Commun. Math. Res.
doi: 10.4208/cmr.2023-0005

Vol. **40**, No. 1, pp. 64-101
February 2024

On a Partially Non-Stationary Vector AR Model with Vector GARCH Noises: Estimation and Testing

Chor-yiu Sin[1], Zichuan Mi[2] and Shiqing Ling[3,*]

[1] *Department of Economics, National Tsing Hua University, Hsinchu 30013, Taiwan, P.R. China.*
[2] *School of Statistics, Shanxi University of Finance and Economics, Taiyuan 030006, P.R. China.*
[3] *Department of Mathematics, The Hong Kong University of Science and Technology, Kowloon, Hong Kong, SAR, P.R. China.*

Received 16 January 2023; Accepted 5 June 2023

Abstract. This paper studies a partially nonstationary vector autoregressive (VAR) model with vector GARCH noises. We study the full rank and the reduced rank quasi-maximum likelihood estimators (QMLE) of parameters in the model. It is shown that both QMLE of long-run parameters asymptotically converge to a functional of two correlated vector Brownian motions. Based these, the likelihood ratio (LR) test statistic for cointegration rank is shown to be a functional of the standard Brownian motion and normal vector, asymptotically. As far as we know, our test is new in the literature. The critical values of the LR test are simulated via the Monte Carlo method. The performance of this test in finite samples is examined through Monte Carlo experiments. We apply our approach to an empirical example of three interest rates.

AMS subject classifications: 62M10, 37M10, 91B84

Key words: Vector AR model, cointegration, full rank estimation, vector GARCH process, partially nonstationary, reduced rank estimation.

---

*Corresponding author. *Email address:* `maling@ust.hk` (S. Ling)

1 Introduction

Since Granger [17] and Engle and Granger [13], the cointegrating time series has been a leading topic in the literature of economics. Numerous economic models, such as consumption function, purchasing power parity, money demand function, hedging ratio of spot and futures exchange rates, and yield curves of different terms of maturities, have been shown to have the cointigrating structure. The partially nonstationary Vector AR model or cointegrating time series models without GARCH effect have been extensively discussed over the past decades, see for example, Phillips and Durlauf [37] and Stock and Watson [43] in early years. Recently, Wang and Phillips [46] proposed a specification test for nonlinear non-stationary models. Kristensen and Rahbek [22] analyzed estimators and tests for a general class of VEC models that allows for asymmetric and non-linear error correction. Wang [45] established a martingale limit theorem for non-linear cointegration systems. Cavaliere *et al.* [9] considered bootstrap tests on the cointegration rank in vector AR models. Liang *et al.* [28] investigated local linear estimation of a nonparametric cointegration model. Cavaliere *et al.* [8] investigated a number of methods for estimating the cointegration rank in integrated vector AR systems with unknown AR order. Cai *et al.* [7] studied a new class of bivariate threshold cointegration models. Johansen and Nielsen [20] studied nonstationary cointegration in the fractionally cointegrated VAR model. Lin *et al.* [29] considered a double-nonlinear cointegration. She and Ling [40] studied a heavy-tailed VEC model. A recent overview on times series cointegration was given by Johansen [19].

Economic and financial time series often exhibit time-varying variances, called ARCH-type volatilities. Since Engle [11] and Bollerslve [4] proposed the ARCH/GARCH models, this kind of time series models have been extensively studied and applied in financial markets, see a nice review in Francq and Zakoian [14]. Ling *et al.* [31–34], Seo [39] established the asymptotic theory of the quasi-maximum likelihood estimator (QMLE) of unit root with the GARCH errors. Li *et al.* [27] investigated vector time series that exhibit both cointegration and time-varying variances. Li and Li [26] studied least absolute deviation estimation for unit root processes with GARCH errors. Chan and Zhang [10] provided an inference procedure for unit-root models with infinite variance GARCH errors. Lange [23] and Lange *et al.* [24] studied estimation and asymptotic inference for the AR-ARCH model. Shinki and Zhang [41] established asymptotic theory for fractionally integrated asymmetric power ARCH models. Zhang and Ling [48] established the asymptotic inference for AR models with heavy-tailed G-GARCH noises. Zhang *et al.* [47] studied an AR(1) model with ARCH(1) errors.

This paper is to study a partially nonstationary VAR model with vector GARCH noises. We study the full rank and the reduced rank QMLE of parameters in the model. It is shown that both QMLE of long-run parameters asymptotically converge to a functional of two correlated vector Brownian motions. Based these, the likelihood ratio (LR) test statistic for the cointegration rank is shown to be a functional of the standard Brownian motion and normal vector, asymptotically. The critical values of the LR test are simulated via the Monte Carlo method. As far as we know, our test is new in the literature.

The remaining part of this paper proceeds as follows. Section 2 gives the model and assumptions. Sections 3 and 4 study the full rank and the reduced rank QMLEs, respectively. Section 5 studies the LR test for the cointegration rank. Section 6 presents the method of simulating critical values. Section 7 reports the results from Monte Carlo experiments and Section 8 gives an illustrative empirical example of three interest rates. Conclusion is in Section 9. All the technical proofs are given in Appendix A. Throughout, $\longrightarrow_{\mathcal{L}}$ denotes convergence in distribution, $O_p(1)$ denotes a series of random numbers that are bounded in probability, and $o_p(1)$ denotes a series of random numbers converging to zero in probability.

2 Model and assumptions

We consider an m-dimensional autoregressive (AR) process $\{Y_t\}$, generated by

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_s Y_{t-s} + \varepsilon_t, \tag{2.1}$$

$$\varepsilon_t = (\varepsilon_{1t},\ldots,\varepsilon_{mt})', \tag{2.2}$$

$$\varepsilon_{it} = \eta_{it}\sqrt{h_{it-1}}, \quad h_{it-1} = a_{i0} + \sum_{j=1}^{q} a_{ij}\varepsilon_{it-j}^2 + \sum_{k=1}^{p} b_{ik}h_{it-1-k}, \tag{2.3}$$

where Φ_j's are constant matrices. In (2.3), $\eta_t = (\eta_{1t},\ldots,\eta_{mt})'$ is a sequence of independently and identically distributed (i.i.d.) random vectors with zero mean and $E(\eta_t\eta_t') = \Gamma \equiv (\sigma_{ij})_{m\times m}$, a positive definite matrix with $\sigma_{ii} = 1$ and $\sigma_{ij} = \sigma_{ji}$. It is easy to see that

$$E(\varepsilon_t|\mathcal{F}_{t-1}) = 0, \quad E(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = V_{t-1} = D_{t-1}\Gamma D_{t-1},$$

where

$$\mathcal{F}_{t-1} = \sigma\{\eta_s, s = t-1, t-2, \ldots\},$$
$$D_{t-1} = \operatorname{diag}\left(\sqrt{h_{1t-1}},\ldots,\sqrt{h_{mt-1}}\right).$$

V_{t-1} is the time-varying covariance matrix with constant correlation. The process ε_t in (2.2)-(2.3) is the multivariate generalized autoregressive conditional heteroskedasticity (GARCH) process proposed by Bollerslev [5] and has been widely used in the literature, for example by Tse [44]. As a referee commented, there is no cross-sectional dependence in the model structure of the multivariate GARCH (2.3). This is a strong restriction. Without this, it will be more difficult to work out the asymptotic theory of the estimated parameters and need to be further explored in the future. For example, ones may consider to extend the results in this paper to other types of GARCH noises such as BEKK or DCC-GARCH noises.

Denote L as the lag operator and define $\Phi(L)=I_m-\sum_{j=1}^s\Phi_jL^j$. We first make the following assumption.

Assumption 2.1. $|\Phi(z)|=0$ implies that either $|z|>1$ or $z=1$.

Define

$$W_t=Y_t-Y_{t-1},\quad \Phi_j^*=-\sum_{k=j+1}^s\Phi_k,\quad C=-\Phi(1)=-\left(I_m-\sum_{j=1}^s\Phi_j\right).$$

By Taylor's formula, $\Phi(L)$ can be decomposed as

$$\Phi(z)=(1-z)I_m-Cz-\sum_{j=1}^{s-1}\Phi_j^*(1-z)z^j. \tag{2.4}$$

Thus, we can reparameterize process (2.1) as

$$W_t=CY_{t-1}+\sum_{j=1}^{s-1}\Phi_j^*W_{t-j}+\varepsilon_t. \tag{2.5}$$

Following Ahn and Reinsel [1] and Johansen [18], we can decompose $C=AB$, where A and B are respectively $m\times r$ and $r\times m$ matrices of rank r. Define $d=m-r$. Denote $B_\perp$ as a $d\times m$ matrix of full rank such that $BB_\perp'=0_{r\times d}$, $\bar{B}=(BB')^{-1}B$ and $\bar{B}_\perp=(B_\perp B_\perp')^{-1}B_\perp$, and $A_\perp$ as an $m\times d$ matrix of full rank such that $A'A_\perp=0_{r\times d}$, $\bar{A}=A(A'A)^{-1}$ and $\bar{A}_\perp=A_\perp(A_\perp'A_\perp)^{-1}$.

We further impose the following condition.

Assumption 2.2. $|A_\perp'(I_m-\sum_{j=1}^{s-1}\Phi_j^*)B_\perp'|\neq 0$.

From the proof of [18, Theorem 4.2], we have the following decomposition:

$$B_\perp Y_t=B_\perp Y_{t-1}+u_{1t},\quad BY_t=u_{2t}, \tag{2.6}$$

where

$$\begin{aligned}
&u_t = (u_{1t}', u_{2t}')' = \psi(L) a_t, \\
&a_t \equiv (\bar{A}_\perp, \bar{A})' \varepsilon_t, \\
&\psi(z) \equiv \left[(\bar{A}_\perp, \bar{A})' \Phi(z) \left(\bar{B}_\perp', \bar{B}'(1-z)^{-1} \right) \right]^{-1},
\end{aligned}$$

where $\psi(z)$ exists by Assumption 2.2, see [18, p. 51]. Thus, $B_\perp Y_t$ is $I(1)$ while BY_t is $I(0)$. Under Assumption 2.1, model (2.1) or (2.5) is not stationary. But since BY_t is stationary if ε_t is stationary, we call process (2.1)-(2.3) a partially nonstationary VAR process with vector GARCH noises. A sufficient condition for the strict stationarity of ε_t is as follows.

Assumption 2.3. For $i = 1,\ldots,m, a_{i0} > 0, a_{i1},\ldots,a_{iq}, b_{i1},\ldots,b_{ip} \geq 0$, and $\sum_{j=1}^{q} a_{ij} + \sum_{k=1}^{p} b_{ik} < 1$.

The next assumption is the necessary and sufficient conditions such that the fourth moment of each component in ε_t is finite, see [34].

