(i)} - \widehat{\sigma}_t^{2(i)})^2},$$

where $\sigma_t^{2(i)}$ denotes the conditional variance at time t in the ith trial and $\widehat{\sigma}_t^2 = \widehat{h}_t(\widehat{\boldsymbol{\theta}}_{1,n})$ is its estimator. The RMSE of estimators $\widetilde{\sigma}_t^2 = \widehat{h}_t(\widetilde{\boldsymbol{\theta}}_{1,n})$ is similarly defined.

Let $\widehat{\mathbf{V}}_{11}$ and $\widetilde{\mathbf{V}}_{11}$ be the estimated asymptotic variance matrices of the estimators $\widehat{\boldsymbol{\theta}}_{1,n}$ and $\widetilde{\boldsymbol{\theta}}_{1,n}$, respectively. We illustrate the relative efficiency (RE), which was proposed by Bloomfield and Watson (1975) and is equal to

$$\mathrm{RE1} = \frac{\det(\widetilde{\mathbf{V}}_{11})}{\det(\widehat{\mathbf{V}}_{11})}, \quad \mathrm{RE2} = \frac{||\widetilde{\mathbf{V}}_{11}||}{||\widehat{\mathbf{V}}_{11}||}, \quad \mathrm{RE3} = \frac{\mathrm{tr}(\widetilde{\mathbf{V}}_{11})}{\mathrm{tr}(\widehat{\mathbf{V}}_{11})},$$

where det() and tr() are the determinant and trace of a matrix, respectively.

5.1. A Monte Carlo Experiment

In the numerical experiments, we generate data from three cases to show the test size (probability of Type I error) and test power. Typically, the measurement equations are set to be of log-linear or linear specification. We generate N=500 data sets, each consisting of n = 1,000 observations. Taking the log-RealGARCH(p, q) model as an example, we generate the data according to the following procedure:

(i) Specify the true value $\boldsymbol{\theta}^\circ$ and initial values $\{\sigma_t, t = 0, -1, \ldots, 1-p\}$ and $\{\ln X_t, t = 0, -1, \ldots, 1-q\}$.

(ii) Simulate the values of z_t and u_t i.i.d. for $t = 1, \ldots, n+500$.

(iii) Set σ_t and $\ln X_t$ iteratively by $\ln\sigma_t^2 = \omega + \sum_{i=1}^{p}\beta_i \ln\sigma_{t-i}^2 + \sum_{j=1}^{q}\gamma_j \ln X_{t-j}$ and $\ln X_t = \mu + \varphi \ln\sigma_t^2 + \tau(z_t) + u_t$.

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

(iv) Set $R_t = \sigma_t z_t$, $t = 1, \ldots, n+500$.
(v) Discard the first 500 observations from the initial samples of $n+500$.
Then, we divide the data into two parts: The first part consists of 900 observations, which are used to fit a log-RealGARCH model and a log-GARCH-X model to obtain the QML estimators. The second part consists of the remaining 100 observations, which are used to perform one-step-ahead forecasting.

Case 1. (The null hypothesis is true.)

The first case is the standard log-RealGARCH model, which has a logarithmic measurement equation:

$$\begin{cases} R_t = \sigma_t z_t, \\ \ln \sigma_t^2 = 0.125 + 0.335 \ln \sigma_{t-1}^2 + 0.514 \ln X_{t-1}, \\ \ln X_t = -0.454 + 0.980 \ln \sigma_t^2 - 0.177 z_t + 0.076(z_t^2 - 1) + u_t, \\ u_t \sim \mathcal{N}(0, 0.244). \end{cases}$$

Table 1 illustrates the result when the model is correctly specified. QML estimators $\widehat{\boldsymbol{\theta}}_{1,n}$ and $\widetilde{\boldsymbol{\theta}}_{1,n}$ are close, which accords with the theoretical result that they have the same limit. The plug-in estimators of the asymptotic variance matrices perform well, as the medians of the standard errors (Column 'Std.') are very close to the RMSEs. The test sizes of the element-wise t-tests and the self-consistency test are close to the significance level (5%). The LBQ tests also have an approximately 5% probability of incorrectly rejecting the null hypothesis.

The RMSEs of $\widehat{\boldsymbol{\theta}}_{1,n}$ are smaller than the RMSEs of $\widetilde{\boldsymbol{\theta}}_{1,n}$, which indicates that the log-RealGARCH estimators are more efficient. This is also shown by the relative efficiency, where RE1: 16.82, RE2: 3.375, and RE3: 3.333 on average. In terms of the in-sample performance, RMSE($\widehat{\sigma}$) is slightly smaller than RMSE($\widetilde{\sigma}$), which indicates that the performance of the log-RealGARCH model is better. However, the partial log-likelihood of the log-GARCH-X model is slightly larger because the partial log-likelihood is the objective function in the QMLE of the log-GARCH-X model, which is always maximized. This is also why we need RMSE($\widehat{\sigma}$) for assessing the in-sample performance. This phenomenon is observed in all simulations and practical analyses (Tables 1–5), which indicates that the log-GARCH-X model may overfit the data. The PLLH and the RMSE of out-of-sample forecasting illustrate that the log-RealGARCH model outperforms the log-GARCH-X model.

Case 2. (The null hypothesis is false.)

The second case is a log-RealGARCH model with a linear measurement equation:

$$\begin{cases} R_t = \sigma_t z_t, \\ \ln \sigma_t^2 = 0.018 + 0.518 \ln \sigma_{t-1}^2 + 0.262 \ln X_{t-1}, \\ X_t = -0.426 + 1.504\sigma_t^2 - 0.165 z_t + 0.155(z_t^2 - 1) + u_t, \\ u_t \sim \mathcal{U}(-0.5, 0.5) \times \sqrt{12 \times 0.016}. \end{cases}$$

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

TABLE 1. Simulation of the log-RealGARCH(1,1) model

	True value	Estimator $\widehat{\boldsymbol{\theta}}_n$			Estimator $\widetilde{\boldsymbol{\theta}}_{1,n}$			$\widehat{\boldsymbol{\theta}}_{1,n} - \widetilde{\boldsymbol{\theta}}_{1,n}$			
	($n = 900$)	Median	Std	RMSE	Median	Std	RMSE	Median	Std	RMSE	Test size (t-test)
ω	0.125	0.118	(0.059)	0.062	0.117	(0.062)	0.066	0.000	(0.021)	0.024	6.4%
β_1	0.335	0.332	(0.042)	0.040	0.330	(0.114)	0.126	0.000	(0.109)	0.120	8.6%
γ_1	0.514	0.513	(0.051)	0.052	0.508	(0.083)	0.087	0.003	(0.067)	0.070	6.4%
μ	−0.454	−0.451	(0.073)	0.079							
φ	0.980	0.975	(0.086)	0.090							
τ_1	−0.177	−0.176	(0.017)	0.017							
τ_2	0.076	0.076	(0.012)	0.012							
σ_u^2	0.244	0.243	(0.011)	0.012							
PLLH	−975.0	−973.6			−973.5						
RMSE(σ)	–	0.043			0.054						
LBQ test size($\widehat{z}_t$)			6.8%			6.8%					
LBQ test size($\lvert\widehat{z}_t\rvert$)			5.6%			5.6%					
LBQ test size($\widehat{u}_t$)			2.8%			–					
LBQ test size($\lvert\widehat{u}_t\rvert$)			6.8%			–					
Test size (self-inconsistency): 7.2%		RE1: 16.82			RE2: 3.375			RE3: 3.333			
Out-of-sample forecasting ($n = 100$)											
PLLH	−108.3	−108.4			−108.5						
RMSE(σ)	–	0.036			0.049						

The median of the standard error is displayed in parentheses.
LBQ: Ljung-Box Q-test (lag = 20).
Each null hypothesis is rejected at a level of greater than 5%.

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

TABLE 2. Estimation of a log-RealGARCH(1,1) model on a log-RealGARCH(1,1) model with linear measurement equation $X_t = \mu + \varphi\sigma_t^2 + \tau_1 z_t + \tau_2(z_t^2 - 1) + u_t$

	True value	Estimator $\widehat{\boldsymbol{\theta}}_n$			Estimator $\widetilde{\boldsymbol{\theta}}_{1,n}$			$\widehat{\boldsymbol{\theta}}_{1,n} - \widetilde{\boldsymbol{\theta}}_{1,n}$			
	($n = 900$)	Median	Std	RMSE	Median	Std	RMSE	Median	Std	RMSE	Test power (t-test)
ω	0.018	−0.096	(0.031)	0.117	0.018	(0.056)	0.093	−0.114	(0.041)	0.163	83.6%
β_1	0.518	0.520	(0.026)	0.026	0.463	(0.303)	0.431	0.054	(0.161)	0.431	32.4%
γ_1	0.262	0.545	(0.073)	0.293	0.272	(0.166)	0.154	0.274	(0.167)	0.299	36.0%
μ	−0.426	0.224	(0.026)	0.649							
φ	1.504	0.686	(0.075)	0.816							
τ_1	−0.165	−0.104	(0.003)	0.061							
τ_2	0.155	0.096	(0.004)	0.059							
σ_u^2	0.016	0.008	(0.000)	0.008							
PLLH0	−1,391	−1,404			−1,401						
RMSE(σ)	−	0.127			0.059						
LBQ test power($\widehat{z}_t$)			4.6%			4.8%					
LBQ test power($\lvert\widehat{z}_t\rvert$)			8.0%			4.8%					
LBQ test power($\widehat{u}_t$)			5.4%			−					
LBQ test power($\lvert\widehat{u}_t\rvert$)			69.8%			−					
Test power (self-inconsistency): 76.4%		RE1: 4,365			RE2: 26.07		RE3: 26.66				
Out-of-sample forecasting ($n = 100$)											
PLLH	−155.4	−155.4			−155.0						
RMSE(σ)	−	0.120			0.059						

The median of the standard error is displayed in parentheses.

LBQ: Ljung-Box Q-test (lag = 20).

Each null hypothesis is rejected at a level of greater than 5%.

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

TABLE 3. Estimation of a log-RealGARCH(1,1) model on a log-RealGARCH(1,1) model with measurement equation $\ln X_t = -0.326 + 0.32 \ln X_{t-1} + 0.37 \ln \sigma_t^2 - 0.087 z_t + 0.134(z_t^2 - 1) + u_t$, $u_t \sim \mathrm{N}(0, 0.289)$

	True value	Estimator $\widehat{\boldsymbol{\theta}}_n$			Estimator $\widetilde{\boldsymbol{\theta}}_{1,n}$			$\widehat{\boldsymbol{\theta}}_{1,n} - \widetilde{\boldsymbol{\theta}}_{1,n}$			
	($n = 900$)	Median	Std	RMSE	Median	Std	RMSE	Median	Std	RMSE	Test power (t-test)
ω	0.116	0.122	(0.045)	0.052	0.11	(0.032)	0.033	0.010	(0.017)	0.027	17.4%
β_1	0.598	0.337	(0.052)	0.274	0.596	(0.077)	0.083	−0.256	(0.100)	0.271	77.4%
γ_1	0.372	0.502	(0.055)	0.140	0.368	(0.063)	0.064	0.128	(0.056)	0.136	64.8%
PLLH0	−1,137	−1,139			−1,135						
RMSE(σ)	–	0.137			0.073						
LBQ test power($\widehat{z}_t$)			6.6%			6.8%					
LBQ test power($\lvert\widehat{z}_t\rvert$)			8.4%			5.8%					
LBQ test power($\widehat{u}_t$)			13.6%			–					
LBQ test power($\lvert\widehat{u}_t\rvert$)			7.2%			–					
Test power (self-inconsistency): 61.2%				RE1: 1.245	RE2: 2.127		RE3: 1.787				
Out-of-sample forecasting ($n = 100$)											
PLLH	−126.0	−126.5			−126.7						
RMSE(σ)	–	0.125			0.064						

The median of the standard error is displayed in parentheses.

LBQ: Ljung-Box Q-test (lag = 20).

Each null hypothesis is rejected at a level of greater than 5%.

