

Statistica Sinica Preprint No: SS-2025-0286

Title	Statistical Inference for Heavy-tailed Vector Arma-garch/Igarch Models
Manuscript ID	SS-2025-0286
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0286
Complete List of Authors	Bibi Cai, Mengya Liu and Shiqing Ling
Corresponding Authors	Shiqing Ling
E-mails	maling@ust.hk
Notice: Accepted author version.	

STATISTICAL INFERENCE FOR HEAVY-TAILED VECTOR ARMA-GARCH/IGARCH MODELS

Bibi Cai, Mengya Liu and Shiqing Ling*

Changsha University of Science and Technology,

Central China Normal University,

and Hong Kong University of Science and Technology

Abstract: This paper develops the statistical procedure for the heavy-tailed vector ARMA-GARCH model. We first study the self-weighted quasi-maximum likelihood estimator (QMLE) of the vector ARMA-GARCH model and establish its consistency and asymptotic normality under a fractional moment condition. Using the self-weighted QMLE, we establish the asymptotic normality of the local QMLE for VARMA model when the noise follows a finite second moment vector GARCH or IGARCH model. Based on the two estimators, we construct the Wald statistics for testing linear and nonlinear restrictions, and especially, we propose a Wald statistic for testing the vector IGARCH model. We also construct two portmanteau tests for checking the adequacy of models. All test statistics are shown to be asymptotically χ^2 distributions as the sample size goes to infinity. Simulation results show that our procedure works well in the finite samples and one real example is given to demonstrate how to build a model by using our

procedure.

Key words and phrases: Asymptotic normality, consistency, model checking, self-weighted QMLE, VARMA model, VGARCH model.

1. Introduction

Since the seminal work of Engle (1982) and Bollerslev (1986), the GARCH model has been widely applied in economics and finance, see Li, Ling and McAleer (2002) and Düker et al. (2025). Various multivariate GARCH-type models have been proposed, for example, BEKK model by Engle and Kroner (1995), constant conditional correlation (CCC) model by Bollerslev (1990) and Jeantheau (1998), and dynamic conditional correlation (DCC) model by Tse and Tsui (2002) and Engle (2002), see also Francq and Zakoïan (2019). The asymptotic theory for several estimators has been well established under specific moment conditions. Jeantheau (1998) demonstrated the strong consistency of QMLE for multivariate GARCH model under logarithmic moment. Under at least finite sixth moment condition, Comte and Lieberman (2003) and Pedersen and Rahbek (2014) derived the asymptotic normality of the QMLE and variance targeting estimator (VTE) of the BEKK model, respectively. Francq, Horváth and Zakoïan (2016) established the asymptotic theory of VTE for (asymmetric) CCC-GARCH

model under the finite fourth moment condition. When the fourth moment does not exist, Pedersen (2016) found that the limiting distribution of VTE for the CCC-GARCH model was multivariate stable. Furthermore, Francq and Zakoïan (2012) relaxed the moment condition and established the asymptotic normality for the CCC asymmetric GARCH (AGARCH) model under strict stationarity condition, see also Hafner and Preminger (2009), Francq and Zakoïan (2016), and Maïnassara, Kadmiri and Sausseureau (2022).

The conditional mean usually is not a constant in economic and financial time series data, and the ARMA model with GARCH-type errors has been commonly used to analyze the real data in practice. For univariate ARMA-GARCH model, Ling and Li (1997b) and Francq and Zakoïan (2004) established the asymptotic normality of the QMLE under the finite fourth moment, and Ling (2007) proposed the global self-weighted QMLE to control the leverage points and derived its consistency and asymptotic normality under fractional moment condition, see also Zhu and Ling (2011) and Zhu and Ling (2015). For vector ARMA-GARCH model, Ling and McAleer (2003) established the consistency and asymptotic normality of QMLE under the finite sixth moment condition, see also McAleer et al. (2008) and McAleer, Hoti and Chan (2009). However, when the vector GARCH pro-

cess has an infinite fourth moment, its tail index, denoted by 2ι , is less than 4 and it has the heavy-tailed feature. Heavy-tailed phenomena have been widely demonstrated in financial markets, see Koedijk and Kool (1992), Francq and Zakoïan (2013), and Yang and Ling (2017), and traditional estimators may fail to have desirable asymptotic properties under such conditions. Zhang and Ling (2026) studied the least square estimator (LSE) of the vector AR model [i.e., model (2.1) with $q = 0$] with ε_t defined by (2.2) in Section 2 and showed that it is $n^{1-1/\iota}$ -consistent when the tail index 2ι is in $(2, 4)$ and inconsistent when in $(0, 2)$, see Lange (2011) and Zhang and Ling (2015). This result shows that the conventional time series analysis procedures would mislead practitioners if the vector ARMA model with the heavy-tailed GARCH noise.

This paper develops the statistical procedure for the heavy-tailed vector ARMA-GARCH model. We first study the self-weighted QMLE of the vector ARMA-GARCH model and establish its consistency and asymptotic normality under a fractional moment condition. Using the self-weighted QMLE, we establish the asymptotic normality of the local QMLE for VARMA model when the noise follows a finite second moment vector GARCH or IGARCH model. Based on the two estimators, we construct the Wald statistics for testing linear and nonlinear restrictions, and especially, we

propose a Wald statistic for testing the vector IGARCH model. We also construct two portmanteau tests for checking the adequacy of models. All test statistics are shown to be asymptotically χ^2 distributions as the sample size goes to infinity. Our proposed method can be extended for other vector GARCH specifications such as the BEKK model of Engle and Kroner (1995). But the conditions of these GARCH-type models depend on each model itself and need to be worked out case by case.

The rest of this paper is organized as follows. Section 2 introduces the model and assumptions. Section 3 studies the self-weighted QMLE, while Section 4 investigates the local QMLE. Section 5 discusses the model diagnostic checking. Section 6 reports simulation. The real example is given in Section 7. All proofs and additional simulation results offered in supplementary material. Throughout this paper, $C > 0$ and $\rho \in (0, 1)$ are generic constants that may vary across different places. The notation $\|\cdot\|$ denotes the Euclidean norm, and the transpose of matrix A is denoted by A' . \xrightarrow{p} and $\xrightarrow{\mathcal{L}}$ denote convergence in probability and in distribution as $n \rightarrow \infty$ respectively.

2. Model and Assumptions

Assume that $\{Y_t : t = 0, \pm 1, \pm 2, \dots\}$ are generated by the m -dimensional VARMA-GARCH model as follows:

$$Y_t - \mu = \sum_{i=1}^p \Phi_i(Y_{t-i} - \mu) + \varepsilon_t + \sum_{i=1}^q \Psi_i \varepsilon_{t-i}, \quad (2.1)$$

$$\varepsilon_t = D_t \eta_t \text{ and } H_t = W + \sum_{i=1}^r A_i \vec{\varepsilon}_{t-i} + \sum_{i=1}^s B_i H_{t-i}, \quad (2.2)$$

where $H_t = (h_{1t}, \dots, h_{mt})'$, $D_t = \text{diag}(h_{1t}^{1/2}, \dots, h_{mt}^{1/2})$, $\vec{\varepsilon}_t = (\varepsilon_{1t}^2, \dots, \varepsilon_{mt}^2)'$, $\eta_t = (\eta_{1t}, \dots, \eta_{mt})'$ is a sequence of i.i.d. random vector, with mean zero and covariance Γ ,

$$\Gamma = \begin{pmatrix} 1 & \sigma_{12} & \cdots & \sigma_{1,m-1} & \sigma_{1m} \\ \sigma_{21} & 1 & \cdots & \sigma_{2,m-1} & \sigma_{2m} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{m,m-1} & 1 \end{pmatrix}, \quad (2.3)$$

in which $\sigma_{ij} = \sigma_{ji}$. Model (2.2) is the ECCC-GARCH model, which has been discussed in Jeantheau (1998), Ling and McAleer (2003), He and Teräsvirta (2004) and Nakatani and Teräsvirta (2009). Model (2.1) reduces to the AR(p) model when $q = 0$. When $q > 0$, model (2.1) has several strong advantages: (i) it can provide a more parsimonious description of linear vector time series model, see Dufour and Pelletier (2022); (ii) The class of VARMA model is closed under linear transformation and temporal ag-

gregation, see Lütkepohl (2005) and Amendola, Niglio and Vitale (2009); (iii) A parsimonious VARMA model leads to improve forecast accuracy compared with VAR model, see Wilms, Basu, Bien and Matteson (2023) and Düker et al. (2025). The true parameter vector is denoted by $\lambda_0 = (\varphi'_0, \delta'_0, \sigma'_0)'$, where $\varphi_0 = \text{vec}(\mu_0, \Phi_{01}, \dots, \Phi_{0p}, \Psi_{01}, \dots, \Psi_{0q})$, $\delta_0 = \text{vec}(W_0, A_{01}, \dots, A_{0r}, B_{01}, \dots, B_{0s})$, and $\sigma_0 = (\sigma_{021}, \dots, \sigma_{0m,1}, \sigma_{032}, \dots, \sigma_{0m,2}, \dots, \sigma_{0m,m-1})'$. The elements of A_i and B_j are nonnegative, $i = 1, \dots, r$, $j = 1, \dots, s$. Denote $\Phi(L) = I_m - \Phi_1 L - \dots - \Phi_p L^p$ and $\Psi(L) = I_m + \Psi_1 L + \dots + \Psi_q L^q$ with I_m being the $m \times m$ identity matrix. The parameter space Θ is assumed to be a compact subspace of $R^{m+(p+q)m^2} \times (0, \infty)^m \times [0, \infty)^{(r+s)m^2} \times (-1, 1)^{m(m-1)/2}$ with λ_0 being an interior point in Θ , and for each $\lambda \in \Theta$, we make the following assumptions.

Assumption 2.1 *All the roots of $|\Phi(L)| = 0$ and $|\Psi(L)| = 0$ lie outside the unit circle.*

Assumption 2.2 *The terms of $\Phi(L)$ and $\Psi(L)$ are left coprime and satisfy other identifiability conditions given in Dunsmuir and Hannan (1976).*

Assumption 2.3 *The matrix Γ is a finite and positive definite symmetric matrix. Each diagonal element of Γ is 1 and $\rho(\Gamma)$ has a positive lower bound over Θ . Each element of W has positive lower and upper bounds over Θ . All the roots of $|I_m - \sum_{i=1}^s B_i L^i| = 0$ lie outside the unit circle.*

Assumption 2.4 $\sum_{i=1}^r A_i L^i$ and $I_m - \sum_{i=1}^s B_i L^i$ are left coprime and satisfy the identifiability conditions given in Jeantheau (1998).

Assumption 2.5 $E \|\varepsilon_t\|^{2\iota} < \infty$, for some $\iota > 0$.

Assumption 2.1 is the stationarity and invertibility condition of model (2.1), under which it follows that

$$\Psi^{-1}(L) = \sum_{k=0}^{\infty} \gamma_k L^k \quad \text{and} \quad \Psi^{-1}(L)\Phi(L) = \sum_{k=0}^{\infty} \gamma_{1k} L^k, \quad (2.4)$$

where $\sup_{\Theta} \|\gamma_k\| = O(\rho^k)$ and $\sup_{\Theta} \|\gamma_{1k}\| = O(\rho^k)$ with $\rho \in (0, 1)$. Assumptions 2.2 and 2.4 are the identifiability conditions in Jeantheau (1998), which may be overly restrictive. Similar identifiability conditions can also be used in Francq and Zakoïan (2010) and Francq and Zakoïan (2012). Dufour and Pelletier (2022) and Wilms, Basu, Bien and Matteson (2023) recently has provided weaker identifiability conditions. It can be shown that the main results of this paper remain valid under these alternative identification conditions.