Assumption 2.4. For $i=1,\ldots,m$, all eigenvalues of $E(A_{it} \otimes A_{it})$ lie inside the unit circle, where $\otimes$ denotes the Kronecker product and

$$A_{it} = \begin{pmatrix}
a_{i1}\eta_{it}^2 & \cdots & a_{iq}\eta_{it}^2 & b_{i1}\eta_{it}^2 & \cdots & b_{ip}\eta_{it}^2 \\
 & I_{q-1} & 0_{(q-1)\times 1} & & 0_{(q-1)\times p} & \\
a_{i1} & \cdots & a_{iq} & b_{i1} & \cdots & b_{ip} \\
 & 0_{(p-1)\times q} & & & I_{p-1} & 0_{(p-1)\times 1}
\end{pmatrix}.$$

We further make an assumption as follows, which is to allow the parameters in (2.1) and those in (2.2)-(2.3) to be estimated separately without altering the asymptotic distributions. Without this assumption, the asymptotic distribution of the estimated parameters can be derived, but is rather complicated. To make it simple, we avoid this case in this paper.

Assumption 2.5. η_t is symmetrically distributed.

In many empirical studies such as those in Section 8 below, model (2.5) is augmented with an unknown constant, that is

$$W_t = CY_{t-1} + \sum_{j=1}^{s-1} \Phi_j^* W_{t-j} + \varepsilon_t + \mu. \tag{2.7}$$

As we can see in all the subsequent sections, the asymptotic theories are essentially the same, with the standard Brownian motion $B_d(u)$ replaced by the standard Brownian bridge $B_d(u) - \int_0^1 B_d(u)du$. For ease of exposition, we focus on model (2.5). Model (2.7) will be briefly discussed in Sections 5 and 6.

3 Full rank estimation

This section considers the full rank estimators for the mean parameters $\varphi \equiv vec[C,\Phi_1^*,\ldots,\Phi_{s-1}^*]$ and the estimators for the variance parameters $\delta \equiv [\delta_1',\delta_2']'$, $\delta_1 \equiv [a_0',a_1',\ldots,a_q',b_1',\ldots,b_p']'$, $a_j \equiv [a_{1j},\ldots,a_{mj}]'$, $b_k \equiv [b_{1k},\ldots,b_{mk}]'$, $j=0,1,\ldots,q$, $k=1,\ldots,p$, and $\delta_2 \equiv \tilde{v}(\Gamma)$, which is obtained from $vec(\Gamma)$ by eliminating the super-diagonal and the diagonal elements of Γ (see [36, p. 27]).

Given $\{Y_t : t=1,\ldots,n\}$, conditional on the initial values $Y_t = 0$ for $t \leq 0$, the log-likelihood function (LF) (with a constant ignored), as a function of the true parameters, can be written as

$$l_F(\varphi,\delta)=\sum_{t=1}^n l_t, \quad l_t=-\frac{1}{2}\varepsilon_t' V_{t-1}^{-1}\varepsilon_t-\frac{1}{2}\ln|V_{t-1}|, \tag{3.1}$$

where

$$V_{t-1}=D_{t-1}\Gamma D_{t-1}, \quad D_{t-1}=\mathrm{diag}\big(\sqrt{h_{1t-1}},\ldots,\sqrt{h_{mt-1}}\big).$$

Further denote $h_{t-1}=(h_{1t-1},\ldots,h_{mt-1})'$, $H_{t-1}=(h_{1t-1}^{-1},\ldots,h_{mt-1}^{-1})'$. The score function, as a function of the true parameters, can be expressed as

$$\begin{aligned}
\nabla_\varphi l_t &= -\frac{1}{2}\nabla_\varphi h_{t-1}\big(\iota-w\big(\varepsilon_t\varepsilon_t' V_{t-1}^{-1}\big)\big)\odot H_{t-1}+(X_{t-1}\otimes I_m)V_{t-1}^{-1}\varepsilon_t,\\
\nabla_\delta l_t &= \begin{pmatrix} -\dfrac{1}{2}\nabla_{\delta_1} h_{t-1}\big(\iota-w\big(\varepsilon_t\varepsilon_t' V_{t-1}^{-1}\big)\big)\odot H_{t-1} \\ -\tilde{v}\big(\Gamma^{-1}-\Gamma^{-1}D_{t-1}^{-1}\varepsilon_t\varepsilon_t' D_{t-1}^{-1}\Gamma^{-1}\big) \end{pmatrix},
\end{aligned}$$

where $X_{t-1}=[Y_{t-1}',W_{t-1}',\ldots,W_{t-s+1}']'$, $\iota=(1,1,\ldots,1)'$ and $w(\chi)$ is a vector containing the diagonal elements of the square matrix χ. For expositional sake, we present the formulae for $\nabla_\varphi h_{t-1}$ and $\nabla_{\delta_1} h_{t-1}$ in (B.1) and (B.2) of Appendix B.

We first find an initial estimator $(\hat{\varphi},\hat{\delta})$. For instance, we may adopt the LSE (least squares estimator) considered in [27] and use the residuals of the VAR for estimating the variance parameters (see, e.g. [32]). Given this initial estimator, we perform a one-step iteration

$$\dot{\varphi}=\hat{\varphi}-\left(\sum_{t=1}^n F_t|_{\hat{\varphi},\hat{\delta}}\right)^{-1}\left(\sum_{t=1}^n \nabla_\varphi l_t|_{\hat{\varphi},\hat{\delta}}\right), \tag{3.2}$$

$$\dot{\delta}=\hat{\delta}-\left(\sum_{t=1}^n S_t|_{\hat{\varphi},\hat{\delta}}\right)^{-1}\left(\sum_{t=1}^n \nabla_\delta l_t|_{\hat{\varphi},\hat{\delta}}\right), \tag{3.3}$$

where, in terms of the true parameters,

$$F_t = -\left(X_{t-1}X'_{t-1} \otimes V_{t-1}^{-1}\right) - \frac{1}{4}(\nabla_{\varphi} h_{t-1}) D_{t-1}^{-2}\left(\Gamma^{-1} \odot \Gamma + I_m\right) D_{t-1}^{-2}\left(\nabla'_{\varphi} h_{t-1}\right), \tag{3.4}$$

and, in terms of the true parameters, $S_t = (S_{ijt})_{2\times 2}$ with

$$\begin{aligned}
S_{11t} &= -\frac{1}{4}(\nabla_{\delta_1} h_{t-1}) D_{t-1}^{-2}\left(\Gamma^{-1} \odot \Gamma + I_m\right) D_{t-1}^{-2}\left(\nabla'_{\delta_1} h_{t-1}\right), \\
S_{12t} &= -(\nabla_{\delta_1} h_{t-1}) D_{t-1}^{-2} \Psi_m \left(I_m \otimes \Gamma^{-1}\right) N_m \tilde{L}'_m, \\
S_{22t} &= -2\tilde{L}_m N_m [\Gamma^{-1} \otimes \Gamma^{-1}] N_m \tilde{L}'_m,
\end{aligned}$$

Ψ_m, N_m and $\tilde{L}_m$ are constant matrices of dimensions mxm^2, m^2xm^2 and $m(m-1)/2$ respectively. See [36, p. 109, pp. 48-49, pp. 96-97]. We reproduce them in Appendix B with $m=2$, see (B.5)-(B.7).

Remark 3.1. Due to the different heteroskedasticity, the algorithm of this one-step iteration is somewhat different from that in [27]. However, the proof of our Theorem 3.1 below is similar to that of [27, Theorem 2]. We simply provide a sketchy proof in Appendix A. In practice, we may repeat the iterative procedure in (3.2)-(3.3), and get an estimator closer to the quasi-maximum likelihood estimator (QMLE), though the asymptotic distribution is not altered.

Recall that $\dot{\varphi} = vec[\dot{C}, \dot{\Phi}_1^*, \ldots, \dot{\Phi}_{s-1}^*]$. In Theorem 3.1 below, we state the asymptotic distribution of $n(\dot{C}-C)\bar{B}'_{\perp}$ (the nonstationary mean parameters), that of $\sqrt{n}vec[(\dot{C}-C)\bar{B}', (\dot{\Phi}_1^* - \Phi_1^*), \ldots, (\dot{\Phi}_{s-1}^* - \Phi_{s-1}^*)]$ (the stationary mean parameters), and that of $\sqrt{n}(\dot{\delta}-\delta)$ (the variance parameters). To facilitate our discussion, we first introduce another set of score function and Hessian function in (3.5)-(3.6), in terms of the true parameters. These score function and Hessian function will also be used in Section 4.2

$$\nabla_{\alpha_2} l_t = -\frac{1}{2}\nabla_{\alpha_2} h_{t-1}\left(\iota - w\left(\varepsilon_t \varepsilon'_t V_{t-1}^{-1}\right)\right) \odot H_{t-1} + (U_{t-1} \otimes I_m) V_{t-1}^{-1} \varepsilon_t, \tag{3.5}$$

$$R_{2t} = -\left(U_{t-1}U'_{t-1} \otimes V_{t-1}^{-1}\right) - \frac{1}{4}(\nabla_{\alpha_2} h_{t-1}) D_{t-1}^{-2}\left(\Gamma^{-1} \odot \Gamma + I_m\right) D_{t-1}^{-2}\left(\nabla'_{\alpha_2} h_{t-1}\right), \tag{3.6}$$

where

$$\alpha_2 = vec[A, \Phi_1^*, \ldots, \Phi_{s-1}^*], \quad U_{t-1} = [Y'_{t-1}B', W'_{t-1}, \ldots, W'_{t-s+1}]'.$$

As in [27, Theorem 2(a)], the asymptotic distribution of $n(\dot{C}-C)\bar{B}'_{\perp}$ is a functional of two correlated vector Brownian motions, though the covariance matrix is different because of a different heteroskedasticity model. To facilitate our discussion on Theorem 3.1, first refer to the heteroskedasticity model (2.3). For $i=1,2,\ldots,m$, let

$$a^{(i)}(z)b^{(i)}(z)^{-1}=\sum_{l=1}^{\infty}\nu_{il}z^l,$$

where

$$a^{(i)}(z)=\sum_{l=1}^{q}a_{il}z^l,\quad b^{(i)}(z)=1-\sum_{l=1}^{p}b_{il}z^l.$$

Denote $\nu_l=(\nu_{1l},\ldots,\nu_{ml})'$, $l=1,2,\ldots$. Let $(W_m'(u),W_m^{*\prime}(u))'$ be a $2m$-dimensional Brownian motion (BM) with the covariance matrix

$$u\Omega\equiv u\begin{pmatrix}(EV_{t-1}) & I_m\\ I_m & \Omega_1^*\end{pmatrix}, \tag{3.7}$$

where

$$\begin{aligned}
&\Omega_1^*=E\left(V_{t-1}^{-1}\right)+(\Delta-\iota\iota')\odot\sum_{l=1}^{\infty}\left(\nu_l\nu_l'\odot E(\Pi_{lt})\right),\\
&\Delta=E\left[w\left(\eta_t\eta_t'\Gamma^{-1}\right)\left(w\left(\eta_t\eta_t'\Gamma^{-1}\right)\right)'\right],\\
&\Pi_{lt}=\left(\varepsilon_{t-l}\varepsilon_{t-l}'\odot H_{t-1}H_{t-1}'\right).
\end{aligned}$$

Thus,

$$B_d(u)\equiv\Omega_{a_1}^{-\frac{1}{2}}[I_d,0_{d\times r}]\Omega_a^{\frac{1}{2}}(EV_{t-1})^{-\frac{1}{2}}W_m(u)$$

is a d-dimensional standard BM, where

$$\Omega_a=E(a_ta_t'),\quad \Omega_{a_1}=[I_d,0_{d\times r}]\Omega_a[I_d,0_{d\times r}]'.$$

We now state our first result as follows.