Corsi (2009) documented that the volatility innovations of the linear RealGARCH model should have appropriately truncated tails to guarantee the positivity of volatilities. Thus, we use the uniform distribution to simulate both z_t and u_t.

When we fit a log-RealGARCH model, the variance equation is correctly specified but the measurement equation is not. Therefore, the log-GARCH-X estimator, namely, $\widetilde{\boldsymbol{\theta}}_{1,n}$, is consistent. However, the log-RealGARCH estimator, namely, $\widehat{\boldsymbol{\theta}}_{1,n}$, does not converge to the coefficients of the true model.

According to Table 2, the log-GARCH-X estimator, namely, $\widetilde{\boldsymbol{\theta}}_{1,n}$, is consistent, and the medians of the standard errors are close to the RMSEs. By contrast, the log-RealGARCH estimator, namely, $\widehat{\boldsymbol{\theta}}_{1,n}$, is inconsistent, and the medians of the standard errors are far from the RMSEs. The LBQ test of $|\widehat{u}_t|$ shows that the model is inadequate in 69.8% of the trials, while the self-consistency test indicates that 76.4% of the trials are self-inconsistent. The PLLHs and the RMSEs of both in-sample data and out-of-sample forecasts show that the log-GARCH-X model outperforms the log-RealGARCH model.

Case 3. (The null hypothesis is false.)

The last case is a log-RealGARCH model with a logarithmic measurement equation. The measurement equation, which includes $\ln X_{t-1}$, is different from that of the standard model. This case is constructed to show the inadequacy of the LBQ test in testing the model adequacy.

$$\begin{cases} R_t = \sigma_t z_t, \\ \ln \sigma_t^2 = 0.116 + 0.598 \ln \sigma_{t-1}^2 + 0.372 \ln X_{t-1}, \\ \ln X_t = -0.326 + 0.32 \ln X_{t-1} + 0.37 \ln \sigma_t^2 - 0.087 z_t + 0.134 (z_t^2 - 1) + u_t, \\ u_t \sim \mathcal{N}(0, 0.289). \end{cases}$$

Table 3 displays the log-RealGARCH and log-GARCH-X models that are fitted on simulation data when the measurement equation is misspecified. The log-GARCH-X estimator, namely, $\widetilde{\boldsymbol{\theta}}_{1,n}$, is consistent, and the medians of the standard errors are close to the RMSEs. By contrast, the log-RealGARCH estimator, namely, $\widehat{\boldsymbol{\theta}}_{1,n}$, is inconsistent, and the median standard errors are far from the RMSEs. In this case, no more than 15% of the trials are rejected by the LBQ test. However, the self-inconsistency test indicates that 61.2% of the trials are self-inconsistent. The RMSEs of both in-sample data and out-of-sample forecasts show that the log-GARCH-X model performs better.

5.2. Application

We consider daily returns and 5-min realized volatility series of Nasdaq 100 and S&P500 time series, which are available at http://realized.oxford-man.ox.ac.uk. The observations cover the period from January 1, 2012 to May 20, 2016, and the number of observations is approximately 1,000. We divide the data into two parts: the subsampled data from January 1, 2012 to December 31, 2015 for fitting models and the remaining observations from January 1, 2016 to May 20, 2016 for the out-of-sample forecasting.

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

TABLE 4. QML estimators of nasdaq 100

Nasdaq		Pair (A)			Pair (B)			Pair (C)	
	$\widehat{\theta}_n$	$\widetilde{\theta}_n$	Diff.	$\widehat{\theta}_n$	$\widetilde{\theta}_n$	Diff.	$\widehat{\theta}_n$	$\widetilde{\theta}_n$	Diff.
ω	0.125*	0.131	−0.006	0.173***	0.095**	0.078	−0.084*	0.131	−0.215**
	(0.070)	(0.098)	(0.025)	(0.041)	(0.047)	(0.347)	(0.043)	(0.098)	(0.085)
β_1	0.335***	0.126	0.209	0.673***	0.117	0.556	0.693***	0.126	0.567**
	(0.041)	(0.149)	(0.129)	(0.076)	(0.129)	(0.643)	(0.078)	(0.149)	(0.240)
γ_1	0.514***	0.645***	−0.134	0.033	1.007***	−0.974***	0.088***	0.645***	−0.557***
	(0.058)	(0.106)	(0.079)	(0.022)	(0.182)	(0.058)	(0.028)	(0.106)	(0.058)
μ	−0.454***			−2.051			−0.771		
	(0.086)			(1.282)			(0.317)		
φ	0.980***			4.359*			2.187***		
	(0.097)			(2.254)			(0.589)		
τ_1	−0.177***			0.105***			0.099***		
	(0.019)			(0.022)			(0.023)		
τ_2	0.076***			0.201***			0.200***		
	(0.011)			(0.018)			(0.019)		
σ_u^2	0.244***			0.396			0.417		
	(0.015)			(0.251)			(0.269)		
PLLH	−976.2	−975.0		−1,028	−975.4		−1,012	−975.0	
LLH	−1,617			−1,888			−1,894		
P-value:									
LBQ($\widehat{z}_t$)	0.298	0.270		0.220	0.285		0.224	0.270	
LBQ($\lvert\widehat{z}_t\rvert$)	0.806	0.570		<0.001***	0.643		<0.001***	0.570	
LBQ($\widehat{u}_t$)	0.181			0.473			0.199		
LBQ($\lvert\widehat{u}_t\rvert$)	0.208			1.000			1.000		
Self-consistency test			0.246			<0.001***			<0.001***
Relative efficiency:									
RE1			44.96			519.2			267.7
RE2			4.203			5.773			4.384
RE3			4.333			6.518			4.967
Out-of-sample forecasting:									
PLLH	−131.8	−132.3		−140.3	−132.9		−135.9	−132.3	
TLLH	−160.7			−223.4			−208.9		
F.E.(X_t)	0.6965			0.7678			0.7492		

The standard error is displayed in parentheses.
***,**,*: significance level at 0.01, 0.05, and 0.1.
LBQ: Ljung-Box Q-test (lag = 20).
F.E.: One-step-ahead forecasting error, calculated by the root mean square.

We fit the following three pairs of models and apply the proposed selection method:

Pair (A): a log-RealGARCH model vs. a log-GARCH-X model;

Pair (B): a linear-RealGARCH model vs. a GARCH-X model; and

Pair (C): a log-RealGARCH model with a linear measurement equation vs. a log-GARCH-X model.

We have derived the asymptotic distribution of the self-consistency test statistic for Pair (A). According to Remark 3.2, the corresponding results for Pair (B) and Pair (C) can be easily obtained. We similarly construct tests for these pairs of models and demonstrate the procedure for model selection.

Table 4 illustrates the QML estimators of the three pairs of models of Nasdaq 100. The stationarity condition, $-1 < \widehat{\beta}_1 + \widehat{\varphi}\widehat{\gamma}_1 < 1$, is satisfied. The absolute

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

TABLE 5. QML estimators of S&P 500

S&P	Pair (A)			Pair (B)			Pair (C)		
	$\widehat{\boldsymbol{\theta}}_n$	$\widetilde{\boldsymbol{\theta}}_n$	Diff.	$\widehat{\boldsymbol{\theta}}_n$	$\widetilde{\boldsymbol{\theta}}_n$	Diff.	$\widehat{\boldsymbol{\theta}}_n$	$\widetilde{\boldsymbol{\theta}}_n$	Diff.
ω	0.040	0.108	−0.068**	0.243***	0.049*	0.194	−0.051	0.108	−0.159***
	(0.053)	(0.071)	(0.027)	(0.053)	(0.026)	(0.371)	(0.060)	(0.071)	(0.041)
β_1	0.371***	0.244***	0.127	0.532***	0.199***	0.333	0.265***	0.244***	0.021
	(0.043)	(0.105)	(0.100)	(0.093)	(0.087)	(0.707)	(0.076)	(0.105)	(0.108)
γ_1	0.459***	0.608***	−0.149**	0.032	0.912***	−0.880***	0.443***	0.608***	−0.165***
	(0.043)	(0.080)	(0.060)	(0.021)	(0.131)	(0.047)	(0.045)	(0.080)	(0.060)
μ	−0.359***			−2.589***			−0.266**		
	(0.075)			(1.822)			(0.114)		
φ	1.010***			5.623***			1.479***		
	(0.075)			(3.314)			(0.236)		
τ_1	−0.119***			−0.150***			−0.134***		
	(0.021)			(0.039)			(0.042)		
τ_2	0.017***			0.318***			0.202***		
	(0.016)			(0.031)			(0.028)		
σ_u^2	0.347***			1.296			1.472		
	(0.019)			(1.028)			(1.263)		
PLLH	−937.4	−934.9		−1,012	−937.2		−940.9	−934.9	
LLH	−1,739			−2,407			−2,394		
P-value:									
LBQ($\widehat{z}_t$)	0.707	0.721		0.334	0.746		0.655	0.721	
LBQ($\lvert\widehat{z}_t\rvert$)	0.919	0.953		<0.001***	0.943		0.691	0.953	
LBQ($\widehat{u}_t$)	0.047**			0.406			0.946		
LBQ($\lvert\widehat{u}_t\rvert$)	0.840			1.000			1.000		
Self-consistency test			<0.001***			<0.001***			<0.001***
Relative efficiency:									
RE1			16.98			14.34			4.368
RE2			3.725			1.955			2.044
RE3			3.437			2.137			1.958
Out-of-sample forecasting:									
PLLH	−116.6	−117.5		−126.8	−118.2		−117.2	−117.5	
TLLH	−202.4			−244.9			−227.5		
F.E.(X_t)	0.8496			0.8689			0.7546		

The standard error is displayed in parentheses.
***,**,*: significance level at 0.01, 0.05, and 0.1.
LBQ: Ljung-Box Q-test (lag = 20).
F.E.: One-step-ahead forecasting error, calculated by the root mean square.

residuals $\{|\widehat{z}_t|\}$ are autocorrelated in the models of Pair (B) and Pair (C), which means that the models are inadequate. The self-consistency test also shows the self-inconsistency of these two pairs. Since the LBQ tests and the self-consistency test do not reject the null hypothesis of each test in Pair (A), the log-RealGARCH model is finally selected.

Then, we examine the performances of these pairs of models. Except for the log-GARCH-X model, the log-RealGARCH model has the largest in-sample PLLH. The REs show the relative efficiency of the log-RealGARCH model. In the out-of-sample forecasting, the log-RealGARCH model has the largest PLLH and the smallest forecasting error of X_t. All these results show that the log-RealGARCH model performs the best among these models, which accords with the model that we selected.

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

Table 5 illustrates the QML estimators of the three pairs of models of S&P500. For Pair (A), the LBQ test of $\{\widehat{u}_t\}$ shows that the model is inadequate at the 5% significance level. For Pair (B), the LBQ test of $\{|\widehat{z}_t|\}$ shows that the residual still has ARCH effects. For Pair (C), four LBQ tests do not reject the autocorrelation. However, the self-consistency tests reject all the pairs at the 0.1% significance level. Therefore, we cannot select a reasonable model among them, and another measurement equation should be considered. Lastly, the LBQ tests of $\{|\widehat{u}_t|\}$ take the value of 1 in Pair (B) and Pair (C), as the LBQ test is not robust to extreme values.

6. CONCLUSIONS

This article provides sufficient and necessary conditions for the strict stationarity of the log-RealGARCH model and the log-GARCH-X model. Under these conditions, strong consistency and asymptotic normality of the quasi-maximum likelihood estimators of these two models are obtained. Based on asymptotic results, we propose a Hausman-type test for determining whether the given measurement equation is self-consistent in the logarithmic realized GARCH model. This method can be easily extended to other types of RealGARCH models, e.g., the linear RealGARCH model, for performing model selection.

Simulations are used to evaluate the small-sample performance of the QML estimators. When the data are generated from a log-RealGARCH model, the log-RealGARCH estimators are more efficient than the log-GARCH-X estimators. The test size of the self-consistency test is close to the nominal significance level. When the data are generated from a linear RealGARCH model and various other models, the test can be easily extended and used to help select the model. The empirical study is based on Nasdaq 100 and S&P 500 data. The standard log-RealGARCH model is selected to fit the Nasdaq 100 well. However, the S&P 500 results reject both the linear RealGARCH model and the log-RealGARCH model.