Let

$$G = \begin{pmatrix} B_1 & \cdots & B_s \\ I_{m(s-1)} & & 0 \end{pmatrix}. \quad (2.5)$$

Since $|zI_{ms \times ms} - G| = 0$ and $|I_m - \sum_{i=1}^s B_i z^i| = 0$ have the same roots, Assumption 2.3 implies that

$$0 \leq \rho(G) < 1, \quad (2.6)$$

where $\rho(G)$ is the spectral radius of matrix G . Furthermore, by the com-

compactness of Θ ,

$$\sup_{\lambda \in \Theta} \rho(G) < 1. \quad (2.7)$$

Under this condition, we have

$$\left(I_m - \sum_{i=1}^s B_i L^i \right)^{-1} = \sum_{k=0}^{\infty} \Gamma_k L^k, \quad \left(I_m - \sum_{i=1}^s B_i L^i \right)^{-1} \left(\sum_{i=1}^r A_i L^i \right) = \sum_{k=1}^{\infty} \Gamma_{1k} L^k, \quad (2.8)$$

where $\sup_{\Theta} \|\Gamma_k\| = O(\rho^k)$ and $\sup_{\Theta} \|\Gamma_{1k}\| = O(\rho^k)$ with $\rho = \rho(G)$. The necessary and sufficient condition for $E\|\varepsilon_t\|^2 < \infty$ is that all the roots of $|I_m - \sum_{i=1}^r A_i L^i - \sum_{i=1}^s B_i L^i| = 0$ lie outside the unit circle, see Ling and McAleer (2003).

3. Self-weighted QMLE for VARMA-GARCH

Given the observations $\{Y_n, \dots, Y_1\}$ with the initial values $\{Y_0, Y_{-1}, Y_{-2}, \dots\}$ which are generated by (2.1)-(2.2), the parametric model can be written as

$$\varepsilon_t(\varphi) = (Y_t - \mu) - \sum_{i=1}^p \Phi_i (Y_{t-i} - \mu) - \sum_{i=1}^q \Psi_i \varepsilon_{t-i}(\varphi), \quad (3.1)$$

$$\eta_t(\lambda) = D_t^{-1}(\lambda) \varepsilon_t(\varphi), \quad H_t(\lambda) = W + \sum_{i=1}^r A_i \vec{\varepsilon}_{t-i}(\varphi) + \sum_{i=1}^s B_i H_{t-i}(\lambda), \quad (3.2)$$

where $\varepsilon_t(\varphi_0) = \varepsilon_{0t}$, $H_t(\lambda_0) = H_{0t}$, $D_t(\lambda_0) = D_{0t}$, and $\eta_t(\lambda_0) = \eta_{0t}$. The

log-quasi-likelihood function based on $\{\varepsilon_t(\varphi) : t = 1, \dots, n\}$ is

$$L_n(\lambda) = \frac{1}{n} \sum_{t=1}^n l_t(\lambda), \quad l_t(\lambda) = -\frac{1}{2} \ln |D_t(\lambda) \Gamma D_t(\lambda)| - \frac{1}{2} \varepsilon_t'(\varphi) (D_t(\lambda) \Gamma D_t(\lambda))^{-1} \varepsilon_t(\varphi). \quad (3.3)$$

The maximizer of $L_n(\lambda)$ on Θ is the QMLE of λ_0 . Since we only have a fractional moment condition as in Assumption 2.5, we introduce the weighted log-quasi-likelihood function:

$$L_{sn}(\lambda) = \frac{1}{n} \sum_{t=1}^n \omega_t l_t(\lambda), \quad (3.4)$$

where $l_t(\lambda)$ is the same as in (3.3). We make the following assumption for ω_t :

Assumption 3.1 $\omega_t = \omega(Y_{t-1}, Y_{t-2}, \dots)$ and ω is a measurable, positive and bounded function on R^{Z_0} with $E[\omega_t \xi_{\rho t-1}^4] < \infty$ and $Z_0 = (0, 1, 2, \dots)$ for any $\rho \in (0, 1)$, where $\xi_{\rho t} = 1 + \sum_{i=0}^{\infty} \rho^i \|Y_{t-i}\|$.

The initial values Y_i are unknown when $i \leq 0$ in practice and need to be replaced by some constants. Denote ε_t , H_t , and ω_t as ε_t^ϵ , H_t^ϵ , and ω_t^ϵ , respectively, with Y_i being a constant not depending on parameters when $i \leq 0$. Thus, the weighted log-quasi-likelihood function (3.4) is modified as follows:

$$\begin{aligned} L_{sn}^\epsilon(\lambda) &= \frac{1}{n} \sum_{t=1}^n \omega_t^\epsilon l_t^\epsilon(\lambda), \\ l_t^\epsilon(\lambda) &= -\frac{1}{2} \ln |D_t^\epsilon(\lambda) \Gamma D_t^\epsilon(\lambda)| - \frac{1}{2} \varepsilon_t^{\epsilon'}(\varphi) (D_t^\epsilon(\lambda) \Gamma D_t^\epsilon(\lambda))^{-1} \varepsilon_t^\epsilon(\varphi), \end{aligned} \quad (3.5)$$

where $D_t^\epsilon = \text{diag}(h_{1t}^{\epsilon 1/2}, \dots, h_{mt}^{\epsilon 1/2})$. To make the initial values Y_i ignorable when $i \leq 0$, we have the following assumption.

Assumption 3.2 $E|\omega_t - \omega_t^\epsilon|^{\iota_0/4} = O(t^{-2})$, where $\iota_0 = \min\{\nu, 1\}$.

We call the maximizer of $L_{sn}^\epsilon(\lambda)$ on Θ the self-weighted QMLE, denoted by $\hat{\lambda}_{sn}$, because the weight ω_t only depends on $\{Y_t\}$ itself. Our result for $\hat{\lambda}_{sn}$ is as follows.

Theorem 3.1 *Suppose that Assumptions 2.1-2.5 and 3.1-3.2 hold. Then*

(i) $\hat{\lambda}_{sn} \xrightarrow{p} \lambda_0$;

(ii) $\sqrt{n}(\hat{\lambda}_{sn} - \lambda_0) \xrightarrow{\mathcal{L}} N(0, \Sigma_0^{-1} \Omega_0 \Sigma_0^{-1})$ if $E \|\eta_t\|^4 < \infty$,

where $\Sigma_0 = E \left[\omega_t \frac{\partial^2 l_t(\lambda_0)}{\partial \lambda \partial \lambda'} \right]$, $\Omega_0 = E \left[\omega_t^2 \frac{\partial l_t(\lambda_0)}{\partial \lambda} \frac{\partial l_t(\lambda_0)}{\partial \lambda'} \right]$.

By direct calculation,

$$\Sigma_0 = E \left[w_t \frac{\partial^2 l_t(\lambda_0)}{\partial \lambda \partial \lambda'} \right] = - \begin{pmatrix} \Sigma_{\tilde{\lambda}0} & \Sigma_{\tilde{\lambda}\sigma 0} \\ \Sigma'_{\tilde{\lambda}\sigma 0} & \frac{1}{2} \omega_t P' P \end{pmatrix},$$

where $\tilde{\lambda} = (\varphi', \delta')'$,

$$\Sigma_{\tilde{\lambda}0} = E \left[w_t \frac{\partial \varepsilon'_{0t}}{\partial \tilde{\lambda}} (D_{0t} \Gamma_0 D_{0t})^{-1} \frac{\partial \varepsilon_{0t}}{\partial \tilde{\lambda}'} \right] + \frac{1}{4} E \left[w_t \frac{\partial H'_{0t}}{\partial \tilde{\lambda}} D_{0t}^{-2} C D_{0t}^{-2} \frac{\partial H_{0t}}{\partial \tilde{\lambda}'} \right],$$

$$\Sigma_{\tilde{\lambda}\sigma 0} = \frac{1}{2} E \left[w_t \frac{\partial H'_{0t}}{\partial \tilde{\lambda}} D_{0t}^{-2} \right] C_1 P,$$

$\partial \varepsilon_{0t} / \partial \tilde{\lambda}' = \partial \varepsilon_t / \partial \tilde{\lambda}'|_{\lambda_0}$, $\partial H_{0t} / \partial \tilde{\lambda}' = \partial H_t / \partial \tilde{\lambda}'|_{\lambda_0}$, $P = (I_m \otimes \Gamma_0^{-1}) \mathcal{K}$, $C_1 =$

(C_{11}, \dots, C_{1m}) , C_{1i} is an $m \times m$ matrix with the (i, i) th component being

1 and the other components zero, $\mathcal{K} = \partial \text{vec}(\Gamma) / \partial \sigma'$ is a constant matrix,

and $C = \Gamma_0^{-1} \odot \Gamma_0 + I_m$, with $A \odot B = (a_{ij} b_{ij})$ for two matrices $A = (a_{ij})$

and $B = (b_{ij})$. In practice, Σ_0 and Ω_0 can be evaluated by

$$\hat{\Sigma}_{sn} = - \begin{pmatrix} \hat{\Sigma}_{\hat{\lambda}_{sn}} & \hat{\Sigma}_{\hat{\lambda}_{sn}\sigma_{sn}} \\ \hat{\Sigma}'_{\hat{\lambda}_{sn}\sigma_{sn}} & \frac{1}{2}\omega_t^\epsilon \hat{P}'_{sn} \hat{P}_{sn} \end{pmatrix}, \quad \hat{\Omega}_{sn} = \frac{1}{n} \sum_{t=1}^n \left[\omega_t^{\epsilon 2} \frac{\partial l_t^\epsilon(\lambda)}{\partial \lambda} \frac{\partial l_t^\epsilon(\lambda)}{\partial \lambda'} \right]_{\hat{\lambda}_{sn}}, \quad (3.6)$$

where $\hat{\Gamma}_{sn} = \Gamma|_{\hat{\lambda}_{sn}}$,

$$\hat{\Sigma}_{\hat{\lambda}_{sn}} = \frac{1}{n} \sum_{t=1}^n \left[\omega_t^\epsilon \frac{\partial \varepsilon_t^{\epsilon'}}{\partial \tilde{\lambda}} (D_t^\epsilon \Gamma D_t^\epsilon)^{-1} \frac{\partial \varepsilon_t^\epsilon}{\partial \tilde{\lambda}'} \right]_{\hat{\lambda}_{sn}} + \frac{1}{4n} \sum_{t=1}^n \left[\omega_t^\epsilon \frac{\partial H_t^{\epsilon'}}{\partial \tilde{\lambda}} D_t^{\epsilon-2} \hat{C}_{sn} D_t^{\epsilon-2} \frac{\partial H_t^\epsilon}{\partial \tilde{\lambda}'} \right]_{\hat{\lambda}_{sn}},$$

$$\hat{\Sigma}_{\hat{\lambda}_{sn}\sigma_{sn}} = \frac{1}{2n} \sum_{t=1}^n \left[\omega_t^\epsilon \frac{\partial H_t^{\epsilon'}}{\partial \tilde{\lambda}} D_t^{\epsilon-2} \right]_{\hat{\lambda}_{sn}} C_1 \hat{P}_{sn}, \quad \hat{P}_{sn} = (I_m \otimes \hat{\Gamma}_{sn}^{-1}) \mathcal{K},$$

$$\hat{C}_{sn} = \hat{\Gamma}_{sn}^{-1} \odot \hat{\Gamma}_{sn} + I_m.$$

Similar to (S.62) in Supplementary Material, we can show that $\hat{\Sigma}_{sn} = \Sigma_0 + o_p(1)$ and $\hat{\Omega}_{sn} = \Omega_0 + o_p(1)$. Using these results, we can consider the Wald test statistic, denoted by W_{sn} , for the following hypothesis:

$$H_0 : \Upsilon(\lambda) = \Upsilon(\lambda_0) \text{ against } H_1 : \Upsilon(\lambda) \neq \Upsilon(\lambda_0),$$

where $\Upsilon(\lambda)$ is a function from $\Theta \rightarrow R^{p_1}$ with a continuous derivative function $\Upsilon^1(\lambda)$. In particular, when $\Upsilon(\lambda) = \Pi\lambda$ and $\Upsilon(\lambda_0) = \lambda_{10}$, it reduces to the usual p_1 linear hypothesis, where Π is a constant matrix. By Theorem 3.1, we have a corollary as follows.