Theorem 3.1. *Suppose Assumptions* 2.1-2.5 *hold. Then*

(a) $n(\dot{C}-C)\bar{B}_\perp'=n\dot{C}\bar{B}_\perp' \longrightarrow_{\mathcal{L}} \Omega_1^{-1}M^*$,

(b) $\sqrt{n}vec\left[(\dot{C}-C)\bar{B}',(\dot{\Phi}_1^*-\Phi_1^*),\ldots,(\dot{\Phi}_{s-1}^*-\Phi_{s-1}^*)\right] \longrightarrow_{\mathcal{L}} N\left(0,\Omega_2^{-1}\Omega_2^*\Omega_2^{-1}\right)$,

(c) $\sqrt{n}(\dot{\delta}-\delta) \longrightarrow_{\mathcal{L}} N\left(0,\Omega_\delta^{-1}\Omega_\delta^*\Omega_\delta^{-1}\right)$,

where

$$\Omega_1=E\left(V_{t-1}^{-1}\right)+\left(\Gamma^{-1}\odot\Gamma+I_m\right)\odot\sum_{l=1}^{\infty}\left(\nu_l\nu_l'\odot E(\Pi_{lt})\right),$$

$$M^* = \left(\int_0^1 B_d(u)dW_m^*(u)'\right)'\left(\int_0^1 B_d(u)B_d(u)'du\right)^{-1}\Omega_{a_1}^{-\frac{1}{2}}\psi_{11}^{-1},$$
$$\psi_{11} = [I_d, 0_{d\times r}]\left(\sum_{k=0}^{\infty}\psi_k\right)[I_d, 0_{d\times r}]',$$
$$\Omega_2 = -E(R_{2t}), \quad \Omega_2^* = E\big(\nabla_{\alpha_2}l_t\nabla'_{\alpha_2}l_t\big),$$
$$\Omega_\delta = -E(S_t), \quad \Omega_\delta^* = E\big(\nabla_\delta l_t\nabla'_\delta l_t\big).$$

Remark 3.2. The results are similar to [27, Theorem 2(a)], but with different definitions of Ω_1 and M^*. When $\eta_t \sim N(0,\Gamma)$, the QMLE here boils down to a MLE with normal error. In this case, the $(\Gamma^{-1}\odot\Gamma + I_m)$ in F_t or S_{11t} in (3.5)-(3.6) may be replaced by $(\Delta - \iota\iota')$ and $\Omega_1^* = \Omega_1$ and $\Omega_2^* = \Omega_2$. The asymptotic distributions in (b) and (c) are simplified, as $\Omega_2^{-1}\Omega_2^*\Omega_2^{-1} = \Omega_2^{*-1} = \Omega_2^{-1}$ and $\Omega_\delta^{-1}\Omega_\delta^*\Omega_\delta^{-1} = \Omega_\delta^{*-1} = \Omega_\delta^{-1}$.

Remark 3.3. When the errors are conditional heteroskedastic, and h_{it-1}'s are not constant, $\dot{C}$ is more efficient than the LSE of C in [1], in the sense discussed in [34].

4 Reduced rank estimation

We rewrite (2.5) in a reduced rank form as follows:

$$W_t = ABY_{t-1} + \sum_{j=1}^{s-1}\Phi_j^* W_{t-j} + \varepsilon_t, \tag{4.1}$$

where A and B are as defined in Section 2. This section considers the reduced rank estimator for $\alpha = [\alpha_1', \alpha_2']'$ with $\alpha_1 \equiv vec[B]$ and $\alpha_2 \equiv vec[A, \Phi_1^*, \ldots, \Phi_{s-1}^*]$. We first adopt Anderson [2,3] or Johansen [18] approach to obtain an initial estimator. The asymptotic properties are shown in Section 4.1. In Section 4.2, we consider the reduced rank QMLE.

4.1 Initial estimator for the "Mean" parameters

This initial estimator is essentially the QMLE which ignores the possible GARCH, i.e. the maximizer of the LF in (3.1) with $V_{t-1}(\varphi,\delta)$ replaced by a constant matrix. Alternatively put, we adopt Anderson [2,3]'s or Johansen [18]'s estimator. Denote this estimator as $\hat{\alpha} = [\hat{\alpha}_1', \hat{\alpha}_2']'$ with $\hat{\alpha}_1 = vec[\hat{B}]$ and $\hat{\alpha}_2 = vec[\hat{A}, \hat{\Phi}_1^*, \ldots, \hat{\Phi}_{s-1}^*]$. Using the arguments in [27, Section 5] (see also [18, Lemma 13.2]), we obtain the asymptotic distribution of the normalized $\hat{\alpha}_1$ ($\hat{B}$ is normalized by $(\hat{B}\bar{B}')^{-1}$) and that of the normalized $\hat{\alpha}_2$ ($\hat{A}$ is normalized by $\hat{B}\bar{B}'$). A sketchy proof can be found in Appendix A.

Theorem 4.1. *Suppose Assumptions* 2.1-2.5 *hold. Then*

(a) $$n\big((\hat{B}\bar{B}')^{-1}\hat{B}-B\big)\bar{B}'_{\perp}=n(\hat{B}\bar{B}')^{-1}\hat{B}\bar{B}'_{\perp} \longrightarrow_{\mathcal{L}} \big(A'(EV_{t-1})^{-1}A\big)^{-1}A'(EV_{t-1})^{-1}(A_{\perp},A)M,$$

(b) $$\sqrt{n}vec\big[\big(\hat{A}(\hat{B}\bar{B}')-A\big),\big(\hat{\Phi}_1^*-\Phi_1^*\big),\dots,\big(\hat{\Phi}_{s-1}^*-\Phi_{s-1}^*\big)\big] \longrightarrow_{\mathcal{L}} N\big(0,\Sigma_2^{-1}\Sigma_2^*\Sigma_2^{-1}\big),$$

where

$$\begin{aligned}
&M=\Omega_a^{\frac{1}{2}}\left(\int_0^1 B_d(u)dB_m(u)'\right)'\left(\int_0^1 B_d(u)B_d(u)'du\right)^{-1}\Omega_{a_1}^{-\frac{1}{2}}\psi_{11}^{-1},\\
&B_m(u)=(EV_{t-1})^{-\frac{1}{2}}W_m(u),\\
&\Sigma_2=E\big(U_{t-1}U_{t-1}'\otimes I_m\big),\quad \Sigma_2^*=E\big(U_{t-1}U_{t-1}'\otimes V_{t-1}\big),
\end{aligned}$$

and the remaining variables are as defined in Theorem 3.1.

Remark 4.1. The normalization factors $(\hat{B}\bar{B}')^{-1}$ and $\hat{B}\bar{B}'$ in (a) and (b) respectively are adopted from [18]. And have been found very useful in deriving a lot of hypothesis testing. Though we allow possible conditional heteroskedasticity, EV_t is a constant matrix and thus the asymptotic distributions in (a) and (b) are exactly the same as those in [18, Lemma 13.2], regardless of the presence of conditional heteroskedasticity (at least of that specified in (2.3)). Because of this, the test for reduced rank in [18] has correct asymptotic size. See also [15,25].

Remark 4.2. As in [1], if the components of Y_t can be arranged so that the last d components are non-cointegrated, then we can impose the structure $B=[I_r,B_0]$. Decompose $\hat{B}=[\hat{B}_1,\hat{B}_2]$, where $\hat{B}_1$ is rxr and $\hat{B}_2$ is rxd. Provided that $\hat{B}_1$ is invertible, it is easy to show that

$$n\big(\hat{B}_1^{-1}\hat{B}_2-B_0\big) \longrightarrow_{\mathcal{L}} \big(A'(EV_{t-1})^{-1}A\big)^{-1}A'(EV_{t-1})^{-1}(A_{\perp},A)MP_{21}^{-1}, \tag{4.2}$$

$$\sqrt{n}vec\big[\big(\hat{A}\hat{B}_1-A\big),\big(\hat{\Phi}_1^*-\Phi_1^*\big),\dots,\big(\hat{\Phi}_{s-1}^*-\Phi_{s-1}^*\big)\big] \longrightarrow_{\mathcal{L}} N\big(0,\Sigma_2^{-1}\Sigma_2^*\Sigma_2^{-1}\big), \tag{4.3}$$

where P_{21} is a $d\times d$ matrix such that $[0_{d\times r},I_d][\bar{B}'_{\perp},\bar{B}']=[P_{21},P_{22}]$. The distribution in (4.2) is exactly the same as that in [1], if their Jordan canonical form applies and $A=\bar{B}'$ up to an rxr invertible matrix.