REFERENCES

Andreou, E. & B.J.M. Werker (2015) Residual-based rank specification tests for AR-GARCH type models. *Journal of Econometrics* 185(2), 305–331.

Barndorff-Nielsen, O.E. & N. Shephard (2002) Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B* (*Statistical Methodology*) 64(2), 253–280.

Billingsley, P. (1995) *Probability and Measure*. Wiley.

Bloomfield, P. & G.S. Watson (1975) The inefficiency of least squares. *Biometrika* 62(1), 121–128.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3), 307–327.

Bougerol, P. & N. Picard (1992) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52(1–2), 115–127.

Brockwell, P.J. & R.A. Davis (2009) *Time Series: Theory and Methods*. Springer.

Corsi, F. (2009) A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7(2), 174–196.

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), 987–1007.
Engle, R.F. & V.K. Ng (1993) Measuring and testing the impact of news on volatility. *The Journal of Finance* 48(5), 1749–1778.
Francq, C. & G. Sucarrat (2017) An equation-by-equation estimator of a multivariate log-GARCH-X model of financial returns. *Journal of Multivariate Analysis* 153, 16–32.
Francq, C., O. Wintenberger, & J.-M. Zakoïan (2013) GARCH models without positivity constraints: Exponential or log GARCH? *Journal of Econometrics* 177(1), 34–46.
Francq, C. & J.-M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH. *Bernoulli* 10(4), 605–637.
Geweke, J. (1986) Modeling the persistence of conditional variances: A comment. *Econometric Reviews* 5(1), 57–61.
Han, H. (2015) Asymptotic properties of GARCH-X processes. *Journal of Financial Econometrics* 13(1), 188–221.
Han, H. & D. Kristensen (2014) Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates. *Journal of Business and Economic Statistics* 32(3), 416–429.
Hansen, P.R. & Z. Huang (2016) Exponential GARCH modeling with realized measures of volatility. *Journal of Business and Economic Statistics* 34(2), 269–287.
Hansen, P.R., Z. Huang, & H.H. Shek (2012) Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* 27(6), 877–906.
Hausman, J.A. (1978) Specification tests in econometrics. *Econometrica* 46(6), 1251–1271.
Halunga, A.G. & C.D. Orme (2009) First-order asymptotic theory for parametric misspecification tests of GARCH models. *Econometric Theory* 25(2), 364–410.
Huang, Z., H. Liu, & T. Wang (2016) Modeling long memory volatility using realized measures of volatility: A realized HAR GARCH model. *Economic Modelling* 52, 812–821.
Lee S.W. & B.E. Hansen (1994) Asymptotic theory for the GARCH (1, 1) quasi-maximum likelihood estimator. *Econometric Theory* 10(1), 29–52.
Lundbergh, S. & T. Teräsvirta (2002) Evaluating GARCH models. *Journal of Econometrics* 110(2), 417–435.
Mcleod, A.I. & W.K. Li (1983) Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis* 4(4), 269–273.
Nelson, D.B. (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59(2), 347–370.
Pantula, S.G. (1986) Modeling the persistence of conditional variances: A comment. *Econometric Reviews* 5(1), 71–74.
Straumann, D. (2005) *Estimation in Conditionally Heteroscedastic Time Series Model*. Lecture Notes in Statistics, vol. 181. Springer.
Straumann, D. & T. Mikosch (2006) Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics* 34(5), 2449–2495.
Sucarrat, G., S. Grønneberg, & A. Escribano (2016) Estimation and inference in univariate and multivariate log-GARCH-X models when the conditional density is unknown. *Computational Statistics and Data Analysis* 100, 582–594.

APPENDIX A: Proofs

The definition of *exponentially fast almost sure convergence* will facilitate the formulation of many of our results. A sequence $\{v_n\}_{n\in\mathbb{Z}}$ of random elements with values in a normed vector space $(B, \|\cdot\|)$ is said to converge to zero exponentially fast a.s. when $n \to \infty$, henceforth, denoted by $v_n \xrightarrow{e.a.s.} 0$, if there exists $c > 1$ with $c^n \|v_n\| \xrightarrow{a.s.} 0$.

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

The following lemma given by Straumann (2005) is very useful in the proof of exponentially fast a.s. convergence.

LEMMA A.1 (Straumann, 2005, Proposition 2.5.1). *Let $\{\xi_n\}_{n\in\mathbb{Z}}$ be a sequence of random variables with $\xi_n \xrightarrow{e.a.s.} 0$ and $\{v_n\}_{n\in\mathbb{Z}}$ a sequence of identically distributed random elements with values in a normed vector space $(B, \|\cdot\|)$. If $E[\ln^+ \|v_0\|] < \infty$, then $\sum_{n=0}^{\infty} \xi_n v_n$ converges a.s. and $\xi_n v_n \xrightarrow{e.a.s.} 0$ as $n \to \infty$.*

A.1. Proofs of Consistency

LEMMA A.2 (Existence of log-moments). *Let $\{(R_t, \ln X_t)\}$ be a stationary log-RealGARCH(p,q) process with the parameter vector $\boldsymbol{\theta}^\circ$. If $E|w_t^\circ|^\eta < \infty$ and $\eta \geq 1$, then*
(i) $E|\ln \sigma_t^2|^\eta < \infty$, (ii) $E|\ln X_t|^\eta < \infty$,
(iii) Furthermore, if Assumption C1 holds, we have $E\|\ln h_t\|_{\Theta_1}^\eta < \infty$.

Proof. (i) Since $\ln \sigma_t^2$ satisfies equation (5), it has an a.s. representation,

$$\ln \sigma_t^2 = \mu_2 + \sum_{l=-\infty}^{\infty} \zeta_l w_{t-l}^\circ, \tag{A.1}$$

with $\mu_2 = \frac{\omega^\circ + \gamma_\bullet^\circ \mu^\circ}{\beta_\bullet^\circ + \varphi^\circ \gamma_\bullet^\circ}$, $\beta_\bullet^\circ = \sum_{j=1}^p \beta_j^\circ$, $\gamma_\bullet^\circ = \sum_{j=1}^q \gamma_j^\circ$, and $\sum_{l=-\infty}^{\infty} |\zeta_l| < \infty$. For any $\eta \geq 1$, by Minkowski inequality, we have

$$\begin{aligned}
\mathsf{E}\left|\sum_{|l|\leq m} \zeta_l w_{t-l}^\circ\right|^\eta &\leq \left(\sum_{|l|\leq m} (\mathsf{E}|\zeta_l w_{t-l}^\circ|^\eta)^{\frac{1}{\eta}}\right)^\eta = \left(\sum_{|l|\leq m} |\zeta_l| (\mathsf{E}|w_{t-l}^\circ|^\eta)^{\frac{1}{\eta}}\right)^\eta \\
&\leq \left(\sum_{|l|\leq m} |\zeta_l|\right)^\eta \sup_l \mathsf{E}|w_{t-l}^\circ|^\eta = \left(\sum_{|l|\leq m} |\zeta_l|\right)^\eta \mathsf{E}|w_{t-l}^\circ|^\eta \\
&\leq \left(\sum_{l=-\infty}^{\infty} |\zeta_l|\right)^\eta \mathsf{E}|w_{t-l}^\circ|^\eta.
\end{aligned}$$

Thus, by dominated convergence theorem, we have $\mathsf{E}\left|\sum_{l=-\infty}^{\infty} \zeta_l w_{t-l}^\circ\right|^\eta < \infty$, and therefore $\mathsf{E}|\ln \sigma_t^2|^\eta < \infty$.

(ii) By the definition of measurement equation and cr-inequality, we have

$$\begin{aligned}
\mathsf{E}|\ln X_t|^\eta &= \mathsf{E}|\mu^\circ + \varphi^\circ \ln \sigma_t^2 + \tau^\circ(z_t) + u_t|^\eta \\
&\leq 3^{\eta-1}\left(|\mu^\circ|^\eta + \mathsf{E}|\varphi^\circ \ln \sigma_t^2|^\eta + \mathsf{E}|w_t^\circ|^\eta\right) < \infty.
\end{aligned}$$

(iii) By Assumption C1, $\ln(h_t(\boldsymbol{\theta}_1))$ is well-defined by equation (7), for each $\boldsymbol{\theta}_1 \in \Theta_1$. Since $\mathsf{E}|w_t^\circ|^\eta < \infty$ and thereby $\mathsf{E}|\ln X_t|^\eta < \infty$, following an argument similar to (i), we have $\mathsf{E}\|\ln h_t\|_{\Theta_1}^\eta < \infty$. ■

LEMMA A.3. *$\ln \widehat{h}_t$ and $\ln h_t$ are random elements in $C(\Theta_1, \mathbb{R})$ defined by equations (7) and (8), respectively. Then $\ln \widehat{h}_t$ and $\ln h_t$, $t \in \mathbb{Z}_+$ are almost surely twice continuous*

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

differentiable in the interior of Θ_1. *If* $\mathsf{E}|w_t^\circ|^\eta < \infty$ *for some* $\eta \geq 1$, *then* $\|\ln \widehat{h}_t - \ln h_t\|_{\Theta_1} \xrightarrow{e.a.s.} 0$.

Proof. By definition equation (7), the unique stationary ergodic solution $\ln h_t$ has the almost sure representation,

$$\ln h_t(\boldsymbol{\theta}_1) = \frac{\omega}{\mathcal{B}_{\boldsymbol{\theta}_1}(1)} + \sum_{j=1}^{\infty} \pi_j(\boldsymbol{\theta}_1) \ln X_{t-j},$$

where $\sum_{j=1}^{\infty} \pi_j(\boldsymbol{\theta}_1) z^j = \mathcal{B}_{\boldsymbol{\theta}_1}^{-1}(z)\mathcal{A}_{\boldsymbol{\theta}_1}(z)$, $|z| \leq 1$. By Cauchy inequalities, there exist $c > 0$ and $0 < \rho < 1$ such that ith derivatives of π_j with respect to $\boldsymbol{\theta}_1$ satisfy $|\pi_j^{(i)}(\boldsymbol{\theta}_1)| \leq c\rho^j$, $j \geq 0$, for every $\boldsymbol{\theta}_1 \in \Theta_1$, $i = 0, 1, 2$. Consequently, $\ln h_t(\boldsymbol{\theta}_1)$, $t \in \mathbb{Z}$, are almost surely twice continuous differentiable in the interior of Θ_1. With respect to $\ln \widehat{h}_t(\boldsymbol{\theta}_1)$, $t \in \mathbb{Z}$, the differentiability can be directly seen from definition equation (8) by induction.

To show $\|\ln \widehat{h}_t - \ln h_t\|_{\Theta_1} \xrightarrow{e.a.s.} 0$, first, let $\varsigma_t = \frac{\omega}{\mathcal{B}_{\boldsymbol{\theta}_1}(1)}$, $t = 1-p, \ldots, 0$. Denote the corresponding $\ln \widehat{h}_t(\boldsymbol{\theta}_1)$ by $\ln \widehat{h}_t^\star(\boldsymbol{\theta}_1)$. It is straightforward to verify that

$$\ln \widehat{h}_t^\star(\boldsymbol{\theta}_1) = \frac{\omega}{\mathcal{B}_{\boldsymbol{\theta}_1}(1)} + \sum_{j=1}^{t-1} \pi_j(\boldsymbol{\theta}_1) \ln X_{t-j},$$

for all $t \geq 1$. Thus,

$$\begin{aligned} |\ln \widehat{h}_t^\star(\boldsymbol{\theta}_1) - \ln h_t(\boldsymbol{\theta}_1)| &\leq \sum_{j=t}^{\infty} |\pi_j(\boldsymbol{\theta}_1)| |\ln X_{t-j}| \\ &\leq c \sum_{j=t}^{\infty} \rho^j |\ln X_{t-j}| = c\rho^t \sum_{k=0}^{\infty} \rho^k |\ln X_{-k}|. \end{aligned}$$

Since $\mathsf{E}|w_t^\circ|^\eta < \infty$, by Lemma A.2, we have $\mathsf{E}|\ln X_t|^\eta < \infty$. Then we have $\mathsf{E}[\ln^+(|\ln X_t|)] < \infty$ by using Lemma 2.5.3 in Straumann (2005). Therefore using Lemma A.1, we have $\|\ln \widehat{h}_t^\star - \ln h_t\|_{\Theta_1} \xrightarrow{e.a.s.} 0$.

