Econometrics **2013**, *1*, 1-31; doi:10.3390/econometrics1010001

OPEN ACCESS

econometrics

ISSN 2225-1146

www.mdpi.com/journal/econometrics

Article

On Diagnostic Checking of Vector ARMA-GARCH Models with Gaussian and Student-t Innovations

Yongning Wang and Ruey S. Tsay *

Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA

* Author to whom correspondence should be addressed; E-Mail: ruey.tsay@chicagobooth.edu; Tel.: +1-773-702-6750; Fax: +1-773-702-0458.

Received: 15 January 2013; in revised form: 9 March 2013 / Accepted: 14 March 2013 / Published: 4 April 2013

Abstract: This paper focuses on the diagnostic checking of vector ARMA (VARMA) models with multivariate GARCH errors. For a fitted VARMA-GARCH model with Gaussian or Student-t innovations, we derive the asymptotic distributions of autocorrelation matrices of the cross-product vector of standardized residuals. This is different from the traditional approach that employs only the squared series of standardized residuals. We then study two portmanteau statistics, called $Q_1(M)$ and $Q_2(M)$, for model checking. A residual-based bootstrap method is provided and demonstrated as an effective way to approximate the diagnostic checking statistics. Simulations are used to compare the performance of the proposed statistics with other methods available in the literature. In addition, we also investigate the effect of GARCH shocks on checking a fitted VARMA model. Empirical sizes and powers of the proposed statistics are investigated and the results suggest a procedure of using jointly $Q_1(M)$ and $Q_2(M)$ in diagnostic checking. The bivariate time series of FTSE 100 and DAX index returns is used to illustrate the performance of the proposed portmanteau statistics. The results show that it is important to consider the cross-product series of standardized residuals and GARCH effects in model checking.

Keywords: Vector autoregressive moving-average process; multivariate GARCH model; asymptotic distribution; portmanteau statistic; model checking; heavy tail; multivariate time series; bootstrap

---

1. Introduction

Model checking, or diagnostic checking, is an important step in statistical modeling, especially in the iterative model-building process of Box and Jenkins [1]. Like other statistical analysis, standardized residuals are often used to check the adequacy of a fitted time series model. Basically, one examines the standardized residual plots for outliers and violations of randomness and performs statistical tests to detect serial dependence in the residual series. In the univariate case, residuals of the fitted model are used to obtain the portmanteau test for autoregressive moving-average (ARMA) processes. See Box and Pierce [2] and Ljung and Box [3] for the commonly used Ljung–Box $Q(m)$ statistic. The test has been extended to multivariate autoregressive processes by Chitturi [4] and multivariate ARMA processes by Hosking [5] and Li and McLeod [6].

Most of the previous studies in time-series model checking assume that the innovations are independent and identically distributed (iid) random variables. This assumption is known to be questionable for the data in economics and finance, especially after the introduction of the generalized autoregressive conditional heteroscedastic (GARCH) models of Engle [7] and Bollerslev [8]. In fact, conditional heteroscedasticity exists not only in time series of asset returns and foreign exchange rates, but also in series of traffic volume of a big city or on the internet. For univariate time series with conditional heteroscedasticity, Diebold examines the impact of GARCH effects on the Bartlett standard errors and the Ljung–Box statistic, and proposes a robust Ljung–Box statistic [9,10]. Ling and Li derive the asymptotic properties of maximum likelihood estimators and the $Q(m)$ statistic for univariate fractionally integrated ARMA-GARCH models [11]. Ling and Li use the sum of squared series of standardized residuals to define a multivariate $Q(m)$ statistic and derive its asymptotic distribution when the time series has ARCH errors [12]. Tse examined the residual-based diagnostics for univariate and multivariate conditional heteroscedasticity models [13].

In this paper, we study the portmanteau statistics for the cross-product vector of the standardized residuals of a vector ARMA model with multivariate GARCH errors. Specifically, we employ the process $vech(\hat{\varepsilon}_t\hat{\varepsilon}_t{}')$ in model checking, where $\hat{\varepsilon}_t$ is the standardized residuals of a fitted vector ARMA-GARCH model and $vech$ denotes the half-stacking operator of a symmetric matrix. The innovations of the GARCH errors follow either a multivariate Gaussian or a multivariate Student-t distribution. Under the commonly used assumption of multivariate GARCH processes, the cross-product vector of the standardized residuals should be serially independent. The proposed portmanteau statistics are aimed at verifying this independence condition. They are more general than the traditional multivariate $Q(m)$ statistics because the latter statistics only employ the squared series $\hat{\varepsilon}_t^2$ of the standardized residuals. For instance, Ling and Li use $\hat{\varepsilon}_t'\hat{\varepsilon}_t = \sum_{i=1}^{k}\hat{\varepsilon}_{it}^2$ to detect conditional heteroscedasticity [12]. The improved performance of the proposed test statistics over the traditional ones is demonstrated by simulation and a real example. Another contribution of the paper is to consider the limiting distribution of the proposed test statistics when the innovations follow a multivariate Student-t distribution. This is highly relevant as most financial time series exhibit certain heavy-tail phenomenon. Furthermore, considering the difficulty in computing the proposed test statistics in real application, we provide a bootstrap approach based on the re-sampled standardized residuals to approximate the (asymptotic) distributions of the sample cross-covariance matrices of the standardized residuals. It is demonstrated by simulated data and the real example that the bootstrap method gives an effective way to obtain the proposed model checking statistics.

The paper is organized as follows. In Section 2, we define the model considered in the paper and state the assumptions used. In Section 3, we consider the maximum likelihood estimation of a vector ARMA-GARCH model with Gaussian innovations and derive the asymptotic distributions of the sample cross-covariance matrices of the standardized residuals and the proposed test statistics. We investigate estimation and model checking for a vector ARMA-GARCH model with multivariate Student-t innovations in Section 4. In Section 5, we introduce the residual-based bootstrap method and justify its validity. We use simulation studies in Section 6 to study the performance of the proposed test statistics in finite samples. An empirical example is analyzed in Section 7 and Section 8 concludes. Finally, complicated proofs are in the Appendix.

2. The Model

Let $Y_t = (y_{1t}, \ldots, y_{kt})'$ be a k-dimensional time series. In this paper, we assume that Y_t follows a stationary and invertible vector autoregressive moving-average, VARMA(p, q), model with shock $a_t = (a_{1t}, \ldots, a_{kt})'$ being a multivariate GARCH(r, s) process with innovation $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{kt})'$. Specifically, we have

$$Y_t = \sum_{i=1}^{p} \Phi_i Y_{t-i} + a_t + \sum_{j=1}^{q} \Theta_j a_{t-j}, \qquad a_t = \Sigma_t^{1/2} \varepsilon_t, \tag{1}$$

$$\Sigma_t = A_0 + \sum_{i=1}^{r} A_i a_{t-i} a'_{t-i} A'_i + \sum_{i=1}^{s} B_i \Sigma_{t-i} B'_i, \quad t = 0, 1, \ldots \tag{2}$$

where p, q, r and s are non-negative integers, $\{\varepsilon_t\}$ are independent and identically distributed random vectors with mean zero and identity covariance matrix. Let $\mathcal{F}_j$ denote the information available at time j, *i.e.*, $\mathcal{F}_j = \sigma\{\varepsilon_j, \varepsilon_{j-1}, \cdots\}$ then we have a_t satisfies $E(a_t|\mathcal{F}_{t-1}) = 0$ and cov$(a_t|\mathcal{F}_{t-1}) = \Sigma_t$, with $\Sigma_t^{1/2}$ being the positive-definite square-root matrix of the conditional covariance matrix Σ_t, and ε_t follows either a multivariate Gaussian or Student-t distribution. Define $\varphi = vec(\Phi_1, ..., \Phi_p, \Theta_1, ..., \Theta_q)$, $\delta = vec(A_0, A_1, \ldots, A_r, B_1, \ldots, B_s)$, and $\lambda = (\varphi', \delta')'$. Denote the true parameter vector by $\lambda_0 = (\varphi'_0, \delta'_0)'$. Here we assume that $E(Y_t) = 0$ for simplicity. The multivariate GARCH model in Equation (2) is a special case of the BEKK model of Engle and Kroner [14]. We use it instead of the general BEKK model purely for simplicity. For more details about multivariate time series models, see Tsay [15] and Lütkepohl [16].

Let $H_t = vec(\Sigma_t)$ be the column-stacking vector of the matrix Σ_t and denote the Kronecker product of matrices A and B by $A \otimes B$. The volatility model in Equation (2) can be written as

$$H_t = vec(A_0) + \sum_{i=1}^{r} (A_i \otimes A_i) vec(a_{t-i} a'_{t-i}) + \sum_{i=1}^{s} (B_i \otimes B_i) H_{t-i}. \tag{3}$$

For the model considered, we make the following assumptions:

Assumption 1. (i).The parameter space Λ is a compact subset of $\mathbb{R}^{(1+p+q+r+s)k^2}$. (ii). The true parameter $\lambda_0 = (\phi_0', \delta_0')'$ is an interior point of the compact set Λ. (iii). λ_0 is identifiably unique with respect to the log-likelihood function.

Assumption 2. Let $\Phi(L) = I - \sum_{i=1}^{p} \Phi_i L^i$ be the matrix polynomial of the autoregressive part of the model, where L is the lag operator such that $LY_t = Y_{t-1}$. All zeros of the polynomial $|\Phi(L)|$ are outside the unit circle, *i.e.*, they are greater than 1 in modules.

Assumption 3. Let $\Theta(L) = I + \sum_{i=1}^{q} \Theta_i L^i$ be the matrix polynomial of the moving-average part of the model. All zeros of the polynomial $|\Theta(L)|$ are outside the unit circle.

Assumption 4. $\Phi(L)$ and $\Theta(L)$ are left coprime and the matrices Φ_p and Θ_q satisfy the condition that $Rank[\Phi_p, \Theta_q] = dim(Y_t)$.

Assumption 5. A_0 is a $k \times k$ positive definite matrix.

Assumption 6. A_i and B_j are arbitrary $k \times k$ matrices such that $\rho(\sum_{i=1}^{r} A_i \otimes A_i + \sum_{j=1}^{s} B_j \otimes B_j) < 1$, where $\rho(A)$ denotes the modulus of the matrix A, *i.e.*, the largest eigenvalue of A in modulus.

Assumption 7. $I_{k^2} - \sum_{i=1}^{r}(A_i \otimes A_i)L^i$ and $\sum_{j=1}^{s}(B_j \otimes B_j)L^j$ are left coprime and satisfy the conditions that $Rank[A_r, B_s] = dim(Y_t)$.

Assumption 8. Let $N = k(k+1)/2$. All eigenvalues of the matrix $\sum_{i=1}^{\infty}(\Psi_i \otimes \Psi_i)(G_k - I_{N^2})$ have modulus smaller than one, where Ψ_i are $N \times N$ matrices defined recursively by:

$$\begin{aligned} &\Psi_0 = I_N, \\ &\Psi_i = -L_k(B_i \otimes B_i)D_k + \sum_{j=1}^{i} L_k(A_j \otimes A_j + B_j \otimes B_j)D_k \Psi_{i-j}, \quad i = 1, 2, ... \end{aligned} \tag{4}$$

$G_k = 2(L_k \otimes D_k^+) + (I_k \otimes K_{kk} \otimes I_k)(D_k \otimes D_k) + I_{N^2}$, and $D_k^+ = (D_k' D_k)^{-1} D_k'$, where L_k, D_k, and $K_{k_1 k_2}$ are the elimination, duplication, and commutation matrices, respectively. Specifically, the dimensions of L_k, D_k and $K_{k_1 k_2}$ are $N \times k^2$, $k^2 \times N$ and $k_1 k_2 \times k_1 k_2$, respectively. These matrices are useful in dealing with the vec operator, and they satisfy

$$vech(\Sigma_t) = L_k vec(\Sigma_t), \quad vec(\Sigma_t) = D_k vech(\Sigma_t), \quad vec(A') = K_{k_1 k_2} vec(A),$$

where A is a $k_1 \times k_2$ matrix.

Assumption 1 is standard. Assumptions 2 and 3 ensure that the VARMA model is stationary and invertible. Assumption 4, which is referred to as the block identifiability condition, is sufficient for the VARMA model to be identifiable. Under Assumption 5 and using properties of the BEKK representation of Engle and Kroner [14], we are guaranteed to have positive definite covariance matrices Σ_t, and A_0 can be written by its Cholesky decomposition $A_0 = CC'$. Furthermore, it can be shown that, under Assumption 6, a_t is strictly stationary and ergodic with $E \parallel a_t \parallel^2 < \infty$. Similar as Assumption 4, Assumption 7 is sufficient for the identifiability of the GARCH process. Based on Theorem 2 of Hafner [17], Assumption 8 is a sufficient and necessary condition for $E \parallel a_t \parallel^4 < \infty$.

3. Diagnostic Checking for VARMA-GARCH Models with Gaussian Innovations

In this section, we assume that the innovations $\{\varepsilon_t\}$ of the GARCH model follow a multivariate Gaussian distribution with mean zero and cov(ε_t) = I_k, the $k \times k$ identity matrix.

3.1. Estimation

Suppose that $Y_1, ..., Y_n$ are a realization of the vector ARMA-GARCH model in Equations (1)-(2). Given $\mathcal{F}_0$, the approximate maximum likelihood estimate (MLE) $\hat{\lambda}_n$ of λ maximizes the conditional log-likelihood function,

$$L_n(\lambda) = \frac{1}{n}\sum_{t=1}^{n} l_t(\lambda), \qquad l_t(\lambda) = -\frac{1}{2}\ln det(\Sigma_t) - \frac{1}{2}a_t'\Sigma_t^{-1}a_t. \tag{5}$$

where

$$a_t = \Theta(L)^{-1}\Phi(L)Y_t \tag{6}$$

and Σ_t follows Equation (2).

To obtain $\hat{\lambda}_n$, we consider the first-order derivatives and the information matrix of the log likelihood function. They can be calculated as follows:

$$\frac{\partial l_t}{\partial \varphi} = \frac{1}{2}(\frac{\partial H_t}{\partial \varphi'})' vec(\Sigma_t^{-1}a_t a_t'\Sigma_t^{-1} - \Sigma_t^{-1}) - (\frac{\partial a_t}{\partial \varphi'})'\Sigma_t^{-1}a_t, \tag{7}$$

$$\frac{\partial l_t}{\partial \delta} = \frac{1}{2}(\frac{\partial H_t}{\partial \delta'})' vec(\Sigma_t^{-1}a_t a_t'\Sigma_t^{-1} - \Sigma_t^{-1}), \tag{8}$$

where

$$\frac{\partial a_t}{\partial \varphi'} = \Theta^{-1}(L)[X_{t-1} \otimes I_k], \quad X_{t-1} = (Y_{t-1}', ..., Y_{t-p}', a_{t-1}', ..., a_{t-q}'),$$

$$\frac{\partial H_t}{\partial \varphi'} = \left(I_{k^2} - \sum_{i=1}^{s}(B_i \otimes B_i)L^i\right)^{-1}\left[\sum_{i=1}^{r}(A_i \otimes A_i)L^i((I_k \otimes a_t + a_t \otimes I_k)\frac{\partial a_t}{\partial \varphi'})\right],$$

$$\frac{\partial H_t}{\partial \delta'} = \left(I_{k^2} - \sum_{i=1}^{s}(B_i \otimes B_i)L^i\right)^{-1}[I_{k^2}, \widetilde{H}_{t-1}^{(1)}, ..., \widetilde{H}_{t-r}^{(1)}, \widetilde{H}_{t-1}^{(2)}, ..., \widetilde{H}_{t-s}^{(2)}],$$

$$\widetilde{H}_{t-i}^{(1)} = (I_k \otimes A_i)\cdot[I_k \otimes (a_{t-i}a_{t-i}'e_1), ..., I_k \otimes (a_{t-i}a_{t-i}'e_k)] + (A_i \otimes I_k)\cdot((a_{t-i}a_{t-i}') \otimes I_k),$$

$$\widetilde{H}_{t-i}^{(2)} = (I_k \otimes A_i)\cdot[I_k \otimes (\Sigma_{t-i}e_1), ..., I_k \otimes (\Sigma_{t-i}e_k)] + (A_i \otimes I_k)\cdot(\Sigma_{t-i} \otimes I_k),$$

and e_j is the j-th unit vector with 1 in the j-th element and 0 elsewhere.