4.2 Reduced rank estimation that incorporates GARCH

This subsection uses the initial estimator $\hat{\alpha}$ in Section 4.1 and $\hat{\delta}$ suggested in Section 3 to obtain a new reduced rank estimation that incorporates GARCH. The LF based on the error-correction form (4.1) is similar to that in (3.1), but now it is a function of (α,δ) instead. Conditional on the initial values $Y_t=0$ for $t\leq 0$, the log-likelihood function, as a function of the true parameters, can be written as

$$l_R(\alpha,\delta)=\sum_{t=1}^{n} l_t, \quad l_t=-\frac{1}{2}\varepsilon_t' V_{t-1}^{-1}\varepsilon_t-\frac{1}{2}\ln|V_{t-1}|. \tag{4.4}$$

As we argue in the proof of Theorem 4.2, α and δ can be estimated separately without altering the asymptotic distributions. In the following, we confine our attention to estimating α. The score function (with respect to α_1 and α_2), as a function of the true parameters, can be expressed as

$$\begin{aligned}
\nabla_{\alpha_1} l_t &= -\frac{1}{2}\nabla_{\alpha_1} h_{t-1}\big(\iota - w\big(\varepsilon_t\varepsilon_t' V_{t-1}^{-1}\big)\big)\odot H_{t-1} + \big(Y_{t-1}\otimes A'\big)V_{t-1}^{-1}\varepsilon_t,\\
\nabla_{\alpha_2} l_t &= -\frac{1}{2}\nabla_{\alpha_2} h_{t-1}\big(\iota - w\big(\varepsilon_t\varepsilon_t' V_{t-1}^{-1}\big)\big)\odot H_{t-1} + (U_{t-1}\otimes I_m)V_{t-1}^{-1}\varepsilon_t,
\end{aligned}$$

where we recall that $U_{t-1}=[Y_{t-1}'B',W_{t-1}',\ldots,W_{t-s+1}']'$. For expositional sake, we present the formulae for $\nabla_{\alpha_1} h_{t-1}$ and $\nabla_{\alpha_2} h_{t-1}$ in (B.3) and (B.4) of Appendix B. Given the initial estimator $(\hat{\alpha},\hat{\delta})$, we perform a one-step iteration

$$\dot{\alpha}_1=\hat{\alpha}_1-\left(\sum_{t=1}^{n} R_{1t}\big|_{\hat{\alpha},\hat{\delta}}\right)^{-1}\left(\sum_{t=1}^{n}\nabla_{\alpha_1} l_t\big|_{\hat{\alpha},\hat{\delta}}\right), \tag{4.5}$$

$$\dot{\alpha}_2=\hat{\alpha}_2-\left(\sum_{t=1}^{n} R_{2t}\big|_{\hat{\alpha},\hat{\delta}}\right)^{-1}\left(\sum_{t=1}^{n}\nabla_{\alpha_2} l_t\big|_{\hat{\alpha},\hat{\delta}}\right), \tag{4.6}$$

where, in terms of the true parameters,

$$\begin{aligned}
R_{1t}=&-\big(Y_{t-1}Y_{t-1}'\otimes A'V_{t-1}^{-1}A\big)\\
&-\frac{1}{4}(\nabla_{\alpha_1} h_{t-1})D_{t-1}^{-2}\big(\Gamma^{-1}\odot\Gamma+I_m\big)D_{t-1}^{-2}\big(\nabla_{\alpha_1}' h_{t-1}\big),
\end{aligned} \tag{4.7}$$

$$\begin{aligned}
R_{2t}=&-\big(U_{t-1}U_{t-1}'\otimes V_{t-1}^{-1}\big)\\
&-\frac{1}{4}(\nabla_{\alpha_2} h_{t-1})D_{t-1}^{-2}\big(\Gamma^{-1}\odot\Gamma+I_m\big)D_{t-1}^{-2}\big(\nabla_{\alpha_2}' h_{t-1}\big).
\end{aligned} \tag{4.8}$$

Because of the symmetry assumption in Assumption 2.5, all the results in Sections 4 and 5 will not be affected if the initial estimator $\tilde{\delta}$ satisfies the condition

$\sqrt{n}(\tilde{\delta}-\delta)=O_p(1)$. Note that $\dot{\alpha}_1=vec(\dot{B})$. To distinguish the reduced rank estimator from the full rank estimator for $\Phi_1^*,\ldots,\Phi_{s-1}^*$, we denote $\dot{\alpha}_2=vec[\dot{A},\ddot{\Phi}_1^*,\ldots,\ddot{\Phi}_{s-1}^*]$. That said, it is clear from the proofs of Theorem 3.1(b) and Theorem 4.2(b) that for $j=1,\ldots,s-1$, $\dot{\Phi}_j^*-\ddot{\Phi}_j^*=o_p(1)$. The asymptotic distribution of the normalized $\dot{\alpha}_1$ (normalized by $(\dot{B}\bar{B}')^{-1}$) and that of the normalized $\dot{\alpha}_2$ (with $\dot{A}$ normalized by $(\dot{B}\bar{B}')$) are given in Theorem 4.2 below.

Theorem 4.2. *Suppose Assumptions* 2.1-2.5 *hold. Then*

(a) $n\big((\dot{B}\bar{B}')^{-1}\dot{B}-B\big)\bar{B}'_\perp=n(\dot{B}\bar{B}')^{-1}\dot{B}\bar{B}'_\perp \ \longrightarrow_{\mathcal{L}}\ \big(A'\Omega_1 A\big)^{-1}A'M^*,$

(b) $\sqrt{n}vec\big[\big(\dot{A}(\dot{B}\bar{B}')-A\big),\big(\ddot{\Phi}_1^*-\Phi_1^*\big),\ldots,\big(\ddot{\Phi}_{s-1}^*-\Phi_{s-1}^*\big)\big] \ \longrightarrow_{\mathcal{L}}\ N\big(0,\Omega_2^{-1}\Omega_2^*\Omega_2^{-1}\big),$

where M^* and the remaining variables are as defined in Theorem 3.1.

Remark 4.3. The normalization factors $(\dot{B}\bar{B}')^{-1}$ and $\dot{B}\bar{B}'$ in (a) and (b) are adopted such that $(\dot{B}\bar{B}')^{-1}\dot{B}-B=O_p(n^{-1})$. This property plays an important role in proving Lemma 5.1, Theorems 5.1 and 5.2, see also for the discussion in [18, Section 13.2]. The asymptotic distribution in (b) is exactly the same as that in Theorem 3.1(b). Heuristically, the asymptotic distribution of the "stationary mean" parameters is unaltered, regardless of we imposing the reduced rank. In contrast, the asymptotic distribution in (a) (that of the "nonstationary mean" parameters) is different from that in Theorem 3.1(a).

Remark 4.4. Decompose $\dot{B}=[\dot{B}_1,\dot{B}_2]$, where $\dot{B}_1$ is rxr and $\dot{B}_2$ is rxd. If the components of Y_t can be arranged as in [1] such that the last d components are non-cointegrated, and $\dot{B}_1$ is invertible, it is easy to show that

$$n(\dot{B}_1^{-1}\dot{B}_2-B_0) \ \longrightarrow_{\mathcal{L}}\ (A'\Omega_1A)^{-1}A'M^*P_{21}^{-1}, \tag{4.9}$$

$$\sqrt{n}vec[(\dot{A}\dot{B}_1-A),(\ddot{\Phi}_1^*-\Phi_1^*),\ldots,(\ddot{\Phi}_{s-1}^*-\Phi_{s-1}^*)] \ \longrightarrow_{\mathcal{L}}\ N(0,\Omega_2^{-1}\Omega_2^*\Omega_2^{-1}), \tag{4.10}$$

where P_{21} is defined around (4.3). The distribution in (4.9) is similar to that in [27], with different definitions of Ω_1 and $W_m^*(u)$.

5 Testing for reduced rank

This section considers the null and the alternative hypotheses

$$H_0:\mathrm{rank}(C)=r<m \quad \text{vs} \quad H_a:\mathrm{rank}(C)=m. \tag{5.1}$$

The likelihood ratio (LR) test statistic is as follows:

$$LR_G \equiv 2\left[l_F(\dot{\varphi},\dot{\delta}) - l_R(\dot{\alpha},\dot{\delta})\right], \tag{5.2}$$

where $l_F(.,.)$ is the (full-rank) LF as defined in (3.1) and $l_R(.,.)$ is the (reduced-rank) LF as defined in (4.4). $\dot{\varphi}$ and $\dot{\alpha}$ are respectively the full rank (see Section 3) and the reduced rank (see Section 4) estimators for φ and α. $\dot{\delta}$ is as defined in Theorem 3.1. The following lemma gives the asymptotic distribution of LR_G.

Lemma 5.1. *Suppose Assumptions* 2.1-2.5 *hold. Then under the null H_0, we have*

$$\begin{aligned} LR_G \ \longrightarrow_{\mathcal{L}} \ \operatorname{tr}\Bigg[& \left(\int_0^1 B_d(u) dV_d^*(u)'\right)' \left(\int_0^1 B_d(u) B_d(u)' du\right)^{-1} \\ & \times \left(\int_0^1 B_d(u) dV_d^*(u)'\right)\Bigg], \end{aligned} \tag{5.3}$$

where

$$\begin{aligned} V_d^*(u) =& \Upsilon B_d(u) \\ & + \left[\left(A_\perp' \Omega_1^{-1} A_\perp\right)^{-\frac{1}{2}} A_\perp' \Omega_1^{-1} \Omega_1^* \Omega_1^{-1} A_\perp \left(A_\perp' \Omega_1^{-1} A_\perp\right)^{-\frac{1}{2}} - \Upsilon\Upsilon'\right]^{\frac{1}{2}} V_d(u), \\ \Upsilon =& \left(A_\perp' \Omega_1^{-1} A_\perp\right)^{\frac{1}{2}} \left(A_\perp' (EV_{t-1}) A_\perp\right)^{-\frac{1}{2}}, \end{aligned}$$

and $(B_d'(u), V_d'(u))'$ is a $2d$-dimensional standard Brownian motion.

Remark 5.1. When V_t is a constant matrix, $\Omega_1^* = \Omega_1 = (EV_{t-1})^{-1}, \Upsilon = I_d, V_d^*(u) = B_d(u)$. The distribution of LR_G is exactly the same as that in [38] and that of a special case in [18].

In principle, the critical value of the distribution in (5.3) can be simulated via Monte Carlo method. However, the number of nuisance parameters equals $d^2 + (1+d)d/2$. We only consider two special cases in Theorems 5.1 and 5.2 below.

Recall the full-rank estimator $\dot{\varphi} = vec[\dot{C}, \dot{\Phi}_1^*, \ldots, \dot{\Phi}_{s-1}^*]$ (see (3.2)). We define another estimator $\ddot{\varphi} \equiv vec[\dot{A}\dot{B}, \dot{\Phi}_1^*, \ldots, \dot{\Phi}_{s-1}^*]$, where we recall that $\dot{\alpha}_1 = vec[\dot{B}]$ is obtained from (4.5) and $\dot{A}$ is obtained from (4.6). Expansions which are similar to those standardly used in the likelihood theory give (see, for instance, [16, Section 5.6])

$$LR_G = (\dot{\varphi} - \ddot{\varphi})' \left(-\sum_{t=1}^n F_t|_{\dot{\varphi},\dot{\delta}}\right) (\dot{\varphi} - \ddot{\varphi}) + o_p(1), \tag{5.4}$$

where in terms of the true parameters, F_t is as defined in (3.4). In $\ddot{\varphi}$, one may use $\ddot{\Phi}_j^*$ (the reduced rank estimator) instead of $\dot{\Phi}_j^*$ (the full rank estimator), since $\sqrt{n}(\dot{\Phi}_j^* - \ddot{\Phi}_j^*) = o_p(1)$. Eq. (5.4) gives us an asymptotically equivalent form of the LR_G, which is computationally easier. We use this form in the Monte Carlo experiments as well as in the empirical example in Sections 7 and 8 below. On the other hand, this form suggests a Hausman-type test, which renders a distribution simpler than that in Lemma 5.1. (The crucial arguments can be found around (A.27) in Appendix A.)