If the initial values $\varsigma_t, t = 1-p, \ldots, 0$ are not all equal to $\frac{\omega}{\mathcal{B}_{\boldsymbol{\theta}_1}(1)}$, we have

$$\ln \widehat{h}_t^\star(\boldsymbol{\theta}_1) - \ln \widehat{h}_t(\boldsymbol{\theta}_1) = \begin{cases} \frac{\omega}{\mathcal{B}_{\boldsymbol{\theta}_1}(1)} - \varsigma_t, & t = 1-p, \ldots, 0, \\ \sum_{i=1}^{p} \beta_i \left(\ln \widehat{h}_{t-i}^\star(\boldsymbol{\theta}_1) - \ln \widehat{h}_{t-i}(\boldsymbol{\theta}_1) \right), & t > 0. \end{cases}$$

Thus, $\{\ln \widehat{h}_t^\star(\boldsymbol{\theta}_1) - \ln \widehat{h}_t(\boldsymbol{\theta}_1)\}$ is a nonpersistent process, which converges to zero exponentially fast a.s. Due to the compactness of Θ_1 and Assumption C1, $\|\frac{\omega}{\mathcal{B}_{\boldsymbol{\theta}_1}(1)}\|_{\Theta_1}$ is bounded, which leads to the uniform convergence with respect to $\boldsymbol{\theta}_1$, i.e., $\|\ln \widehat{h}_t^\star - \ln \widehat{h}_t\|_{\Theta_1} \xrightarrow{e.a.s.} 0$. Therefore, $\|\ln \widehat{h}_t - \ln h_t\|_{\Theta_1} \xrightarrow{e.a.s.} 0$. ■

LEMMA A.4. *Under the assumptions in Lemma A.3, we have* $\left\| \frac{1}{\widehat{h}_t} - \frac{1}{h_t} \right\|_{\Theta_1} \xrightarrow{e.a.s.} 0$ *and* $\|\widehat{g}_t - g_t\|_{\Theta_1 \times \Theta_2} \xrightarrow{e.a.s.} 0$.

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

Proof. By Lemma A.3, we have $\|\ln\widehat{h}_t - \ln h_t\|_{\Theta_1} \xrightarrow{e.a.s.} 0$. As soon as $\|\ln\widehat{h}_t - \ln h_t\|_{\Theta_1} \le 1$ a.s., for any fixed $\boldsymbol{\theta}_1$ we have

$$\left|\frac{1}{\widehat{h}_t(\boldsymbol{\theta}_1)} - \frac{1}{h_t(\boldsymbol{\theta}_1)}\right| = \left|\exp(-\ln\widehat{h}_t(\boldsymbol{\theta}_1)) - \exp(-\ln h_t(\boldsymbol{\theta}_1))\right|$$
$$= \exp(-\ln h_t(\boldsymbol{\theta}_1))\left|1 - (\exp(\ln h_t(\boldsymbol{\theta}_1) - \ln\widehat{h}_t(\boldsymbol{\theta}_1)))\right|$$
$$\le \exp(|\ln h_t(\boldsymbol{\theta}_1)| + 1)\left|\ln h_t(\boldsymbol{\theta}_1) - \ln\widehat{h}_t(\boldsymbol{\theta}_1)\right|.$$

Then $\left\|\frac{1}{\widehat{h}_t} - \frac{1}{h_t}\right\|_{\Theta_1} \le \|\exp(|\ln h_t| + 1)\|_{\Theta_1}\|\ln h_t - \ln\widehat{h}_t\|_{\Theta_1}$. Since $\|\ln h_t - \ln\widehat{h}_t\|_{\Theta_1} \xrightarrow{e.a.s.} 0$ together with

$$\mathsf{E}[\ln^+(\|\exp(|\ln h_t| + 1)\|_{\Theta_1})] = \mathsf{E}[\|\ln^+(\exp(|\ln h_t| + 1))\|_{\Theta_1}] = 1 + \mathsf{E}\|\ln h_t\|_{\Theta_1} < \infty,$$

we have $\left\|\frac{1}{\widehat{h}_t} - \frac{1}{h_t}\right\|_{\Theta_1} \xrightarrow{e.a.s.} 0$ with an application of Lemma A.1.

Then we turn to the second statement. By calculation,

$$\widehat{g}_t - g_t = -\varphi(\ln\widehat{h}_t - \ln h_t) - \tau_1 R_t(\widehat{h}_t^{-1/2} - h_t^{-1/2}) - \tau_2 R_t^2(\widehat{h}_t^{-1} - h_t^{-1}).$$

If $\mathsf{E}|w_t^\circ| < \infty$, we have $\mathsf{E}[\ln\sigma_t^2] < \infty$ by Lemma A.2, and thereby $\mathsf{E}[\ln^+\sigma_t] < \infty$. Thus together with $\mathsf{E}[\ln^+|z_t|] < \infty$, we have $\mathsf{E}[\ln^+|R_t|] \le \mathsf{E}[\ln^+|\sigma_t|] + \mathsf{E}[\ln^+|z_t|] < \infty$. With an application of Lemma A.1, we have $\|\widehat{g}_t - g_t\|_{\Theta_1\times\Theta_2} \xrightarrow{e.a.s.} 0$. ■

LEMMA A.5. *If $h_1(\boldsymbol{\theta}_1) = h_1(\boldsymbol{\theta}_1^\circ)$ and $g_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = g_1(\boldsymbol{\theta}_1^\circ, \boldsymbol{\theta}_2^\circ)$ a.s. and assumptions C1 and C2 hold, then $\boldsymbol{\theta} = \boldsymbol{\theta}^\circ$.*

Proof. First, we show that for any $\boldsymbol{\theta}_1 \in \Theta_1$, the relation $h_1(\boldsymbol{\theta}_1) = \sigma_1^2$ a.s. implies $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^\circ$. Note that we have

$$\mathcal{B}_{\boldsymbol{\theta}_1}(B)\ln h_t(\boldsymbol{\theta}_1) = \omega + \mathcal{A}_{\boldsymbol{\theta}_1}(B)\ln X_t.$$

If $\ln h_1(\boldsymbol{\theta}_1) = \ln h_1(\boldsymbol{\theta}_1^\circ)$ a.s., by stationarity we have $\ln h_t(\boldsymbol{\theta}_1) = \ln h_t(\boldsymbol{\theta}_1^\circ)$ for all t, and thus we have almost surely

$$\left\{\frac{\mathcal{A}_{\boldsymbol{\theta}_1}(B)}{\mathcal{B}_{\boldsymbol{\theta}_1}(B)} - \frac{\mathcal{A}_{\boldsymbol{\theta}_1^\circ}(B)}{\mathcal{B}_{\boldsymbol{\theta}_1^\circ}(B)}\right\}\ln X_t = \frac{\omega}{\mathcal{B}_{\boldsymbol{\theta}_1}(1)} - \frac{\omega}{\mathcal{B}_{\boldsymbol{\theta}_1^\circ}(1)}.$$

Since $\ln X_t$ is a nontrivial random variable, we have $\frac{\mathcal{A}_{\boldsymbol{\theta}_1}(B)}{\mathcal{B}_{\boldsymbol{\theta}_1}(B)} = \frac{\mathcal{A}_{\boldsymbol{\theta}_1^\circ}(B)}{\mathcal{B}_{\boldsymbol{\theta}_1^\circ}(B)}$, and then by Assumption C1, we have $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^\circ$.

Second, if $g_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = u_1$ a.s., it implies

$$\mu - \mu^\circ + (\varphi - \varphi^\circ)\ln\sigma_1^2 + (\tau(z_1) - \tau^\circ(z_1)) = 0,$$

and then by C2, we have $(\mu, \varphi, \tau_1, \tau_2) = (\mu^\circ, \varphi^\circ, \tau_1^\circ, \tau_2^\circ)$.

Finally, since $\sigma_u^{2\circ} = \mathrm{var}[g_1(\boldsymbol{\theta}_1^\circ, \boldsymbol{\theta}_2^\circ)]$, the conclusion follows. ■

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

LEMMA A.6. *If assumptions C1–C3 hold, $\boldsymbol{\theta}^\circ$ is the unique minimum of L on Θ, where*

$$L(\boldsymbol{\theta}) := E[l_t(\boldsymbol{\theta})] = E[\ln h_t(\boldsymbol{\theta}_1) + \frac{R_t^2}{h_t(\boldsymbol{\theta}_1)} + \ln(\sigma_u^2) + \frac{g_t^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\sigma_u^2}].$$

Proof. First, we minimize on σ_u^2 and get $\sigma_u^2 = \mathsf{E}[g_t^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)]$. By Assumption C3, we have $\mathsf{E}[\frac{R_t^2}{h_t(\boldsymbol{\theta}_1)}] = \mathsf{E}[\frac{\sigma_t^2}{h_t(\boldsymbol{\theta}_1)}]$. Then we have

$$\min L(\boldsymbol{\theta}) = \min \mathsf{E}[\ln h_t(\boldsymbol{\theta}_1) + \frac{\sigma_t^2}{h_t(\boldsymbol{\theta}_1)} + \ln(\mathsf{E}[g_t^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)]) + 1].$$

Indeed, $(\boldsymbol{\theta}_1^\circ, \boldsymbol{\theta}_2^\circ)$ is both the minimizer of $\mathsf{E}[\ln h_t(\boldsymbol{\theta}_1) + \frac{\sigma_t^2}{h_t(\boldsymbol{\theta}_1)}]$ and $\mathsf{E}[g_t^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)]$. Since $\mathsf{E}[\ln h_t(\boldsymbol{\theta}_1) + \frac{\sigma_t^2}{h_t(\boldsymbol{\theta}_1)}]$ is minimized if and only if $h_t(\boldsymbol{\theta}_1) = \sigma_t^2 = h_t(\boldsymbol{\theta}_1^\circ)$, then Lemma A.5 leads to $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^\circ$. Note that $g_t(\boldsymbol{\theta}_1^\circ, \boldsymbol{\theta}_2^\circ) = u_t$, which is independent of both $\mathcal{F}_{t-1}$ and R_t. We have

$$\mathsf{E}[g_t(\boldsymbol{\theta}_1^\circ, \boldsymbol{\theta}_2^\circ) g_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)] = \mathsf{E}[u_t \ln X_t] = \sigma_u^{2\circ}, \forall (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \Theta_1 \times \Theta_2.$$

Then $\mathsf{E}[g_t^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)]$ and $\mathsf{E}[g_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) - g_t(\boldsymbol{\theta}_1^\circ, \boldsymbol{\theta}_2^\circ)]^2$ reach the minimum at the same $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Thus, $\mathsf{E}[g_t^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)]$ is minimized if and only if $g_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = g_t(\boldsymbol{\theta}_1^\circ, \boldsymbol{\theta}_2^\circ)$. Then Lemma A.5 leads to $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^\circ$. At last, from $\sigma_u^2 = \mathsf{E}[g_t^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)]$ and $g_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = g_t(\boldsymbol{\theta}_1^\circ, \boldsymbol{\theta}_2^\circ)$, we have $\sigma_u^2 = \mathsf{E}[g_t^2(\boldsymbol{\theta}_1^\circ, \boldsymbol{\theta}_2^\circ)] = \sigma_u^{2\circ}$. ■

Proof of Theorem 3.1

Proof. Under standard arguments, e.g., the proof of Theorem 4.1 in Francq et al. (2013), the consistency is obtained by showing the following intermediate results,
(*i*) $\|\widehat{L}_n - L_n\|_\Theta \xrightarrow{a.s.} 0$;
(*ii*) if $h_1(\boldsymbol{\theta}_1) = h_1(\boldsymbol{\theta}_1^\circ)$ and $g_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = g_1(\boldsymbol{\theta}_1^\circ, \boldsymbol{\theta}_2^\circ)$ a.s. then $\boldsymbol{\theta} = \boldsymbol{\theta}^\circ$;
(*iii*) $\boldsymbol{\theta}^\circ$ is the unique minimum of L on Θ;
(*iv*) any $\boldsymbol{\theta} \neq \boldsymbol{\theta}^\circ$ has a neighborhood $V(\boldsymbol{\theta})$ such that
$\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}^* \in V(\boldsymbol{\theta})} \widetilde{L}_n(\boldsymbol{\theta}^*) \geq \mathsf{E}[l_t(\boldsymbol{\theta}^\circ)]$, a.s.