Differentiating Equations (7) and (8) conditional on $\mathcal{F}_{t-1}$ and taking expectation, we have that at $\lambda = \lambda_0$,

$$E\left[\frac{\partial^2 l_t}{\partial\varphi\partial\varphi'}|\mathcal{F}_{t-1}\right] = -\frac{1}{2}(\frac{\partial H_t}{\partial\varphi'})'(\Sigma_t^{-1}\otimes\Sigma_t^{-1})(\frac{\partial H_t}{\partial\varphi'}) - (\frac{\partial a_t}{\partial\varphi'})'\Sigma_t^{-1}(\frac{\partial a_t}{\partial\varphi'}), \quad (9)$$

$$E\left[\frac{\partial^2 l_t}{\partial\delta\partial\delta'}|\mathcal{F}_{t-1}\right] = -\frac{1}{2}(\frac{\partial H_t}{\partial\delta'})'(\Sigma_t^{-1}\otimes\Sigma_t^{-1})(\frac{\partial H_t}{\partial\delta'}), \quad (10)$$

$$E\left[\frac{\partial^2 l_t}{\partial\varphi\partial\delta'}|\mathcal{F}_{t-1}\right] = O. \quad (11)$$

The following theorems provide the asymptotic properties of the information matrix and MLE $\hat{\lambda}_n$ of the model.

Theorem 3.1. Suppose $\{Y_t\}$ and $\{a_t\}$ are generated by (1) and (2) with ε_t being multivariate Gaussian with mean zero and identity covariance matrix. Assume that Assumptions 1–6 hold, then at $\lambda = \lambda_0$,

$$-\frac{1}{n}\sum_{t=1}^{n}\left[\begin{array}{cc}\partial^2 l_t/(\partial\varphi\partial\varphi') & \partial^2 l_t/(\partial\delta\partial\varphi') \\ \partial^2 l_t/(\partial\varphi\partial\delta') & \partial^2 l_t/(\partial\delta\partial\delta')\end{array}\right] \to_{a.s.} \left[\begin{array}{cc}\Omega_\varphi & O \\ O & \Omega_\delta\end{array}\right] \equiv \Omega$$

as $n \to \infty$, where $\to_{a.s.}$ denotes convergence with probability 1, and Ω_φ and Ω_δ are positive definite matrices given by

$$\Omega_\varphi = E[\frac{1}{2}(\frac{\partial H_t}{\partial\varphi'})'(\Sigma_t^{-1}\otimes\Sigma_t^{-1})(\frac{\partial H_t}{\partial\varphi'}) + (\frac{\partial a_t}{\partial\varphi'})'\Sigma_t^{-1}(\frac{\partial a_t}{\partial\varphi'})],$$
$$\Omega_\delta = E[\frac{1}{2}(\frac{\partial H_t}{\partial\delta'})'(\Sigma_t^{-1}\otimes\Sigma_t^{-1})(\frac{\partial H_t}{\partial\delta'})].$$

Theorem 3.2. Under the assumptions of Theorem 3.1, the following results hold:
(a). There exists a MLE $\hat{\lambda}_n$ satisfying the equation $\partial L_n(\lambda)/\partial\lambda = 0$ and $\hat{\lambda}_n \to^p \lambda_0$ as $n \to \infty$.
(b). $\sqrt{n}(\hat{\lambda}_n - \lambda_0) \to_{\mathcal{L}} N(0, \Omega_0^{-1})$ as $n \to \infty$, where $\to_{\mathcal{L}}$ denotes convergence in distribution, $\Omega_0 = diag\{\Omega_{\varphi 0}, \Omega_{\delta 0}\}$, and $\Omega_{\varphi 0}$ and $\Omega_{\delta 0}$ are values of Ω_φ and Ω_δ at $\lambda = \lambda_0$. Further, the information matrices $\Omega_{\varphi 0}$ and $\Omega_{\delta 0}$ can be estimated consistently and separately by

$$\Omega_\varphi = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{1}{2}(\frac{\partial H_t}{\partial\varphi'})'(\Sigma_t^{-1}\otimes\Sigma_t^{-1})(\frac{\partial H_t}{\partial\varphi'}) + (\frac{\partial a_t}{\partial\varphi'})'\Sigma_t^{-1}(\frac{\partial a_t}{\partial\varphi'})\right], \quad (12)$$

$$\Omega_\delta = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{1}{2}(\frac{\partial H_t}{\partial\delta'})'(\Sigma_t^{-1}\otimes\Sigma_t^{-1})(\frac{\partial H_t}{\partial\delta'})\right], \quad (13)$$

where the terms are evaluated at $\lambda = \hat{\lambda}_n$.

The proofs of Theorems 3.1 and 3.2 are similar to those of Theorems 3.1 and 3.2 in Ling and Li [11] and are omitted.

3.2. Asymptotic Distributions of Sample Matrices and Diagnostic Checking

In this subsection, we investigate the asymptotic distributions of sample autocovariance matrices of the standardized residuals and propose two portmanteau statistics, namely $Q_1(M)$ and $Q_2(M)$, for checking a fitted VARMA-GARCH model when the innovations are Gaussian, where M is a pre-specified positive integer. We start with the $Q_1(M)$ statistic.

For the innovation $\varepsilon_t = \Sigma_t^{-1/2} a_t$, define its lag-$m$ sample autocorrelation matrix by $\rho_m = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t \varepsilon'_{t-m}$, where we make use of the condition $E(\varepsilon_t) = 0$ and cov$(\varepsilon_t) = I_k$. Let $vec(\rho) = (vec'(\rho_1), ..., vec'(\rho_M))'$. We have the following lemma concerning the distribution of $vec(\rho)$.

Lemma 3.1. If the time series Y_t is generated by the VARMA-GARCH model in Equations (1)-(2) with standard multivariate Gaussian innovations, then under Assumptions 1–6,

$$\sqrt{n} vec(\rho_m) \rightarrow_{\mathcal{L}} N(0, I_{k^2}).$$

Proof. Since $n \cdot vec(\rho_m) = \sum_{t=m+1}^{n} vec(\varepsilon_t \varepsilon'_{t-m}) \equiv \sum_{t=m+1}^{n} vec(W_{tm})$, and noting that, for $m > 0$, $\{W_{tm}, \mathcal{F}_{t-1}\}$ is a martingale difference, it can be shown by Central Limit Theorem for martingale difference and the Cramer–Wold device that the asymptotic distribution of $\sqrt{n} vec(\rho_m)$ is $N(0, E[vec(W_{tm})vec'(W_{tm})])$. The covariance matrix can be obtained as $E[vec(W_{tm})vec'(W_{tm})] = E\{[(\varepsilon_{t-m}\varepsilon'_{t-m}) \otimes (\varepsilon_t \varepsilon'_t)]\}$ and $E[\varepsilon_t \varepsilon'_t | \mathcal{F}_{t-1}] = $ cov$(\varepsilon_t) = I_k$. Therefore, $E[vec(W_{tm})vec'(W_{tm})] = I_{k^2}$. □

Let $\hat{a}_t = \sum_{i=1}^{p} \widehat{\Phi}_i Y_{t-i} - \sum_{j=1}^{q} \widehat{\Theta}_j a_{t-j}$ be the residual of a fitted model, where $\widehat{\Phi}_i$ and $\widehat{\Theta}_j$ are MLE of Φ_i and Θ_j, respectively. Also, let $\hat{\mu}_t$ and $\widehat{\Sigma}_t$ be the MLE of the mean vector of Y_t and the conditional covariance matrix of a_t. Denote by $\widehat{\Sigma}_t^{1/2}$ the positive definite square-root matrix of $\widehat{\Sigma}_t$. We estimate the innovation ε_t by $\hat{\varepsilon}_t = \widehat{\Sigma}_t^{-1/2} \hat{a}_t$. The lag-$m$ autocorrelation matrix of standardized residuals is defined as $\hat{\rho}_m = n^{-1} \sum_{t=m+1}^{n} (\hat{\varepsilon}_t - \tilde{\varepsilon})(\hat{\varepsilon}_{t-m} - \tilde{\varepsilon})'$, where $\tilde{\varepsilon} = n^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t$.

To find the asymptotic distribution of $\sqrt{n} vec(\hat{\rho}_m)$, we need Lemma 2 below that provides some properties of $E[\frac{\partial vec(\rho_m)}{\partial \lambda'}]$.

Lemma 3.2. Under the assumptions of Lemma 3.1, we have that at $\lambda = \lambda_0$,

$$E[\frac{\partial L_n}{\partial \varphi} vec'(\rho_m)] = -E[(\frac{\partial a_t}{\partial \varphi'})'(\varepsilon'_{t-m} \otimes \Sigma_t^{-1/2})], \quad E[\frac{\partial L_n}{\partial \delta} vec'(\rho_m)] = O,$$

and

$$E[(\frac{\partial L_n}{\partial \lambda'})' vec'(\rho_m)] = -E[(\frac{\partial vec(\rho_m)}{\partial \lambda'})'].$$

Define $\Upsilon' = (\Upsilon'_1, ..., \Upsilon'_M)$, where $\Upsilon_m = [-E\{(\varepsilon_{t-m} \otimes \Sigma_t^{-1/2})(\frac{\partial a_t}{\partial \varphi'})\}, O']$. Since $\hat{\lambda}_n - \lambda_0 = (n\Omega_0)^{-1}(\partial l / \partial \lambda) + o_p(n^{-1/2})$, and by Lemma 2, the asymptotic covariance between $n^{1/2}(\hat{\lambda}_n - \lambda_0)$ and $n^{1/2} vec(\rho)$ is given by $\Omega_0^{-1}\Upsilon'$. Using a standard Taylor's expansion, we have

$$vec(\hat{\rho}) = vec(\rho) + \frac{\partial vec(\rho)}{\partial \lambda'}(\hat{\lambda} - \lambda) + o_p(n^{-1/2}).$$

It is then straightforward to obtain the following theorem given the asymptotic distribution of $vec(\hat{\rho})$.

Theorem 3.3. Suppose the time series Y_t is generated by (1)-(2) with standard multivariate Gaussian innovations. Then, under Assumptions 1–6, $\sqrt{n} vec(\hat{\rho})$ is asymptotically normal with mean 0 and covariance matrix

$$V_1 = I_M \otimes I_{k^2} - \Upsilon \Omega_0^{-1} \Upsilon'.$$

Note that when $k = 1$, Theorem 3 reduces to Theorem 4.1 of Ling and Li [11].

For conditional heteroscedastic data, it is important to explore the asymptotic properties of autocovariance matrices of the cross-product vector of standardized residuals. To this end, we consider the $Q_2(M)$ statistic below. For the innovation ε_t and the standardized residual $\hat{\varepsilon}_t$, define the mean-corrected matrix processes $c_t = \varepsilon_t\varepsilon_t' - I_k$ and $\hat{c}_t = \hat{\varepsilon}_t\hat{\varepsilon}_t' - \tilde{c}$, where $\tilde{c} = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_t'$ is the sample covariance matrix of the standardized residuals. Next, define the lag-m autocovariance matrices of the cross-product vectors of ε_t and $\hat{\varepsilon}_t$ as follows:

$$\begin{aligned}
\widetilde{r}_m &= \frac{1}{n}\sum_{t=m+1}^{n} vec(c_t)vec'(c_{t-m}), \qquad \hat{\widetilde{r}}_m = \frac{1}{n}\sum_{t=m+1}^{n} vec(\hat{c}_t)vec'(\hat{c}_{t-m}) \\
r_m &= \frac{1}{n}\sum_{t=m+1}^{n} vech(c_t)vech'(c_{t-m}), \qquad \hat{r}_m = \frac{1}{n}\sum_{t=m+1}^{n} vech(\hat{c}_t)vech'(\hat{c}_{t-m}).
\end{aligned}$$

Let $vec(r) = (vec'(r_1), ..., vec'(r_M))'$, then we have the following lemma.

Lemma 3.3. If the time series Y_t is generated by the VARMA-GARCH model (1)–(2) with standard multivariate Gaussian innovation ε_t. Then, under Assumptions 1–6,

$$\sqrt{n}vec(\widetilde{r}_m) \to_{\mathcal{L}} N[0, (I_{k^2} + K_{k,k}) \otimes (I_{k^2} + K_{k,k})].$$

Proof. From the definition, $n \cdot vec(\widetilde{r}_m) = \sum_{t=m+1}^{n} vec(U_{tm})$, where $U_{tm} = vec(c_t)vec'(c_{t-m})$. For $m > 0$, it can be shown, similar to Lemma 1, that the asymptotic distribution of $\sqrt{n}vec(\widetilde{r}_m)$ is $N(0, E[vec(U_{tm})vec'(U_{tm})])$. The covariance matrix can be calculated via $E[vec(U_{tm})vec'(U_{tm})]$ = $E\{[vec(c_{t-m}) \otimes vec(c_t)][vec'(c_{t-m}) \otimes vec'(c_t)]\} = E\{[vec(c_{t-m})vec'(c_{t-m})] \otimes [vec(c_t)vec'(c_t)]\}$, and $E[vec(c_t) \otimes vec'(c_t)|\mathcal{F}_{t-1}] = \text{cov}[vec(\varepsilon_t\varepsilon_t')] = I_{k^2} + K_{k,k}$. Therefore, $E[vec(U_{tm})vec'(U_{tm})]$ = $(I_{k^2} + K_{k,k}) \otimes (I_{k^2} + K_{k,k})$. $\square$

Using $vec(r_m) = (D_k^+ \otimes D_k^+)vec(\widetilde{r}_m)$ and $D_k^+ K_{k,k} = D_k^+$, we obtain the following corollary for $\sqrt{n}vec(r_m)$.

Corollary 3.1. If the time series Y_t is generated by the VARMA-GARCH model (1)–(2) with the standard multivariate Gaussian innovation. Then, under Assumptions 1–6,

$$\sqrt{n}vec(r_m) \to_{\mathcal{L}} N[0, 4(D_k^+ D_k^{+'}) \otimes (D_k^+ D_k^{+'})].$$

To find the asymptotic distribution of $\sqrt{n}vec(\hat{r}_m)$, we also need the following lemma for $E[(\frac{\partial L_n}{\partial \lambda'})'vec'(r_m)]$ and $E[\frac{\partial vec(r_m)}{\partial \lambda'}]$.

Lemma 3.4. Under the assumptions of Lemma 3.3, we have that at $\lambda = \lambda_0$,

$$\begin{aligned}
E\left[\frac{\partial L_n}{\partial \lambda}vec'(r_m)\right] &= \frac{1}{n}\sum_{t=m+1}^{n} E\left\{\left(\frac{\partial vec(\Sigma_t^{1/2})}{\partial \lambda'}\right)'(\Sigma_t^{-1/2} \otimes I_k + I_k \otimes \Sigma_t^{-1/2})[vech'(c_{t-m}) \otimes D_k^{+'}]\right\} \\
&\equiv \mathbf{X}_m', \qquad (14)
\end{aligned}$$

and

$$E\left[(\frac{\partial L_n}{\partial \lambda'})'vec'(r_m)\right] = -E\left[(\frac{\partial vec(r_m)}{\partial \lambda'})'\right].$$

Similar as Theorem 3.3, the following theorem provides the asymptotic distribution of $vec(\hat{r}_m)$.

Theorem 3.4. Suppose the time series Y_t is generated by (1)-(2) with the standard multivariate Gaussian innovation ε_t. Then, under Assumptions 1–6, $\sqrt{n}vech(\hat{r})$ is asymptotically normal with mean 0 and covariance matrix

$$V_2 = 4I_M \otimes (D_k^+ D_k^{+'}) \otimes (D_k^+ D_k^{+'}) - \mathbf{X}\Omega_0^{-1}\mathbf{X}',$$

where $\mathbf{X}' = (\mathbf{X}_1', ..., \mathbf{X}_M')$ with $\mathbf{X}_m$ being defined in (14).