We define a Hausman-type test statistic as follows:

$$H_G \equiv (\dot{\varphi} - \ddot{\varphi})' \left(-\sum_{t=1}^{n} F_t^H \big|_{\dot{\varphi},\dot{\delta}} \right) (\dot{\varphi} - \ddot{\varphi}), \tag{5.5}$$

where, in terms of the true parameters,

$$F_t^H = -\Big(X_{t-1}X_{t-1}' \otimes A_\perp \left(A_\perp' \Omega_1^{-1} \Omega_1^* \Omega_1^{-1} A_\perp \right)^{-1} A_\perp' \Big). \tag{5.6}$$

The following theorem gives the asymptotic distribution of H_G.

Theorem 5.1. *Suppose the assumptions in Lemma* 5.1 *hold. Then*

$$H_G \ \longrightarrow_{\mathcal{L}} \ \operatorname{tr}\left\{ \left[\zeta \left(I_d - \Lambda_d^H \right)^{\frac{1}{2}} + \Phi \Lambda_d^{H\frac{1}{2}} \right]' \left[\zeta (I_d - \Lambda_d^H)^{\frac{1}{2}} + \Phi \Lambda_d^{H\frac{1}{2}} \right] \right\}, \tag{5.7}$$

where Λ_d^H is a diagonal matrix containing the d eigenvalues of $(I_d - \Upsilon^H \Upsilon^{H'})$,

$$\Upsilon^H = \left(A_\perp' \Omega_1^{-1} \Omega_1^* \Omega_1^{-1} A_\perp \right)^{-\frac{1}{2}} \left(A_\perp' \Omega_1^{-1} A_\perp \right) \left(A_\perp' (EV_{t-1}) A_\perp \right)^{-\frac{1}{2}},$$

$\Phi \sim N(0, I_d)$ and independent of

$$\zeta = \left[\int_0^1 B_d(u) B_d(u)' du \right]^{-\frac{1}{2}} \int_0^1 B_d(u) dB_d(u)',$$

$B_d(u)$ is a d-dimensional standard Brownian motion.

Remark 5.2. When V_t is a constant matrix, $\Omega_1^* = \Omega_1 = (EV_{t-1})^{-1}$. The distribution of H_G is exactly the same as that in [38] and that of a special case in [18].

When $\Omega_1^* = \Omega_1$, the distribution of LR_G can be simplified as follows.

Theorem 5.2. *If the assumptions in Lemma* 5.1 *hold and* $\Omega_1^*=\Omega_1$, *then*

$$\begin{aligned} LR_G &= H_G + o_p(1) \\ &\longrightarrow_{\mathcal{L}} \ \mathrm{tr}\left\{\left[\zeta(I_d-\Lambda_d)^{\frac{1}{2}}+\Phi\Lambda_d^{\frac{1}{2}}\right]'\left[\zeta(I_d-\Lambda_d)^{\frac{1}{2}}+\Phi\Lambda_d^{\frac{1}{2}}\right]\right\}, \end{aligned} \tag{5.8}$$

where $\Lambda_d=\Lambda_d^H$ *is a diagonal matrix containing the d eigenvalues of* $(I_d-\mathrm{YY}')$, Y *is as defined in Lemma* 5.1 *while* Φ *and* ζ *are as defined in Theorem* 5.1.

When $\Omega_1^*=\Omega_1$, we have $\mathrm{Y}^H=\mathrm{Y}$ and $\Lambda_d^H=\Lambda_d$. Thus, the distributions in both theorems in Theorems 5.1 and 5.2, are the same in this case.

We close this section with variants of Theorems 5.1 and 5.2, in which the AR model contains a constant term. Modifying upon (4.1),

$$W_t=ABY_{t-1}+\sum_{j=1}^{s-1}\Phi_j^*W_{t-j}+\varepsilon_t+\mu, \tag{5.9}$$

where μ is unknown but we do know $B_\perp\mu=0$. This is a model used in many empirical examples, including that in [38]. Denote the corresponding Hausman-type test and LR test as $H_{G\mu}$ and $LR_{G\mu}$, respectively.

The following two corollaries can be obtained straightforwardly from Theorems 5.1 and 5.2, respectively. The proofs are thus omitted.

Corollary 5.1. *Suppose the assumptions in Theorem* 5.1 *hold. Then the Hausman-type test statistic (with an unknown constant estimated)*

$$H_{G\mu} \ \longrightarrow_{\mathcal{L}} \ \mathrm{tr}\left\{\left[\bar{\zeta}\left(I_d-\Lambda_d^H\right)^{\frac{1}{2}}+\Phi\Lambda_d^{H\frac{1}{2}}\right]'\left[\bar{\zeta}\left(I_d-\Lambda_d^H\right)^{\frac{1}{2}}+\Phi\Lambda_d^{H\frac{1}{2}}\right]\right\}, \tag{5.10}$$

where Λ_d^H *is as defined in Theorem* 5.1, $\Phi\sim N(0,I_d)$ *and independent of*

$$\bar{\zeta}=\left[\int_0^1\bar{B}_d(u)\bar{B}_d(u)'du\right]^{-\frac{1}{2}}\int_0^1\bar{B}_d(u)dB_d(u)', \quad \bar{B}\equiv\left[B_d(u)-\int_0^1 B_d(u)du\right],$$

and $B_d(u)$ *is a d-dimensional standard Brownian motion.*

Corollary 5.2. *Suppose the assumptions in Theorem* 5.2 *hold. Then the LR test statistic (with an unknown constant estimated)*

$$LR_{G\mu} \ \longrightarrow_{\mathcal{L}} \ \mathrm{tr}\left\{\left[\bar{\zeta}(I_d-\Lambda_d)^{\frac{1}{2}}+\Phi\Lambda_d^{\frac{1}{2}}\right]'\left[\bar{\zeta}(I_d-\Lambda_d)^{\frac{1}{2}}+\Phi\Lambda_d^{\frac{1}{2}}\right]\right\}, \tag{5.11}$$

where Λ_d *is as defined in Theorem* 5.2, Φ *and* $\bar{\zeta}$ *are as defined in Corollary* 5.1.

6 Simulating critical values and estimating nuisance parameters

As an illustration, this section simulates and tabulates the critical values of test statistics for the cases $d=1$ and $d=2$. Cases of higher dimensions can be done similarly. Then we show how to estimate the nuisance parameters in practice.

When there is no unknown constant in the mean part, LR_G and H_G are asymptotically distributed as

$$\operatorname{tr}\left\{\left[\zeta\left(I_d-\Lambda_d^H\right)^{\frac{1}{2}}+\Phi\Lambda_d^{H\frac{1}{2}}\right]'\left[\zeta\left(I_d-\Lambda_d^H\right)^{\frac{1}{2}}+\Phi\Lambda_d^{H\frac{1}{2}}\right]\right\}, \tag{6.1}$$

$$\operatorname{tr}\left\{\left[\zeta(I_d-\Lambda_d)^{\frac{1}{2}}+\Phi\Lambda_d^{\frac{1}{2}}\right]'\left[\zeta(I_d-\Lambda_d)^{\frac{1}{2}}+\Phi\Lambda_d^{\frac{1}{2}}\right]\right\}, \tag{6.2}$$

respectively. When there is an unknown constant in the mean part, $LR_{G\mu}$ and $H_{G\mu}$ are asymptotically distributed as

$$\operatorname{tr}\left\{\left[\bar{\zeta}\left(I_d-\Lambda_d^H\right)^{\frac{1}{2}}+\Phi\Lambda_d^{H\frac{1}{2}}\right]'\left[\bar{\zeta}\left(I_d-\Lambda_d^H\right)^{\frac{1}{2}}+\Phi\Lambda_d^{H\frac{1}{2}}\right]\right\}, \tag{6.3}$$

$$\operatorname{tr}\left\{\left[\bar{\zeta}(I_d-\Lambda_d)^{\frac{1}{2}}+\Phi\Lambda_d^{\frac{1}{2}}\right]'\left[\bar{\zeta}(I_d-\Lambda_d)^{\frac{1}{2}}+\Phi\Lambda_d^{\frac{1}{2}}\right]\right\}, \tag{6.4}$$

respectively. If $\Lambda_d^H=\Lambda_d$, the distributions of (6.1) and (6.3) reduce to those of (6.2) and (6.4), respectively.

The critical values of the distribution in (6.1)-(6.4) can be simulated via Monte Carlo method. For $d=1$, we denote $\Lambda_1^H=\lambda_1$ or $\Lambda_1=\lambda_1$; while for $d=2$, we denote $\Lambda_2^H=\operatorname{diag}(\lambda_1,\lambda_2)$ or $\Lambda_2=\operatorname{diag}(\lambda_1,\lambda_2)$. For each independent replication, Φ is generated from a d-dimensional standard normal distribution, while the n (the sample size) ϵ_s's are generated from n i.i.d. d-dimensional standard normal distribution, which is also independent of that of Φ, $y_t\equiv\sum_{s=1}^n\epsilon_s$. When one considers (6.1)-(6.2),

$$\zeta\equiv\left(n^{-2}\sum_{t=1}^n y_{t-1}y_{t-1}'\right)^{-\frac{1}{2}}\left(n^{-1}\sum_{t=1}^n y_{t-1}\epsilon_t'\right).$$

When one considers (6.3)-(6.4),

$$\bar{\zeta}\equiv\left(n^{-2}\sum_{t=1}^n(y_{t-1}-\bar{y})(y_{t-1}-\bar{y})'\right)^{-\frac{1}{2}}\left(n^{-1}\sum_{t=1}^n(y_{t-1}-\bar{y})\epsilon_t'\right),\quad \bar{y}\equiv n^{-1}\sum_{t=1}^n y_{t-1}.$$

We simulate the critical values with $d=1$ and $d=2$ and (λ_1, λ_2) ranging from 0.0 to 0.9. For intermediate values of (λ_1,λ_2), the critical values could be obtained by interpolation. The simulated critical values with 100,000 replications and $n=$ 2,000 are tabulated in Tables 1-3.

When one applies Theorem 5.1 or Corollary 5.1, the d eigenvalues of

$$I_d-\left(A'_\perp\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}A_\perp\right)^{-\frac{1}{2}}\left(A'_\perp\Omega_1^{-1}A_\perp\right)\left(A'_\perp(EV_{t-1})A_\perp\right)^{-1}$$
$$\times\left(A'_\perp\Omega_1^{-1}A_\perp\right)\left(A'_\perp\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}A_\perp\right)^{-\frac{1}{2}}$$

needs to be estimated. On the other hand, when one applies Theorem 5.2 or Corollary 5.2, the d eigenvalues of

$$I_d-\left(A'_\perp\Omega_1^{-1}A_\perp\right)^{\frac{1}{2}}\left(A'_\perp(EV_{t-1})A_\perp\right)^{-1}\left(A'_\perp\Omega_1^{-1}A_\perp\right)^{\frac{1}{2}}.$$

Table 1: 90% simulated critical values.