Corollary 3.2 *If assumptions of Theorem 3.1 hold and $E \|\eta_{0t}\|^4 < \infty$, then it follows that*

$$W_{sn} = n(\Upsilon(\hat{\lambda}_{sn}) - \Upsilon(\lambda_0))' (\Upsilon^1(\hat{\lambda}_{sn}) \hat{\Sigma}_{sn}^{-1} \hat{\Omega}_{sn} \hat{\Sigma}_{sn}^{-1} \Upsilon^1(\hat{\lambda}_{sn}))^{-1} (\Upsilon(\hat{\lambda}_{sn}) - \Upsilon(\lambda_0)) \xrightarrow{\mathcal{L}} \chi_{p_1}^2, \quad (3.7)$$

under H_0 , where $\hat{\Sigma}_{sn}$ and $\hat{\Omega}_{sn}$ are defined in (3.6).

Note that λ_0 is assumed to be an interior point in Θ . W_{sn} cannot be applied for testing whether the coefficients in the GARCH part are zero since zero is on the boundary of parametric space. To test these zero parameters, we expect that the approach in Francq and Zakoïan (2009) can be extended for VGARCH models and the related research are in progress.

Similar to Ling (2007) and Zhu and Ling (2011), many weights satisfy Assumptions 3.1-3.2. When tail index $2\iota \in (1, \infty)$, the weight can be chosen as

$$\omega_t = \left(\max \left\{ 1, C^{-1} \sum_{k=1}^{\infty} \frac{1}{k^9} \|Y_{t-k}\| I\{\|Y_{t-k}\| > C\} \right\} \right)^{-4}, \quad (3.8)$$

for some $C > 0$. When $q = s = 0$ (AR-ARCH model), for any $2\iota > 0$, the weight can be selected as

$$\omega_t = \left(\max \left\{ 1, C^{-1} \sum_{k=1}^{p+r} \frac{1}{k^9} \|Y_{t-k}\| I\{\|Y_{t-k}\| > C\} \right\} \right)^{-4}. \quad (3.9)$$

When $2\iota \in (0, 1]$ and $q > 0$ or $s > 0$, the weight ω_t needs to be modified as follows:

$$\omega_t = \left(\max \left\{ 1, C^{-1} \sum_{k=1}^{\infty} \frac{1}{k^{1+8/\iota_1}} \|Y_{t-k}\| I\{\|Y_{t-k}\| > C\} \right\} \right)^{-4}, \quad (3.10)$$

where ι_1 is any constant such that $0 < \iota_1 < \iota$. These weights satisfy Assumptions 3.1-3.2. They can downweight the quasi-information matrices with large points (in absolute value) such that the magnitudes of elements

are not larger than C^4 , but take full advantage of all matrices without these points. The sufficient and necessary condition of Assumption 2.5 is similar to Theorem 2.1 of Ling (2007). The choice of weight function ω_t depends on tail index 2ι . Theoretical index of the underlying vector GARCH process is the same as the tail index of Y_t , see Remark 3.4 of Zhang and Ling (2026). We can use Hill's estimator to estimate the tail index of $\{Y_t\}$ which may provide some guidelines for choice 2ι . Simulation results show that C can be chosen by the 90% or 95%-quantile of data $\{\|Y_1\|, \dots, \|Y_n\|\}$.

4. Local QMLE for VARMA-GARCH/IGARCH

The self-weighted QMLE in Section 3 may not be efficient, although it can reduce the moment condition of ε_t . Using $\hat{\lambda}_{sn}$ in Theorem 3.1 as an initial estimator of λ_0 , we can obtain the local QMLE by the following one-step iteration:

$$\hat{\lambda}_n = \hat{\lambda}_{sn} - \left[\sum_{t=1}^n \frac{\partial^2 l_t^e(\hat{\lambda}_{sn})}{\partial \lambda \partial \lambda'} \right]^{-1} \sum_{t=1}^n \frac{\partial l_t^e(\hat{\lambda}_{sn})}{\partial \lambda}. \quad (4.1)$$

We need one more assumption to get the asymptotic normality of $\hat{\lambda}_n$:

Assumption 4.1 *All the roots of $|I_m - \sum_{i=1}^r A_i L^i - \sum_{j=1}^s B_j L^j| = 0$ lie outside the unit circle or $|I_m - \sum_{i=1}^r A_i L^i - \sum_{j=1}^s B_j L^j| = 0$ have some roots on the unit circle and others lie outside the unit circle.*

Ling and McAleer (2003) showed that all the roots of $|I_m - \sum_{i=1}^r A_i L^i -$

$\sum_{j=1}^s B_j L^j| = 0$ lie outside the unit circle is the sufficient condition for $E \|\varepsilon_t\|^2 < \infty$. When $|I_m - \sum_{i=1}^r A_i L^i - \sum_{j=1}^s B_j L^j| = 0$ have some roots on the unit circle and others lie outside the unit circle, model (2.2) is called the vector IGARCH (VIGARCH) model and Zhang and Ling (2026) showed that $E \|\varepsilon_t\|^{2-\tau} < \infty$ for any $\tau \in (0, 1)$.

Theorem 4.1 *Suppose that Assumptions 2.1-2.4, 3.1-3.2, and 4.1 hold. If $E \|\eta_{0t}\|^4 < \infty$, and $\hat{\lambda}_n$ is obtained by (4.1), then*

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) \xrightarrow{\mathcal{L}} N(0, \Sigma^{-1} \Omega \Sigma^{-1}),$$

where $\Sigma = E \left[\frac{\partial^2 l_t(\lambda_0)}{\partial \lambda \partial \lambda'} \right]$ and $\Omega = E \left[\frac{\partial l_t(\lambda_0)}{\partial \lambda} \frac{\partial l_t(\lambda_0)}{\partial \lambda'} \right]$.

The comparison of the efficiency of the local QMLE and the self-weighted QMLE in Theorem 3.1 is similar to Ling (2007) and Zhu and Ling (2011), and we can show that the local QMLE is more efficient than the self-weighted QMLE.

Let $\hat{\Sigma}_n$ and $\hat{\Omega}_n$ be defined as $\hat{\Sigma}_{sn}$ and $\hat{\Omega}_{sn}$ in (3.6) with $w_t^\varepsilon = 1$ and $\hat{\lambda}_{sn}$ replaced by $\hat{\lambda}_n$, we can estimate Σ and Ω by $\hat{\Sigma}_n$ and $\hat{\Omega}_n$, respectively. Using Lemmas A.5-A.7 in supplementary material, we can show that $\hat{\Sigma}_n = \Sigma + o_p(1)$ and $\hat{\Omega}_n = \Omega + o_p(1)$. Using this result, we can test the null hypothesis H_0 , as in Corollary 3.2, by the Wald test and have the following corollary:

Corollary 4.2 *Under assumptions of Theorem 4.1 and under H_0 , it follows*

that

$$W_n = n(\Upsilon(\hat{\lambda}_n) - \Upsilon(\lambda_0))' (\Upsilon^1(\hat{\lambda}_n) \hat{\Sigma}_n^{-1} \hat{\Omega}_n \hat{\Sigma}_n^{-1} \Upsilon^1'(\hat{\lambda}_n))^{-1} (\Upsilon(\hat{\lambda}_n) - \Upsilon(\lambda_0)) \xrightarrow{\mathcal{L}} \chi_{p_1}^2. \quad (4.2)$$

It is interesting to test if model (2.2) is a VIGARCH process. Denote $\Upsilon(\lambda) = |I_m - \sum_{i=1}^r A_i - \sum_{j=1}^s B_j|$. Note that model (2.2) is a VIGARCH process with a unit root 1 if and only if $\Upsilon(\lambda_0) = 0$. Thus, we consider the following hypothesis:

$$H_0 : \Upsilon(\lambda) = 0 \text{ against } H_1 : \Upsilon(\lambda) \neq 0. \quad (4.3)$$

In this case, denote $\Lambda = I_m - \sum_{i=1}^r A_i - \sum_{j=1}^s B_j$. $\Upsilon^1(\lambda) = \text{vec}\{\text{adj}(\Lambda')\}'$ ($0_{m^2 \times (2m+(p+q)m^2)}, -I_{m^2 \times m^2}, \dots, -I_{m^2 \times m^2}, 0_{m^2 \times (m(m-1)/2)}$), where the number of $-I_{m^2 \times m^2}$ is $(r+s)$, and the test statistics in (3.7) and (4.2) would be

$$W_{sn} = n (\det(\Lambda))_{\hat{\lambda}_{sn}}^2 \left(\Upsilon^1(\hat{\lambda}_{sn}) \hat{\Sigma}_{sn}^{-1} \hat{\Omega}_{sn} \hat{\Sigma}_{sn}^{-1} \Upsilon^1'(\hat{\lambda}_{sn}) \right)^{-1} \xrightarrow{\mathcal{L}} \chi_1^2, \quad (4.4)$$

and

$$W_n = n (\det(\Lambda))_{\hat{\lambda}_n}^2 \left(\Upsilon^1(\hat{\lambda}_n) \hat{\Sigma}_n^{-1} \hat{\Omega}_n \hat{\Sigma}_n^{-1} \Upsilon^1'(\hat{\lambda}_n) \right)^{-1} \xrightarrow{\mathcal{L}} \chi_1^2, \quad (4.5)$$

respectively, as $n \rightarrow \infty$.

5. Model Diagnostic Checking

Model diagnostic checking is an important step in applications, see Li (2004). Denote

$$\varsigma_t(\lambda) = \Gamma^{-1/2}\eta_t(\lambda), \quad (5.1)$$

where the square root needs to be understood in the sense of the Cholesky factorization, that is, $\Gamma^{1/2}(\Gamma^{1/2})' = \Gamma$. Then, $\varsigma_{0t} \equiv \varsigma_t(\lambda_0) = \Gamma_0^{-1/2}\eta_{0t}$ and $\{\varsigma_{0t}\}$ is a sequence of i.i.d. random vectors with mean zero and covariance I_m . As in Ling and Li (1997a), we assume that $E(\varsigma_{0,it}^3) = 0$ and $\varsigma_{0,it}$ have the same finite fourth-order moment, $i = 1, \dots, m$; $\varsigma_{0,it}$ and $\varsigma_{0,jt}$ are mutually uncorrelated up to the fourth order for $i \neq j, i, j = 1, \dots, m$, where $\varsigma_{0,it}$ is the i th component of ς_{0t} . The main problem is that Y_t only has a fractional moment and hence the statistic in Ling and Li (1997a) cannot be directly used.