Again when $k = 1$, Theorem 3.4 is a generalization of Theorem 4.2 of Ling and Li [11]. By Theorems 3.3 and 3.4, we have

$$Q_1(M) = T \cdot vec'(\hat{\rho}) \cdot \hat{V}_1^{-1} \cdot vec(\hat{\rho}) \sim \chi^2(Mk^2) \tag{15}$$

$$Q_2(M) = T \cdot vec'(\hat{r}) \cdot \hat{V}_2^{-1} \cdot vec(\hat{r}) \sim \chi^2(M[\frac{k(k+1)}{2}]^2) \tag{16}$$

where $\hat{V}_1 = I_M \otimes I_{k^2} - \hat{\mathbf{\Upsilon}}\hat{\Omega}_0^{-1}\hat{\mathbf{\Upsilon}}'$, $\hat{V}_2 = 4I_M \otimes (D_k^+ D_k^{+'}) \otimes (D_k^+ D_k^{+'}) - \hat{\mathbf{X}}\hat{\Omega}_0^{-1}\hat{\mathbf{X}}'$, $\hat{\Omega}_0 = diag(\hat{\Omega}_\varphi, \hat{\Omega}_\delta)$. $\hat{\mathbf{\Upsilon}} = (\hat{\mathbf{\Upsilon}}_1', ..., \hat{\mathbf{\Upsilon}}_M')'$, where $\hat{\mathbf{\Upsilon}}_m = \frac{1}{n}\sum_{t=m+1}^n (\hat{\varepsilon}_{t-m} \otimes \Sigma_t^{-1/2}(\partial a_t/\partial\lambda'))$ is estimated at $\lambda = \hat{\lambda}_n$. The same applies to $\hat{\mathbf{X}} = (\hat{\mathbf{X}}_1', ..., \hat{\mathbf{X}}_M')'$, where

$$\hat{\mathbf{X}}_m = \frac{1}{n}\sum_{t=m+1}^n (vec(\hat{c}_{t-m}) \otimes I_{k^2})(I_k \otimes \Sigma_t^{-1/2} a_t a_t' + \Sigma_t^{-1/2} a_t a_t' \otimes I_k)\frac{\partial vec(\Sigma_t^{-1/2})}{\partial\lambda'}.$$

The statistics $Q_1(M)$ and $Q_2(M)$ can be used jointly to test the simultaneous significance of $\hat{\rho}_i$ and $\hat{r}_i$, $i = 1, 2, ..., M$. In addition, the statistic $Q_2(1) = \frac{T}{4} \cdot vec(\hat{r}) \cdot [(D_k^+ D_k^{+'}) \otimes (D_k^+ D_k^{+'})]^{-1} \cdot vec'(\hat{r}) \sim \chi^2(\frac{k(k+1)}{2})$ can be used to test whether a fitted VARMA model has GARCH innovations.

4. Diagnostic Checking for VARMA-GARCH Models with Multivariate Student-t Innovations

The heavy-tail phenomenon is commonly seen in financial data. To properly describe this phenomenon, multivariate Student-t distributions are often employed in volatility modeling of multiple asset returns. It is, then, desirable to investigate model checking of a fitted vector ARMA-GARCH model with (standardized) multivariate Student-t innovations. Specifically, in this section, we assume that the probability density function of ε_t is

$$f(\varepsilon_t) = \frac{\Gamma((\nu+k)/2)}{[\pi(\nu-2)]^{k/2}\Gamma(\nu/2)}[1 + \frac{\varepsilon_t\varepsilon_t'}{\nu-2}]^{-(\nu+k)/2}, \tag{17}$$

where ν is a positive number denoting the degrees of freedom. Note that $E(\varepsilon_t) = 0$ and cov$(\varepsilon_t) = I_k$. We further assume that $\nu > 4$ so that components of the Student-t distribution have a finite fourth moment.

The standardized multivariate Student-t distribution can also be written as

$$\varepsilon_t = \sqrt{\frac{(v-2)\zeta_t}{\xi_t}}u_t,$$

where u_t is uniformly distributed on the unit sphere surface in R^k, ζ_t is a chi-square random variable with k degrees of freedom, ξ_t is a gamma variate with mean ν and variance 2ν, and u_t, ζ_t and ξ_t are mutually independent, see Fiorentini *et al.* [19] for more details. We use this expression in some of the derivations below.

4.1. Estimation

We begin with estimation. Let $Y_1, \ldots, Y_n$ be a realization of the VARMA-GARCH model (1)-(2) with innovations following the standardized multivariate Student-t distribution of (17). Similarly as the case of Gaussian innovations, we denote the parameter vector of the model by $\theta = (\varphi', \delta', \eta)' = (\lambda', \eta)'$, where $\eta = \frac{1}{\nu}$. Ignoring the constant, the log-likelihood function of the data is

$$L_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l_t(\theta), \qquad l_t(\theta) = c(\eta) + d_t(\lambda) + g_t(\lambda, \eta) \tag{18}$$

where

$$\begin{aligned}
c(\eta) &= \ln[\Gamma(\frac{k\eta+1}{2\eta})] - \ln[\Gamma(\frac{1}{2\eta})] - \frac{k}{2}\ln(\frac{1-2\eta}{\eta}),\\
d_t(\lambda) &= -\frac{1}{2}\ln(det(\Sigma_t(\lambda))),\\
g_t(\lambda,\eta) &= -(\frac{k\eta+1}{2\eta})\ln[1+\frac{\eta}{1-2\eta}(\varepsilon_t'\varepsilon_t)].
\end{aligned}$$

The first-order derivatives and the information matrix are calculated as

$$\begin{aligned}
\frac{\partial d_t(\lambda)}{\partial\lambda} &= -\frac{1}{2}vec'(\Sigma_t^{-1})\frac{\partial H_t}{\partial\lambda'}, \qquad \frac{\partial d_t}{\partial\eta} = 0,\\
\frac{\partial g_t(\lambda,\eta)}{\partial\varphi} &= \frac{k\eta+1}{2(1-2\eta+\eta(\varepsilon_t'\varepsilon_t))}[2a_t'\Sigma_t^{-1}(\frac{\partial a_t}{\partial\varphi'}) + vec'(\Sigma_t^{-1}a_ta_t'\Sigma_t^{-1})\frac{\partial H_t}{\partial\varphi'}]\\
\frac{\partial g_t(\lambda,\eta)}{\partial\delta} &= \frac{k\eta+1}{2(1-2\eta+\eta(\varepsilon_t'\varepsilon_t))}[vec'(\Sigma_t^{-1}a_ta_t'\Sigma_t^{-1})\frac{\partial H_t}{\partial\delta'}]\\
\frac{\partial g_t(\lambda,\eta)}{\partial\eta} &= -\frac{k\eta+1}{2\eta(1-2\eta)}\frac{\varepsilon_t'\varepsilon_t}{1-2\eta+\eta(\varepsilon_t'\varepsilon_t)} + \frac{1}{2\eta}\ln[1+\frac{\eta}{1-2\eta}(\varepsilon_t'\varepsilon_t)]\\
\frac{\partial c(\eta)}{\partial\eta} &= \frac{k}{2\eta(1-2\eta)} - \frac{1}{2\eta^2}[\Xi(\frac{k\eta+1}{2\eta}) - \Xi(\frac{1}{2\eta})], \qquad \frac{\partial c(\eta)}{\partial\lambda} = 0,
\end{aligned}$$

where $\Xi(x) = \partial\ln\Gamma(x)/\partial x$ is the Gauss-ψ function, or di-gamma function. Then we can find the information matrix by

$$\begin{aligned}
E\left[\frac{\partial^2 d_t}{\partial\lambda\partial\lambda'}\right] &= -\frac{k+\nu}{2(k+\nu+2)}E[(\frac{\partial H_t}{\partial\lambda'})'(\Sigma_t^{-1}\otimes\Sigma_t^{-1})(\frac{\partial H_t}{\partial\lambda'})],\\
E\left[\frac{\partial^2 g_t}{\partial\varphi\partial\varphi'}\right] &= -\frac{\nu(k+\nu)}{(\nu-2)(k+\nu+2)}E[(\frac{\partial a_t}{\partial\varphi'})'\Sigma_t^{-1}(\frac{\partial a_t}{\partial\varphi'})]\\
&\quad + \frac{1}{2(k+\nu+2)}E\{[(\frac{\partial H_t}{\partial\varphi'})'vec(\Sigma_t^{-1})][vec'(\Sigma_t^{-1})(\frac{\partial H_t}{\partial\varphi'})]\},\\
E\left[\frac{\partial^2 g_t}{\partial\delta\partial\delta'}\right] &= \frac{1}{2(k+\nu+2)}E\{[(\frac{\partial H_t}{\partial\delta'})'vec(\Sigma_t^{-1})][vec'(\Sigma_t^{-1})(\frac{\partial H_t}{\partial\delta'})]\},\\
E\left[\frac{\partial^2 g_t}{\partial\eta\partial\lambda'}\right] &= \frac{(k+2)\nu^2}{(\nu-2)(k+\nu)(k+\nu+2)}E[vec'(\Sigma_t^{-1})(\frac{\partial H_t}{\partial\lambda'})],\\
E\left[\frac{\partial^2}{\partial\eta\partial\eta'}(g_t+c(\eta))\right] &= -\frac{\nu^4}{4}[\Xi'(\frac{\nu}{2}) - \Xi'(\frac{k+\nu}{2})] + \frac{k\nu^4[\nu^2+k(\nu-4)-8]}{2(\nu-2)^2(k+\nu)(k+\nu+2)}.
\end{aligned}$$

The following theorems, corresponding to Theorems 3.1 and 3.2, provide the asymptotic properties of the information matrix and MLE $\hat{\theta}_n$ for a GARCH Student-t model.

Theorem 4.1. Suppose the k-dimensional time series $\{Y_t\}$ and $\{a_t\}$ are generated by (1)–(2) with innovations ε_t following the distribution in (17). Assume Assumptions 1–6 hold, then at $\lambda = \lambda_0$

$$-\frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix} \partial^2 l_t/(\partial\varphi\partial\varphi') & \partial^2 l_t/(\partial\delta\partial\varphi') & \partial^2 l_t/(\partial\eta\partial\varphi') \\ \partial^2 l_t/(\partial\varphi\partial\delta') & \partial^2 l_t/(\partial\delta\partial\delta') & \partial^2 l_t/(\partial\eta\partial\delta') \\ \partial^2 l_t/(\partial\varphi\partial\eta) & \partial^2 l_t/(\partial\delta\partial\eta) & \partial^2 l_t/(\partial\eta\partial\eta) \end{bmatrix} \to_{a.s.} \begin{bmatrix} \Omega_{\varphi\varphi} & O & \Omega_{\varphi\eta} \\ O & \Omega_{\delta\delta} & \Omega_{\delta\eta} \\ \Omega'_{\varphi\eta} & \Omega'_{\delta\eta} & \Omega_{\eta\eta} \end{bmatrix} \equiv \Omega_*$$

as $n \to \infty$ and Ω_* is positive definite, where

$$\begin{aligned}
\Omega_{\varphi\varphi} &= \frac{k+\nu}{2(k+\nu+2)}E[(\frac{\partial H_t}{\partial\varphi'})'(\Sigma_t^{-1}\otimes\Sigma_t^{-1})(\frac{\partial H_t}{\partial\varphi'})] + \frac{\nu(k+\nu)}{(\nu-2)(k+\nu+2)}E[(\frac{\partial a_t}{\partial\varphi'})'\Sigma_t^{-1}(\frac{\partial a_t}{\partial\varphi'})] \\
&- \frac{1}{2(k+\nu+2)}E\{[(\frac{\partial H_t}{\partial\varphi'})'vec(\Sigma_t^{-1})][vec'(\Sigma_t^{-1})(\frac{\partial H_t}{\partial\varphi'})]\}, \\
\Omega_{\varphi\eta} &= -\frac{(k+2)\nu^2}{(\nu-2)(k+\nu)(k+\nu+2)}E[vec'(\Sigma_t^{-1})(\frac{\partial H_t}{\partial\varphi'})], \\
\Omega_{\delta\eta} &= -\frac{(k+2)\nu^2}{(\nu-2)(k+\nu)(k+\nu+2)}E[vec'(\Sigma_t^{-1})(\frac{\partial H_t}{\partial\delta'})], \\
\Omega_{\delta\delta} &= \frac{k+\nu}{2(k+\nu+2)}E[(\frac{\partial H_t}{\partial\delta'})'(\Sigma_t^{-1}\otimes\Sigma_t^{-1})(\frac{\partial H_t}{\partial\delta'})] \\
&- \frac{1}{2(k+\nu+2)}E\{[(\frac{\partial H_t}{\partial\delta'})'vec(\Sigma_t^{-1})][vec'(\Sigma_t^{-1})(\frac{\partial H_t}{\partial\delta'})]\}, \\
\Omega_{\eta\eta} &= \frac{\nu^4}{4}[\Xi'(\frac{\nu}{2}) - \Xi'(\frac{k+\nu}{2})] - \frac{k\nu^4[\nu^2+k(\nu-4)-8]}{2(\nu-2)^2(k+\nu)(k+\nu+2)}.
\end{aligned}$$

Theorem 4.2. Under the assumptions of Theorem 4.1, the following results hold:
(a). There exists a MLE $\hat{\theta}_n$ satisfying the equation $\partial L_n(\theta)/\partial\theta = 0$ and $\hat{\theta}_n \to^p \theta_0$ as $n \to \infty$.
(b). $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_{\mathcal{L}} N(0, \Omega_{0*}^{-1})$ as $n \to \infty$, where

$$\Omega_{0*} = \begin{bmatrix} \Omega_{\varphi\varphi 0} & O & \Omega_{\varphi\eta 0} \\ O & \Omega_{\delta\delta 0} & \Omega_{\delta\eta 0} \\ \Omega'_{\varphi\eta 0} & \Omega'_{\delta\eta 0} & \Omega_{\eta\eta 0} \end{bmatrix},$$

and $\Omega_{\varphi\varphi 0}, \Omega_{\varphi\eta 0}, \Omega_{\delta\delta 0}, \Omega_{\delta\eta 0}$ and $\Omega_{\eta\eta 0}$ are values of $\Omega_{\varphi\varphi}, \Omega_{\varphi\eta}, \Omega_{\delta\delta}, \Omega_{\delta\eta}$ and $\Omega_{\eta\eta}$ at $\theta = \theta_0$. Further, the information matrices $\Omega_{\varphi\varphi 0}, \Omega_{\varphi\eta 0}, \Omega_{\delta\delta 0}, \Omega_{\delta\eta 0}$ and $\Omega_{\eta\eta 0}$ can be estimated consistently by

$$\begin{aligned}
\Omega_{\varphi\varphi 0} &= \frac{k+\nu}{2(m+\nu+2)}\sum_{t=1}^{n}[(\frac{\partial H_t}{\partial\varphi'})'(\Sigma_t^{-1}\otimes\Sigma_t^{-1})(\frac{\partial H_t}{\partial\varphi'})] + \frac{\nu(k+\nu)}{(\nu-2)(k+\nu+2)}\sum_{t=1}^{n}[(\frac{\partial a_t}{\partial\varphi'})'\Sigma_t^{-1}(\frac{\partial a_t}{\partial\varphi'})] \\
&- \frac{1}{2(k+\nu+2)}\sum_{t=1}^{n}\{[(\frac{\partial H_t}{\partial\varphi'})'vec(\Sigma_t^{-1})][vec'(\Sigma_t^{-1})(\frac{\partial H_t}{\partial\varphi'})]\} \\
\Omega_{\varphi\eta 0} &= -\frac{(k+2)\nu^2}{(\nu-2)(k+\nu)(k+\nu+2)}\sum_{t=1}^{n}[vec'(\Sigma_t^{-1})(\frac{\partial H_t}{\partial\varphi'})], \\
\Omega_{\delta\eta 0} &= -\frac{(k+2)\nu^2}{(\nu-2)(k+\nu)(k+\nu+2)}\sum_{t=1}^{n}[vec'(\Sigma_t^{-1})(\frac{\partial H_t}{\partial\delta'})], \\
\Omega_{\delta\delta 0} &= \frac{k+\nu}{2(k+\nu+2)}\sum_{t=1}^{n}[(\frac{\partial H_t}{\partial\delta'})'(\Sigma_t^{-1}\otimes\Sigma_t^{-1})(\frac{\partial H_t}{\partial\delta'})] \\
&- \frac{1}{2(k+\nu+2)}\sum_{t=1}^{n}\{[(\frac{\partial H_t}{\partial\delta'})'vec(\Sigma_t^{-1})][vec'(\Sigma_t^{-1})(\frac{\partial H_t}{\partial\delta'})]\}, \\
\Omega_{\eta\eta 0} &= \frac{\nu^4}{4}[\Xi'(\frac{\nu}{2}) - \Xi'(\frac{k+\nu}{2})] - \frac{k\nu^4[\nu^2+k(\nu-4)-8]}{2(\nu-2)^2(k+\nu)(k+\nu+2)}.
\end{aligned}$$

4.2. Diagnostic Checking

Turn to model checking for a fitted VARMA-GARCH model with Student-t innovations. We derive in this subsection the asymptotic distributions of autocovariance matrices of constructed processes of the standardized residuals and the portmanteau statistics. The derivations are similar to those in the Gaussian case, and ρ and r are defined in the same way as those of subsection 3.2. First, we study the asymptotic distribution of $\hat{\rho}$ and obtain the corresponding portmanteau statistic. It can be shown that Lemma 3.1 continues to hold when the innovations are Student-t. Also, using calculation similar to that of Lemma 3.2, we obtain the following lemma concerning $E[\frac{\partial L_n}{\partial\lambda}vec'(\rho)]$ and $E[\frac{\partial vec(\rho)}{\partial\lambda'}]$.