H_G or LR_G	(λ_1, λ_2)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
d=1		2.995	2.978	2.964	2.941	2.914	2.883	2.845	2.811	2.782	2.746
d=2	0.0	10.479	10.386	10.312	10.234	10.119	10.003	9.906	9.796	9.680	9.551
	0.1		10.295	10.217	10.125	10.018	9.919	9.808	9.679	9.561	9.455
	0.2			10.116	10.028	9.916	9.819	9.691	9.579	9.453	9.322
	0.3				9.931	9.816	9.693	9.565	9.442	9.310	9.163
	0.4					9.707	9.576	9.440	9.313	9.176	9.018
	0.5						9.444	9.310	9.177	9.030	8.866
	0.6							9.153	9.015	8.857	8.698
	0.7								8.847	8.688	8.520
	0.8									8.526	8.345
	0.9										8.166
$H_{G\mu}$ or $LR_{G\mu}$	(λ_1, λ_2)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
d=1		6.588	6.327	6.051	5.767	5.457	5.113	4.715	4.272	3.794	3.270
d=2	0.0	15.842	15.578	15.270	14.955	14.630	14.277	13.913	13.537	13.132	12.712
	0.1		15.264	14.977	14.650	14.326	13.984	13.616	13.219	12.808	12.393
	0.2			14.661	14.341	13.985	13.630	13.259	12.861	12.457	12.034
	0.3				14.000	13.658	13.310	12.919	12.515	12.087	11.666
	0.4					13.306	12.961	12.572	12.152	11.709	11.264
	0.5						12.595	12.191	11.774	11.309	10.850
	0.6							11.780	11.356	10.897	10.412
	0.7								10.921	10.456	9.956
	0.8									9.963	9.466
	0.9										8.942

Table 2: 95% simulated critical values.

H_G or LR_G	(λ_1,λ_2)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
d=1		4.153	4.140	4.138	4.108	4.083	4.043	4.013	3.963	3.920	3.867
d=2	0.0	12.286	12.237	12.158	12.073	11.987	11.887	11.789	11.676	11.559	11.446
	0.1		12.140	12.071	11.987	11.902	11.818	11.692	11.578	11.434	11.284
	0.2			11.973	11.879	11.791	11.691	11.566	11.433	11.293	11.141
	0.3				11.752	11.669	11.570	11.432	11.296	11.158	11.010
	0.4					11.557	11.438	11.310	11.171	11.024	10.847
	0.5						11.322	11.176	11.049	10.854	10.693
	0.6							11.035	10.894	10.713	10.529
	0.7								10.719	10.555	10.353
	0.8									10.342	10.144
	0.9										9.932
$H_{G\mu}$ or $LR_{G\mu}$	(λ_1,λ_2)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
d=1		8.167	7.932	7.656	7.373	7.049	6.679	6.251	5.763	5.220	4.576
d=2	0.0	18.064	17.783	17.508	17.201	16.877	16.516	16.137	15.739	15.332	14.908
	0.1		17.530	17.212	16.897	16.564	16.206	15.833	15.439	15.004	14.561
	0.2			16.917	16.583	16.246	15.894	15.495	15.094	14.650	14.178
	0.3				16.253	15.906	15.539	15.147	14.736	14.277	13.806
	0.4					15.556	15.153	14.747	14.345	13.888	13.372
	0.5						14.741	14.352	13.919	13.470	12.958
	0.6							13.924	13.484	13.018	12.479
	0.7								13.031	12.553	11.975
	0.8									12.045	11.429
	0.9										10.861

For this, we need to consistently estimate EV_{t-1}, $A_\perp$, Ω_1 and Ω_1^*.

EV_{t-1} can be consistently estimated by $n^{-1}\sum_{t=1}^n \dot{V}_{t-1}$. On the other hand, by the term definition of $A_\perp$ (see around (2.2) above), it can be consistently estimated by $(I_m - cc'\dot{A}(\dot{A}'cc'\dot{A})^{-1}\dot{A}')c_\perp$, where $c=(I_r,0_{rxd})'$ and $c_\perp=(0_{d\times r},I_d)'$, see [18, p. 48] for a similar estimator. It is not difficult to see that

$$
\begin{aligned}
&\left(\Gamma^{-1}\odot\Gamma+I_m\right)\odot\sum_{l=1}^{\infty}\left(\nu_l\nu_l'\odot E(\Pi_{lt})\right)\\
&=E\left(\sum_{l=1}^{\infty}\mathrm{diag}(\nu_l\odot\varepsilon_{t-l})D_t^{-2}\left(\Gamma^{-1}\odot\Gamma+I_m\right)D_t^{-2}\sum_{l=1}^{\infty}\mathrm{diag}(\nu_l\odot\varepsilon_{t-l})\right).
\end{aligned}
$$

Denote $\xi_t=\sum_{l=1}^{\infty}\mathrm{diag}(\nu_l\odot\varepsilon_{t-l})$. With the initial values $\dot{\xi}_0=\dot{\xi}_{-1}=\cdots=\dot{\xi}_{-p+1}=0$, for $t=1,\ldots,n$, we can recursively compute

Table 3: 99% simulated critical values.

H_G or LR_G	(λ_1,λ_2)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
d=1		7.018	6.941	6.939	6.931	6.929	6.895	6.842	6.839	6.774	6.718
d=2	0.0	16.278	16.144	16.041	15.986	15.895	15.802	15.716	15.623	15.530	15.435
	0.1		16.105	15.991	15.920	15.806	15.643	15.552	15.482	15.337	15.247
	0.2			15.898	15.812	15.647	15.556	15.405	15.319	15.191	15.023
	0.3				15.702	15.609	15.471	15.318	15.202	15.021	14.870
	0.4					15.510	15.374	15.231	15.087	14.928	14.747
	0.5						15.298	15.115	14.954	14.820	14.612
	0.6							14.993	14.809	14.622	14.480
	0.7								14.668	14.435	14.259
	0.8									14.255	14.064
	0.9										13.770
$H_{G\mu}$ or $LR_{G\mu}$	(λ_1,λ_2)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
d=1		11.690	11.469	11.291	11.027	10.727	10.293	9.817	9.316	8.655	7.789
d=2	0.0	22.745	22.477	22.192	21.901	21.541	21.236	20.842	20.390	19.987	19.498
	0.1		22.311	22.013	21.670	21.346	20.978	20.528	20.137	19.640	19.157
	0.2			21.739	21.389	21.052	20.695	20.250	19.815	19.281	18.757
	0.3				21.093	20.787	20.360	19.921	19.441	18.904	18.329
	0.4					20.385	20.004	19.548	19.072	18.549	17.907
	0.5						19.545	19.122	18.639	18.047	17.429
	0.6							18.673	18.199	17.607	16.943
	0.7								17.617	17.093	16.376
	0.8									16.478	15.794
	0.9										15.178

$$\dot{\xi}_t=\sum_{l=1}^{q}\mathrm{diag}(\dot{a}_l\odot\dot{\varepsilon}_{t-l})+\sum_{l=1}^{p}\dot{\xi}_{t-l}\mathrm{diag}(\dot{b}_l).$$

Ω_1 and Ω_1^* can be consistently estimated by

$$\frac{1}{n}\sum_{t=1}^{n}\dot{V}_{t-1}^{-1}+\frac{1}{n}\sum_{t=1}^{n}\dot{\xi}_t\dot{D}_{t-1}^{-2}\big(\dot{\Gamma}^{-1}\odot\dot{\Gamma}+I_m\big)\dot{D}_{t-1}^{-2}\dot{\xi}_t,$$

$$\frac{1}{n}\sum_{t=1}^{n}\dot{V}_{t-1}^{-1}+\frac{1}{n}\sum_{t=1}^{n}\dot{\xi}_t\dot{D}_{t-1}^{-2}(\dot{\Delta}-\iota\iota')\dot{D}_{t-1}^{-2}\dot{\xi}_t,$$

respectively, where

$$\dot{\Delta}=n^{-1}\sum_{t=1}^{n}w\big(\dot{\eta}_t\dot{\eta}_t'\dot{\Gamma}^{-1}\big)w\big(\dot{\eta}_t\dot{\eta}_t'\dot{\Gamma}^{-1}\big)'.$$

7 Monte Carlo experiments

This section examines the performance of the test statistic LR_G in the finite samples through Monte Carlo experiments. We consider the case with $\operatorname{rank}(C)=1$ under the null H_0. With $\Gamma=I_m,\varepsilon_t$ is generated by the following model:

$$\varepsilon_{it}=\eta_{it}\sqrt{h_{it}},\quad h_{it}=0.1+0.3\varepsilon_{it-1}^2+0.6h_{it-1},\quad \eta_{it}\sim i.i.d.\,N(0,1).$$

A tri-variate $AR(1)$ model is considered and C in the error-correction form (4.1) is

$$\begin{aligned}
&DGP(a) && C=AB, \quad A=(-0.4,0.12,0.12)', \quad B=(1.0,-2.5,0.0).\\
&DGP(b) && C=\kappa I_3, \quad \kappa=-0.1.\\
&DGP(c) && C=\kappa I_3, \quad \kappa=-0.5.
\end{aligned}$$

For each DGP, the sample sizes $n=200,400,800$ are considered. Reduced rank estimation with and without GARCH are used. The empirical means and standard deviations of the estimated A and B for DGP(a) (the null model) are reported in Table 4. The biases of the estimators with GARCH are comparable to those without GARCH, if not smaller than. The standard deviations and the mean squared errors are definitely smaller, even when the sample size is as small as 200. That

Table 4: Empirical means and standard deviations for DGP(a).

			$A_1=-0.4$	$A_2=0.12$	$A_3=0.12$	$B_2=-2.5$	$B_3=0.0$
n=200	*No GARCH*	Mean	-0.4026	0.1244	0.1219	-2.5038	-0.0007
		SD	0.0334	0.0343	0.0265	0.0717	0.0395
	With GARCH	Mean	-0.4019	0.1219	0.1209	-2.5001	-0.0010
		SD	0.0229	0.0265	0.0204	0.0589	0.0322
n=400	*No GARCH*	Mean	-0.4016	0.1228	0.1220	-2.5012	-0.0002
		SD	0.0225	0.0268	0.0179	0.0351	0.0188
	With GARCH	Mean	-0.4010	0.1211	0.1210	-2.5011	-0.0002
		SD	0.0149	0.0176	0.0137	0.0283	0.0134
n=800	*No GARCH*	Mean	-0.4009	0.1210	0.1207	-2.5007	0.0000
		SD	0.0165	0.0213	0.0126	0.0164	0.0095
	With GARCH	Mean	-0.4002	0.1203	0.1204	-2.5006	-0.0002
		SD	0.0103	0.0125	0.0094	0.0122	0.0065

number of replications $=1{,}000$.