To solve this problem, we define a weighted quadratic form of standardized residuals as follows:

$$\varsigma_{\omega t}(\lambda) \equiv \omega_t[\varsigma_t'(\lambda)\varsigma_t(\lambda) - m] = \omega_t[\varepsilon_t'(\varphi)(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\varepsilon_t(\varphi) - m]. \quad (5.2)$$

If the models (2.1) and (2.2) are correct, then the autocorrelation function of $\varsigma_{\omega t}(\lambda_0)$

$$R_l = \frac{E\varsigma_{\omega t}(\lambda_0)\varsigma_{\omega t-l}(\lambda_0)}{E\varsigma_{\omega t}^2(\lambda_0)} = 0, \quad (5.3)$$

for all $l \geq 1$. Let $\varsigma_{\omega t}^e(\lambda)$ be defined in the same way as $\varsigma_{\omega t}$ with ω_t , $\varepsilon_t(\varphi)$, and $D_t(\lambda)$ being replaced by ω_t^e , $\varepsilon_t^e(\varphi)$, and $D_t^e(\lambda)$. Thus R_l can be estimated by its sample autocorrelation function

$$\hat{R}_{sn,l} = \frac{\sum_{t=l+1}^n [\varsigma_{\omega t}^e(\hat{\lambda}_{sn}) - \bar{\varsigma}_n] [\varsigma_{\omega t-l}^e(\hat{\lambda}_{sn}) - \bar{\varsigma}_n]}{\sum_{t=1}^n [\varsigma_{\omega t}^e(\hat{\lambda}_{sn}) - \bar{\varsigma}_n]^2}, \quad (5.4)$$

where $\bar{\varsigma}_n = \sum_{t=1}^n \varsigma_{\omega t}^e(\hat{\lambda}_{sn})/n$. Denote $\varsigma_{0\omega,t} = \varsigma_{\omega t}(\lambda_0)$. If models (2.1)-(2.2) are correctly specified, then $\hat{R}_{sn,l}$ should be very small. Let $\hat{R}_{sn} = (\hat{R}_{sn,1}, \dots, \hat{R}_{sn,M})'$ for a given positive number M . The following theorem gives its limiting distribution.

Theorem 5.1 *If assumptions of Theorem 3.1 are satisfied and $E \|\eta_{0t}\|^4 < \infty$, then*

$$\sqrt{n} \hat{R}_{sn} \xrightarrow{\mathcal{L}} N\left(0, \left(mkE\omega_t^2\right)^{-2} A\right), \quad \text{as } n \rightarrow \infty, \quad (5.5)$$

where $k = E(\varsigma_{0,it}^4) - 1$, $A = VZV'$, $V = (I_M, X\Sigma_0^{-1})$, Σ_0 is defined as in

Theorem 3.1, $X = (X_1, \dots, X_M)'$, $Z = E(Z_t Z_t')$,

$$X_l = E \left[\omega_t \frac{\partial \text{vec}(D_{0t} \Gamma_0 D_{0t})'}{\partial \lambda} \text{vec}\{(D_{0t} \Gamma_0 D_{0t})^{-1} \varsigma_{0\omega,t-l}\} \right],$$

$$Z_t = \left(\varsigma_{0\omega t} \varsigma_{0\omega,t-1}, \dots, \varsigma_{0\omega t} \varsigma_{0\omega,t-M}, \omega_t \frac{\partial l_t(\lambda_0)}{\partial \lambda'} \right)'.$$

Using Theorem 5.1, we can construct the statistic as follows:

$$Q_{sn}(M) = n \left(mkE\omega_t^2 \right)^2 \hat{R}'_{sn} A^{-1} \hat{R}_{sn}, \quad (5.6)$$

and it is asymptotically $\chi^2(M)$ distributed as $n \rightarrow \infty$. $Q_{sn}(M)$ is used as a portmanteau statistic for checking adequacy of models (2.1)-(2.2). In practice, X , Σ_0 , and Z can be replaced respectively by the corresponding sample averages. The constant $(mkE\omega_t^2)$ can be replaced by $\sum_{t=1}^n [\varsigma_{\omega t}^\epsilon(\hat{\lambda}_{sn}) - \bar{\varsigma}_n] ^2 / n$.

When model (2.2) has a finite second moment or is an VIGARCH process, using local QMLE $\hat{\lambda}_n$, the portmanteau test in Ling and Li (1997a) can be applied for checking the adequacy of model (2.1)-(2.2). Specifically, let $\varsigma_t^\epsilon(\lambda)$ be defined in the same way as $\varsigma_t(\lambda)$ with $\varepsilon_t(\varphi)$ and $D_t(\lambda)$ being replaced by $\varepsilon_t^\epsilon(\varphi)$ and $D_t^\epsilon(\lambda)$. The lag l sum of square (standardized) residual sample autocorrelations can be defined as

$$\hat{R}_{n,l} = \frac{\sum_{t=l+1}^n [\varsigma_t^{\epsilon'}(\hat{\lambda}_n)\varsigma_t^\epsilon(\hat{\lambda}_n) - m] [\varsigma_{t-l}^{\epsilon'}(\hat{\lambda}_n)\varsigma_{t-l}^\epsilon(\hat{\lambda}_n) - m]}{\sum_{t=1}^n [\varsigma_t^{\epsilon'}(\hat{\lambda}_n)\varsigma_t^\epsilon(\hat{\lambda}_n) - m]^2}. \quad (5.7)$$

Let $\hat{R}_n = (\hat{R}_{n,1}, \dots, \hat{R}_{n,M})'$. Using Lemmas A.2-A.3 and a similar argument as Theorem in Ling and Li (1997a), we can show that the statistic

$$Q_n(M) = n\hat{R}_n' B^{-1} \hat{R}_n \quad (5.8)$$

will be asymptotically $\chi^2(M)$ distributed, where

$$B = I_M + X_1(k\Sigma^{-1} + \Sigma^{-1}\Omega\Sigma^{-1})X_1'/(mk)^2,$$

$$X_{1l} = E\left\{\frac{\partial \text{vec}(D_{0t}\Gamma_0 D_{0t}')}{\partial \lambda} \text{vec}\left[(D_{0t}\Gamma_0 D_{0t}')^{-1}(\varsigma_{0,t-l}^\epsilon \varsigma_{0,t-l} - m)\right]\right\},$$

$X_1 = (X_{11}, \dots, X_{1M})'$, and Σ and Ω are defined in Theorem 4.1. In practice,

X_1 , Σ , and Ω can be replaced respectively by the corresponding sample averages. The constant mk can be replaced by $\sum_{t=1}^n [\zeta_t^{\epsilon'}(\hat{\lambda}_n)\zeta_t^\epsilon(\hat{\lambda}_n) - m]^2/n$.

6. Simulation

In this section, we conduct simulation experiments in a bivariate case to assess the performance of the self-weighted QMLE ($\hat{\lambda}_{sn}$) and the local QMLE ($\hat{\lambda}_n$) in the finite sample. We generate 1000 replications of sample size $n = 1000, 5000, 10000$ by the following VAR(1) - GARCH(1,1) model:

$$Y_t - \mu = \Phi(Y_{t-1} - \mu) + \varepsilon_t, \quad (6.1)$$

$$\varepsilon_t = D_t \eta_t, \quad H_t = W + A\varepsilon_{t-1} + BH_{t-1}, \quad (6.2)$$

with a full matrix A and a diagonal matrix B . The innovation η_t is a sequence of i.i.d. multivariate Gaussian distribution and standardized t_5 distribution, respectively, with mean 0 and covariance matrix Γ . The parameter matrices in the ARMA part are set to be

$$\mu = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \text{ and } \Phi = \begin{pmatrix} 0.5 & 0.05 \\ 0.05 & 0.5 \end{pmatrix}, \quad (6.3)$$

while the parameter matrices in the GARCH part are set to be

$$W = \begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix}, A = \begin{pmatrix} 0.225 & 0.2 \\ 0.2 & 0.225 \end{pmatrix}, B = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix} \text{ and } \sigma_{21} = 0.3, \quad (6.4)$$

for the case $E \|\varepsilon_t\|^2 < \infty$,

$$W = \begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix}, A = \begin{pmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{pmatrix}, B = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix} \text{ and } \sigma_{21} = 0.3, \quad (6.5)$$

for the VIGARCH case,

$$W = \begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix}, A = \begin{pmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{pmatrix}, B = \begin{pmatrix} 0.55 & 0 \\ 0 & 0.55 \end{pmatrix} \text{ and } \sigma_{21} = 0.3, \quad (6.6)$$

for the case $E \|\varepsilon_t\|^{2\iota} = \infty$ with $\iota < 1$.

Tables 1-6 list the sample biases (Bias), the asymptotic standard deviations (AD), and the sample standard deviations (SD) of $\hat{\lambda}_{sn}$ (SQMLE) and $\hat{\lambda}_n$ (LQMLE). ADs are the estimated asymptotic variance in Theorem 3.1 and 4.1. The weight ω_t is chosen as in (3.8) with C being the 90%-quantile of data $\{\|Y_1\|, \dots, \|Y_n\|\}$. Considering that the asymptotic normality of the local QMLE requires finite variance and IGARCH case, we hereafter exclude all simulation results concerning the LQMLE (including its AD and SD) for case where $E \|\varepsilon_t\|^{2\iota} = \infty$ with $\iota < 1$. From Tables 1-6, we find that: (i) in all cases, the AD and SD of both estimators are close to each other, which indicates our methods are reliable; (ii) the local QMLE has smaller AD and SD than those of self-weighted QMLE; (iii) Bias, AD, and SD for both estimators become smaller as sample size n increases.

Next, we assess the finite sample performance of the Wald tests W_{sn} and W_n as in (3.7) and (4.2). We generate 1000 replications of sample

Table 1: Bias and standard deviations of estimators when $E \|\varepsilon_t\|^2 < \infty$ and