Lemma 4.1. Under the assumptions of Lemma 3.5, we have that at $\theta = \theta_0$,

$$\begin{aligned}
E\left[\frac{\partial L_n}{\partial\varphi}vec'(\rho_m)\right] &= \frac{-\nu}{\nu-2}E\left[(\frac{\partial a_t}{\partial\varphi'})'(\varepsilon_{t-m}'\otimes\Sigma_t'^{-1/2})\right], \\
E\left[(\frac{\partial l}{\partial\delta'})'vec'(\rho_m)\right] &= E\left[\frac{\partial l}{\partial\eta}vec'(\rho_m)\right] = O, \\
E\left[(\frac{\partial l}{\partial\theta'})'vec'(\rho_m)\right] &= \frac{-\nu}{\nu-2}E\left[(\frac{\partial vec(\rho_m)}{\partial\theta'})'\right].
\end{aligned}$$

Furthermore, define $\boldsymbol{\Upsilon}_*' = (\boldsymbol{\Upsilon}_{1*}', ..., \boldsymbol{\Upsilon}_{M*}')$, $\boldsymbol{\Upsilon}_{k*} = [-E[(\varepsilon_{t-m}\otimes\Sigma_t^{-1/2})(\frac{\partial a_t}{\partial\varphi'})], O']$. Since $E[\partial vec(\rho_m)/\partial\theta'] = -\boldsymbol{\Upsilon}_{m*}$, we can derive the distribution of $\hat{\rho}$.

Theorem 4.3. Suppose the time series Y_t is generated under the assumptions of Theorem 4.1, then $\sqrt{n}vec(\hat{\rho})$ is asymptotically normal with mean 0 and covariance matrix

$$V_{1*} = I_M\otimes I_{k^2} - \frac{\nu+2}{\nu-2}\boldsymbol{\Upsilon}_*\Omega_{0*}^{-1}\boldsymbol{\Upsilon}_*'.$$

The asymptotic distribution of lag-m autocovariance matrix of the product vector of squared standardized residuals is derived in the following lemma.

Lemma 4.2. Under the assumptions of Theorem 4.1, we have

$$\sqrt{n}vec(\widetilde{r}_m) \to_{\mathcal{L}} N(0, W \otimes W),$$

where $W = \frac{\nu-2}{\nu-4}(I_{k^2} + K_{k,k}) - \frac{2}{\nu-4}vec(I_k)vec'(I_k)$.

Proof. The proof of Lemma 4.2 is similar to that of Lemma 3.3. Here we have

$$\begin{aligned} var(vec(\varepsilon_t\varepsilon_t')) &= E[vec(\frac{\nu-2}{\xi_t}\zeta_t u_t u_t' - I_k)vec'(\frac{\nu-2}{\xi_t}\zeta_t u_t u_t' - I_k)] \\ &= (\nu-2)^2 E[(\frac{\zeta_t}{\xi_t})^2]E[vec(u_t u_t')vec'(u_t u_t')] - vec(I_k)vec'(I_k) \\ &= \frac{k(k+2)(\nu-2)}{\nu-4}\frac{1}{k(k+2)}[I_{k^2} + K_{k,k} + vec(I_k)vec'(I_k)] - vec(I_k)vec'(I_k) \\ &= \frac{\nu-2}{\nu-4}(I_{k^2} + K_{k,k}) - \frac{2}{\nu-4}vec(I_k)vec'(I_k). \qquad \Box. \end{aligned}$$

Notice that as $\nu \to \infty$, $var(vec(\varepsilon_t\varepsilon_t')) \to I_{k^2} + K_{k,k}$, which is the covariance matrix for the Gaussian case of Section 3. It is straightforward to find the distribution of $\sqrt{n}vec(r_m)$.

Corollary 4.1. If the time series Y_t is generated by the VARMA-GARCH model (1)–(2) with standardized multivariate Student-t innovation, then under Assumptions 1–6,

$$\sqrt{n}vec(r_m) \to_{\mathcal{L}} N(0, \Delta \otimes \Delta),$$

where $\Delta = (\frac{2}{\nu-4})^2((\nu-2)(D_k^+ D_k^{+'}) - vech(I_k)vech'(I_k))$.

Next, similar to Lemma 3.4, we study the relation between $E[(\frac{\partial L_n}{\partial\theta'})'vec'(r_m)]$ and $E[\frac{\partial vec(r_m)}{\partial\theta'}]$.

Lemma 4.3. Under the assumptions of Lemma 3.5, we have that at $\theta = \theta_0$,

$$\begin{aligned} E\left[(\frac{\partial L_n}{\partial\lambda'})'vec'(r_m)\right] &= \frac{1}{n}\sum_{t=m+1}^{n} E\left\{\left(\frac{\partial vec(\Sigma_t^{1/2})}{\partial\lambda'}\right)'(\Sigma_t^{-1/2}\otimes I_k + I_k\otimes\Sigma_t^{-1/2})[vech'(c_{t-m})\otimes D_k^{+'}]\right\} \\ &+ \frac{1}{n}\frac{1}{\nu-2}\sum_{t=m+1}^{n} E\{\left(\frac{\partial vec(\Sigma_t^{1/2})}{\partial\lambda'}\right)'(\Sigma_t^{-1/2}\otimes I_k + I_k\otimes\Sigma_t^{-1/2}) \\ &\cdot\ [vech'(c_{t-m})\otimes(vec(I_k)vech'(I_k))]\} \\ &\equiv \widetilde{\mathbf{X}}_{m*}' + \widetilde{\mathbf{Z}}_{m*}'. \end{aligned}$$

Next, define $\mathbf{X}_{m*} = [\widetilde{\mathbf{X}}_{m*}, O]$, $\mathbf{X}_*' = (\mathbf{X}_{1*}', \ldots, \mathbf{X}_{M*}')$, $\mathbf{Z}_{m*} = [\widetilde{\mathbf{Z}}_{m*}, O]$, $\mathbf{Z}_*' = (\mathbf{Z}_{1*}', \ldots, \mathbf{Z}_{M*}')$.

Proof. From the calculation of $\frac{\partial L_n}{\partial \lambda}$, we have

$$E[\frac{\partial L_n}{\partial \lambda} vec'(r_m)] = \frac{1}{n} \sum_{t=m+1}^{n} E \frac{k\eta + 1}{2(1 - 2\eta + \eta(\varepsilon_t' \varepsilon_t))} (\frac{\partial H_t}{\partial \lambda'})' (\Sigma_t^{-1/2} \otimes \Sigma_t^{-1/2})$$
$$\cdot vec(c_t)(vech'(c_{t-m}) \otimes vech'(c_t))]$$
$$= \frac{1}{n} \sum_{t=m+1}^{n} E\left[\left(\frac{\partial vec(\Sigma_t^{1/2})}{\partial \lambda'} \right)' (\Sigma_t^{-1/2} \otimes I_k + I_k \otimes \Sigma_t^{-1/2}) \right]$$
$$\cdot [vech'(c_{t-m}) \otimes \frac{k\eta + 1}{2(1 - 2\eta + \eta(\varepsilon_t' \varepsilon_t))} (vec(c_t) vech'(c_t))]$$
$$= \frac{1}{n} \sum_{t=m+1}^{n} E\{ \left(\frac{\partial vec(\Sigma_t^{1/2})}{\partial \lambda'} \right)' (\Sigma_t^{-1/2} \otimes I_k + I_k \otimes \Sigma_t^{-1/2})$$
$$\cdot [vech'(c_{t-m}) \otimes (D_k^{+'} + \frac{1}{\nu - 2} vec(I_k) vech'(I_k))]\} = \widetilde{\mathbf{X}}_{m*}' + \widetilde{\mathbf{Z}}_{m*}'. \quad \square$$

We can also find that $E[\partial vec(r_m)/\partial \theta] = -\mathbf{X}_{m*}$ and obtain the following portmanteau statistic for diagnostic checking.

Theorem 4.4. Suppose the time series Y_t is generated under the assumptions of Theorem 4.1, then $\sqrt{n} vec(\hat{r})$ is asymptotically normal with mean 0 and covariance matrix

$$V_{2*} = (\frac{2}{\nu - 4})^2 \cdot I_M \otimes (\Psi \otimes \Psi) - (\mathbf{X}_* \Omega_{0*}^{-1} \mathbf{X}_*' - \mathbf{Z}_* \Omega_{0*}^{-1} \mathbf{Z}_*'),$$

where $\Psi = (\nu - 2)(D_k^+ D_k^{+'}) - vech(I_k) vech'(I_k)$.

By Theorems 4.3 and 4.4, we know that

$$Q_1(M) = n \cdot vec'(\hat{\rho}) \hat{V}_{1*}^{-1} vec(\hat{\rho}) \sim \chi^2(Mk^2), \tag{19}$$

and

$$Q_2(M) = n \cdot vec'(\hat{r}) \hat{V}_{2*}^{-1} vec(\hat{r}) \sim \chi^2(M[\frac{k(k+1)}{2}]^2) \tag{20}$$

where $\hat{V}_{1*} = I_M \otimes I_{k^2} - \frac{\nu+2}{\nu-2} \hat{\mathbf{\Upsilon}}_* \hat{\Omega}_{0*}^{-1} \hat{\mathbf{\Upsilon}}_*'$, $\hat{V}_{2*} = (\frac{2}{\hat{\nu}-4})^2 \cdot I_M \otimes (\Psi \otimes \Psi) - (\hat{\mathbf{X}}_* \hat{\Omega}_{0*}^{-1} \hat{\mathbf{X}}_*' - \hat{\mathbf{Z}}_* \hat{\Omega}_{0*}^{-1} \hat{\mathbf{Z}}_*')$,

$$\hat{\Omega}_{0*} = \begin{bmatrix} \hat{\Omega}_{\varphi\varphi 0} & O & \hat{\Omega}_{\varphi\eta 0} \\ O & \hat{\Omega}_{\delta\delta 0} & \hat{\Omega}_{\delta\eta 0} \\ \hat{\Omega}_{\varphi\eta 0}' & \hat{\Omega}_{\delta\eta 0}' & \hat{\Omega}_{\eta\eta 0} \end{bmatrix},$$

where $\hat{\mathbf{\Upsilon}}_* = (\hat{\mathbf{\Upsilon}}_{1*}', \ldots, \hat{\mathbf{\Upsilon}}_{M*}')'$, $\hat{\mathbf{\Upsilon}}_{m*} = \frac{1}{n} \sum_{t=m+1}^{n} (\varepsilon_{t-m} \otimes \Sigma_t^{-1/2} (\partial a_t / \partial \lambda'), O)$ is estimated at $\lambda = \hat{\lambda}_n$. $\hat{\mathbf{X}}_* = (\hat{\mathbf{X}}_{1*}', \ldots, \hat{\mathbf{X}}_{M*}')'$, $\hat{\mathbf{X}}_{m*} = \frac{1}{n} \sum_{t=m+1}^{n} (vech(c_{t-m}) \otimes D_k^{+'}) [(I_k \otimes \Sigma_t^{-1/2} a_t a_t' + \Sigma_t^{-1/2} a_t a_t' \otimes I_k) \frac{\partial vec(\Sigma_t^{-1/2})}{\partial \lambda'}, O]'$, $\hat{\mathbf{Z}}_* = (\hat{\mathbf{Z}}_{1*}', \ldots, \hat{\mathbf{Z}}_{M*}')'$, $\hat{\mathbf{Z}}_{m*} = \frac{1}{n} \sum_{t=m+1}^{n} \frac{1}{\hat{\nu}-2} [vech(c_{t-m}) \otimes (vec(I_k) vech'(I_k))] [(I_k \otimes \Sigma_t^{-1/2} a_t a_t' + \Sigma_t^{-1/2} a_t a_t' \otimes I_k) \frac{\partial vec(\Sigma_t^{-1/2})}{\partial \lambda'}, O]'$ are also estimated at $\lambda = \hat{\lambda}_n$.

5. Residual-based Bootstrap Approximation for Model Checking Statistics

As discussed in previous sections, the model checking statistics $Q_1(M)$ and $Q_2(M)$ depend on the estimation of covariance matrices V_1 and V_2, which are obtained by the first and second order derivatives of log-likelihood function, while it is complicated to implement the procedure in real application. Alternatively, in this section we approximate the test statistics under both the null and alternative hypothesis by bootstrap. To be specific, we consider constructing pseudo time series based on i.i.d. random draws from the fitted standardized residuals with a discrete distribution. For simplicity, our bootstrap procedure and asymptotic properties are based on the case of Gaussian innovations in this section, and the approach could be easily extended for the Student-t error GARCH model.

5.1. Residual-based Bootstrap Procedure

By resampling the standardized residuals of the fitted model, we use the residual-based bootstrap to approximate the (asymptotic) distributions of sample standardized residual autocorrelation matrix and standardized residual cross-product autocovariance matrix, *i.e.*, $vec(\hat{\rho})$ and $vec(\hat{r})$, and thus find the (empirical) p-values of the model checking statistics $Q_1(M)$ and $Q_2(M)$. Particularly, the innovations ε_t are mimicked by i.i.d. random draws with replacement from the estimated errors $\hat{\varepsilon}_t$. Define $\Sigma_{\hat{\varepsilon}}$ as the sample covariance matrix of $\hat{\varepsilon}_t$, *i.e.*, $\Sigma_{\hat{\varepsilon}} \equiv n^{-1}\sum_{t=1}^{n}(\hat{\varepsilon}_t - \widetilde{\varepsilon})(\hat{\varepsilon}_t - \widetilde{\varepsilon})'$, where $\widetilde{\varepsilon}$ is the sample mean of $\hat{\varepsilon}_t$, as defined in subsection 3.2. The fitted standardized residuals are normalized by $\breve{\varepsilon}_t = \Sigma_{\hat{\varepsilon}}^{-1/2}(\hat{\varepsilon}_t - \widetilde{\varepsilon})$. To generate each bootstrap pseudo-series, the residual sample ε_t^* is i.i.d. random draw from the normalized version $\{\breve{\varepsilon}_t\}$. To construct the bootstrap sample $\{Y_t^*\}_{t=1}^{n}$, we replace the unknown parameter λ with its estimator $\hat{\lambda}_n$. The procedure allows the bootstrap residual data $\{\varepsilon_t^*\}_{t=1-d}^{n}$ to satisfy the model hypothesized under the null, irrespective of whether $\{\varepsilon_t\}_{t=1-d}^{n}$ follow the model under the null hypothesis. Hence the bootstrap statistics possess the same asymptotic distribution under the maintain hypothesis, *i.e.*, $H_0 \cup H_a$. As the following steps, we introduce the residual-based bootstrap algorithm to approximate the test statistics.