Table 5: Rejection frequency of testing $H_0: r=1$.

	size = 0.05				size = 0.10			
	(DGP)	(a)	(b)	(c)	(DGP)	(a)	(b)	(c)
$n=200$	LR_{NG}	0.077	0.006	0.361	LR_{NG}	0.151	0.020	0.545
	LR_G	0.079	0.023	0.641	LR_G	0.135	0.076	0.767
$n=400$	LR_{NG}	0.071	0.022	0.972	LR_{NG}	0.127	0.062	0.995
	LR_G	0.078	0.114	0.997	LR_G	0.139	0.230	0.998
$n=800$	LR_{NG}	0.058	0.160	1.000	LR_{NG}	0.107	0.329	1.000
	LR_G	0.051	0.603	1.000	LR_G	0.107	0.774	1.000

number of replications $=1{,}000$.

said, as pointed out by a referee, one should be careful interpreting the standard deviations of the estimators for B_2 and B_3, since their distributions have fat tails.

Rejection frequencies are summarized in Table 5, where LR_{NG} is Johansen test or Reinsel-Ahn test. Both LR_{NG} and LR_G are of the reasonably correct finite-sample size, even when the number of observations is as small as 200. Both tests slightly over-reject when the sample size is 200 or 400, and the over-rejections are comparable. Moreover, it is clear that LR_G is more powerful than LR_{NG}.

8 An empirical example

In this section, we fit our model to the logarithms of three US monthly interest rates. The series are the federal funds rate, the 90-day treasury bill rate, and the one-year treasury bill rate, from January 1960 to December 1979 and thus we have 240 observations. We first estimate a $VAR(s)$ with different order s, where $s=1,\ldots,6$. $VAR(4)$ attains the lowest AIC. The residuals are then applied to a test for multivariate heteroskedasticity, along the lines in [30] (see also [21], who apply a similar test to an entirely different set of data and find no heteroskedasticity). The χ_R^2 test statistics with different numbers of terms R are reported in Table 6.

Table 6: Test statistics for multivariate heteroskedasticity.

R	1	2	3	4	5	6	7	8	9	10
	30.35	40.08	49.77	51.08	67.34	93.83	108.56	134.11	138.06	139.30

Table 6 clearly shows that the hypothesis of homoskedasticity is rejected and suggests that ε_t is not i.i.d. In view of this, apart from Johansen estimation, we also perform the full-rank as well as the reduced-rank estimation elucidated in Sections 3-4, which incorporate a GARCH(1,1) model. Results of the LR tests for reduced rank are summarized in Table 7. While we confine our discussion to the case that $s=4$, only for completeness we report other cases of different s.

Table 7 shows the hypothesis that $r=0$ is rejected by both tests. While LR_{NG} can hardly reject or only marginally rejects the null of $r=1$, our LR_G clearly rejects it. As with the empirical findings in Stock and Watson [42] and Reinsel and Ahn [38] who use the same dataset as ours (the former used the levels while the latter used the logarithms), the LR_{NG} does not reject the null of $r=2$. Due to its high power, LR_G rejects the null of $r=2$. All in all, unlike LR_{NG}, the LR_G strongly rejects that the reduced rank is 1. Judging from the LR_G, there is some evidence that the rank is 3, i.e. the interest rates are stationary. Similar results are found when we try GARCH models with different orders.

Table 7: LR test statistics.

	$H_0:r=0$		$H_0:r=1$		$H_0:r=2$	
s	LR_{NG}	LR_G	LR_{NG}	LR_G	LR_{NG}	LR_G
1	69.70(.000)	71.17(.000)	16.21(.084)	53.97(.000)	0.107(.939)	42.048(.000)
2	49.07(.000)	29.08(.003)	13.02(.210)	16.32(.009)	0.757(.801)	1.519(.259)
3	34.86(.019)	11.89(.381)	11.35(.320)	8.10(.147)	0.692(.814)	4.317(.056)
4	**46.43(.000)**	**52.25(.000)**	**15.60(.101)**	**31.01(.000)**	**1.044(.748)**	**33.044(.000)**
5	47.26(.000)	56.77(.000)	13.25(.197)	19.54(.003)	1.231(.714)	18.168(.000)
6	38.65(.006)	56.77(.000)	13.43(.188)	19.54(.003)	1.085(.741)	18.168(.000)

p-values are in brackets.

9 Conclusions

This paper studied a partially nonstationary AR model with vector GARCH noises. The asymptotic theory of the full rank and reduced rank QMLEs for the model were established. Based on the two estimators, the LR and the modified LR tests are constructed for testing the cointegration rank and their asymptotic distributions are derived. The simulation results show that our test for the reduced rank has substantial improvement upon the conventional LR test suggested in [38]. We also apply our approach to an empirical example of three interest rates. In contrast to the empirical results in the existing literature, our LR test shows evidence that the US monthly interest rates are stationary. This result is more in line with the common belief that the US interest rates are controllable under the stabilization mechanism of the US Federal Reserve Board.

Appendix A. Technical proofs

Consider the one-step iteration in (3.2)-(3.3). Denote

$$\begin{aligned}\bar{Q}^* &= \mathrm{diag}(Q\otimes I_m, I_{(s-1)m^2}),\\ \bar{D}^* &= \mathrm{diag}\big(nI_{dm}, \sqrt{n}I_{rm+(s-1)m^2}\big),\end{aligned}$$

where $Q'=[B'_\perp, B']$. We first give the following lemma which is about the normalized score $\bar{D}^{*-1}\bar{Q}^*\nabla_\varphi l_t$ and the normalized Hessian $\bar{D}^{*-1}\bar{Q}^*F_t\bar{Q}^*\bar{D}^{*-1}$. $n^{-1/2}\nabla_\delta l_t$ and $n^{-1}S_t$ are also considered. The proof is similar to that of [27, Lemma 1] and thus it is omitted.

Lemma A.1. *Suppose Assumptions* 2.1$-$2.5 *hold. Then*

(a) $$\sum_{t=1}^n \bar{D}^{*-1}\bar{Q}^*\nabla_\varphi l_t \ \longrightarrow_{\mathcal{L}} \ \left\{vec\left[\left(\int_0^1 B_d(u)dW_m^*(u)'\right)'\Omega_{a_1}^{\frac{1}{2}}\psi'_{11}\right]', \big[N\big(0,\Omega_2^*\big)\big]'\right\}',$$

(b) $$-\sum_{t=1}^n \bar{D}^{*-1}\bar{Q}^*F_t\bar{Q}^*\bar{D}^{*-1} \longrightarrow_{\mathcal{L}} \ \mathrm{diag}\left\{\left[\psi_{11}\Omega_{a_1}^{\frac{1}{2}}\int_0^1 B_d(u)B_d(u)'du\Omega_{a_1}^{\frac{1}{2}}\psi'_{11}\otimes\Omega_1\right], \Omega_2\right\},$$

(c) $$n^{-\frac{1}{2}}\sum_{t=1}^n\nabla_\delta l_t \ \longrightarrow_{\mathcal{L}} \ N(0,\Omega_\delta^*), \quad -n^{-1}\sum_{t=1}^n S_t \ \longrightarrow_p \ \Omega_\delta,$$

where all the variables are as defined in Theorem 3.1.

Proof of Theorem 3.1. Following the lines in [27, Section 4], the full rank estimator $(\dot{\varphi},\dot{\delta})$ admits an asymptotic expansion such that

$$\bar{D}^*\bar{Q}^{*'-1}(\dot{\varphi}-\varphi) = -\left(\sum_{t=1}^n \bar{D}^{*-1}\bar{Q}^*F_t\bar{Q}^{*'}\bar{D}^{*-1}\right)^{-1}\left(\sum_{t=1}^n \bar{D}^{*-1}\bar{Q}^*\nabla_\varphi l_t\right) + o_p(1), \quad \text{(A.1)}$$

$$\sqrt{n}(\dot{\delta}-\delta) = -\left(\sum_{t=1}^n n^{-1}S_t\right)^{-1}\left(\sum_{t=1}^n n^{-1/2}\nabla_\delta l_t\right) + o_p(1). \quad \text{(A.2)}$$

Theorem 3.1 then follows from (A.1)-(A.2) and Lemma A.1. $\square$

Proof of Theorem 4.1. From the proof of [18, Lemma 13.2], in our notation,

$$\begin{aligned} n(\hat{B}\bar{B}')^{-1}\hat{B}\bar{B}'_{\perp} = &\left(A'(EV_{t-1})^{-1}A\right)^{-1}A'(EV_{t-1})^{-1}\left(n^{-1}\sum_{t=1}^{n}\varepsilon_t Z'_{1t-1}\right)\\ &\times\left(n^{-2}\sum_{t=1}^{n}Z_{1t-1}Z'_{1t-1}\right)^{-1}+o_p(1), \end{aligned}$$

where $Z_{1t-1}\equiv B_{\perp}Y_{t-1}$. By the definition of a_t (see around (2.6)), $\varepsilon_t=(A_{\perp},A)a_t$. Therefore, by the arguments similar to those for proving Lemma A.1(a),

$$\begin{aligned} n^{-1}\sum_{t=1}^{n}\varepsilon_t Z'_{1t-1} &= (A_{\perp},A)\left(n^{-1}\sum_{t=1}^{n}a_t Z'_{1t-1}\right)\\ &\longrightarrow_{\mathcal{L}} \ (A_{\perp},A)\Omega_a^{\frac{1}{2}}\left[\int_0^1 B_d(u)dB_m(u)'\right]'\Omega_{a_1}^{\frac{1}{2}}\Psi'_{11}. \end{aligned}$$

On the other hand, by Lemma A.1(b),

$$n^{-2}\sum_{t=1}^{n}Z_{1t-1}Z'_{1t-1} \ \longrightarrow_{\mathcal{L}} \ \Psi_{11}\Omega_{a_1}^{\frac{1}{2}}\left[\int_0^1 B_d(u)B_d(u)'du\right]\Omega_{a_1}^{\frac{1}{2}}\Psi'_{11}.$$

Therefore, part (a) is proved. The proof of part (b) is straightforward and thus it is omitted. This completes the proof. □

The following lemma is useful for proving Theorem 4.2.

Lemma A.2. *Under the assumptions in Theorem* 4.2*, it follows that*

(a) $(\hat{B}\bar{B}')^{-1}(\dot{B}-\hat{B})=O_p(n^{-\frac{1}{2}})$,

(b) $\hat{A}(\dot{B}\bar{B}')=\hat{A}(\hat{B}\bar{B}')+O_p(n^{-\frac{1}{2}})=A+O_p(n^{-\frac{1}{2}})$,

(c) $(\dot{B}\bar{B}')^{-1}\hat{B}P_1=(\hat{B}\bar{B}')^{-1}\hat{B}P_1+O_p(n^{-\frac{3}{2}})=BP_1+O_p(n^{-1})$,

(d) $(\dot{B}\bar{B}')^{-1}\hat{B}P_2=(\hat{B}\bar{B}')^{-1}\hat{B}P_2+O_p(n^{-\frac{1}{2}})=BP_2+O_p(n^{-\frac{1}{2}})$.