$$\eta_t \sim \mathcal{N}(0, \Gamma)$$

parameter	Estimator	n=1000			n=5000			n=10000		
		Bias	AD	SD	Bias	AD	SD	Bias	AD	SD
μ_1	SQMLE	-0.0006	0.0355	0.0362	0.0000	0.0159	0.0157	-0.0002	0.0112	0.0111
	LQMLE	-0.0008	0.0353	0.0358	-0.0001	0.0158	0.0156	-0.0003	0.0112	0.0111
μ_2	SQMLE	-0.0011	0.0355	0.0346	-0.0005	0.0159	0.0156	-0.0007	0.0112	0.0107
	LQMLE	-0.0014	0.0352	0.0342	-0.0007	0.0158	0.0155	-0.0007	0.0112	0.0107
Φ_{11}	SQMLE	-0.0024	0.0301	0.0314	0.0003	0.0135	0.0137	0.0003	0.0096	0.0100
	LQMLE	-0.0025	0.0289	0.0300	0.0002	0.0130	0.0131	0.0002	0.0092	0.0095
Φ_{21}	SQMLE	-0.0005	0.0300	0.0309	0.0002	0.0134	0.0134	0.0001	0.0095	0.0096
	LQMLE	-0.0006	0.0288	0.0292	0.0000	0.0129	0.0129	0.0001	0.0091	0.0091
Φ_{12}	SQMLE	0.0019	0.0299	0.0304	0.0002	0.0134	0.0136	-0.0004	0.0095	0.0091
	LQMLE	0.0021	0.0288	0.0290	0.0002	0.0129	0.0131	-0.0004	0.0091	0.0088
Φ_{22}	SQMLE	-0.0027	0.0302	0.0299	-0.0001	0.0135	0.0138	-0.0001	0.0096	0.0096
	LQMLE	-0.0029	0.0290	0.0286	0.0000	0.0130	0.0129	0.0000	0.0092	0.0093
W_1	SQMLE	0.0016	0.0133	0.0140	-0.0001	0.0058	0.0058	0.0001	0.0041	0.0040
	LQMLE	0.0014	0.0131	0.0139	-0.0001	0.0057	0.0057	0.0001	0.0041	0.0039
W_2	SQMLE	0.0021	0.0133	0.0141	0.0003	0.0058	0.0060	0.0001	0.0041	0.0041
	LQMLE	0.0019	0.0131	0.0139	0.0002	0.0057	0.0059	0.0001	0.0041	0.0041
A_{11}	SQMLE	-0.0027	0.0428	0.0441	-0.0011	0.0194	0.0191	-0.0008	0.0137	0.0134
	LQMLE	-0.0028	0.0408	0.0426	-0.0011	0.0185	0.0183	-0.0007	0.0131	0.0130
A_{21}	SQMLE	0.0046	0.0414	0.0438	-0.0006	0.0185	0.0185	0.0001	0.0131	0.0134
	LQMLE	0.0044	0.0395	0.0413	-0.0004	0.0177	0.0176	0.0002	0.0126	0.0129
A_{12}	SQMLE	0.0012	0.0408	0.0434	0.0006	0.0186	0.0180	0.0006	0.0131	0.0133
	LQMLE	0.0013	0.0391	0.0412	0.0007	0.0178	0.0174	0.0006	0.0125	0.0126
A_{22}	SQMLE	-0.0031	0.0430	0.0456	-0.0021	0.0194	0.0191	-0.0001	0.0137	0.0136
	LQMLE	-0.0033	0.0411	0.0434	-0.0019	0.0185	0.0183	-0.0001	0.0131	0.0131
B_{11}	SQMLE	-0.0025	0.0577	0.0606	0.0007	0.0259	0.0252	-0.0003	0.0183	0.0180
	LQMLE	-0.0024	0.0563	0.0607	0.0007	0.0253	0.0249	-0.0003	0.0179	0.0176
B_{22}	SQMLE	-0.0057	0.0582	0.0605	0.0008	0.0259	0.0267	-0.0001	0.0183	0.0184
	LQMLE	-0.0052	0.0569	0.0592	0.0007	0.0253	0.0261	-0.0001	0.0179	0.0183
σ_{21}	SQMLE	0.0008	0.0293	0.0296	0.0000	0.0131	0.0132	0.0002	0.0093	0.0096
	LQMLE	0.0009	0.0287	0.0289	0.0000	0.0129	0.0128	0.0001	0.0091	0.0094

size $n = 1000, 5000, 10000$ from the VAR(1)-GARCH(1,1) models (6.1) and (6.2), where η_t is same as before. When considering the Wald test for the ARMA part, the parameter matrices in the ARMA part are set to be

$$\mu = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \text{ and } \Phi = \begin{pmatrix} 0.5 & \kappa \\ 0.05 & 0.5 \end{pmatrix}, \quad (6.7)$$

and the parameter matrices in the GARCH part are the same as in (6.4) (i.e. the case $E \|\varepsilon_t\|^2 < \infty$), (6.5) (i.e. the VIGARCH case), and (6.6) (i.e. the case $E \|\varepsilon_t\|^{2\iota} = \infty$ with $\iota < 1$). $\kappa = 0, 0.05, 0.1$ in all cases. We

Table 2: Bias and standard deviations of estimators when $E \|\varepsilon_t\|^2 < \infty$ and

$$\eta_t \sim t_5$$

parameter	Estimator	n=1000			n=5000			n=10000		
		Bias	AD	SD	Bias	AD	SD	Bias	AD	SD
μ_1	SQMLE	-0.0004	0.0349	0.0358	-0.0001	0.0164	0.0169	-0.0003	0.0116	0.0114
	LQMLE	-0.0008	0.0348	0.0360	-0.0001	0.0164	0.0170	-0.0003	0.0116	0.0115
μ_2	SQMLE	-0.0008	0.0349	0.0368	0.0004	0.0163	0.0165	0.0000	0.0116	0.0114
	LQMLE	-0.0013	0.0349	0.0369	0.0004	0.0163	0.0165	0.0000	0.0116	0.0114
Φ_{11}	SQMLE	-0.0035	0.0364	0.0379	-0.0010	0.0172	0.0181	-0.0003	0.0123	0.0122
	LQMLE	-0.0042	0.0348	0.0360	-0.0010	0.0165	0.0172	-0.0004	0.0117	0.0118
Φ_{21}	SQMLE	-0.0013	0.0363	0.0391	-0.0007	0.0170	0.0178	-0.0002	0.0122	0.0122
	LQMLE	-0.0019	0.0346	0.0374	-0.0005	0.0162	0.0170	-0.0001	0.0116	0.0116
Φ_{12}	SQMLE	-0.0013	0.0361	0.0373	0.0004	0.0170	0.0182	0.0002	0.0121	0.0121
	LQMLE	-0.0014	0.0343	0.0358	0.0003	0.0162	0.0177	0.0004	0.0115	0.0115
Φ_{22}	SQMLE	-0.0036	0.0367	0.0392	-0.0004	0.0170	0.0176	-0.0006	0.0122	0.0124
	LQMLE	-0.0039	0.0350	0.0368	-0.0007	0.0162	0.0169	-0.0003	0.0117	0.0120
W_1	SQMLE	0.0020	0.0162	0.0182	0.0010	0.0080	0.0087	-0.0001	0.0058	0.0060
	LQMLE	0.0016	0.0160	0.0176	0.0009	0.0079	0.0087	-0.0001	0.0056	0.0058
W_2	SQMLE	0.0022	0.0162	0.0192	0.0006	0.0080	0.0086	0.0006	0.0059	0.0065
	LQMLE	0.0019	0.0160	0.0191	0.0005	0.0079	0.0084	0.0005	0.0058	0.0063
A_{11}	SQMLE	0.0001	0.0678	0.0798	0.0036	0.0356	0.0447	0.0005	0.0251	0.0268
	LQMLE	-0.0002	0.0632	0.0727	0.0030	0.0330	0.0408	0.0003	0.0233	0.0250
A_{21}	SQMLE	0.0099	0.0663	0.0780	0.0031	0.0329	0.0362	0.0026	0.0238	0.0267
	LQMLE	0.0099	0.0614	0.0722	0.0030	0.0307	0.0337	0.0024	0.0221	0.0246
A_{12}	SQMLE	0.0035	0.0635	0.0721	0.0021	0.0323	0.0342	0.0018	0.0238	0.0273
	LQMLE	0.0036	0.0589	0.0679	0.0022	0.0301	0.0317	0.0016	0.0222	0.0254
A_{22}	SQMLE	0.0059	0.0714	0.0966	-0.0010	0.0337	0.0370	0.0001	0.0244	0.0253
	LQMLE	0.0057	0.0664	0.0883	-0.0013	0.0313	0.0339	0.0003	0.0229	0.0239
B_{11}	SQMLE	-0.0095	0.0803	0.0931	-0.0063	0.0405	0.0449	-0.0008	0.0298	0.0321
	LQMLE	-0.0095	0.0774	0.0898	-0.0058	0.0390	0.0432	-0.0008	0.0287	0.0309
B_{22}	SQMLE	-0.0165	0.0807	0.0959	-0.0032	0.0407	0.0443	-0.0034	0.0298	0.0315
	LQMLE	-0.0171	0.0778	0.0957	-0.0028	0.0392	0.0426	-0.0035	0.0287	0.0304
σ_{21}	SQMLE	0.0001	0.0315	0.0329	0.0005	0.0145	0.0150	-0.0001	0.0103	0.0105
	LQMLE	0.0002	0.0309	0.0324	0.0006	0.0141	0.0145	-0.0003	0.0101	0.0103

fit each replication by a VAR(1)-GARCH(1,1) model with self-weighted QMLE and local QMLE, respectively, and then use W_{sn} and W_n to detect the null hypothesis that $\kappa = 0$ in the model (6.1). The significance level is set to be 0.05 in all cases. Table 7 reports the percentages of the empirical size ($\kappa = 0$) and the power of the Wald tests, respectively.

To assess the performance of the Wald tests (4.4) and (4.5) for testing VIGARCH model, we generate data from the VAR(1)-GARCH(1,1) models (6.1) and (6.2) where the parameters as (6.3), (6.4), (6.5) and (6.6), and η_t

Table 3: Bias and standard deviations of estimators for VIGARCH case

and $\eta_t \sim \mathcal{N}(0, \Gamma)$

parameter	Estimator	n=1000			n=5000			n=10000		
		Bias	AD	SD	Bias	AD	SD	Bias	AD	SD
μ_1	SQMLE	-0.0007	0.0435	0.0448	0.0000	0.0193	0.0190	-0.0002	0.0137	0.0136
	LQMLE	-0.0007	0.0434	0.0445	0.0000	0.0193	0.0190	-0.0002	0.0136	0.0136
μ_2	SQMLE	-0.0013	0.0434	0.0422	-0.0007	0.0193	0.0189	-0.0008	0.0137	0.0131
	LQMLE	-0.0015	0.0433	0.0420	-0.0008	0.0193	0.0189	-0.0009	0.0136	0.0131
Φ_{11}	SQMLE	-0.0024	0.0300	0.0311	0.0002	0.0135	0.0135	0.0003	0.0096	0.0099
	LQMLE	-0.0024	0.0287	0.0297	0.0002	0.0129	0.0130	0.0003	0.0091	0.0094
Φ_{21}	SQMLE	-0.0006	0.0293	0.0301	0.0002	0.0131	0.0131	0.0002	0.0093	0.0094
	LQMLE	-0.0006	0.0279	0.0282	0.0000	0.0125	0.0125	0.0001	0.0088	0.0089
Φ_{12}	SQMLE	0.0021	0.0292	0.0293	0.0001	0.0131	0.0132	-0.0004	0.0093	0.0090
	LQMLE	0.0023	0.0279	0.0279	0.0002	0.0125	0.0128	-0.0004	0.0088	0.0085
Φ_{22}	SQMLE	-0.0027	0.0302	0.0293	0.0000	0.0135	0.0136	-0.0001	0.0096	0.0095
	LQMLE	-0.0030	0.0288	0.0282	0.0000	0.0129	0.0127	-0.0001	0.0091	0.0092
W_1	SQMLE	0.0022	0.0147	0.0153	-0.0001	0.0064	0.0061	0.0002	0.0045	0.0044
	LQMLE	0.0021	0.0144	0.0153	-0.0001	0.0062	0.0060	0.0002	0.0044	0.0043
W_2	SQMLE	0.0026	0.0147	0.0157	0.0004	0.0064	0.0067	0.0002	0.0045	0.0044
	LQMLE	0.0025	0.0144	0.0154	0.0003	0.0063	0.0066	0.0002	0.0044	0.0044
A_{11}	SQMLE	-0.0019	0.0466	0.0478	-0.0008	0.0210	0.0208	-0.0007	0.0149	0.0146
	LQMLE	-0.0029	0.0443	0.0457	-0.0011	0.0200	0.0197	-0.0007	0.0142	0.0139
A_{21}	SQMLE	0.0053	0.0394	0.0416	-0.0004	0.0175	0.0172	0.0001	0.0124	0.0127
	LQMLE	0.0043	0.0373	0.0392	-0.0004	0.0167	0.0164	0.0002	0.0118	0.0121
A_{12}	SQMLE	0.0015	0.0389	0.0413	0.0007	0.0176	0.0169	0.0005	0.0124	0.0125
	LQMLE	0.0012	0.0370	0.0392	0.0005	0.0167	0.0162	0.0005	0.0118	0.0118
A_{22}	SQMLE	-0.0022	0.0469	0.0490	-0.0023	0.0210	0.0206	0.0001	0.0149	0.0149
	LQMLE	-0.0033	0.0446	0.0467	-0.0021	0.0200	0.0197	0.0000	0.0142	0.0141
B_{11}	SQMLE	-0.0022	0.0473	0.0489	0.0004	0.0212	0.0200	-0.0005	0.0150	0.0147
	LQMLE	-0.0019	0.0460	0.0485	0.0006	0.0206	0.0196	-0.0004	0.0145	0.0142
B_{22}	SQMLE	-0.0050	0.0474	0.0498	0.0010	0.0212	0.0215	-0.0001	0.0150	0.0149
	LQMLE	-0.0041	0.0461	0.0485	0.0009	0.0206	0.0208	-0.0001	0.0145	0.0147
σ_{21}	SQMLE	0.0011	0.0294	0.0295	0.0002	0.0132	0.0133	0.0002	0.0094	0.0097
	LQMLE	0.0009	0.0287	0.0288	0.0000	0.0129	0.0128	0.0001	0.0091	0.0094

is same as before. The percentages of the empirical size (i.e. the VIGARCH case) and the power (i.e. the case $E \|\varepsilon_t\|^2 < \infty$ and the case $E \|\varepsilon_t\|^{2\iota} = \infty$ with $\iota < 1$) of the Wald tests are reported in Table 8, respectively.