Step 1. Define $d = \max(p, q, r, s)$, and let $(\varepsilon_{1-d}^*, ..., \varepsilon_0^*)$ be some starting values. For the checked VARMA(p, q)-GARCH(r, s) model, generate Y_t^* by

$$Y_t^* = \sum_{i=1}^{p}\hat{\Phi}_i Y_{t-i}^* + a_t^* + \sum_{i=1}^{q}\hat{\Theta}_i a_{t-i}^*, \qquad a_t^* = \Sigma_t^{*1/2}\varepsilon_t^* \tag{21}$$

$$\Sigma_t^* = \hat{A}_0 + \sum_{i=1}^{r}\hat{A}_i a_{t-i}^* a_{t-i}^{*'}\hat{A}_i' + \sum_{i=1}^{s}\hat{B}_i\Sigma_{t-i}^*\hat{B}_i', \quad t = 0, 1, \ldots, n \tag{22}$$

where ε_t^* are i.i.d. random draw from $\{\breve{\varepsilon}_t\}$.

Step 2. Based on the bootstrap pseudo-series $\{Y_t^*\}_{t=1}^n$, let $\hat{\lambda}_n^*$ be the MLE of the parameters of the model under the null hypothesis, and denote by $\hat{\varepsilon}_t^*$ the corresponding estimated standardized residuals. The bootstrap analogues of $\hat{\rho}_m$ and $\hat{r}_m$ are obtained by

$$\hat{\rho}_m^* = n^{-1} \sum_{t=m+1}^{n} (\hat{\varepsilon}_t^* - \widetilde{\varepsilon}^*)(\hat{\varepsilon}_{t-m}^* - \widetilde{\varepsilon}^*)'$$
$$\hat{r}_m^* = n^{-1} \sum_{t=m+1}^{n} vech(\hat{c}_t^* - \widetilde{c}^*)vech(\hat{c}_{t-m}^* - \widetilde{c}^*)', \qquad m = 1, ..., M$$

where $\widetilde{\varepsilon}^* = n^{-1}\sum_{t=1}^n \hat{\varepsilon}_t^*$, $\hat{c}_t^* = \hat{\varepsilon}_t^*\hat{\varepsilon}_t^{*\prime}$, $\widetilde{c}^* = n^{-1}\sum_{t=1}^n \hat{c}_t^*$, and M is the pre-specified largest lag tested for zero autocorrelations in standardized residual and its cross-product vector. Define $vec(\hat{\rho}^*) = (vec'(\hat{\rho}_1^*), ..., vec'(\hat{\rho}_M^*))'$ and $vec(\hat{r}^*) = (vec'(\hat{r}_1^*), ..., vec'(\hat{r}_M^*))'$.

Step 3. Repeat **Step 1** and **Step 2** for B times. Denote the $vec(\hat{\rho}^*)$ and $vec(\hat{r}^*)$ obtained in the b^{th} iteration by $vec(\hat{\rho}^{*(b)})$ and $vec(\hat{r}^{*(b)})$, where $b = 1, 2, \ldots, B$.

Step 4. Approximate V_1 by $\hat{V}_1^*$, the covariance matrix of $\{\sqrt{n}vec(\hat{\rho}^{*(b)})\}_{b=1}^B$. Calculate B bootstrap model checking statistics $Q_1^{*(b)}(M)$ by $\{Q_1^{*(b)}(M) = n \cdot vec(\hat{\rho}^{*(b)})'\hat{V}_1^{*-1}vec(\hat{\rho}^{*(b)})\}_{b=1}^B$. Similar procedures are followed to obtain the covariance matrix of $\{\sqrt{n}vec(\hat{r}^{*(b)})\}_{b=1}^B$, denoted by $\hat{V}_2^*$, and $\{Q_2^{*(b)}(M)\}_{b=1}^B$.

Step 5. Let $Q_1(M) = n \cdot vec(\hat{\rho})'\hat{V}_1^{*-1}vec(\hat{\rho})$, $Q_2(M) = n \cdot vec(\hat{r})'\hat{V}_2^{*-1}vec(\hat{r})$. The model is not adequate if

$$Q_1(M) > q_\alpha^{1*}, \quad \text{or} \quad Q_2(M) > q_\alpha^{2*}$$

where q_α^{1*} and q_α^{2*} are the $(1 - \alpha)$th-quantiles of the distributions of $\{Q_1^{*(b)}(M)\}_{b=1}^B$ and $\{Q_2^{*(b)}(M)\}_{b=1}^B$, respectively.

5.2. Properties and Validity of the Residual-based Bootstrap Process

In this subsection, we introduce some basic properties of the residual-based bootstrap procedure and justify its validity. As a first step, we consider transforming the VARMA(p, q)-GARCH(r, s) BEKK model to a truncated version of VAR(∞)-ARCH(∞) model. Define $\eta = \max(r, s)$.

For $i > r$ or $j > s$, let $A_i = B_j = O_{k\times k}$, and denote $R_t = vec(a_t a_t')$, $\widetilde{R}_t = [R_t', O_{1\times(2\eta-1)k^2}]'$, $\widetilde{A}_0 = [O_{1\times k^2}, vec(A_0)', O_{1\times(2(\eta-1)k^2)}]'$, $J_{i,\eta} = [O_{k^2\times(i-1)k^2}, I_{k^2}, O_{k^2\times(2\eta-i)k^2}]$, $G_t = [R_t', H_t', R_{t-1}', H_{t-1}', ..., R_{t-\eta+1}', H_{t-\eta+1}']'$, where $H_t = vec(\Sigma_t)$.

$$\widetilde{A} = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ A_1\otimes A_1 & B_1\otimes B_1 & \dots & A_{\eta-1}\otimes A_{\eta-1} & B_{\eta-1}\otimes B_{\eta-1} & A_\eta\otimes A_\eta & B_\eta\otimes B_\eta \\ 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & 0 \end{bmatrix} \tag{23}$$

Consider the recursive relationship $G_t = \widetilde{A}_0 + \widetilde{A}G_{t-1} + \widetilde{R}_t = \sum_{j=0}^{\infty}\widetilde{A}^j\widetilde{A}_0 + \sum_{j=0}^{\infty}\widetilde{A}^j\widetilde{R}_{t-j}$, and notice that the processes $\{\varepsilon_t\}$ and $\{\Sigma_t\}$ are unobserved for $t \le 0$ in practice, we have the following lemma concerning the transformation of VARMA(p,q)-GARCH(r,s) BEKK model.

Lemma 5.1. Given $\mathcal{F}_0$ and under Assumptions 3–7, model (1)-(3) has the following form of transformation

$$a_t = Y_t + \sum_{i=1}^{t-1}\Pi_i Y_{t-i}, \quad a_t = \Sigma_t^{1/2}\varepsilon_t, \tag{24}$$

$$\begin{aligned} H_t &= J_{2,\eta}\big[\sum_{j=0}^{\infty}\widetilde{A}^j\widetilde{A}_0 + \sum_{j=0}^{\infty}\widetilde{A}^j\widetilde{R}_{t-j}\big] = J_{2,\eta}(\sum_{j=0}^{\infty}\widetilde{A}^j\widetilde{A}_0) + \sum_{j=0}^{t-1}[J_{2,\eta}\widetilde{A}^j J_{1,\eta}' R_{t-j}] \\ &= J_{2,\eta}(\sum_{j=0}^{\infty}\widetilde{A}^j\widetilde{A}_0) + \sum_{j=1}^{t-1}[J_{2,\eta}\widetilde{A}^j J_{1,\eta}' R_{t-j}] \equiv C_0 + \sum_{j=1}^{t-1}C_j R_{t-j}. \end{aligned} \tag{25}$$

Based on the above transformation, we provide the bootstrap counterparts and the corresponding log-likelihood function.

$$a_t^* = Y_t^* + \sum_{i=1}^{t-1}\Pi_i Y_{t-i}^*, \quad a_t^* = \Sigma_t^{*1/2}\varepsilon_t^*, \tag{26}$$

$$H_t^* = C_0 + \sum_{j=1}^{t-1}C_j R_{t-j}^*, \tag{27}$$

$$L_n^*(\lambda) = \frac{1}{n}\sum_{t=1}^{n} l_t^*(\lambda), \quad l_t^*(\lambda) = -\frac{1}{2}\ln det(\Sigma_t^*) - \frac{1}{2}a_t^{*'}\Sigma_t^{*-1}a_t^*. \tag{28}$$

Lemma 5.2. Under Assumptions 1–8, for all $\lambda \in \Lambda$,

$$E^*|\frac{1}{n}\sum_{i=1}^{n}[\ln det(\Sigma_t^*) - \ln det(\Sigma_t)]| = o_p(1), \tag{29}$$

$$E^*|\frac{1}{n}\sum_{i=1}^{n}[a_t^{*'}\Sigma_t^{*-1}a_t^* - a_t'\Sigma_t^{-1}a_t]| = o_p(1), \tag{30}$$

where $E^*(X)$ denotes the bootstrap expectation of the random variable X, *i.e.*, $E^*(X) = E(X|\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_n)$ and $P^*(X \leq x) = P(X \leq x|\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_n)$.

Lemma 5.3. Under Assumptions 1–8,

$$E^*\Big(\sup_{\lambda_1,\lambda_2\in\Lambda} \frac{1}{|\lambda_1-\lambda_2|} |\ln det\Sigma_t^*(\lambda_1) - \ln det\Sigma_t^*(\lambda_2)|\Big) = K_t \tag{31}$$

$$E^*\Big(\sup_{\lambda_1,\lambda_2\in\Lambda} \frac{1}{|\lambda_1-\lambda_2|} |a_t^{*'}(\lambda_1)\Sigma_t^{*-1}(\lambda_1)a_t^*(\lambda_1) - a_t(\lambda_2)'\Sigma_t^{-1}(\lambda_2)a_t(\lambda_2)|\Big) = K_t \tag{32}$$

where K_t is a sequence of $O_p(1)$ random variables and $E^*(\cdot)$ is defined similarly as in **Lemma 5.2**.

Based on Lemmas 5.2 and 5.3, we have the following propositions, which establishes the asymptotic properties of the MLE $\hat{\lambda}_n^*$ of the bootstrap procedure (21) – (22).

Proposition 5.1. Under Assumptions 1–8 and under $H_0 \cup H_a$, $\hat{\lambda}_n^* - \hat{\lambda}_n = o_{p^*}(1)$, where $o_{p^*}(1)$ means that for all $\epsilon > 0$, $P\{\| \hat{\lambda}_n^* - \hat{\lambda}_n \| > \epsilon|\mathcal{D}_n\} \to_p 0$ with $\mathcal{D}_n = \{\check{\varepsilon}_t\}_{1-d}^n$.

Proof. To show the proposition, we need first show that

$$E^* \sup_{\lambda\in\Lambda} |L_n^*(\lambda) - L_n(\lambda)| = o_p(1). \tag{33}$$

It follows from Lemma 5.2 that $E^*|L_n^*(\lambda) - L_n(\lambda)| = o_p(1)$. Since Λ is a compact set, we have the equicontinuity condition for $L_n^*(\lambda) - L_n(\lambda)$ by Lemma 5.3, which implies

$$\begin{aligned} E^* \sup_{\lambda_1,\lambda_2\in\Lambda} \Big(\frac{1}{|\lambda_1-\lambda_2|}\Big[&|\ln det\Sigma_t^*(\lambda_1) - \ln det\Sigma_t^*(\lambda_2)| \\ &+|a_t^{*'}(\lambda_1)\Sigma_t^{*-1}(\lambda_1)a_t^*(\lambda_1) - a_t(\lambda_2)'\Sigma_t^{-1}(\lambda_2)a_t(\lambda_2)|\Big]\Big) = K_t \end{aligned}$$

Thus, (33) holds. Therefore by standard arguments we conclude that $\hat{\lambda}_n^* - \hat{\lambda}_n = o_{p^*}(1)$. $\square$.

Proposition 5.2. Under Assumptions 1–8 and under $H_0 \cup H_a$, as $n \to \infty$, $\sqrt{n}(\hat{\lambda}_n^* - \hat{\lambda}_n) \to_{\mathcal{L}^*} N(0, \Omega_0^{-1})$ in probability, that is for each point $x \in \mathbb{R}^{dim(\lambda)}$, $P\{\sqrt{n}(\hat{\lambda}_n^* - \hat{\lambda}_n) < x|\mathcal{D}_n\} \to_p \Phi_{dim(\lambda)}(x, \Omega_0^{-1})$, where $\Phi_{dim(\lambda)}(\cdot, \Omega_0^{-1})$ denotes the CDF of the $dim(\lambda)$-variate Gaussian distribution $N(0, \Omega_0^{-1})$.

Proof. To show the proposition, we shall show that $\Omega^{*-1}(\lambda) - \Omega^{-1}(\lambda) = o_{p^*}(1)$, where $\Omega^*(\lambda) = diag\{\Omega_\phi^*(\lambda), \Omega_\delta^*(\lambda)\}$,

$$\begin{aligned} \Omega_\phi^*(\lambda) &= E^*[\frac{1}{2}(\frac{\partial H_t^*}{\partial\varphi'})'(\Sigma_t^{*-1} \otimes \Sigma_t^{*-1})(\frac{\partial H_t^*}{\partial\varphi'}) + (\frac{\partial a_t}{\partial\varphi'})'\Sigma_t^{*-1}(\frac{\partial a_t}{\partial\varphi'})], \\ \Omega_\delta^*(\lambda) &= E^*[\frac{1}{2}(\frac{\partial H_t^*}{\partial\delta'})'(\Sigma_t^{*-1} \otimes \Sigma_t^{*-1})(\frac{\partial H_t^*}{\partial\delta'})]. \end{aligned}$$

Proceeding as the proof of Proposition 5.1, we obtained that $\Omega^*(\lambda) - \Omega(\lambda) = o_{p^*}(1)$ uniformly in a neighborhood of $\hat{\lambda}_n$, say $\mathcal{N}(\hat{\lambda}_n)$, which implies that $\Omega^*(\lambda)$ is a positive definite matrix for all $\lambda \in \mathcal{N}(\lambda)$.

Hence by Slutsky's theorem, we have $\Omega^{*-1}(\lambda) - \Omega^{-1}(\lambda) = o_{p^*}(1)$, which completes the proof. $\square$.

For the fitted standardized residual series $\{\hat{\varepsilon}_t^*\}$ obtained by the bootstrap time series, it follows by a Taylor's expansion that

$$vec(\hat{\rho}^*) = vec(\rho) + \frac{\partial vec(\rho)}{\partial \lambda'}(\hat{\lambda}^* - \lambda) + o_{p^*}(n^{-1/2})$$
$$vec(\hat{r}^*) = vec(r) + \frac{\partial vec(r)}{\partial \lambda'}(\hat{\lambda}^* - \lambda) + o_{p^*}(n^{-1/2})$$

Then we have the following results for the fitted residuals of the bootstrap processes.

Proposition 5.3. Under Assumptions 1–8 and under $H_0 \cup H_a$, as $n \to \infty$,

$$\sqrt{n}(vec(\hat{\rho}^*)) \to_{\mathcal{L}^*} N(0, V_1); \qquad \sqrt{n}(vec(\hat{r}^*)) \to_{\mathcal{L}^*} N(0, V_2)$$

in probability.