By Lemma A.3, we have $\|\ln \widehat{h}_t - \ln h_t\|_{\Theta_1} \xrightarrow{e.a.s.} 0$. The proof of step (i) then follows from Lemma.A.4.

Step (ii) is proved in Lemma A.5.

For step (iii), Lemma A.6 proves that $\boldsymbol{\theta}^\circ$ is the unique minimum of L on Θ.

In step (iv), we prove $\mathsf{E}[l_t(\boldsymbol{\theta}^\circ)] < \infty$. Equation (7) shows that $\{l_t\}$ is stationary ergodic. Lemma A.2 entails that $\mathsf{E}\left|\ln(h_t(\boldsymbol{\theta}_1^\circ))\right| < \infty$. It follows that $\mathsf{E}[l_t(\boldsymbol{\theta}^\circ)] < \infty$. The rest of the proof of (iv) is identical to the proof of Theorem 2.1(iv) of Francq and Zakoïan (2004). ■

A.2. Proofs of Asymptotic Normality

LEMMA A.7. *Suppose that the assumptions in Theorem 3.1 are satisfied. Then, for any $\eta \geq 1$, there exists a compact set $\check{\Theta}_1(\eta)$, and $\boldsymbol{\theta}_1^\circ \in \check{\Theta}_1(\eta)$, such that*
(*i*) $E\|\sigma_t^2 / h_t(\boldsymbol{\theta}_1)\|^\eta_{\check{\Theta}_1(\eta)} < \infty$;

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

(*ii*) $E\|\ln h_t(\boldsymbol{\theta}_1)\|^{\eta}_{\breve{\Theta}_1(\eta)} < \infty$;
(*iii*) $E\|g_t^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\|^{\eta}_{\breve{\Theta}_1(\eta)\times\Theta_2} < \infty$.

Proof. By the representation (7), we have

$$\begin{aligned}
&\ln\sigma_t^2 - \ln h_t(\boldsymbol{\theta}_1) \\
&= \mathcal{B}^{-1}_{\boldsymbol{\theta}_1^\circ}(B)\left(\omega^\circ + \mathcal{A}_{\boldsymbol{\theta}_1^\circ}(B)\ln X_t\right) - \mathcal{B}^{-1}_{\boldsymbol{\theta}_1}(B)\left(\omega + \mathcal{A}_{\boldsymbol{\theta}_1}(B)\ln X_t\right) \\
&= \mathcal{B}^{-1}_{\boldsymbol{\theta}_1^\circ}(B)\left(\omega^\circ - \omega + \mathcal{A}_{\boldsymbol{\theta}_1^\circ - \boldsymbol{\theta}_1}(B)\ln X_t\right) + \mathcal{B}^{-1}_{\boldsymbol{\theta}_1^\circ - \boldsymbol{\theta}_1}(B)\left(\omega + \mathcal{A}_{\boldsymbol{\theta}_1}(B)\ln X_t\right).
\end{aligned}$$

Similarly to the proof of Lemma 7.3 in Francq et al. (2013), by Assumption C4, for any η, there exists a compact set $\breve{\Theta}_1(\eta)$, such that

$$\mathsf{E}[\exp(\eta\|\ln\sigma_t^2 - \ln h_t(\boldsymbol{\theta}_1))\|_{\breve{\Theta}_1(\eta)}] < \infty.$$

Consequently,

$$\mathsf{E}\|\frac{\sigma_t^2}{h_t(\boldsymbol{\theta}_1)} \vee \frac{h_t(\boldsymbol{\theta}_1)}{\sigma_t^2}\|^{\eta}_{\breve{\Theta}_1(\eta)} < \infty.$$

Thus the assertion is proved.

(ii) Note that $\mathsf{E}|\ln X_t|^{\eta} < \infty$ for any $\eta > 0$ by Assumption C4. Following the argument in the proof of Lemma A.2(iii), we have $\mathsf{E}\|\ln h_t(\boldsymbol{\theta}_1)\|^{\eta}_{\breve{\Theta}_1(\eta)} < \infty$.

(iii)By definition equation (8), the result follows directly by (i) and (ii). ■

LEMMA A.8. *Suppose that the assumptions in Lemma A.7 are satisfied such that* $\breve{\Theta}_1(\eta)$ *exists. Let* $\breve{\Theta}(\eta) = \breve{\Theta}_1(\eta) \times \Theta_2$. *Then*
(*i*) $E\|\nabla_1(\ln h_t)\|^{\eta}_{\breve{\Theta}_1(\eta)} < \infty$;
(*ii*) $E\|\nabla_1 g_t\|^{\eta}_{\breve{\Theta}(\eta)} < \infty$, *provided that* $E|z_t|^{2\eta} < \infty$;
(*iii*) $E\|\nabla_2 g_t\|^{\eta}_{\breve{\Theta}(\eta)} < \infty$, *provided that* $E|z_t|^{2\eta} < \infty$.

Proof. (i) Random element $\nabla_1(\ln h_t)$ is the stationary solution of the equation

$$\begin{aligned}
\nabla_1(\ln h_t) = &\sum_{i=1}^{p}\beta_i\nabla_1(\ln h_{t-i}) \\
&+ (1, \ln h_{t-1}, \ldots, \ln h_{t-p}, \ln X_{t-1}, \ldots, \ln X_{t-q})^T. \qquad \textbf{(A.2)}
\end{aligned}$$

Assumption C1 entails that $\nabla_1(\ln h_t)$ is a linear combination of $\ln X_{t-i}$ for $i \geq 1$. By Lemma A.7(ii), we have $\mathsf{E}\|\nabla_1(\ln h_t)\|^{\eta}_{\breve{\Theta}_1(\eta)} < \infty$.

For (ii) and (iii), we have

$$\begin{aligned}
\nabla_1 g_t &= -\varphi\nabla_1(\ln h_t) - \tau_1 R_t \nabla_1 h_t^{-1/2} - \tau_2 R_t^2 \nabla_1 h_t^{-1} \\
&= (-\varphi + \frac{1}{2}\tau_1\frac{R_t}{h_t^{1/2}} + \tau_2\frac{R_t^2}{h_t})\nabla_1(\ln h_t) \qquad \textbf{(A.3)}
\end{aligned}$$

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

and

$$\nabla_2 g_t = (-1, -\ln h_t, -R_t h_t^{-1/2}, 1 - R_t^2 h_t^{-1})^T. \tag{A.4}$$

Note that $R_t = \sigma_t z_t$, and z_t is independent of both $\sigma_t / h_t^{1/2}$ and $\nabla_1(\ln h_t)$ by Assumption C3. The result follows by $\mathsf{E}|z_t|^{2\eta} < \infty$, Lemma A.7, and Cauchy-Schwarz inequality. ■

LEMMA A.9. *Suppose that the assumptions in Lemma A.7 are satisfied such that* $\breve{\Theta}_1(\eta)$ *exists. If* $\mathsf{E}|z_t|^{2\eta} < \infty$, *restricted on the vector space with norm* $\|\cdot\|_{\breve{\Theta}_1(\eta)}$ *or* $\|\cdot\|_{\breve{\Theta}_1(\eta)\times\Theta_2}$, *we have*

(*i*) $(1 - \frac{R_t^2}{\widehat{h}_t})\nabla_1(\ln \widehat{h}_t) - (1 - \frac{R_t^2}{h_t})\nabla_1(\ln h_t) \xrightarrow{e.a.s.} 0$;

(*ii*) $\widehat{g}_t^2 - g_t^2 \xrightarrow{e.a.s.} 0$;

(*iii*) $\nabla_1 \widehat{g}_t^2 - \nabla_1 g_t^2 \xrightarrow{e.a.s.} 0$ *and* $\nabla_2 \widehat{g}_t^2 - \nabla_2 g_t^2 \xrightarrow{e.a.s.} 0$.

Proof. (i) Analogously to the proof of Theorem 4.2 in Francq et al. (2013), we have $\nabla_1(\ln \widehat{h}_t) - \nabla_1(\ln h_t) \xrightarrow{e.a.s.} 0$. Since

$$\begin{aligned}
&\|\frac{R_t^2}{\widehat{h}_t}\nabla_1(\ln \widehat{h}_t) - \frac{R_t^2}{h_t}\nabla_1(\ln h_t)\|_{\breve{\Theta}_1(\eta)} \\
\le\ &\|\frac{R_t^2}{\widehat{h}_t} - \frac{R_t^2}{h_t}\|_{\breve{\Theta}_1(\eta)}\|\nabla_1(\ln h_t)\|_{\breve{\Theta}_1(\eta)} + \|\frac{R_t^2}{h_t}\|_{\breve{\Theta}_1(\eta)}\|\nabla_1(\ln \widehat{h}_t) - \nabla_1(\ln h_t)\|_{\breve{\Theta}_1(\eta)} \\
&+\|\frac{R_t^2}{\widehat{h}_t} - \frac{R_t^2}{h_t}\|_{\breve{\Theta}_1(\eta)}\|\nabla_1(\ln \widehat{h}_t) - \nabla_1(\ln h_t)\|_{\breve{\Theta}_1(\eta)},
\end{aligned}$$

a sufficient condition (see Lemma A.1) for the exponentially fast almost sure convergence is $\mathsf{E}[\ln^+ \|\frac{R_t^2}{h_t}\|_{\breve{\Theta}_1(\eta)}] < \infty$ and $\mathsf{E}[\ln^+ \|\nabla_1(\ln h_t)\|_{\breve{\Theta}_1(\eta)}] < \infty$. It is implied by the existence of power moment $\mathsf{E}\|\frac{R_t^2}{h_t}\|^{\eta}_{\breve{\Theta}_1(\eta)} < \infty$ and $\mathsf{E}\|\nabla_1(\ln h_t)\|^{\eta}_{\breve{\Theta}_1(\eta)} < \infty$, for some $\eta > 0$. Thus, the result follows from Lemma A.7.

(ii) Analogously to the argument of (i), the conclusion holds by using Lemma A.4.

(iii) Similarly to equations (A.3) and (A.4), we have

$$\nabla_1 \widehat{g}_t = (-\varphi + \frac{1}{2}\tau_1 \frac{R_t}{\widehat{h}_t^{1/2}} + \tau_2 \frac{R_t^2}{\widehat{h}_t})\nabla_1(\ln \widehat{h}_t)$$

and

$$\nabla_2 \widehat{g}_t = (-1, -\ln \widehat{h}_t, -\frac{R_t}{\widehat{h}_t^{1/2}}, 1 - \frac{R_t^2}{\widehat{h}_t})^T.$$

Analogously to the argument of (i), $\nabla_1 \widehat{g}_t - \nabla_1 g_t \xrightarrow{e.a.s.} 0$ and $\nabla_2 \widehat{g}_t - \nabla_2 g_t \xrightarrow{e.a.s.} 0$ by Lemma A.7. Following the argument of (i) again, the conclusion holds by using Lemma A.4. ■

Letting $\triangle_{11} = \partial^2 / \partial \boldsymbol{\theta}_1^T \boldsymbol{\theta}_1$, $\triangle_{12} = \partial^2 / \partial \boldsymbol{\theta}_1^T \partial \boldsymbol{\theta}_2$ and $\triangle_{22} = \partial^2 / \partial \boldsymbol{\theta}_2^T \partial \boldsymbol{\theta}_2$, we have the following lemma for the second derivative of the log-likelihood function:

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

LEMMA A.10. *Suppose that the assumptions in Lemma A.7 are satisfied such that* $\breve{\Theta}_1(\eta)$ *exists. If* $E|z_t|^4 < \infty$, *then the following moment conditions hold,*
(*i*) $E\|\triangle_{11}(\ln h_t)\|^{\eta}_{\breve{\Theta}_1(\eta)} < \infty$, *for any* $\eta > 0$;
(*ii*) $E\|\triangle_{ij}(g_t)\|^2_{\breve{\Theta}_1(2)\times\Theta_2} < \infty$, *for* $i, j = 1, 2$;
(*iii*) $E\|\triangle_{ij}(g_t^2)\|_{\breve{\Theta}_1(2)\times\Theta_2} < \infty$, *for* $i, j = 1, 2$.