From Tables 7 and 8, we can find that: (i) for the ARMA part, all the sizes are close to the nominal ones, and a little bit less than the nominal ones when $n=10000$; (ii) for testing VIGARCH model, the sizes of W_{sn} and W_n approach the nominal ones as n increases, while the size of W_n is a little larger than the nominal one when $\eta_t \sim N(0, \Gamma)$; (iii) for all tests, the powers

Table 4: Bias and standard deviations of estimators for VIGARCH case

and $\eta_t \sim t_5$

parameter	Estimator	n=1000			n=5000			n=10000		
		Bias	AD	SD	Bias	AD	SD	Bias	AD	SD
μ_1	SQMLE	-0.0002	0.0400	0.0415	0.0000	0.0190	0.0196	-0.0004	0.0135	0.0133
	LQMLE	-0.0004	0.0401	0.0417	-0.0001	0.0190	0.0197	-0.0004	0.0135	0.0133
μ_2	SQMLE	-0.0006	0.0401	0.0424	0.0005	0.0189	0.0192	-0.0001	0.0135	0.0133
	LQMLE	-0.0008	0.0402	0.0424	0.0004	0.0189	0.0192	0.0000	0.0135	0.0133
Φ_{11}	SQMLE	-0.0033	0.0373	0.0390	-0.0012	0.0179	0.0189	-0.0003	0.0128	0.0128
	LQMLE	-0.0038	0.0355	0.0367	-0.0010	0.0170	0.0178	-0.0003	0.0121	0.0122
Φ_{21}	SQMLE	-0.0011	0.0362	0.0391	-0.0005	0.0171	0.0181	-0.0002	0.0122	0.0124
	LQMLE	-0.0014	0.0344	0.0370	-0.0004	0.0162	0.0171	-0.0001	0.0116	0.0117
Φ_{12}	SQMLE	-0.0019	0.0358	0.0370	0.0005	0.0170	0.0185	0.0002	0.0122	0.0122
	LQMLE	-0.0016	0.0339	0.0353	0.0003	0.0161	0.0177	0.0005	0.0115	0.0114
Φ_{22}	SQMLE	-0.0037	0.0375	0.0390	-0.0004	0.0177	0.0182	-0.0007	0.0126	0.0130
	LQMLE	-0.0039	0.0357	0.0365	-0.0007	0.0168	0.0175	-0.0004	0.0120	0.0124
W_1	SQMLE	0.0019	0.0169	0.0185	0.0009	0.0083	0.0090	0.0000	0.0060	0.0062
	LQMLE	0.0017	0.0166	0.0181	0.0008	0.0081	0.0089	-0.0001	0.0058	0.0060
W_2	SQMLE	0.0020	0.0168	0.0197	0.0007	0.0083	0.0089	0.0005	0.0061	0.0066
	LQMLE	0.0018	0.0166	0.0194	0.0007	0.0081	0.0086	0.0005	0.0060	0.0065
A_{11}	SQMLE	0.0004	0.0759	0.0876	0.0038	0.0394	0.0478	0.0007	0.0278	0.0299
	LQMLE	-0.0004	0.0711	0.0809	0.0029	0.0367	0.0440	0.0006	0.0260	0.0278
A_{21}	SQMLE	0.0087	0.0634	0.0731	0.0028	0.0314	0.0343	0.0022	0.0227	0.0261
	LQMLE	0.0079	0.0587	0.0680	0.0028	0.0294	0.0325	0.0019	0.0212	0.0239
A_{12}	SQMLE	0.0025	0.0600	0.0671	0.0019	0.0308	0.0329	0.0017	0.0226	0.0256
	LQMLE	0.0027	0.0560	0.0637	0.0021	0.0287	0.0305	0.0014	0.0211	0.0240
A_{22}	SQMLE	0.0058	0.0784	0.0926	-0.0006	0.0375	0.0417	0.0001	0.0271	0.0283
	LQMLE	0.0045	0.0731	0.0847	-0.0014	0.0350	0.0378	0.0002	0.0255	0.0265
B_{11}	SQMLE	-0.0070	0.0682	0.0769	-0.0051	0.0345	0.0379	-0.0010	0.0254	0.0275
	LQMLE	-0.0077	0.0658	0.0753	-0.0048	0.0332	0.0364	-0.0009	0.0244	0.0262
B_{22}	SQMLE	-0.0132	0.0689	0.0789	-0.0028	0.0344	0.0376	-0.0026	0.0252	0.0262
	LQMLE	-0.0133	0.0667	0.0782	-0.0025	0.0331	0.0357	-0.0025	0.0243	0.0253
σ_{21}	SQMLE	-0.0001	0.0316	0.0329	0.0005	0.0145	0.0151	-0.0002	0.0103	0.0105
	LQMLE	0.0000	0.0309	0.0323	0.0006	0.0141	0.0145	-0.0003	0.0101	0.0103

of all the tests increase as the value of n increases; (iv) for both parts, the powers of Wald tests when $\eta_t \sim \mathcal{N}(0, \Gamma)$ are larger than the powers of Wald tests when $\eta_t \sim t_5$.

To assess the performance of our portmanteau tests $Q_{sn}(M)$ and $Q_n(M)$ in the finite samples, we generate 1000 replications of sample size $n = 1000, 5000, 10000$ from four different models. The null model is model (6.1) and (6.2) with the parameter sets (6.3), (6.4), (6.5) and (6.6), i.e. the case $E \|\varepsilon_t\|^2 < \infty$, the VIGARCH case, and the case $E \|\varepsilon_t\|^{2\iota} = \infty$ with $\iota < 1$,

Table 5: Bias and standard deviations of estimators when $E \|\varepsilon_t\|^{2\iota} = \infty$ and $\eta_t \sim \mathcal{N}(0, \Gamma)$

parameter	Estimator	n=1000			n=5000			n=10000		
		Bias	AD	SD	Bias	AD	SD	Bias	AD	SD
μ_1	SQMLE	-0.0014	0.0674	0.0736	0.0000	0.0288	0.0287	-0.0001	0.0204	0.0203
μ_2	SQMLE	-0.0023	0.0673	0.0662	-0.0005	0.0288	0.0283	-0.0012	0.0204	0.0202
Φ_{11}	SQMLE	-0.0028	0.0300	0.0310	0.0000	0.0135	0.0136	0.0003	0.0096	0.0099
Φ_{21}	SQMLE	-0.0006	0.0292	0.0299	0.0001	0.0131	0.0129	0.0000	0.0093	0.0093
Φ_{12}	SQMLE	0.0018	0.0291	0.0292	0.0001	0.0131	0.0131	-0.0003	0.0093	0.0089
Φ_{22}	SQMLE	-0.0029	0.0301	0.0295	-0.0001	0.0135	0.0132	0.0000	0.0096	0.0095
W_1	SQMLE	0.0039	0.0215	0.0207	0.0003	0.0087	0.0084	0.0006	0.0062	0.0062
W_2	SQMLE	0.0045	0.0216	0.0215	0.0009	0.0087	0.0090	0.0004	0.0062	0.0061
A_{11}	SQMLE	-0.0013	0.0445	0.0456	-0.0004	0.0201	0.0201	-0.0003	0.0142	0.0140
A_{21}	SQMLE	0.0055	0.0374	0.0391	-0.0001	0.0167	0.0162	0.0002	0.0118	0.0120
A_{12}	SQMLE	0.0022	0.0370	0.0386	0.0011	0.0168	0.0163	0.0005	0.0118	0.0118
A_{22}	SQMLE	-0.0022	0.0447	0.0465	-0.0017	0.0201	0.0199	0.0003	0.0142	0.0142
B_{11}	SQMLE	-0.0013	0.0400	0.0408	-0.0002	0.0180	0.0177	-0.0005	0.0127	0.0125
B_{22}	SQMLE	-0.0026	0.0400	0.0408	0.0006	0.0179	0.0176	-0.0002	0.0127	0.0130
σ_{21}	SQMLE	0.0018	0.0296	0.0295	0.0004	0.0133	0.0135	0.0002	0.0094	0.0097

Table 6: Bias and standard deviations of estimators when $E \|\varepsilon_t\|^{2\iota} = \infty$ and $\eta_t \sim t_5$

parameter	Estimator	n=1000			n=5000			n=10000		
		Bias	AD	SD	Bias	AD	SD	Bias	AD	SD
μ_1	SQMLE	-0.0003	0.0508	0.0538	0.0001	0.0240	0.0251	-0.0006	0.0171	0.0169
μ_2	SQMLE	-0.0010	0.0509	0.0546	0.0007	0.0240	0.0242	-0.0002	0.0172	0.0170
Φ_{11}	SQMLE	-0.0037	0.0375	0.0390	-0.0012	0.0180	0.0190	-0.0004	0.0129	0.0129
Φ_{21}	SQMLE	-0.0018	0.0365	0.0400	-0.0005	0.0172	0.0184	-0.0001	0.0123	0.0126
Φ_{12}	SQMLE	-0.0026	0.0360	0.0373	0.0005	0.0171	0.0188	0.0002	0.0123	0.0122
Φ_{22}	SQMLE	-0.0036	0.0378	0.0394	-0.0006	0.0178	0.0183	-0.0007	0.0128	0.0132
W_1	SQMLE	0.0020	0.0197	0.0208	0.0007	0.0096	0.0106	0.0000	0.0069	0.0071
W_2	SQMLE	0.0020	0.0196	0.0216	0.0009	0.0096	0.0100	0.0006	0.0071	0.0075
A_{11}	SQMLE	0.0006	0.0729	0.0836	0.0029	0.0376	0.0442	0.0006	0.0265	0.0285
A_{21}	SQMLE	0.0074	0.0597	0.0704	0.0029	0.0299	0.0326	0.0020	0.0216	0.0251
A_{12}	SQMLE	0.0028	0.0569	0.0627	0.0021	0.0293	0.0313	0.0018	0.0215	0.0244
A_{22}	SQMLE	0.0044	0.0749	0.0857	-0.0005	0.0357	0.0394	0.0002	0.0259	0.0269
B_{11}	SQMLE	-0.0053	0.0586	0.0644	-0.0035	0.0301	0.0323	-0.0011	0.0219	0.0235
B_{22}	SQMLE	-0.0091	0.0599	0.0673	-0.0023	0.0296	0.0323	-0.0020	0.0216	0.0222
σ_{21}	SQMLE	0.0001	0.0318	0.0330	0.0006	0.0146	0.0150	-0.0003	0.0104	0.0106

denoted by Model 1. The alternative models are as follows:

Model 2 : $Y_t - \mu = \Phi_1(Y_{t-1} - \mu) + \Phi_2(Y_{t-2} - \mu) + \varepsilon_t,$

Model 3 : $Y_t - \mu = \Phi(Y_{t-1} - \mu) + \varepsilon_t + \Theta\varepsilon_{t-1},$

where ε_t is generated by model (6.2) with parameter matrices in (6.4),

Table 7: Size and power of the Wald tests for ARMA part

		$E \ \varepsilon_t\ ^2 < \infty$						VIGARCH						$E \ \varepsilon_t\ ^{2\iota} = \infty$		
		$\kappa = 0.0$		$\kappa = 0.05$		$\kappa = 0.1$		$\kappa = 0.0$		$\kappa = 0.05$		$\kappa = 0.1$		$\kappa = 0.0$	$\kappa = 0.05$	$\kappa = 0.1$
η_t	n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_{sn}	W_{sn}
$\mathcal{N}(0, \Gamma)$	1000	4.6	5.8	41.2	44.6	92.1	93.8	4.8	5.3	42.7	46.7	93.3	95.1	5	42.2	93.5
	5000	4.5	5.4	96.2	97.2	100	100	5.6	5.4	96.9	97.3	100	100	5	96.6	100
	10000	4	4.3	99.9	100	100	100	4.6	4	100	100	100	100	4	100	100
t_5	1000	5.4	5.4	30.9	32.8	77.7	80.8	5.2	5.8	30.9	33.1	78.3	81.7	5.2	28.9	77.6
	5000	5.1	5.8	83.3	84.6	99.8	99.8	5.1	5.7	84	84.7	99.8	99.8	5.4	84	99.8
	10000	3.7	4.1	97.7	98.4	99.9	100	4.2	4.3	97.5	98.3	99.9	99.9	3.4	97.6	99.9

Table 8: Size and power of Wald tests for VIGARCH errors

		Size with VIGARCH				Power with $E \ \varepsilon_t\ ^2 < \infty$				Power with $E \ \varepsilon_t\ ^{2\iota} = \infty$	
		$\mathcal{N}(0, \Gamma)$		t_5		$\mathcal{N}(0, \Gamma)$		t_5		$\mathcal{N}(0, \Gamma)$	t_5
n		W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_{sn}
1000		4.9	6.4	3.1	5.2	72	75.3	30.7	33	54.3	11.2
5000		4.3	5.7	5.1	7.2	100	100	84.5	86.4	99.6	64.8
10000		5.1	7	4.7	6.2	100	100	96.5	97.3	100	92.9

(6.5) and (6.6), i.e. the case $E \|\varepsilon_t\|^2 < \infty$, the VIGARCH case, and the case $E \|\varepsilon_t\|^{2\iota} = \infty$ with $\iota < 1$. The parameters in Model 2 are set to be

$$\mu = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \Phi_1 = \begin{pmatrix} 0.2 & 0.05 \\ 0.05 & 0.2 \end{pmatrix} \text{ and } \Phi_2 = \begin{pmatrix} 0.5 & 0.05 \\ 0.05 & 0.5 \end{pmatrix},$$

and the parameters in Model 3 are set to be

$$\mu = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \Phi = \begin{pmatrix} 0.2 & 0.05 \\ 0.05 & 0.2 \end{pmatrix} \text{ and } \Theta = \begin{pmatrix} 0.7 & 0.05 \\ 0.05 & 0.7 \end{pmatrix}.$$

Model 4: $Y_t - \mu = \Phi(Y_{t-1} - \mu) + \varepsilon_t,$

$$\varepsilon_t = D_t \eta_t, H_t = W + A_1 \vec{\varepsilon}_{t-1} + A_2 \vec{\varepsilon}_{t-2} + B H_{t-1}.$$

where the parameter matrices in the ARMA part are the same as (6.3),

with the parameter matrices in the GARCH part are set to be

$$W = \begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix}, A_1 = \begin{pmatrix} 0.2 & 0.05 \\ 0.05 & 0.2 \end{pmatrix}, A_2 = \begin{pmatrix} 0.2 & 0.05 \\ 0.05 & 0.2 \end{pmatrix}, B = \begin{pmatrix} 0.4 & 0.05 \\ 0.05 & 0.4 \end{pmatrix}, \quad (6.8)$$

for the case $E \|\varepsilon_t\|^2 < \infty$,

$$W = \begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix}, A_1 = \begin{pmatrix} 0.25 & 0.05 \\ 0.05 & 0.25 \end{pmatrix}, A_2 = \begin{pmatrix} 0.2 & 0.05 \\ 0.05 & 0.2 \end{pmatrix}, B = \begin{pmatrix} 0.4 & 0.05 \\ 0.05 & 0.4 \end{pmatrix}, \quad (6.9)$$

for the VIGARCH case,

$$W = \begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix}, A_1 = \begin{pmatrix} 0.25 & 0.05 \\ 0.05 & 0.25 \end{pmatrix}, A_2 = \begin{pmatrix} 0.25 & 0.05 \\ 0.05 & 0.25 \end{pmatrix}, B = \begin{pmatrix} 0.4 & 0.05 \\ 0.05 & 0.4 \end{pmatrix}, \quad (6.10)$$

for the case $E \|\varepsilon_t\|^{2\iota} = \infty$ with $\iota < 1$, and $\sigma_{21} = 0.3$ for all cases, η_t is the same as that of model (6.2).

All simulated data from Models 1, 2, 3, and 4 are fitted by VAR(1)-GARCH(1,1) model. In the simulation, M is taken to be 6 for all cases. The results are summarized in Table 9, and the sizes correspond to the Model 1. From the table, we can show that: (i) the sizes of Q_{sn} and Q_n are a little larger than their nominal ones and close to the nominal ones as sample size n increase, except for some cases; (ii) all the powers increase when n increases; (iii) the powers when $\eta \sim \mathcal{N}(0, \Gamma)$ are greater than the powers when $\eta_t \sim t_5$.

In summary, our simulation results in this section show that our esti-

Table 9: Size and power of portmanteau tests

		$E \ \varepsilon_t\ ^2 < \infty$				<i>VIGARCH</i>				$E \ \varepsilon_t\ ^{2l} = \infty$	
		$\mathcal{N}(0, \Gamma)$		t_5		$\mathcal{N}(0, \Gamma)$		t_5		$\mathcal{N}(0, \Gamma)$	t_5
	n	Q_{sn}	Q_n	Q_{sn}	Q_n	Q_{sn}	Q_n	Q_{sn}	Q_n	Q_{sn}	Q_{sn}
Model 1	1000	6.9	7.3	9.7	5.6	6	6.7	7.5	5.4	6.6	7.2
	5000	4.9	5.2	9.4	5.5	5.1	5.5	8.1	5.4	5.6	7.4
	10000	5.7	6	7.5	6.1	5.8	5.8	6.3	5.8	6.2	6.5
Model 2	1000	50.8	57.2	26.2	25.4	40.9	47.7	23.1	22.4	36	20.7
	5000	99.7	100	71.8	69.5	98.9	99.6	62	61.1	97.9	55.9
	10000	100	100	94.8	91.4	100	100	90.1	86.57	100	85
Model 3	1000	22.7	21.2	16	10.6	16.9	16.7	13.7	10.3	15.5	12.1
	5000	77.1	67.5	56.1	31.3	66.3	57.3	44.7	26.6	53.5	34.4
	10000	97.4	94.6	81.4	48.6	92.8	88.3	70.8	41.5	84.1	56.1
Model 4	1000	76.2	81.5	33.7	29.7	72.4	77.8	30.9	26.7	86.7	35.7
	5000	100	100	88.7	77.8	100	100	84.5	74.2	100	91.4
	10000	100	100	98.6	94.3	100	100	98.2	92.3	100	99.3

mation and testing procedures are reliable and have a good performance in the finite samples. We also conduct the simulations for VARMA(1,1)-GARCH(1,1) model, and the settings and the simulation results are provided in supplementary material due to the page limit.

7. Application

This section studies a two-dimensional financial time series: the daily return in percentage of the Exxon Mobil Corporation (XOM) and the Chevron Corporation (CVX) close prices, as seen in Figure 1. The XOM and CVX data sets are taken from March 19, 2019 to January 8, 2024 with a length of 1209 observations. Such series have also been studied in the literature, see Horváth and Rice (2015) and Noureldin, Shephard and Sheppard (2014).

First, we estimate the tail index of $\{Y_t\}$ by Hill's estimator $\hat{\alpha}_Y(k)$, where

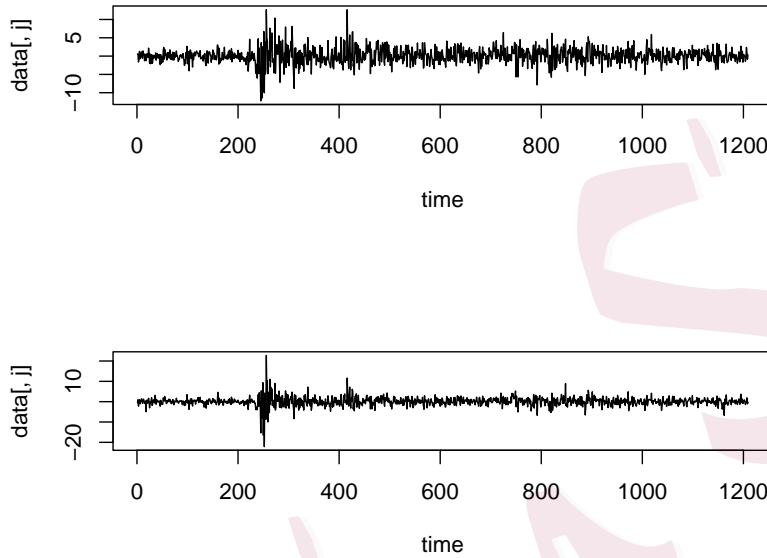


Figure 1: The daily return of XOM/CVX. Data source: Yahoo Finance, <https://finance.yahoo.com/>.

$$\hat{\alpha}_Y(k) = \left[\frac{1}{k} \sum_{i=1}^k \log \frac{\|Y\|_{(n-i+1)}}{\|Y\|_{(n-k)}} \right]^{-1} \quad (6.11)$$

with $\{\|Y\|_{(t)}\}_{t=1}^n$ being the ascending order statistics of $\{\|Y\|_t\}_{t=1}^n$. The plot of $\{\hat{\alpha}_Y(k)\}_{k=1}^{50}$ is given in Figure 2, from which we can see that the tail of $\{Y_t\}$ most likely lies in $[2, 4)$. It seems to be reasonable to use the our procedure to fit a model for this dataset. We now use the t -test to check if the each mean of two time series is zero. It turns out that their p -values