Proof. Denote the covariance matrix (in the bootstrap sense) of $\sqrt{n}vec(\hat{\rho}^*)$ and $\sqrt{n}vec(\hat{r}^*)$ as V_1^* and V_2^*, respectively. Then we have

$$V_1^* = n^{-1}E^*\Big[\frac{\partial vec(\rho^*)}{\partial \lambda}\Omega_0^{*-1}\frac{\partial L_n^*}{\partial \lambda}(\frac{\partial L_n^*}{\partial \lambda})'\Omega_0^{*-1}(\frac{\partial vec(\rho^*)}{\partial \lambda})'\Big]$$
$$V_2^* = n^{-1}E^*\Big[\frac{\partial vec(r^*)}{\partial \lambda}\Omega_0^{*-1}\frac{\partial L_n^*}{\partial \lambda}(\frac{\partial L_n^*}{\partial \lambda})'\Omega_0^{*-1}(\frac{\partial vec(r^*)}{\partial \lambda})'\Big]$$

Using similar method in Proposition 5.1 and the result of Proposition 5.2, we could show that $V_1^* - V_1 = o_{p^*}(1)$ and $V_2^* - V_2 = o_{p^*}(1)$ uniformly in a neighborhood of $\hat{\lambda}_n$. Therefore, the results of the proposition hold. $\square$.

Based on Proposition 5.3, we have the following theorem concerning the asymptotic distributions of the bootstrap model checking statistics.

Theorem 5.1. Under Assumptions 1–8 and under $H_0 \cup H_a$, as $n \to \infty$,

$$n(vec(\hat{\rho}^*))'V_1^{*-1}(vec(\hat{\rho}^*)) \to_{\mathcal{L}^*} \chi^2(Mk^2);$$
$$n(vec(\hat{r}^*))'V_2^{*-1}(vec(\hat{r}^*)) \to_{\mathcal{L}^*} \chi^2(M[\frac{k(k+1)}{2}]^2)$$

in probability.

The results of Theorem 5.1 allow us to implement the bootstrap procedure for model checking. In particular, define $q^1_{n,1-\alpha}$, $q^2_{n,1-\alpha}$ and $q^{*1}_{n,1-\alpha}$, $q^{*2}_{n,1-\alpha}$ to satisfy

$$P\{n(vec(\hat{\rho}))'V_1^{-1}(vec(\hat{\rho})) > q^1_{n,1-\alpha}\} = \alpha; \qquad P\{n(vec(\hat{r}))'V_2^{-1}(vec(\hat{r})) > q^2_{n,1-\alpha}\} = \alpha,$$
$$P^*\{n(vec(\hat{\rho}^*))'V_1^{*-1}(vec(\hat{\rho}^*)) > q^{*1}_{n,1-\alpha}\} = \alpha; \qquad P^*\{n(vec(\hat{r}^*))'V_2^{*-1}(vec(\hat{r}^*)) > q^{*2}_{n,1-\alpha}\} = \alpha$$

Then it follows that $q^i_{n,1-\alpha} \to \chi^2_{1-\alpha}(Mk^2)$ as $n \to \infty$, whereas $q^{*i}_{n,1-\alpha} \to^P \chi^2_{1-\alpha}(Mk^2)$, $i = 1, 2$. In real application, we run the bootstrap algorithm for a large number of B times to estimate V_1^* and V_2^*. Particularly,

$$\hat{V}_1^* = \frac{1}{B}\sum_{b=1}^{B}[vec(\hat{\rho}^{*(b)} - \bar{\rho}^*)][vec'(\hat{\rho}^{*(b)} - \bar{\rho}^*)],$$

$$\hat{V}_2^* = \frac{1}{B}\sum_{b=1}^{B}[vec(\hat{r}^{*(b)} - \bar{r}^*)][vec'(\hat{r}^{*(b)} - \bar{r}^*)]$$

where $\bar{\rho}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\rho}^{*(b)}$, $\bar{r}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{r}^{*(b)}$. Furthermore, define $Q_1^{*(b)}(M) = n(vec(\hat{\rho}^{*(b)}))'\hat{V}_1^{*-1} \cdot(vec(\hat{\rho}^{*(b)}))$, $Q_2^{*(b)}(M) = n(vec(\hat{r}^{*(b)}))'\hat{V}_2^{*-1}(vec(\hat{r}^{*(b)}))$, then $q^{i*}_{1-\alpha}$ are approximated by the values $q^{*i,B}_{1-\alpha}$ $(i = 1, 2)$, which satisfy

$$\frac{1}{B}\sum_{b=1}^{B} 1_{\left(Q_1^*(b)(M) > q^{*1,B}_{1-\alpha}\right)} = \alpha, \quad \frac{1}{B}\sum_{b=1}^{B} 1_{\left(Q_2^*(b)(M) > q^{*2,B}_{1-\alpha}\right)} = \alpha,$$

where $1_{(\cdot)}$ denotes the indicator function.

6. Simulation Study

In this section, we conduct some simulation studies with four objectives. First, we demonstrate the effect of conditional heteroscedasticity on the distributions of autocorrelations of standardized residuals and their cross-product series. Second, we study the empirical size and power of the proposed portmanteau statistics. Thirdly, we compare the performance of the proposed statistics with those of Ling and Li [12] to illustrate the contributions of using the cross-product vector of standardized residuals in model checking. Finally, we examine the effect of our bootstrap approximation method by constructing the distribution of the statistics via bootstrap under both the null and alternative hypotheses. In our simulations, we use bivariate time series as examples.

To study the effect of GARCH shocks on the distributions of autocorrelations of standardized residuals and their squared series, we employ a simple bivariate time series $Y_t = a_t$, where a_t follows either a pure 2-dimensional Gaussian white noise or a BEKK(1,1) model with Gaussian innovations and parameters given below:

$$A_0 = I, \quad A_1 = \begin{bmatrix} 0.05 & -0.1 \\ 0.1 & 0.05 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 0.9 & -0.3 \\ 0.3 & 0.9 \end{bmatrix}. \tag{34}$$

For the pure white noise series, we estimate the sample covariance matrix of Y_t to obtain the standardized residuals. For the GARCH series, we fit a BEKK(1,1) model to obtain the standardized residuals. Figure 1 shows the empirical densities of the (1,1) element of the lag-1 autocorrelation matrix of the standardized residuals and their cross-product series. The plots on the top panel are for the case of pure Gaussian noise whereas those in the bottom panel are for the case of GARCH shocks. Also, the plots in the left panel are for the standardized residuals and those in the right panel are for the cross-product series of the standardized residuals. The results are based on 1000 time series and the sample sizes used are 500, 1000, and 3000. In each figure, the black, blue, and red curves are for sample size 500, 1000, and 3000, respectively. These plots show clearly that (a) the conditional heteroscedasticity has substantial impact on the sample distribution of residual serial correlations, (b) the sample size also affects significantly the residual serial correlations, and (c) the conditional heteroscedasticity seems to affect the autocorrelations of the standardized residuals more than their squared series.

To investigate the effect of excess kurtosis on the residual serial correlations, we employ a GARCH(1,1) model. The GARCH shocks follow a BEKK(1,1) model with parameters in Equation (34), but the innovations ε_t are multivariate Student-t with degrees of freedom 7, 15, 50, respectively. Again, we consider the empirical densities of the (1,1) element of the lag-1 autocorrelation matrix of the standardized residuals and their cross-product series. Figure 2 shows the results based on 1000 time series, each has 3000 observations. From the plots, we observe that (a) the tail-thickness has marked impact on the sample distributions of the serial correlations of standardized residuals and their cross-product series and (b), as expected, when the degrees of freedom ν increases the densities approach those in Figure 1 for the Gaussian innovations with 3000 observations.

Next, we study the empirical size and power of the proposed test statistics $Q_1(M)$ and $Q_2(M)$. Three VAR(p)-GARCH(r, s) models are used in the simulation. Denote the parameters of a model by $vec(\Phi_1, \cdots, \Phi_p)$-$vec(A_0, A_1, \cdots, A_r, B_1, \cdots, B_s)$. The first model employed is a VAR(1)-GARCH(1,1) model with parameters (.2, .3, $-$.6, 1.1)-(1, 0, 0, 1, .5, 0, 0, .5, .5, 0, 0, .5). The second one is a VAR(1)-GARCH(3,1) model with parameters (.2, .3, $-$.6, 1.1, $-$.3, 0, 0, $-$.3)-(1, 0, 0, 1, .2, 0, 0, .2, .2, .2, 0, .2, .2, 0, .1, .2, .85, 0, 0, .85). The third model is VAR(2)-GARCH(3,1) with parameters (.2, .3, $-$.6, 1.1, $-$.3, 0, 0, $-$.3)-(1, 0, 0, 1, .5, 0, 0, .5, .5, .2, 0, .5, .4, 0, .1, .4, .2, 0, 0, .2). For each VAR-GARCH model, we consider Gaussian and Student-t innovations. The degrees of freedom for the multivariate Student-t innovations are 6, 7, and 7, respectively, for Model 1, 2, and 3. We choose these models to show various degrees of conditional heteroscedasticity and tail thickness.

The sample sizes used are 500 and 1000. For each (model, sample size, innovation) combination, we generate 1000 realizations. For each realization $\{Y_t | t = 1, \ldots, n\}$, we fit a VAR(1)-GARCH(1,1) to obtain the standardized residuals, assuming that the distributional type of the innovations is known. Using the standardized residuals and their cross-product series, we compute the proposed portmanteau test statistics $Q_1(M)$ and $Q_2(M)$.

For Model 1, the fitted model is properly specified so that we can obtain the empirical distributions of the proposed test statistics. We then use the asymptotic 5% critical values to tabulate the empirical sizes of the two test statistics. For Models 2 and 3, the fitted model is mis-specified and we use the results to study the power of the proposed test statistics. Note that asymptotic 5% critical values are used in the power study. The simulation results are given in Table 1. For simplicity, we only report the results for $M = 1$.

Furthermore, to demonstrate the importance of using the cross-product vector of the standardized residuals, not simply their squared series, we consider another data generating model. Specifically, Model 4 is

$$\begin{aligned}
a_t &= \Sigma_t^{1/2}\varepsilon_t, \quad \varepsilon_t \sim N(0, I_2) \quad \text{or} \quad \varepsilon_t \sim t(0, I_2, 7), \\
\sigma_{11,t} &= 0.01 + 0.04a_{1,t-1}^2 + 0.04\sigma_{11,t-1}, \\
\sigma_{22,t} &= 0.01 + 0.04a_{2,t-1}^2 + 0.04\sigma_{22,t-1}, \\
\sigma_{12,t} &= 0.04a_{1,t-1}a_{2,t-1} + 0.04\sigma_{12,t-1} + 0.49a_{1,t-2}a_{2,t-2},
\end{aligned} \tag{35}$$

where $a_{i,t}$ is the i-th element of a_t and $\sigma_{ij,t}$ denotes the (i, j) element of the covariance matrix Σ_t. In this particular case, the standardized residuals follow a GARCH(1,1) model whereas the cross-product process $\{a_{1,t}a_{2,t}\}$ has a lag-2 autocorrelation. Thus, the goodness of fit test of Ling and Li [12] is unlikely to reject a fitted GARCH(1,1) model. We generate data from Model 4, fit a GARCH(1,1) model to the data, and compare the power between the proposed $Q_2(M)$ portmanteau statistic and the corresponding test of Ling and Li [12]. The sample sizes used are 500 and 1000, and the results are also given in Table 1.

From Table 1, we make the following observations. First, the empirical sizes of the proposed test statistics $Q_1(M)$ and $Q_2(M)$ seem reasonable. Second, the proposed test statistics also have decent power and, as expected, the power of the tests increases as the sample size increases. Third, the power of $Q_1(M)$ is high for Model 3. This is understandable because the order of the VAR model is mis-specified. Fourth, for Model 2, the power of $Q_2(M)$ is higher than that of $Q_1(M)$ because in this case only the GARCH order is mis-specified. Finally, the results of Model 4 demonstrate clearly the contribution of using the cross-product vector of the standardized residuals in model checking. For this particular model, the test of Ling and Li [12] fails to detect the model inadequacy because the serial dependence of the conditional heteroscedasticity is in the cross-product series $a_{1t}a_{2t}$. The proposed $Q_2(M)$ statistics, on the other hand, has good power in detecting the model inadequacy. In real applications, the serial dependence in the conditional heteroscedasticity is typically unknown and it pays to use the more general test statistic $Q_2(M)$ proposed in the paper.

Finally, we examine the performance of the proposed residual-based bootstrap procedure by comparing the distribution of test statistics obtained by bootstrap with their asymptotic distributions, *i.e.*, the chi-square distributions in Theorem 5.1. Based on 1000 simulated data under Models 1, 2, and 3 with Gaussian error, we fit a VAR(1)-GARCH(1,1) model and approximate the distributions of $Q_1(M)$ and $Q_2(M)$ via 1000 residual-based bootstrap. Particularly, the results of $M = 15$ are reported in Table 2. Although Models 2 and 3 are mis-specified, the (asymptotic) distributions of the test statistics obtained by the re-sampled standardized residuals are approximately the same under both the null and alternative hypotheses, as we discussed in Section 5.

Table 1. Empirical sizes and power of $Q_1(M)$ and $Q_2(M)$ statistics for some VAR-GARCH models and empirical power of the proposed $Q_2(M)$ and that of Ling and Li [12] for Model 4. The latter is denoted by LL. Both Gaussian and Student-t innovations are used. The sample sizes used are 500 and 1000, and the results are based on 1000 realizations. Only results of M = 1 are reported.

		Model 1 (size)		Model 2 (power)		Model 3 (power)		Model 4 (power)	
ε_t	n	$Q_1(M)$	$Q_2(M)$	$Q_1(M)$	$Q_2(M)$	$Q_1(M)$	$Q_2(M)$	$Q_2(M)$	LL
Gaussian	500	0.030	0.061	0.143	0.336	0.983	0.517	0.528	0.067
	1000	0.022	0.053	0.231	0.645	0.999	0.682	0.686	0.053
Student-t	500	0.041	0.068	0.210	0.331	0.987	0.443	0.459	0.062
	1000	0.035	0.059	0.276	0.593	0.999	0.628	0.663	0.051

Figure 1. Empirical densities of the (1,1) element of the lag-1 autocorrelation matrix of the standardized residuals and their squared series. The shocks a_t are either pure Gaussian white noises or a BEKK(1,1) process with Gaussian innovations and parameters given in Equation (34). The results are based on 1000 time series. The sample sizes are 500, 1000, and 3000.