Proof. Random element $\triangle_{11}(\ln h_t)$ is the stationary solution of the equation

$$\begin{aligned}\triangle_{11}(\ln h_t) = &\sum_{i=1}^{p}\beta_i\triangle_{11}(\ln h_{t-i})\\ &+ (0, \nabla_1(\ln h_{t-1}), \ldots, \nabla_1(\ln h_{t-p}), \mathbf{0}_q)\\ &+ (0, \nabla_1(\ln h_{t-1}), \ldots, \nabla_1(\ln h_{t-p}), \mathbf{0}_q)^T,\end{aligned}$$

where $\nabla_1(\ln h_t)$ is defined by equation (A.2). Following an argument like Lemma A.8(i), we have $E\|\triangle_{11}(\ln h_t)\|^{\eta}_{\breve{\Theta}_1(\eta)} < \infty$.

(ii) Note that

$$\triangle_{11}(g_t) = (-\varphi + \frac{1}{2}\tau_1\frac{R_t}{h_t^{1/2}} + \tau_2\frac{R_t^2}{h_t})\triangle_{11}(\ln h_t) - (\frac{1}{4}\tau_1\frac{R_t}{h_t^{1/2}} + \tau_2\frac{R_t^2}{h_t})\nabla_1(\ln h_t)\nabla_1(\ln h_t)^T,$$

$$\triangle_{12}(g_t) = (\mathbf{0}_{p+1}, -\nabla_1(\ln h_t), \frac{1}{2}\frac{R_t}{h_t^{1/2}}\nabla_1(\ln h_t), \frac{R_t^2}{h_t}\nabla_1(\ln h_t))^T,$$

and

$$\triangle_{22}(g_t) = \mathbf{O}_{4\times 4}.$$

Since $R_t = \sigma_t z_t$, and z_t is independent of $\sigma_t/h_t^{1/2}$, $\nabla_1(\ln h_t)$, and $\triangle_{11}(\ln h_t)$, the result follows by Lemma A.8(i) and Cauchy-Schwarz inequality.

(iii) Since

$$\triangle_{ij}(g_t^2) = 2\nabla_i g_t(\nabla_j g_t)^T + 2g_t\triangle_{ij}(g_t), \tag{A.5}$$

the result follows by Lemma A.7(iii), Lemma A.8(ii) and (iii), and Cauchy-Schwarz inequality. ■

LEMMA A.11. *Suppose that the assumptions in Lemma A.10 are satisfied such that* $\breve{\Theta}_1(2)$ *exists. Restricted on the vector space with norm* $\|\cdot\|_{\breve{\Theta}_1(2)}$ *or* $\|\cdot\|_{\breve{\Theta}_1(2)\times\Theta_2}$, *we have*
(*i*) $(1 - \frac{R_t^2}{\widehat{h}_t})\triangle_{11}(\ln\widehat{h}_t) - (1 - \frac{R_t^2}{h_t})\triangle_{11}(\ln h_t) \xrightarrow{e.a.s.} 0$;
(*ii*) $\frac{R_t^2}{\widehat{h}_t}\nabla_1(\ln\widehat{h}_t)\nabla_1(\ln\widehat{h}_t)^T - \frac{R_t^2}{h_t}\nabla_1(\ln h_t)\nabla_1(\ln h_t)^T \xrightarrow{e.a.s.} 0$;
(*iii*) $\triangle_{ij}(\widehat{g}_t^2) - \triangle_{ij}(g_t^2) \xrightarrow{e.a.s.} 0$, *for* $i, j = 1, 2$.

Proof. (i) Analogously to the proof of Theorem 4.2 in Francq et al. (2013), we have $\triangle_{11}(\ln\widehat{h}_t) - \triangle_{11}(\ln h_t) \xrightarrow{e.a.s.} 0$. Following the argument in Lemma A.9(i), we get (i) and

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

(ii) by Lemma A.8(i) and Lemma A.10(i).
(iii) The proof is similar to the proof of Lemma A.9. Since $\nabla_1(\ln h_t)$ and $\triangle_{11}(\ln h_t)$ admit the moment of order 2 on $\breve{\Theta}_1(2)$, we have $\triangle_{ij}\widehat{g}_t - \triangle_{ij}g_t \xrightarrow{e.a.s.} 0$. Then by equation (A.5), using the result in Lemma A.9, and following an argument similar in Lemma A.9(i), we get the result. ■

LEMMA A.12. *Under the assumptions of Theorem 3.2, we have* $\mathbf{H} = E[l_t''(\boldsymbol{\theta}^\circ)]$ *is positive definite.*

Proof. We prove the result by contradiction. Lemma A.10 shows that $\mathsf{E}[l_t''(\boldsymbol{\theta}^\circ)]$ exists. Assume for a $\breve{\boldsymbol{\theta}}_1 = (\breve{\omega}, \breve{\beta}_1, \ldots, \breve{\beta}_p, \breve{\gamma}_1, \ldots, \breve{\gamma}_q)^T$ and $\breve{\boldsymbol{\theta}}_2 = (\breve{\mu}, \breve{\varphi}, \breve{\tau}_1, \breve{\tau}_2)^T$ such that $\breve{\boldsymbol{\theta}}_1^T \mathbf{A}\breve{\boldsymbol{\theta}}_1 + (\breve{\boldsymbol{\theta}}_2^T, \breve{\boldsymbol{\theta}}_2^T)\mathbf{B}(\breve{\boldsymbol{\theta}}_2^T, \breve{\boldsymbol{\theta}}_2^T)^T = 0$, where $\mathbf{A} = \mathsf{E}^\circ[\nabla_1(\ln h_t)\nabla_1(\ln h_t)^T]$ and $\mathbf{B} = \mathsf{E}^\circ[\nabla_{12}g_t\nabla_{12}g_t^T]$. Since $\mathbf{A}$ and $\mathbf{B}$ are positive semidefinite matrix, then we have

$$\breve{\boldsymbol{\theta}}_1^T \nabla_1(\ln h_t)|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^\circ} = 0, \text{ a.s.}$$

and

$$\breve{\boldsymbol{\theta}}_1^T \nabla_1 g_t + \breve{\boldsymbol{\theta}}_2^T \nabla_2 g_t|_{(\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T) = (\boldsymbol{\theta}_1^{\circ T}, \boldsymbol{\theta}_2^{\circ T})} = 0, \text{ a.s..}$$

These two equations result in

$$\breve{\omega} + \sum_{i=1}^{p} \breve{\beta}_i \ln \sigma_{t-i}^2 + \sum_{j=1}^{q} \breve{\gamma}_j \ln X_{t-j} = 0, \text{ a.s.} \quad \text{and}$$

$$\breve{\mu} + \breve{\varphi} \ln \sigma_t^2 + \breve{\tau}_1 z_t + \breve{\tau}_2 z_t^2 = 0, \text{ a.s.}$$

Then following the argument in the proof of Lemma A.5 we have $\breve{\boldsymbol{\theta}}_1 = \mathbf{0}$ and $\breve{\boldsymbol{\theta}}_2 = \mathbf{0}$. ■

Proof of Theorem 3.2

Proof. A Taylor expansion gives $L_n'(\widehat{\boldsymbol{\theta}}_n) - L_n'(\boldsymbol{\theta}^\circ) = L_n''(\breve{\boldsymbol{\theta}}_n)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^\circ)$, where $\breve{\boldsymbol{\theta}}_n$ satisfies $\|\breve{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^\circ\| \le \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^\circ\|$, almost surely. The strategy of proof and the argument employed are classical and similar to those in the proof of Theorem 4.2 in Francq et al. (2013) and Theorem 7.1 in Straumann and Mikosch (2006). The asymptotic normality is obtained by showing the following results,
(i) $\sqrt{n}L_n'(\boldsymbol{\theta}^\circ) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{J})$, as $n \to \infty$;
(ii) $\|L_n''(\breve{\boldsymbol{\theta}}_n) - \mathbf{H}\|$ converges a.s. to 0 for any sequence $(\breve{\boldsymbol{\theta}}_n)$ converges a.s. to $\boldsymbol{\theta}^\circ$ and $\mathbf{H}$ is invertible;
(iii) $\sqrt{n}\|\widehat{L}_n'(\widehat{\boldsymbol{\theta}}_n) - L_n'(\widehat{\boldsymbol{\theta}}_n)\|$ converges a.s. to 0.

For step (i), we need to show that the sequence $\{l_t'(\boldsymbol{\theta}^\circ)\}_{t\in\mathbb{Z}}$ is a martingale difference sequence with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$. Observe that $\nabla_{12}g_t\left(\boldsymbol{\theta}_1^\circ, \boldsymbol{\theta}_2^\circ\right)$ is conditionally independent of u_t, thus $\mathsf{E}^\circ[\nabla_{12}g_t u_t|\mathcal{F}_{t-1}] = \mathbf{0}$. In addition, $\frac{\nabla_1 h_t}{\sigma_t^2}$ is $\mathcal{F}_{t-1}$-measurable, and $\mathsf{E}[\frac{\nabla_1 h_t z_t^2}{\sigma_t^2}|\mathcal{F}_{t-1}] = \frac{\nabla_1 h_t}{\sigma_t^2}\mathsf{E}[z_t^2|\mathcal{F}_{t-1}] = \frac{\nabla_1 h_t}{\sigma_t^2}$. Since $\sigma_u^2 = \mathsf{E}[u_t^2]$, we have

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

$\mathsf{E}[l_t'(\boldsymbol{\theta}^\circ)|\mathcal{F}_{t-1}] = \mathbf{0}$. Then, $\{l_t'(\boldsymbol{\theta}^\circ)\}$ is a stationary and ergodic zero-mean martingale difference sequence with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$. By the cental limit theorem for square-integrable stationary and ergodic martingale difference sequences, we have

$$\sqrt{n}L_n'(\boldsymbol{\theta}^\circ) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathsf{E}[l_t'(\boldsymbol{\theta}^\circ)l_t'(\boldsymbol{\theta}^\circ)^T]),$$

as $n \to \infty$, whenever $\mathsf{E}[l_t'(\boldsymbol{\theta}^\circ)l_t'(\boldsymbol{\theta}^\circ)^T]$ exists. The square-integrability of $l_t'(\boldsymbol{\theta}^\circ)$ is shown in Lemma A.8.

Turning to step (ii), we have $l_t'' =$

$$\begin{bmatrix} (1-\frac{R_t^2}{h_t})\triangle_{11}(\ln h_t) + \frac{R_t^2}{h_t}\nabla_1(\ln h_t)\nabla_1(\ln h_t)^T + \frac{1}{\sigma_u^2}\triangle_{11}(g_t^2) & \bullet & \bullet \\ \frac{1}{\sigma_u^2}\triangle_{12}(g_t^2) & \frac{1}{\sigma_u^2}\triangle_{22}(g_t^2) & \bullet \\ -\frac{1}{\sigma_u^4}\nabla_1(g_t^2) & -\frac{1}{\sigma_u^4}\nabla_2(g_t^2)\ \frac{2g_t^2-\sigma_u^2}{\sigma_u^6} & \end{bmatrix}.$$

Choose $\check{\Theta} := \check{\Theta}_1(2) \times \Theta_2$ given in Lemma A.10, a neighbor of $\boldsymbol{\theta}^\circ$, such that $\sup_{\boldsymbol{\theta}\in\check{\Theta}} \mathsf{E}[l_t''(\boldsymbol{\theta})]$ exists. For sufficient large n, $\widehat{\boldsymbol{\theta}}_n \in \check{\Theta}$ by consistency. Then using the inequality

$$\|L_n''(\check{\boldsymbol{\theta}}_n) - \mathbf{H}\| \le \|L_n''(\boldsymbol{\theta}) - \mathsf{E}[l_t''(\boldsymbol{\theta})]\|_{\check{\Theta}} + \|\mathsf{E}[l_t''(\widehat{\boldsymbol{\theta}}_n)] - \mathbf{H}\|,$$

we find that the first term converges a.s. to 0 by Lemma A.11, and the second term converges a.s. to 0 by the twice continuous differentiability of $\mathsf{E}[l_t(\boldsymbol{\theta})]$. Consequently, $\|L_n''(\check{\boldsymbol{\theta}}_n) - \mathbf{H}\|$ converges a.s. to 0 as $\widehat{\boldsymbol{\theta}}_n$ converges a.s. to $\boldsymbol{\theta}^\circ$. Lemma A.12 shows that $\mathbf{H}$ is invertible.