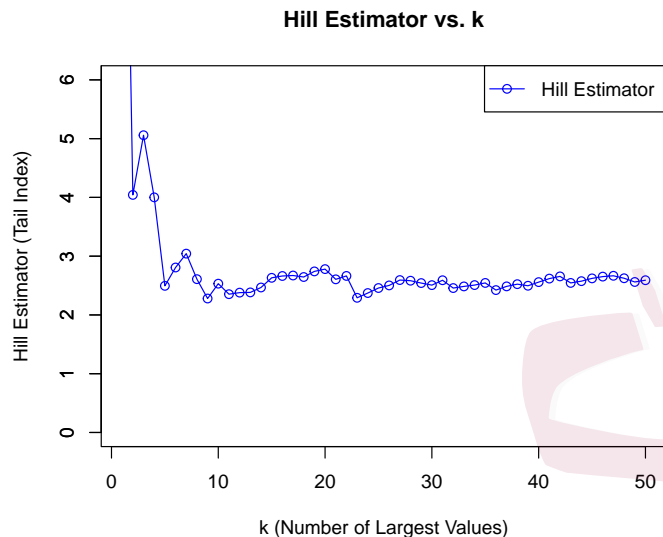


Figure 2: Hill’s estimators Tail index $\{\hat{\alpha}_Y(k)\}$ for $\{y_t\}$.

are 0.3104 and 0.3706, and the null hypothesis cannot be rejected at 5% significance level. We fix Model (2.2) as a GARCH(1,1) process. Based on AIC or BIC, we consider the following VAR(1)-GARCH(1,1) model to fit the dataset:

$$Y_t = \Phi Y_{t-1} + \varepsilon_t,$$

$$\varepsilon_t = D_t \eta_t, \quad H_t = W + A \vec{\varepsilon}_{t-1} + B H_{t-1}.$$

We first apply the self-weighted QMLE with the weight w_t in (3.8) to estimate the parameters and the results are as follows:

$$\begin{aligned} \Phi &= \begin{pmatrix} 0.1662_{(0.0666)} & -0.1748_{(0.0774)} \\ 0.1452_{(0.0648)} & -0.1553_{(0.0737)} \end{pmatrix}, \\ W &= \begin{pmatrix} 0.1610_{(0.0466)} \\ 0.1668_{(0.0537)} \end{pmatrix}, A = \begin{pmatrix} 0.0323_{(0.0214)} & 0.0561_{(0.0268)} \\ 0.0225_{(0.0272)} & 0.0520_{(0.0298)} \end{pmatrix}, \\ B &= \begin{pmatrix} 0.8781_{(0.0288)} & 0 \\ 0 & 0.8679_{(0.0369)} \end{pmatrix}, \sigma_{21} = 0.8703_{(0.0096)}. \end{aligned}$$

The values in the bracket are the standard deviations. We then conduct the portmanteau test $Q_{sn}(6)$ and $Q_{sn}(12)$ to check the model adequacy and their p-value are 0.3338 and 0.7644, respectively, which indicates that the fitted model is adequate at the 5% significance level. Next, we use the self-weighted QMLE as the initial values to obtain the local QMLE. The results are as follows:

$$\begin{aligned} \Phi &= \begin{pmatrix} 0.1360_{(0.0632)} & -0.1504_{(0.0750)} \\ 0.1290_{(0.0632)} & -0.1463_{(0.0742)} \end{pmatrix}, \\ W &= \begin{pmatrix} 0.1840_{(0.0503)} \\ 0.2163_{(0.0716)} \end{pmatrix}, A = \begin{pmatrix} 0.0429_{(0.0233)} & 0.0692_{(0.0258)} \\ 0.0308_{(0.0395)} & 0.1029_{(0.0432)} \end{pmatrix}, \\ B &= \begin{pmatrix} 0.8530_{(0.0278)} & 0 \\ 0 & 0.8025_{(0.0363)} \end{pmatrix}, \sigma_{21} = 0.8678_{(0.0095)}. \end{aligned}$$

The p-values for the portmanteau tests $Q_n(6)$ and $Q_n(12)$ are 0.1941 and 0.3219, respectively, that is, the fitted model is adequate at the 5% significance level. Furthermore, we conduct the Wald test, denoted as W_n , to check whether the coefficients in ARMA part are significantly different from zero, i.e. $H_0 : \Phi_{ij} = 0$ against $H_a : \Phi_{ij} \neq 0$. Almost all of the p-values are less than 0.05, indicating that almost all elements in Φ are significantly different from zero. Finally, we use the Wald test w_n to check whether the model is VIGARCH model and its p-value being 0.0646. The fitted VGARCH model is most likely a VIGARCH model, which has tail index of 2. This is consistent with our Hill estimator in Figure 2.

Supplementary Material

All the proofs and additional simulation results are offered in supplementary material.

Acknowledgements

Shiqing Ling research was partially supported by Hong Kong Research Grants Commission Grants (16303118, 16301620, 16300621, 16500522 and SRFS2223-6S02).

References

- Amendola, A., Niglio, M. and Vitale, C. (2009). Temporal aggregation and closure of VARMA models: some new results. In *Data Analysis and Classification: Proceedings of the 6th Conference of the Classification and Data Analysis Group of the Società Italiana di Statistica*, 435–443. Springer Berlin Heidelberg.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327.
- Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The review of economics and statistics* **72**, 498–505.
- Comte, F. and Lieberman, O. (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* **84**, 61–84.
- Dufour, J. M. and Pelletier, D. (2022). Practical methods for modeling weak VARMA processes: identification, estimation and specification with a macroeconomic application. *Journal of Business & Economic Statistics* **40**, 1140–1152.
- Düker, M. C., Matteson, D. S., Tsay, R. S. and Wilms, I. (2025). Vector AutoRegressive Moving Average Models: A Review. *Wiley Interdisciplinary Reviews: Computational Statistics* **17**, e70009.
- Dunsmuir, W. and Hannan, E. J. (1976). Vector linear time series models. *Advances in Applied Probability* **8**, 339–364.

REFERENCES

- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society* **50**, 987–1007.
- Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of business & economic statistics* **20**, 339–350.
- Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric theory* **11**, 122–150.
- Franco, C. and Zakoian, J. M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**, 605–637.
- Franco, C. and Zakoian, J. M. (2009). Testing the nullity of GARCH coefficients: correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association* **104**, 313–324.
- Franco, C. and Zakoian, J. M. (2010). QML estimation of a class of multivariate GARCH models without moment conditions on the observed process. *unpublished working paper*.
- Franco, C. and Zakoian, J. M. (2012). QML estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory* **28**, 179–206.
- Franco, C. and Zakoian, J. M. (2013). Estimating the marginal law of a time series with applications to heavy-tailed distributions. *Journal of Business & Economic Statistics* **31**, 412–425.

REFERENCES

- Francq, C. and Zakoian, J. M. (2016). Estimating multivariate volatility models equation by equation. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **78**, 613–635.
- Francq, C., Horváth, L. and Zakoian, J. M. (2016). Variance targeting estimation of multivariate GARCH models. *Journal of Financial Econometrics* **14**, 353–382.
- Francq, C. and Zakoian, J. M. (2019). *GARCH models: structure, statistical inference and financial applications*. 2nd Edition. John Wiley & Sons.
- Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* **14**, 70–86.
- Hafner, C. M. and Preminger, A. (2009). Asymptotic theory for a factor GARCH model. *Econometric Theory* **25**, 336–363.
- He, C. and Teräsvirta, T. (2004). An extended constant conditional correlation garch model and its fourth-moment structure. *Econometric Theory* **20**, 904–926.
- Hill, J. B. (2015). Tail index estimation for a filtered dependent time series. *Statistica Sinica* **25**, 609–629.
- Horváth, L. and Rice, G. (2015). Testing for independence between functional time series. *Journal of Econometrics* **189**, 371–382.
- Koedijk, K. G. and Kool, C. J. (1992). Tail estimates of East European exchange rates. *Journal of Business & Economic Statistics* **10**, 83–96.

REFERENCES

-
- Lange, T. (2011). Tail behavior and OLS estimation in AR-GARCH models. *Statistica Sinica* **21**, 1191–1200.
- Li, W.K. (2004). *Diagnostic Checks in Time Series*. Chapman and Hall/CRC.
- Li, W. K., Ling, S. and McAleer, M. (2002). Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys* **16**, 245–269.
- Ling, S. (2007). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics* **140**, 849–873.
- Ling, S. and Li, W. K. (1997a). Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors. *Journal of Time Series Analysis* **18**, 447–464.
- Ling, S. and Li, W. K. (1997b). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of Time Series Analysis* **92**, 1184–1194.
- Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* **19**, 280–310.
- Ling, S. and Peng, L. (2004). Hill's estimator for the tail index of an ARMA model. *Journal of Statistical Planning and Inference* **123**, 279–293.
- Lumsdaine, R. L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH (1, 1) and covariance stationary GARCH (1, 1) models. *Econometrica: Journal of the Econometric Society* **64**, 575–596.

REFERENCES

-
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin-Germany: Springer-Verlag.
- Mainassara, Y. B., Kadmiri, O. and Sausseureau, B. (2022). Estimation of multivariate asymmetric power GARCH models. *Journal of Multivariate Analysis* **192**, 105073.
- McAleer, M., Hoti, S. and Chan, F. (2009). Structure and asymptotic theory for multivariate asymmetric conditional volatility. *Econometric Reviews* **28**, 422–440.
- McAleer, M., Chan, F., Hoti, S. and Lieberman, O. (2008). Generalized autoregressive conditional correlation. *Econometric Theory* **24**, 1554–1583.
- Nakatani, T. and Teräsvirta, T. (2009). Testing for volatility interactions in the constant conditional correlation GARCH model. *The Econometrics Journal* **12**, 147–163.
- Noureldin, D., Shephard, N. and Sheppard, K. (2014). Multivariate rotated ARCH models. *Journal of Econometrics* **179**, 16–30.
- Pedersen, R. S. and Rahbek, A. (2014). Multivariate variance targeting in the BEKK–GARCH model. *The Econometrics Journal* **17**, 24–55.
- Pedersen, R. S. (2016). Targeting estimation of CCC-GARCH models with infinite fourth moments. *Econometric Theory* **32**, 498–531.
- Resnick, S. and Stărică, C. (1995). Consistency of Hill’s estimator for dependent data. *Journal of Applied probability* **32**, 139–167.
- Tse, Y. K. and Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional

REFERENCES

- heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics* **20**, 351–362.
- Wilms, I., Basu, S., Bien, J., and Matteson, D. S. (2023). Sparse identification and estimation of large-scale vector autoregressive moving averages. *Journal of the American Statistical Association* **118**, 571–582.
- Yang, Y. and Ling, S. (2017). Self-weighted LAD-based inference for heavy-tailed threshold autoregressive models. *Journal of Econometrics* **197**, 368–381.
- Zhang, R. and Ling, S. (2015). Asymptotic inference for AR models with heavy-tailed G-GARCH noises. *Econometric Theory* **31**, 880–890.
- Zhang, R. and Ling, S. (2026). Inference for vector AR models with heavy-tailed vector GARCH noises. *Science China Mathematics*, 1–16.
- Zhu, K. and Ling, S. (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA–GARCH/IGARCH models. *Annals of Statistics* **39**, 2131–2163.
- Zhu, K. and Ling, S. (2015). LADE-based inference for ARMA models with unspecified and heavy-tailed heteroscedastic noise. *Journal of the American Statistical Association* **110**, 784–794.

School of Mathematics and Statistics

Changsha University of Science and Technology

REFERENCES

E-mail: mabibicai@ust.hk

School of Mathematics and Statistics

Central China Normal University

E-mail: liumy@ccnu.edu.cn

Department of Mathematics

Hong Kong University of Science and Technology, Hong Kong

E-mail: maling@ust.hk