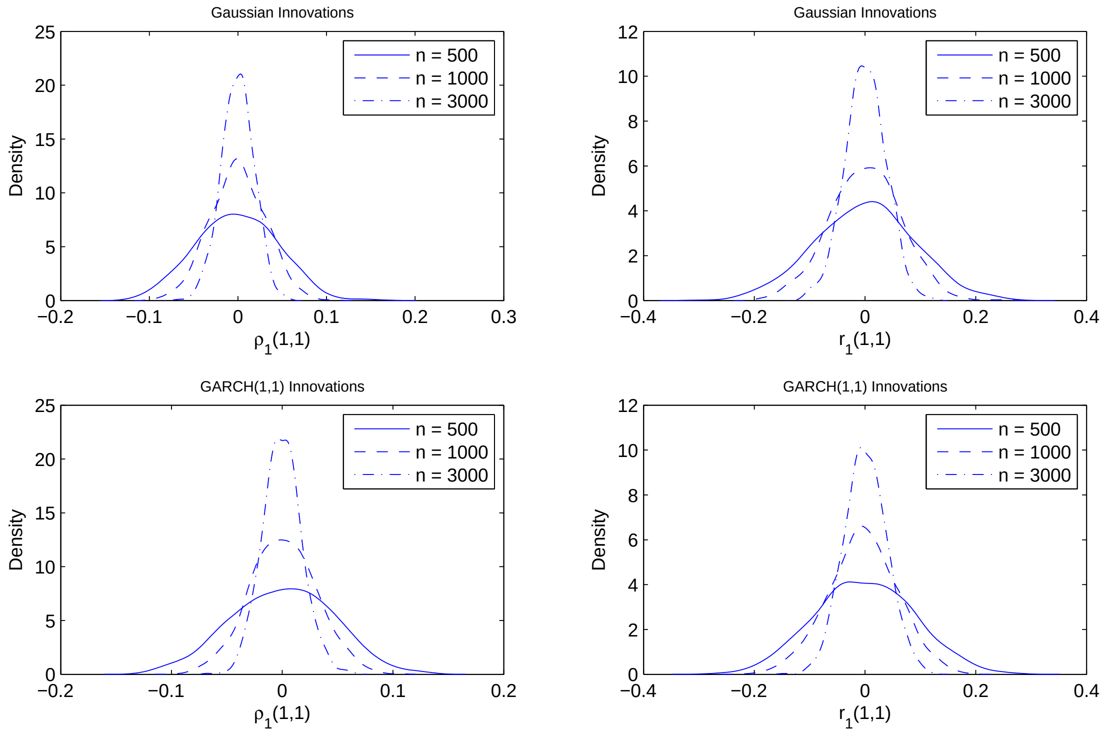


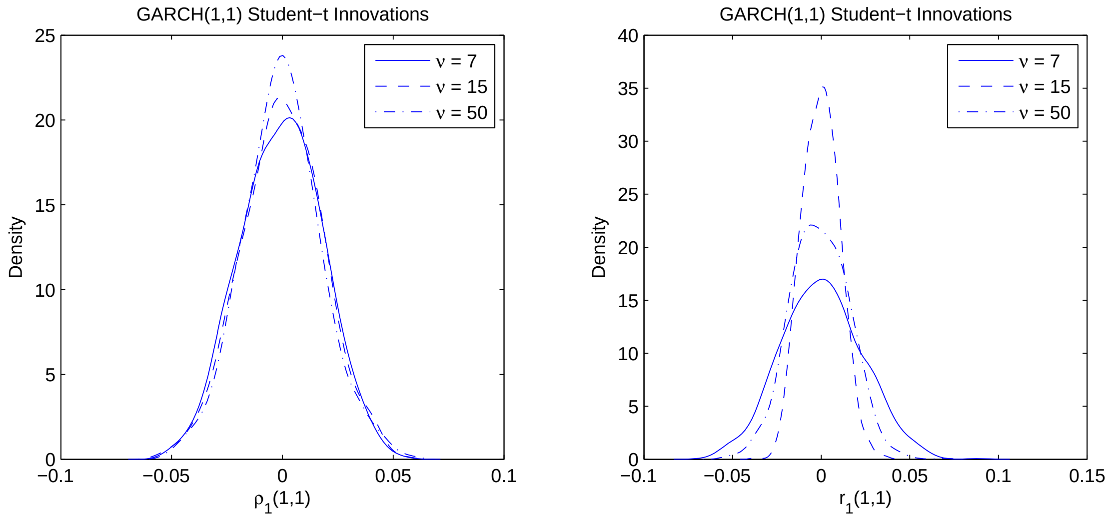
Table 2. Mean, variance, and quantiles of portmanteau statistics $Q_1^*(M)$ and $Q_2^*(M)$ obtained by the residual-based bootstrap approach for some VAR-GARCH Gaussian error models, and that of the corresponding chi-square distributions. The sample size used is 1000, and 1000 bootstrap samples are run for each model. Results of M = 15 are reported.

Model		Mean	Variance	Quantile: 1%	5%	10%	90%	95%	99%
	χ^2_{60}	60	120	37.48	43.19	46.46	74.40	79.08	88.38
	χ^2_{135}	135	270	99.74	109.16	114.42	156.44	163.12	176.14
1	$Q_1^*(M)$	60.11	118.73	37.78	43.59	46.35	74.01	80.07	86.14
	$Q_2^*(M)$	136.99	280.39	98.16	109.34	116.05	158.50	165.40	178.68
2	$Q_1^*(M)$	60.15	119.89	36.08	42.96	46.17	74.11	79.26	87.95
	$Q_2^*(M)$	137.12	282.20	98.83	109.75	114.77	158.43	164.84	178.52
3	$Q_1^*(M)$	60.08	115.74	38.14	44.18	47.71	74.14	77.96	86.51
	$Q_2^*(M)$	137.36	285.68	97.31	108.58	115.17	159.42	165.06	179.60

7. Application

In this section, we apply the proposed portmanteau tests to check the adequacy of a fitted VARMA-GARCH model for the returns of two well-known stock indices. The data consist of 1756

Figure 2. Empirical densities of the (1,1) element of the lag-1 autocorrelation matrix of the standardized residuals and their squared series. The time series Y_t is a GARCH(1,1) process following an BEKK(1,1) model with multivariate Student-t innovations and parameters given in Equation (34). The results are based on 1000 time series with sample size 3000. The degrees of freedom are 7, 15, and 50.



daily closing values of FTSE 100 Index and Deutsche Borse Ag German Stock Index (DAX) from January 3, 2006 to December 31, 2012. We focus on the return series $r_t = (r_{1t}, r_{2t})'$, where r_{1t} is the return on FTSE 100 index and r_{2t} on DAX index. Some preliminary analysis indicates that there exist some serial and cross-sectional correlations in the returns. Thus, a VAR(1) model is entertained. Significant autocorrelations of the cross-product vector of residuals of the fitted VAR(1) model suggest the existence of conditional heteroscedasticity in the returns. We then entertain three VAR-GARCH models and compute the proposed portmanteau test statistics $Q_1(M)$ and $Q_2(M)$ with $M = 15$. These test statistics are computed via two procedures. The first procedure uses the first and second order derivatives of log-likelihood function whereas the second procedure uses the residual-based bootstrap method. The model checking results are given in Table 3.

Table 3. Diagnostic tests for three models fitted to daily returns of FTSE 100 and DAX indices. The sample period is from January 3, 2006 to December 31, 2012. For Models 2 and 3, the degrees of freedom for the Student-t distribution are 7.00 and 7.02, respectively. Portmanteau statistics obtained by the first and second order derivatives of the log-likelihood function and the residual-based bootstrap procedure are reported and denoted by $Q_i(15)$ and $\hat{Q}_i^*(15)$ $(i = 1, 2)$, respectively. For each model, 1000 bootstrap samples are generated.

	Model	Log-likelihood	$Q_1(15)$	$Q_2(15)$	$\hat{Q}_1^*(15)$	$\hat{Q}_2^*(15)$
1	VAR(1)-GARCH(1,1) with Gaussian	-5641.0	52.90	172.65	55.67	170.18
2	VAR(1)-ARCH(1) with Student-t	-5658.3	48.94	235.14	46.65	237.90
3	VAR(1)-GARCH(1,1) with Student-t	-5638.8	43.00	157.75	45.13	158.97

For all the three entertained models, the portmanteau test statistics obtained by the bootstrap method and derivatives of the log-likelihood function are reasonably close. They provide the same diagnostic inference for the fitted models using the 5% critical values. Specifically, for the proposed $Q_1(15)$ and $Q_2(15)$, the asymptotic χ^2 distributions have degrees of freedom 60 and 135, respectively. The corresponding 5% critical values are 79.08 and 163.12, respectively. Therefore, the proposed portmanteau tests reject Models 1 and 2 of Table 3. For Model 3, the Q statistics of the two tests are 43.00 and 157.75, respectively. Thus, Model 3 cannot be rejected by the proposed test statistics. Parameters of the fitted VAR(1)-GARCH(1,1) model with Student-t innovations are given in Table 4.

Next, for comparison purpose, we also compute the test statistic of Ling and Li [12] for Model 2 of Table 3, *i.e.*, the VAR(1)-ARCH(1) model with Student-t innovations. As mentioned before, those authors employ the sum of squared series of the standardized residuals, $\hat{\varepsilon}_t'\hat{\varepsilon}_t$ to obtain their $Q_2(M)$ statistics. In this particular case, $Q_2(15) = 19.45$ with p-value 0.194. Therefore, based on the test statistics of Ling and Li [12], one could not reject Model 2 of Table 3 for the daily return series of FTSE 100 and DAX indices. This is in contrast with the proposed $Q_2(M)$ statistics that rejects Model 2. From the parameter estimates of Model 3, shown in Table 4, there exist some significant coefficients in the higher-order volatility coefficient matrices. The estimation result, thus, provides some support for rejecting Model 2 of Table 3.

Table 4. Estimated coefficients of the VAR(1) model with GARCH(1,1) Student-t innovations for FTSE 100 and DAX return bivariate time series data from January 3rd, 2006 to December 31st, 2012.

Parameters	Estimates	Std. errors
$\Phi_1(1,1)$	0.0306	0.0016
$\Phi_1(2,1)$	0.0014	0.0027
$\Phi_1(1,2)$	0.0358	0.0025
$\Phi_1(2,2)$	-0.0233	0.0038
$C(1,1)$	1.2495	0.0000
$C(2,1)$	-0.0280	0.0057
$C(2,2)$	1.1952	0.0000
$A_{11}(1,1)$	-0.0128	0.0026
$A_{11}(2,1)$	-0.0076	0.0105
$A_{11}(1,2)$	0.0419	0.0101
$A_{11}(2,2)$	0.0531	0.0526
$B_{11}(1,1)$	0.9584	0.0319
$B_{11}(2,1)$	-0.0106	0.0103
$B_{11}(1,2)$	0.0165	0.0116
$B_{11}(2,2)$	0.9606	0.0340
ν	7.0154	1.3909

In summary, the simple example considered in this application demonstrates the importance of using the cross-product vector of standardized residuals in checking a fitted VARMA-GARCH model.

Overlooking the cross-dependence in the standardized residuals may lead to erroneous conclusion. Moreover, the residual-based bootstrap method approximates the portmanteau statistics well, and provides a more straightforward approach to perform model checking in a real application.

8. Conclusion

In conclusion, we have derived the asymptotic distributions of standardized residual autocorrelation and autocovariance functions of squared and cross product of standardized residuals for vector ARMA model with multivariate GARCH innovations. Moreover, we propose two portmanteau statistics for a joint procedure to diagnose VARMA-GARCH models. Both cases of multivariate GARCH Gaussian and Student-t innovations are explored. To make the model checking tests easily implemented in real application, we also provide a residual-based bootstrap approach. Simulation is used to show the difference in the residual autocorrelations and autocovariances between VARMA model with and without GARCH effect. Empirical sizes and powers calculated based on three models suggest that our portmanteau statistics are useful in model checking. A larger power of our test is shown by simulated data compared with previous model checking test, in detecting conditional heteroscedastic data with strong autocorrelation in cross product between elements in the GARCH part. Simulation results also show that the test statistics obtained by bootstrap approximate the theoretical limiting distributions well. By estimation via both the first and second order derivatives of Log-likelihood function and the bootstrap procedure, an empirical example is analyzed to illustrate the importance of considering GARCH effect and heavy tail property for multivariate index or stock return data, and also other financial dataset. In future research, we expect our result be implemented to find other forms of the portmanteau statistic, and different methods of bootstrap for VARMA-GARCH models.

References

1. Box, G.E.P.; Jenkins, G. *Time Series Analysis, Forecasting and Control*. HoldenDay: San Francisco, CA, USA, 1976.
2. Box, G.E.P.; Pierce, D.A. Distribution of the residual autocorrelations in autoregressive integrated moving average time series models. *J. Am. Stat. Assoc.* **1970**, *65*, 1509-1526.
3. Ljung, G.M.; Box, G.E.P. On a measure of lack of fit in time series models. *Biometrika* **1978**, *65*, 297-303.
4. Chitturi, R.V. Distribution of residual autocorrelations in multiple autoregressive schemes. *J. Am. Stat. Assoc.* **1974**, *69*, 928-934.
5. Hoskings, J.R.M. The multivariate portmanteau statistic. *J. Am. Stat. Assoc.* **1980**, *75*, 602-607.
6. McLeod, A.I.; Li, W.K. Distribution of the residual autocorrelations in multivariate ARMA time series models. *J. R. Stat. Soc., Series B (Methodol.)* **1981**, *43*, 231-239.
7. Engle, R.E. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **1982**, *50*, 987-1007.
8. Bollerslev, T. Generalized autoregressive conditional hereroscedasticity. *J. Econometrics* **1986**, *31*, 307-327.

9. Diebold, F.X. Testing for Serieal Correlation in the Presence of ARCH. Proceedings of the American Statistical Association, Business and Economic Statistics Section, Washington, DC, USA, 1987, pp. 323-328.
10. Diebold, F.X. *Empirical Modeling of Exchange Rate Dynamics*. Springer-Verlag: New York, NY, USA, 1988.
11. Ling, S.Q.; Li, W.K. On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Am. Stat. Assoc.* **1997**, *92*, 1184-1194.
12. Ling, S.Q.; Li, W.K. Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors. *J. Time Ser. Anal.* **1997**, *18*, 447-464.
13. Tse, Y.K. Residual-based diagnostics for conditional heteroscedasticity models. *J. Econometrics* **2002**, *5*, 358-373.
14. Engle, R.E.; Kroner, K.F. Multivariate simultaneous generalized ARCH. *Economet. Theor.* **1995**, *11*, 122-150.
15. Tsay, R.S. *Analysis of Financial Time Series*, 3rd ed. Wiley: Hoboken, NJ, USA, 2010.
16. Lütkepohl, H. *A New Introduction to Multiple Time Series Analysis*. Springer Verlag: New York, NY, USA, 2005.
17. Hafner, C.M. Fourth moment structure of multivariate GARCH models. *J. Financ. Economet.* **2003**, *1*, 26-54.
18. Hidalgo, J.; Zaffaroni, P. A goodness-of-fit test for ARCH(∞) models. *J. Econometrics* **2007**, *141*, 835-875.
19. Fiorentini, G.; Sentana, E.; Calzolari, G. Multivariate models with student t innovations. *J. Bus. Econ. Stat.* **2003**, *21*, 532-546.
20. Tiao, G.C.; Tsay, R.S. Model specification in multivariate time series (with discussion). *J. R. Stat. Soc., Series B (Methodol.)* **1989**, *51*, 157-213.
21. Tsay, R.S. Conditional heteroscedastic time series models. *J. Am. Stat. Assoc.* **1987**, *82*, 590-604.
22. Tsay, R.S.; Tiao, G.C. Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. *Ann. Stat.* **1990**, *18*, 220-250.
23. Tsay, R.S. Model checking via parametric bootstraps in time series analysis. *Appl. Stat.* **1992**, *41*, 1-15.