For step (iii), we only need to prove $\widehat{l}_n' - l_n' \xrightarrow{e.a.s.} 0$ on normed vector space $\left(C(\check{\Theta}, \mathbb{R}^{p+q+6}), \|\cdot\|_{\check{\Theta}}\right)$, where

$$l_t' = \left((1-\frac{R_t^2}{h_t})\nabla_1(\ln h_t)^T + \frac{1}{\sigma_u^2}\nabla_1(g_t^2)^T,\ \frac{1}{\sigma_u^2}\nabla_2(g_t^2)^T,\ \frac{1}{\sigma_u^2} - \frac{g_t^2}{\sigma_u^4}\right)^T,$$

and

$$\widehat{l}_t' = \left((1-\frac{R_t^2}{\widehat{h}_t})\nabla_1(\ln \widehat{h}_t)^T + \frac{1}{\sigma_u^2}\nabla_1(\widehat{g}_t^2)^T,\ \frac{1}{\sigma_u^2}\nabla_2(\widehat{g}_t^2)^T,\ \frac{1}{\sigma_u^2} - \frac{\widehat{g}_t^2}{\sigma_u^4}\right)^T.$$

It is proved in Lemma A.9 and thus asymptotic normality is proved.

Then we prove $\mathbf{J}$ is a singular matrix if and only if u_t^2 degenerates. Assume there is a vector $\check{\boldsymbol{\theta}} = (\check{\boldsymbol{\theta}}_1^T, \check{\boldsymbol{\theta}}_2^T, \check{\sigma}_u^2)^T \neq \mathbf{0}$ such that

$$(\check{\boldsymbol{\theta}}_1^T, \check{\boldsymbol{\theta}}_2^T, \check{\sigma}_u^2)\mathbf{J}(\check{\boldsymbol{\theta}}_1^T, \check{\boldsymbol{\theta}}_2^T, \check{\sigma}_u^2)^T = 0. \tag{A.6}$$

Then we have $(\check{\boldsymbol{\theta}}_1^T, \check{\boldsymbol{\theta}}_2^T, \check{\sigma}_u^2)l_t'(\boldsymbol{\theta}^\circ) = 0$ a.s.. If $\check{\sigma}_u^2 = 0$, then by Lemma A.12, we have $(\check{\boldsymbol{\theta}}_1^T, \check{\boldsymbol{\theta}}_2^T)^T = \mathbf{0}$. If $\check{\sigma}_u^2 \neq 0$, without loss of generality, we can assume $\check{\sigma}_u^2 = \sigma_u^{2\circ}$. Then

$$(1-z_t^2)\sigma_u^{2\circ}\check{\boldsymbol{\theta}}_1^T\nabla_1(\ln h_t) + \check{\boldsymbol{\theta}}_1^T\nabla_1(g_t^2) + \check{\boldsymbol{\theta}}_2^T\nabla_2(g_t^2)|_{(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=(\boldsymbol{\theta}_1^\circ,\boldsymbol{\theta}_2^\circ)} = u_t^2 - \sigma_u^{2\circ}.$$

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

Let $\check{c}_t = \check{\boldsymbol{\theta}}_1^T \nabla_1 (\ln h_t)|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^\circ}$. By equations (A.3) and (A.4), we have

$$\left((1-z_t^2)\sigma_u^{2\circ} + 2u_t(-\varphi^\circ + \frac{1}{2}\tau_1^\circ z_t + \tau_2^\circ z_t^2)\right)\check{c}_t$$
$$-2u_t\left(\check{\mu} + \check{\varphi}\ln\sigma_t^2 + \check{\tau}_1 z_t + \check{\tau}_2(z_t^2 - 1)\right) = u_t^2 - \sigma_u^{2\circ}.$$

Since z_t and u_t are both independent to $\mathcal{F}_{t-1}$, we have both

$$\left((1-z_t^2)\sigma_u^{2\circ} + 2u_t(-\varphi^\circ + \frac{1}{2}\tau_1^\circ z_t + \tau_2^\circ z_t^2)\right)\check{c}_t - 2u_t\check{\varphi}\ln\sigma_t^2$$

and

$$2u_t\left(\check{\mu} + \check{\tau}_1 z_t + \check{\tau}_2(z_t^2 - 1)\right) + u_t^2 - \sigma_u^{2\circ}$$

are independent of $\mathcal{F}_{t-1}$. Note that $\check{c}_t$ and $\ln\sigma_t^2$ are $\mathcal{F}_{t-1}$-measurable. When $(\check{c}_t, \ln\sigma_t^2)$ takes distinct values $(c(1), h(1))$ and $(c(2), h(2))$, such that $(c(1), h(1)) \neq (c(2), h(2))$, we have

$$\left((1-z_t^2)\sigma_u^{2\circ} + 2u_t(-\varphi^\circ + \frac{1}{2}\tau_1^\circ z_t + \tau_2^\circ z_t^2)\right)c(1) - 2u_t\check{\varphi}h(1)$$
$$= \left((1-z_t^2)\sigma_u^{2\circ} + 2u_t(-\varphi^\circ + \frac{1}{2}\tau_1^\circ z_t + \tau_2^\circ z_t^2)\right)c(2) - 2u_t\check{\varphi}h(2).$$

Note that u_t and z_t are mutually independent. Considering the linear term of u_t, we have $2(-\varphi^\circ + \frac{1}{2}\tau_1^\circ z_t + \tau_2^\circ z_t^2)(c(1) - c(2)) - 2\check{\varphi}(h(1) - h(2)) = 0$. By Assumption C2, we have $c(1) = c(2)$ and $\check{\varphi} = 0$. Thus $\check{c}_t$ is a constant. By the definition of $\check{c}_t$ and equation (A.2), $\check{c}_t$ is stationary, and thereby we have $\check{c}_t \equiv \check{c}$, which is not dependent on t. Then

$$\left((1-z_t^2)\sigma_u^{2\circ} + 2u_t(-\varphi^\circ + \frac{1}{2}\tau_1^\circ z_t + \tau_2^\circ z_t^2)\right)\check{c}$$
$$= 2u_t\left(\check{\mu} + \check{\tau}_1 z_t + \check{\tau}_2(z_t^2 - 1)\right) + u_t^2 - \sigma_u^{2\circ}.$$

This is a quadratic equation in u_t. Since u_t is independent of z_t, we can conclude that $\check{c} = 0$, $\check{\mu} = \check{\tau}_1 = \check{\tau}_2 = 0$, and $u_t^2 = \sigma_u^{2\circ}$. Thus $u_t = \sigma_u^\circ$ or $-\sigma_u^\circ$ with probability 0.5, which means that u_t^2 degenerates. ■

A.3. Proof of Theorem 4.1

Proof. It is easy to show that the conditions in Theorems 3.3 and 3.4 also hold. We combine l_t and $l_t^{(GX)}$ together by defining

$$l_t^{(C)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma_u^2, \boldsymbol{\theta}_3) = l_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma_u^2) + l_t^{(GX)}(\boldsymbol{\theta}_3),$$

and its estimator

$$\widehat{l}_t^{(C)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma_u^2, \boldsymbol{\theta}_3) = \widehat{l}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma_u^2) + \widehat{l}_t^{(GX)}(\boldsymbol{\theta}_3).$$

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

Then $(\widehat{\boldsymbol{\theta}}_n, \widetilde{\boldsymbol{\theta}}_{1,n})$ is the minimizer of $\sum_{t=1}^n \widehat{l}_t^{(C)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma_u^2, \boldsymbol{\theta}_3)$, because it minimizes both $\sum_{t=1}^n \widehat{l}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma_u^2)$ and $\sum_{t=1}^n \widehat{l}_t^{(GX)}(\boldsymbol{\theta}_3)$. Thus the consistency is automatically satisfied.

The asymptotic normality is easy to obtain. Let $\mathbf{H}^{(C)} = \mathsf{E}^\circ[l_t^{''(C)}]$ and $\mathbf{J}^{(C)} = \mathsf{E}^\circ[l_t^{'(C)}(l_t^{'(C)})^T]$. We then derive the asymptotic covariance $\mathbf{U} = \mathbf{H}^{-1(C)}\mathbf{J}^{(C)}\mathbf{H}^{-1(C)}$. Using the block matrix inversion formula, we have

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{G}^{-1} & \bullet & \mathbf{0} \\ \mathbf{B}_{22}^{-1}\mathbf{B}_{21}\mathbf{G}^{-1} & \mathbf{B}_{22}^{-1} + \mathbf{B}_{22}^{-1}\mathbf{B}_{21}\mathbf{G}\mathbf{B}_{21}^T\mathbf{B}_{22}^{-1} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0}^T & (\sigma_u^{2\circ})^2 \end{bmatrix}. \tag{A.7}$$

Then

$$\mathbf{H}^{(C)} = \begin{bmatrix} \mathbf{H} & \mathbf{O} \\ \mathbf{O} & \mathbf{A} \end{bmatrix}, \quad \text{and} \quad \mathbf{H}^{-1(C)} = \begin{bmatrix} \mathbf{H}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}^{-1} \end{bmatrix},$$

where $\mathbf{O}$ is a zero matrix.

Next we derive the expression of $\mathbf{J}^{(C)}$. Let $\nabla_3 = \partial/\partial\boldsymbol{\theta}_3$. Since $\nabla_1 l_t^{(C)} = \nabla_1 l_t$ and $\nabla_2 l_t^{(C)} = \nabla_2 l_t$, we have

$$\mathsf{E}^\circ[\nabla_3 l_t^{(C)} \nabla_2 (l_t^{(C)})^T] = \mathbf{O}, \quad \text{and}$$
$$\mathsf{E}^\circ[\nabla_3 l_t^{(C)} \nabla_1 (l_t^{(C)})^T] = \mathsf{E}^\circ[\nabla_3 l_t^{(C)} \nabla_3 (l_t^{(C)})^T] = \mathsf{E}[z_t^4 - 1]\mathbf{A}.$$

Then we have

$$\mathbf{J}^{(C)} = \begin{bmatrix} \mathbf{F} + 2\mathbf{A} + 2\mathbf{B}_{11} & \bullet & \bullet & \bullet \\ 2\mathbf{B}_{21} & 2\mathbf{B}_{22} & \bullet & \bullet \\ -2\frac{\mathsf{E}[u_t^3]}{(\sigma_u^{2\circ})^3}\mathsf{E}^\circ[\nabla_1 g_t^T] & -2\frac{\mathsf{E}[u_t^3]}{(\sigma_u^{2\circ})^3}\mathsf{E}^\circ[\nabla_2 g_t^T] & \frac{\mathsf{E}[u_t^4] - (\sigma_u^{2\circ})^2}{(\sigma_u^{2\circ})^4} & \bullet \\ \mathbf{F} + 2\mathbf{A} & \mathbf{O} & \mathbf{O} & \mathbf{F} + 2\mathbf{A} \end{bmatrix}.$$

Finally, with some calculation, we have the expression of $\mathbf{U}$ in Theorem 4.1. ■