Appendix

Proof of Lemma 3.2. From the calculation of $\partial L_n/\partial\lambda$ and iterated expectation, we have

$$
\begin{aligned}
E[\frac{\partial L_n}{\partial\varphi}vec'(\rho_m)] &= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=m+1}^{n}E[\frac{1}{2}(\frac{\partial H_t}{\partial\varphi'})'vec(\Sigma_t^{-1}a_ta_t'\Sigma_t^{-1}-\Sigma_t^{-1})(\varepsilon_{s-m}'\otimes\varepsilon_s') \\
&- (\frac{\partial a_t}{\partial\varphi'})'\Sigma_t^{-1}a_t(\varepsilon_{s-m}'\otimes\varepsilon_s')] \\
&= -\frac{1}{n}\sum_{t=1}^{n}E\left[(\frac{\partial a_t}{\partial\varphi'})'\Sigma_t^{-1}a_t(\varepsilon_{t-m}'\otimes\varepsilon_t')\right] \\
&= -\frac{1}{n}\sum_{t=1}^{n}E\left[(\frac{\partial a_t}{\partial\varphi'})'\Sigma_t^{-1/2}\varepsilon_t(\varepsilon_{t-m}'\otimes\varepsilon_t')\right] \\
&= -\frac{1}{n}\sum_{t=1}^{n}E\left[(\frac{\partial a_t}{\partial\varphi'})'(1\otimes\Sigma_t^{-1/2}\varepsilon_t)(\varepsilon_{t-m}'\otimes\varepsilon_t')\right] \\
&= -\frac{1}{n}\sum_{t=1}^{n}E\left[(\frac{\partial a_t}{\partial\varphi'})'(\varepsilon_{t-m}'\otimes\Sigma_t^{-1/2})\right].
\end{aligned}
$$

$$
E[\frac{\partial L_n}{\partial\delta}vec'(\rho_m)] = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=m+1}^{n}E[\frac{1}{2}(\frac{\partial H_t}{\partial\delta'})'vec(\Sigma_t^{-1}a_ta_t'\Sigma_t^{-1}-\Sigma_t^{-1})(\varepsilon_{s-m}'\otimes\varepsilon_s')] = O.
$$

Next, we compute $\partial vec(\rho_m)/\partial\lambda'$ as

$$
\frac{\partial vec(\rho_m)}{\partial\lambda'} = \frac{1}{n}\sum_{t=m+1}^{n}\frac{\partial vec(\varepsilon_m\varepsilon_{t-m}')}{\partial\lambda} = \frac{1}{n}\sum_{t=m+1}^{n}[(I_k\otimes\varepsilon_t)\frac{\partial\varepsilon_{t-m}}{\partial\lambda'}+(\varepsilon_{t-m}\otimes I_k)\frac{\partial\varepsilon_t}{\partial\lambda'}]
$$

where $\partial\varepsilon_t/\partial\lambda' = \Sigma_t^{-1/2}(\partial a_t/\partial\lambda') + (a_t'\otimes I_k)(\partial vec(\Sigma_t^{-1/2})/\partial\lambda')$. Therefore, we have

$$
\begin{aligned}
\frac{\partial vec(\rho_m)}{\partial\lambda'} &= \frac{1}{n}\sum_{t=m+1}^{n}\{(I_k\otimes\varepsilon_t)[\Sigma_{t-m}^{-1/2}(\partial a_{t-m}/\partial\lambda')+(a_{t-m}'\otimes I_k)(\partial vec(\Sigma_{t-m}^{-1/2})/\partial\lambda')] \\
&+ (\varepsilon_{t-m}\otimes I_k)[\Sigma_t^{-1/2}(\partial a_t/\partial\lambda')+(a_t'\otimes I_k)(\partial vec(\Sigma_t^{-1/2})/\partial\lambda')]\} \\
&\doteq \frac{1}{n}\sum_{t=m+1}^{n}(\varepsilon_{t-m}\otimes\Sigma_t^{-1/2}(\partial a_t/\partial\lambda')),
\end{aligned}
$$

where $\doteq$ indicates asymptotic equivalence as $n\to\infty$. The prior asymptotic equivalence holds because the other three terms converge to zero. It follows that

$$
E\left[\frac{\partial vec(\rho_m)}{\partial\lambda'}\right] = \frac{1}{n}\sum_{t=m+1}^{n}E\left[(\varepsilon_{t-m}\otimes\Sigma_t^{-1/2})\frac{\partial a_t}{\partial\lambda'}\right].
$$

Consequently, the results of Lemma 3.2 hold. $\square$.

Proof of Lemma 3.4. For simplicity, let $\Sigma_t^* = \Sigma_t^{-1}a_t a_t'\Sigma_t^{-1} - \Sigma_t^{-1}$. From the calculation of $\frac{\partial L_n}{\partial \lambda}$, we have

$$
\begin{aligned}
E[\frac{\partial L_n}{\partial \varphi} vec'(r_m)] &= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=m+1}^{n} E[\frac{1}{2}(\frac{\partial H_t}{\partial \varphi'})' vec(\Sigma_t^*)(vech'(c_{s-m}) \otimes vech'(c_s)) \\
&- (\frac{\partial a_t}{\partial \varphi'})'\Sigma_t^{-1}a_t(vech'(c_{s-m}) \otimes vech'(c_s))] \\
&= \frac{1}{n}\sum_{t=1}^{n} E[\frac{1}{2}(\frac{\partial H_t}{\partial \varphi'})' vec(\Sigma_t^*)(vech'(c_{t-m}) \otimes vech'(c_t))]. \\
E[\frac{\partial L_n}{\partial \delta} vec'(r_m)] &= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=m+1}^{n} E[\frac{1}{2}(\frac{\partial H_t}{\partial \delta'})' vec(\Sigma_t^*)(vech'(c_{s-m}) \otimes vech'(c_s))] \\
&= \frac{1}{n}\sum_{t=1}^{n} E[\frac{1}{2}(\frac{\partial H_t}{\partial \delta'})' vec(\Sigma_t^*)(vech'(c_{t-m}) \otimes vech'(c_t))].
\end{aligned}
$$

Then

$$
\begin{aligned}
&E[\frac{\partial L_n}{\partial \lambda} vec'(r_m)] = \frac{1}{n}\sum_{t=m+1}^{n} E[\frac{1}{2}(\frac{\partial H_t}{\partial \lambda'})' vec(\Sigma_t^*)(vech'(c_{t-m}) \otimes vech'(c_t))] \\
&= \frac{1}{n}\sum_{t=m+1}^{n} E[\frac{1}{2}(\frac{\partial H_t}{\partial \lambda'})'(\Sigma_t^{-1/2} \otimes \Sigma_t^{-1/2}) vec(c_t)(vech'(c_{t-m}) \otimes (vech'(c_t)))] \\
&= \frac{1}{n}\sum_{t=m+1}^{n} E\frac{1}{2}[\left(\frac{\partial vec(\Sigma_t^{1/2})}{\partial \lambda'}\right)' (\Sigma_t^{-1/2} \otimes I_k + I_k \otimes \Sigma_t^{-1/2}) vech'(c_{t-m}) \otimes (vec(c_t)vec'(c_t)D_k^{+'})] \\
&= \frac{1}{n}\sum_{t=m+1}^{n} E\{\left(\frac{\partial vec(\Sigma_t^{1/2})}{\partial \lambda'}\right)' (\Sigma_t^{-1/2} \otimes I_k + I_k \otimes \Sigma_t^{-1/2})[vech'(c_{t-m}) \otimes D_k^{+'}]\} \\
&= \mathbf{X}_m'.
\end{aligned}
$$

Next, we compute $\partial vec(r_m)/\partial \lambda'$:

$$
\begin{aligned}
&\frac{\partial vec(r_m)}{\partial \lambda'} = \frac{1}{n}\sum_{t=m+1}^{n} \frac{\partial [vec(vech(c_t)vech'(c_{t-m}))]}{\partial \lambda'} \\
&= \frac{1}{n}\sum_{t=m+1}^{n} [(I_{k^2} \otimes vech(c_t))\frac{\partial vech(c_{t-m})}{\partial \lambda'} + (vech(c_{t-m}) \otimes I_{k^2})\frac{\partial vech(c_t)}{\partial \lambda'}] \\
&\doteq \frac{1}{n}\sum_{t=m+1}^{n} (vech(c_{t-m}) \otimes I_{k^2})\frac{\partial vech(c_t)}{\partial \lambda'} \\
&= \frac{1}{n}\sum_{t=m+1}^{n} (vech(c_{t-m}) \otimes I_{k^2}) D_k^{+}\frac{\partial vec(c_t)}{\partial \lambda'} \\
&= \frac{1}{n}\sum_{t=m+1}^{n} (vech(c_{t-m}) \otimes D_k^{+})[(I_k \otimes \Sigma_t^{-1/2}a_t a_t')\frac{\partial vec(\Sigma_t^{-1/2})}{\partial \lambda'} \\
&+ (\Sigma_t^{-1/2} \otimes \Sigma_t^{-1/2})\frac{\partial a_t a_t'}{\partial \lambda'} + (\Sigma_t^{-1/2}a_t a_t' \otimes I_k)\frac{\partial vec(\Sigma_t^{-1/2})}{\partial \lambda'}].
\end{aligned}
$$

Since

$$\frac{\partial vec(\Sigma_t^{-1/2})}{\partial \lambda'} = -[\Sigma_t^{-1/2} \otimes \Sigma_t^{-1/2}]\frac{\partial vec(\Sigma_t^{1/2})}{\partial \lambda'},$$

it follows that

$$E[\frac{\partial vec(r_m)}{\partial \lambda'}] = -\frac{1}{n}\sum_{t=m+1}^{n} E\left\{[vech(c_{t-m}) \otimes D_k^{+}](\Sigma_t^{-1/2} \otimes I_k + I_k \otimes \Sigma_t^{-1/2})\left(\frac{\partial vec(\Sigma_t^{1/2})}{\partial \lambda'}\right)\right\},$$

which is $-\mathbf{X}_m$ defined in (14). Therefore, the results of Lemma 3.4 hold. $\square$.

Proof of Lemma 5.2. To show the above lemma, we first express H_t as a function of entire past of the cross product of innovations ε_t.

$$\begin{aligned}
H_t &= C_0 + \sum_{j=0}^{t-1} C_j R_{t-j} = C_0 + \sum_{j=0}^{t-1} C_j[\Sigma_{t-j}^{1/2} \otimes \Sigma_{t-j}^{1/2}]vec(\varepsilon_{t-j}\varepsilon_{t-j}') \\
&= C_0 + \sum_{j=0}^{t-1} C_j G(H_{t-j})vec(\varepsilon_{t-j}\varepsilon_{t-j}') \\
H_t^* &= C_0 + \sum_{j=0}^{t-1} C_j R_{t-j}^* = C_0 + \sum_{j=0}^{t+d-1} C_j[\Sigma_{t-j}^{*1/2} \otimes \Sigma_{t-j}^{*1/2}]vec(\varepsilon_{t-j}^*\varepsilon_{t-j}^{*'}) \\
&= C_0 + \sum_{j=0}^{t-1} C_j G(H_{t-j}^*)vec(\varepsilon_{t-j}^*\varepsilon_{t-j}^{*'})
\end{aligned}$$

Thus, one can write $h_t^*(\lambda) = \ln det(\Sigma_t^*)$ as $h_t^*(\lambda) = h(\varepsilon_{t-1}^*\varepsilon_{t-1}^{*'}, \varepsilon_{t-2}^*\varepsilon_{t-2}^{*'}, ..., \varepsilon_{1-d}^*\varepsilon_{1-d}^{*'}; \lambda)$. Let $h_{t,K}^*(\lambda) = h(\varepsilon_{t-1}^*\varepsilon_{t-1}^{*'}, ..., \varepsilon_{t-2}^*\varepsilon_{t-2}^{*'}, ..., \varepsilon_{t-K}^*\varepsilon_{t-K}^{*'}, 0, ..., 0; \lambda)$ be the truncated version. Since $G(H_t) = \Sigma_t^{1/2} \otimes \Sigma_t^{1/2} \equiv (\sum_{i=1}^{k} J_{i,k}' \cdot H_t \cdot e_i)^{1/2} \otimes (\sum_{i=1}^{k} J_{i,k}' \cdot H_t \cdot e_i)^{1/2}$ is a continuous differentiable function in all its argument with finite second moments, then so is $h_t^*(\lambda)$. By similar arguments as **Lemma 6.13** of Hidalgo and Zaffaroni [18], we have

$$E^*|\frac{1}{n}\sum_{t=1}^{n}[h_{t,K}^*(\hat{\lambda}) - h_{t,K}(\lambda_0)]| = O_p(\frac{K^{1/2}}{n^{1/2}}).$$

and the truncation error approaches zero fast enough, *i.e.*,

$$E^*|\frac{1}{n}\sum_{t=1}^{n}[h_t^*(\lambda) - h_{t,K}^*(\lambda)]| \to 0$$

as $K \to \infty$. By choosing K large enough, the result of (29) holds. The proof of (30) proceeds similarly. $\square$.

Proof of Lemma 5.3. We begin with (31). By the mean value theorem,

$$|\ln det\Sigma_t^*(\lambda_1) - \ln det\Sigma_t^*(\lambda_2)| \leq C|\lambda_1 - \lambda_2| \cdot |\frac{\partial H_t^{*\prime}}{\partial \lambda'}(\bar{\lambda}) \cdot vec(\bar{\Sigma}_t^{*-1}(\bar{\lambda}))|$$

where C is a constant and $\bar{\lambda}$ is an intermediate point between λ_1 and λ_2. By the first-order derivatives, we have

$$\frac{\partial a_t^*}{\partial \varphi'} = \Theta^{-1}(L)[X_{t-1}^* \otimes I_k], \quad X_{t-1}^* = (Y_{t-1}^{*'}, ..., Y_{t-p}^{*'}, a_{t-1}^{*'}, ..., a_{t-q}^{*'}),$$
$$\frac{\partial H_t}{\partial \varphi'} = \big(I_{k^2} - \sum_{i=1}^{s}(B_i \otimes B_i)L^i\big)^{-1}\left[\sum_{i=1}^{r}(A_i \otimes A_i)L^i((I_k \otimes a_t^* + a_t^* \otimes I_k)\frac{\partial a_t^*}{\partial \varphi'})\right],$$
$$\frac{\partial H_t^*}{\partial \delta'} = \big(I_{k^2} - \sum_{i=1}^{s}(B_i \otimes B_i)L^i\big)^{-1}[I_{k^2}, \widetilde{H}_{t-1}^{*(1)}, ..., \widetilde{H}_{t-r}^{(1)}, \widetilde{H}_{t-1}^{*(2)}, ..., \widetilde{H}_{t-s}^{*(2)}],$$
$$\widetilde{H}_{t-i}^{*(1)} = (I_k \otimes A_i) \cdot [I_k \otimes (a_{t-i}^* a_{t-i}^{*'} e_1), ..., I_k \otimes (a_{t-i}^* a_{t-i}^{*'} e_k)] + (A_i \otimes I_k) \cdot ((a_{t-i}^* a_{t-i}^{*'}) \otimes I_k),$$
$$\widetilde{H}_{t-i}^{*(2)} = (I_k \otimes A_i) \cdot [I_k \otimes (\Sigma_{t-i}^* e_1), ..., I_k \otimes (\Sigma_{t-i}^* e_k)] + (A_i \otimes I_k) \cdot (\Sigma_{t-i}^* \otimes I_k),$$

By Assumptions 2 and 3, the following expansion holds

$$\| \frac{\partial a_t^*}{\partial \phi} \|_F \leq c_1 + c_2 \sum_{i=1}^{t-1} \rho_1^i \| Y_{t-i}^* \| \equiv K_{1t}$$

where $\| \cdot \|_F$ and $\| \cdot \|$ denote Frobenius and Euclidean norms, respectively. c_1, c_2 and ρ_1 are constants, and $0 < \rho_1 < 1$. K_{1t} is a strictly stationary time series with $E^* K_{1t} < \infty$. Then it follows that

$$\sup_{\lambda \in \Lambda} \| \frac{\partial H_t^*}{\partial \phi} \| \leq c_3 + c_4 \sum_{i=1}^{t-1} \rho_2^i \| Y_{t-1}^* \|^2 \equiv K_{2t}$$

where c_3, c_4 and ρ_2 are constants, and $0 < \rho_2 < 1$. K_{2t} is a strictly stationary time series with $E^* K_{2t} < \infty$. Similarly, it could be obtained that

$$\sup_{\lambda \in \Lambda} \| \frac{\partial H_t^*}{\partial \delta} \| \equiv K_{3t}$$

where K_{3t} is a strictly stationary time series with $E^* K_{3t} < \infty$. Moreover, since $\Sigma_t^* > 0$, $\rho(\Sigma_t^*)$ has a lower bound. Then,

$$\| vec(\Sigma_t^{*-1}(\bar{\lambda})) \| = \| \Sigma_t^{*-1}(\bar{\lambda}) \|_F = O(1)$$

Therefore, it follows that $E^* \sup_{\lambda \in \Lambda} \| \frac{\partial H_t^*}{\partial \lambda}(\bar{\lambda}) \| = K_t$ and $E^* \sup_{\lambda \in \Lambda} \| vec\big(\Sigma_t^*(\bar{\lambda})^{-1}\big) \| = K_t$. So the left side of (31) is K_t. Following the similar arguments, we could show that (32) holds. $\square$.