Proof of Corollary 4.1

Proof. Let

$$\mathbf{J}^{(C)}(1) = \begin{bmatrix} 2\mathbf{B}_{11} & \bullet & \bullet & \bullet \\ 2\mathbf{B}_{21} & 2\mathbf{B}_{22} & \bullet & \bullet \\ -2\frac{\mathsf{E}[u_t^3]}{(\sigma_u^{2\circ})^3}\mathsf{E}^\circ[\nabla_1 g_t^T] & -2\frac{\mathsf{E}[u_t^3]}{(\sigma_u^{2\circ})^3}\mathsf{E}^\circ[\nabla_2 g_t^T] & \frac{\mathsf{E}[u_t^4] - (\sigma_u^{2\circ})^2}{(\sigma_u^{2\circ})^4} & \bullet \\ 2\mathbf{A} & \mathbf{O} & \mathbf{O} & 2\mathbf{A} \end{bmatrix},$$

and $\mathbf{J}^{(C)}(2) = \mathbf{J}^{(C)} - \mathbf{J}^{(C)}(1)$. Let $\mathbf{E} = [\mathbf{I}\,\mathbf{O}\,\mathbf{O} - \mathbf{I}]$ be a $(1+p+q) \times (7+2p+2q)$ matrix, where $\mathbf{I}$ is the identity matrix. We have

$$\begin{aligned} \Sigma &= \mathbf{E}\mathbf{H}^{-1(C)}\mathbf{J}^{(C)}\mathbf{H}^{-1(C)}\mathbf{E}^T \\ &= \mathbf{E}\mathbf{H}^{-1(C)}\mathbf{J}^{(C)}(1)\mathbf{H}^{-1(C)}\mathbf{E}^T + \mathbf{E}\mathbf{H}^{-1(C)}\mathbf{J}^{(C)}(2)\mathbf{H}^{-1(C)}\mathbf{E}^T \\ &= -2\mathbf{G}^{-1} + 2\mathbf{A}^{-1} + \mathbf{G}^{-1}\mathbf{F}\mathbf{G}^{-1} + \mathbf{A}^{-1}\mathbf{F}\mathbf{A}^{-1} - 2\mathbf{A}^{-1}\mathbf{F}\mathbf{G}^{-1} \\ &= \mathbf{V}_{11} + \mathbf{W} - 2\mathbf{A}^{-1}(\mathbf{F} + 2\mathbf{A})\mathbf{G}^{-1}, \end{aligned}$$

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

where $\mathbf{V}_{11} = \mathbf{G}^{-1}(\mathbf{F}+2\mathbf{G})\mathbf{G}^{-1}$ and $\mathbf{W} = \mathbf{A}^{-1}(\mathbf{F}+2\mathbf{A})\mathbf{A}^{-1}$.

To show the rank of Σ, we rewrite Σ as,

$$\begin{aligned}\Sigma &= (\mathbf{A}^{-1}-\mathbf{G}^{-1})(\mathbf{F}+2\mathbf{A})(\mathbf{A}^{-1}-\mathbf{G}^{-1})+2\mathbf{G}^{-1}(\mathbf{G}-\mathbf{A})\mathbf{G}^{-1}\\ &= \mathbf{G}^{-1}(\mathbf{G}-\mathbf{A})[\mathbf{A}^{-1}(\mathbf{F}+2\mathbf{A})\mathbf{A}^{-1}+2(\mathbf{G}-\mathbf{A})^{+}](\mathbf{G}-\mathbf{A})\mathbf{G}^{-1}.\end{aligned}$$

Noting that $\mathbf{G}-\mathbf{A} = \mathbf{B}_{11}-\mathbf{B}_{21}^T\mathbf{B}_{22}^{-1}\mathbf{B}_{21}$ and that $\mathbf{B}$ is a positive semidefinite matrix, thus $\mathbf{G}-\mathbf{A}$ is a positive semidefinite matrix. Since $\mathbf{A}^{-1}(\mathbf{F}+2\mathbf{A})\mathbf{A}^{-1}$ is a positive definite matrix, we have $\operatorname{rank}(\Sigma) = \operatorname{rank}(\mathbf{G}-\mathbf{A})$. As $\dim(\mathbf{G}-\mathbf{A}) = 1+p+q$, in order to prove $\operatorname{rank}(\mathbf{G}-\mathbf{A}) = p+q$, we need to show that $\dim(\ker(\mathbf{G}-\mathbf{A})) = 1$, where $\ker(\cdot)$ is the kernel of a matrix (also known as null space).

First, we prove $\dim(\ker(\mathbf{B})) = 1$. Assume there is a vector $(\check{\boldsymbol{\theta}}_1^T, \check{\boldsymbol{\theta}}_2^T)^T \neq \mathbf{0}$ such that $(\check{\boldsymbol{\theta}}_1^T, \check{\boldsymbol{\theta}}_2^T)\mathbf{B}(\check{\boldsymbol{\theta}}_1^T, \check{\boldsymbol{\theta}}_2^T)^T = 0$. Since $\mathbf{B}$ is a positive semidefinite matrix, we have $\check{\boldsymbol{\theta}}_1^T\nabla_1 g_t + \check{\boldsymbol{\theta}}_2^T\nabla_2 g_t|_{(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=(\boldsymbol{\theta}_1^\circ,\boldsymbol{\theta}_2^\circ)} = 0$, a.s. By equations (A.2), (A.3), and (A.4), we have

$$\begin{aligned}&\check{\boldsymbol{\theta}}_1^T(-\varphi^\circ+\frac{1}{2}\tau_1^\circ z_t+\tau_2^\circ z_t^2)\mathcal{B}_{\boldsymbol{\theta}_1^\circ}^{-1}(B)(1,\ln\sigma_{t-1}^2,\ldots,\ln\sigma_{t-p}^2,\ln X_{t-1},\ldots,\ln X_{t-q})^T\\ &+\check{\boldsymbol{\theta}}_2^T(-1,-\ln\sigma_t^2,-z_t,1-z_t^2)^T = 0.\end{aligned}$$

It is obvious that $\check{\boldsymbol{\theta}}_1 \neq \mathbf{0}$, otherwise we have $\check{\boldsymbol{\theta}}_1 = \mathbf{0}$ and thereby $\check{\boldsymbol{\theta}}_2 = \mathbf{0}$, which makes a contradiction. We then investigate the coefficients of z_t and z_t^2. We have for all t almost surely,

$$\left(\check{\boldsymbol{\theta}}_1^T\frac{1}{2}\tau_1^\circ\mathcal{B}_{\boldsymbol{\theta}_1^\circ}^{-1}(B)(1,\ln\sigma_{t-1}^2,\ldots,\ln\sigma_{t-p}^2,\ln X_{t-1},\ldots,\ln X_{t-q})^T-\check{\tau}_1\right)z_t = 0$$

and

$$\left(\check{\boldsymbol{\theta}}_1^T\tau_2^\circ\mathcal{B}_{\boldsymbol{\theta}_1^\circ}^{-1}(B)(1,\ln\sigma_{t-1}^2,\ldots,\ln\sigma_{t-p}^2,\ln X_{t-1},\ldots,\ln X_{t-q})^T-\check{\tau}_2\right)z_t^2 = 0.$$

Note that both coefficients of z_t and z_t^2 are $\mathcal{F}_{t-1}$-measurable. We conclude that $\check{\boldsymbol{\theta}}_1^T(1,\ln\sigma_{t-1}^2,\ldots,\ln\sigma_{t-p}^2,\ln X_{t-1},\ldots,\ln X_{t-q})^T$ is constant. Thus all the elements of $\check{\boldsymbol{\theta}}_1$ except the first one are zero. By some calculation, we have that $\ker(\mathbf{B})$ is spanned by vector $(\check{\boldsymbol{\theta}}_1^\star, \check{\boldsymbol{\theta}}_2^\star)$, where $\check{\boldsymbol{\theta}}_1^\star = (\sum_{i=1}^p\beta_i, 0,\ldots,0)$ and $\check{\boldsymbol{\theta}}_2^\star = (\tau_2^\circ-\varphi^\circ, 0, \frac{\tau_1^\circ}{2}, \tau_2^\circ)$. Thus $\dim(\ker(\mathbf{B})) = 1$.

Finally, we prove $\dim(\ker(\mathbf{G}-\mathbf{A})) = \dim(\ker(\mathbf{B}))$. On the one hand, we assume there is a vector $\check{\boldsymbol{\theta}}_1 \neq \mathbf{0}$ such that $\check{\boldsymbol{\theta}}_1^T(\mathbf{G}-\mathbf{A})\check{\boldsymbol{\theta}}_1 = 0$. Let $\check{\boldsymbol{\theta}}_2 = \mathbf{B}_{22}^{-1/2}\mathbf{B}_{21}\check{\boldsymbol{\theta}}_1$ and let $\mathbf{B}^\diamondsuit =$

$$\begin{bmatrix}\mathbf{B}_{11} & \mathbf{B}_{12}\mathbf{B}_{22}^{-1/2}\\ \mathbf{B}_{22}^{-1/2}\mathbf{B}_{21} & \mathbf{I}\end{bmatrix} = \begin{bmatrix}\mathbf{I} & \mathbf{O}\\ \mathbf{O} & \mathbf{B}_{22}^{-1/2}\end{bmatrix}\begin{bmatrix}\mathbf{B}_{11} & \mathbf{B}_{12}\\ \mathbf{B}_{21} & \mathbf{B}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{I} & \mathbf{O}\\ \mathbf{O} & \mathbf{B}_{22}^{-1/2}\end{bmatrix},$$

Then we have $(\check{\boldsymbol{\theta}}_1^T, \check{\boldsymbol{\theta}}_2^T)\mathbf{B}^\diamondsuit(\check{\boldsymbol{\theta}}_1^T, \check{\boldsymbol{\theta}}_2^T)^T = \check{\boldsymbol{\theta}}_1^T(\mathbf{G}-\mathbf{A})\check{\boldsymbol{\theta}}_1 = 0$. Thus $\dim(\ker(\mathbf{G}-\mathbf{A})) \leq \dim(\ker(\mathbf{B}^\diamondsuit))$.

On the other hand, assume there is a vector $(\check{\boldsymbol{\theta}}_1^T, \check{\boldsymbol{\theta}}_2^T)^T \neq \mathbf{0}$ such that $(\check{\boldsymbol{\theta}}_1^T, \check{\boldsymbol{\theta}}_2^T)\mathbf{B}^\diamondsuit(\check{\boldsymbol{\theta}}_1^T, \check{\boldsymbol{\theta}}_2^T)^T = 0$. Since

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press

$$\mathbf{B}^{\diamondsuit} = \begin{bmatrix} \mathbf{O} & \mathbf{B}_{12}\mathbf{B}_{22}^{-1/2} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{B}_{22}^{-1/2}\mathbf{B}_{21} & \mathbf{I} \end{bmatrix} + \begin{bmatrix} \mathbf{G}-\mathbf{A} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix},$$

we have $\breve{\boldsymbol{\theta}}_1^T(\mathbf{G}-\mathbf{A})\breve{\boldsymbol{\theta}}_1 = 0$. Thus $\dim(\ker(\mathbf{G}-\mathbf{A})) \geq \dim(\ker(\mathbf{B}^{\diamondsuit}))$.

In conclusion, $\dim(\ker(\mathbf{G}-\mathbf{A})) = \dim(\ker(\mathbf{B}^{\diamondsuit})) = \dim(\ker(\mathbf{B})) = 1$. ■

A.4. Proof of Theorem 4.2

Proof. Using the formula of $\mathbf{H}^{-1}$ in equation (A.7), we have $\mathbf{V}_{11} = \mathbf{G}^{-1}(\mathbf{F}+2\mathbf{G})\mathbf{G}^{-1}$. Compared with $\mathbf{W} = \mathbf{A}^{-1}(\mathbf{F}+2\mathbf{A})\mathbf{A}^{-1}$, using $\mathbf{G} \geq \mathbf{A}$, we have $\mathbf{V}_{11} \leq \mathbf{W}$. ■

https://doi.org/10.1017/S0266466618000348 Published online by Cambridge University Press