Proof. (a) We first denote $D_{\alpha_1}=\mathrm{diag}(nI_{rd},\sqrt{n}I_{r^2})$ and $\hat{Q}^{**}=\mathcal{Q}(I_m\otimes(\hat{B}\bar{B}')')$, with $\mathcal{Q}=(Q\otimes I_r)$, where we recall that $Q'=[B'_{\perp},B']$. Also denote $\hat{\alpha}_1=vec(\hat{B})$, $\check{\alpha}_1=vec((\hat{B}\bar{B}')^{-1}\hat{B})$ and $\dot{\alpha}_1=vec(\dot{B})$. $\hat{\alpha}_2,\check{\alpha}_2,\dot{\alpha}_2$ are defined accordingly.$\hat{\alpha}$, $\check{\alpha}$, $\dot{\alpha}$ are also defined accordingly. Denote $P\equiv Q^{-1}$. Since $\hat{Q}^{**\prime-1}=(P'\otimes I_r)(I_m\otimes(\hat{B}\bar{B}')^{-1})$, we have

$$\left(I_m\otimes(\hat{B}\bar{B}')^{-1}\right)(\dot{\alpha}_1-\hat{\alpha}_1)=\mathcal{Q}'D_{\alpha_1}^{-1}D_{\alpha_1}(P'\otimes I_r)\left(I_m\otimes(\hat{B}\bar{B}')^{-1}\right)(\dot{\alpha}_1-\hat{\alpha}_1)$$

$$=\mathcal{Q}'D_{\alpha_1}^{-1}\big[D_{\alpha_1}\hat{Q}^{**\prime-1}(\dot{\alpha}_1-\hat{\alpha}_1)\big].$$

As $\mathcal{Q}'D_{\alpha_1}^{-1}=O(n^{-1/2})$, it suffices to show $D_{\alpha_1}\hat{Q}^{**\prime-1}(\dot{\alpha}_1-\hat{\alpha}_1)=O_p(1)$. Since $\hat{Q}^{**}=\mathcal{Q}(I_m\otimes(\hat{B}\bar{B}')')$, by (4.5),

$$\begin{aligned}
D_{\alpha_1}\hat{Q}^{**\prime-1}(\dot{\alpha}_1-\hat{\alpha}_1)=&-\left[\sum_{t=1}^n D_{\alpha_1}^{-1}\hat{Q}^{**}(R_{1t}|_{\hat{\alpha},\hat{\delta}})\hat{Q}^{**\prime}D_{\alpha_1}^{-1}\right]^{-1}\\
&\times\left[\sum_{t=1}^n D_{\alpha_1}^{-1}\hat{Q}^{**}(\nabla_{\alpha_1}l_t|_{\hat{\alpha},\hat{\delta}})\right]\\
=&-\left[\sum_{t=1}^n D_{\alpha_1}^{-1}\mathcal{Q}(R_{1t}|_{\check{\alpha},\hat{\delta}})\mathcal{Q}'D_{\alpha_1}^{-1}\right]^{-1}\\
&\times\left[\sum_{t=1}^n D_{\alpha_1}^{-1}\mathcal{Q}(\nabla_{\alpha_1}l_t|_{\check{\alpha},\hat{\delta}})\right].
\end{aligned}\tag{A.3}$$

By Theorem 4.1, $n(\check{\alpha}_1-\alpha_1)=O_p(1)$, $\sqrt{n}(\check{\alpha}_2-\alpha_2)=O_p(1)$, which are in addition to $\sqrt{n}(\hat{\delta}-\delta)=O_p(1)$. It is not difficult to see that

$$\sum_{t=1}^n D_{\alpha_1}^{-1}\mathcal{Q}(R_{1t}|_{\check{\alpha},\dot{\delta}})\mathcal{Q}'D_{\alpha_1}^{-1}=\sum_{t=1}^n D_{\alpha_1}^{-1}\mathcal{Q}R_{1t}\mathcal{Q}'D_{\alpha_1}^{-1}+o_p(1).\tag{A.4}$$

On the other hand, by a Taylor expansion and (A.4), with R_{1t}^* and l_t^* being evaluated at a mid-point of $(\check{\alpha},\dot{\delta})$ and (α,δ),

$$\begin{aligned}
\sum_{t=1}^n D_{\alpha_1}^{-1}\mathcal{Q}(\nabla_{\alpha_1}l_t|_{\check{\alpha},\dot{\delta}})=&\sum_{t=1}^n D_{\alpha_1}^{-1}\mathcal{Q}\nabla_{\alpha_1}l_t+\sum_{t=1}^n D_{\alpha_1}^{-1}\mathcal{Q}(R_{1t}^*)(\check{\alpha}_1-\alpha_1)\\
&+\sum_{t=1}^n D_{\alpha_1}^{-1}\mathcal{Q}(\nabla_{\alpha_1\alpha_2'}l_t^*)(\check{\alpha}_2-\alpha_2)\\
=&\sum_{t=1}^n D_{\alpha_1}^{-1}\mathcal{Q}_1\nabla_{\alpha_1}l_t+\left[\sum_{t=1}^n D_{\alpha_1}^{-1}\mathcal{Q}R_{1t}\mathcal{Q}'D_{\alpha_1}^{-1}+o_p(1)\right]\\
&\times\frac{1}{n}D_{\alpha_1}\big(P'\otimes I_r\big)[n(\check{\alpha}_1-\alpha_1)]\\
&+\left[\frac{1}{\sqrt{n}}\sum_{t=1}^n D_{\alpha_1}^{-1}\mathcal{Q}\big(\nabla_{\alpha_1\alpha_2'}l_t^*\big)\right]\sqrt{n}(\check{\alpha}_2-\alpha_2).
\end{aligned}\tag{A.5}$$

It is not difficult to see that $(\sum_{t=1}^n D_{\alpha_1}^{-1}\mathcal{Q}(\nabla_{\alpha_1\alpha_2'}l_t^*))/\sqrt{n}$ is $O_p(1)$. So is the RHS of (A.5). By Lemma (A.1)(a)-(b), (A.3)-(A.5), $D_{\alpha_1}\hat{Q}^{**\prime-1}(\dot{\alpha}_1-\hat{\alpha}_1)=O_p(1)$. Thus, (a) holds.

(b) By the $\sqrt{n}$-consistency of $\hat{A}(\hat{B}\bar{B}')$ for A, and (a) of this lemma,

$$\hat{A}(\dot{B}\bar{B}')=\hat{A}(\hat{B}\bar{B}')+\hat{A}(\hat{B}\bar{B}')(\hat{B}\bar{B}')^{-1}(\dot{B}-\hat{B})\bar{B}'=\hat{A}(\hat{B}\bar{B}')+O_p(1)O_p\big(n^{-\frac{1}{2}}\big).$$

Thus, (b) holds.

(c) and (d). Denote $\check{B}=(\hat{B}\bar{B}')^{-1}\hat{B}$.

$$(\dot{B}\bar{B}')^{-1}\hat{B}=\big[(\hat{B}\bar{B}')^{-1}\dot{B}\bar{B}'\big]^{-1}(\hat{B}\bar{B}')^{-1}\hat{B}=\big[(\hat{B}\bar{B}')^{-1}\dot{B}\bar{B}'\big]^{-1}\check{B}. \tag{A.6}$$

Using the formula $dF^{-1}=-F^{-1}(dF)F^{-1}$ for the $r\times r$ matrix F with $F(x)=[x\bar{B}]^{-1}$, and applying a Taylor expansion to $[(\hat{B}\bar{B}')^{-1}\dot{B}\bar{B}']^{-1}$ around $\check{B}\bar{B}'$, we have

$$\big[(\hat{B}\bar{B}')^{-1}\dot{B}\bar{B}'\big]^{-1}=[\check{B}\bar{B}']^{-1}-[B^*\bar{B}']^{-1}\big[(\hat{B}\bar{B}')^{-1}\dot{B}-\check{B}\big]\bar{B}'[B^*\bar{B}']^{-1},$$

where B^* lies between $(\hat{B}\bar{B}')^{-1}\dot{B}$ and $\check{B}$. Therefore, the RHS of (A.6) equals

$$\begin{aligned}&\big[(\hat{B}\bar{B}')^{-1}\hat{B}\bar{B}'\big]^{-1}(\hat{B}\bar{B}')^{-1}\hat{B}\\&-[B^*\bar{B}']^{-1}\big[(\hat{B}\bar{B}')^{-1}\dot{B}-\check{B}\big]\bar{B}'[B^*\bar{B}']^{-1}\check{B}\\=&(\hat{B}\bar{B}')^{-1}\hat{B}-[B^*\bar{B}']^{-1}\big[(\hat{B}\bar{B}')^{-1}\dot{B}-\check{B}\big]\bar{B}'[B^*\bar{B}']^{-1}\check{B}.\end{aligned} \tag{A.7}$$

By (a) of this lemma, $(\hat{B}\bar{B}')^{-1}\dot{B}-\check{B}=O_p(n^{-1/2})$. From this, we can show that $[B^*\bar{B}']^{-1}=O_p(1)$. $\bar{B}$ and $\check{B}$ are also $O_P(1)$. By (A.7), (d) holds. By Theorem 4.1, $\check{B}P_1=O_p(n^{-1})$ because $BP_1=0$. By (A.7),

$$\begin{aligned}&[(\hat{B}\bar{B}')^{-1}\hat{B}\bar{B}']^{-1}(\hat{B}\bar{B}')^{-1}\hat{B}P_1-[B^*\bar{B}']^{-1}[(\hat{B}\bar{B}')^{-1}\dot{B}-\check{B}]\bar{B}'[B^*\bar{B}']^{-1}\check{B}P_1\\&=(\hat{B}\bar{B}')^{-1}\hat{B}P_1+O_p\big(n^{-\frac{3}{2}}\big).\end{aligned}$$

Thus, (c) holds. This completes the proof. □

Proof of Theorem 4.2. Denote

$$\begin{aligned}\bar{D}^{**}&\equiv\operatorname{diag}\big(nI_{rd},\sqrt{n}I_{rm+(s-1)m^2}\big),\\\bar{Q}^{**}&\equiv\operatorname{diag}\big((B_\perp\otimes I_r),I_{rm+(s-1)m^2}\big).\end{aligned}$$

Using Assumptions 2.1-2.5 and the arguments around [27, Eq. (5.3)], we can show that

$$n^{-\frac{1}{2}}\bar{D}^{**-1}\bar{Q}^{**}\left(\sum_{t=1}^n\nabla^2_{\alpha\delta'}l_t\right)=o_p(1).$$

Thus, α and δ can be estimated separately without altering the asymptotic distributions. In the following, we confine our attention to estimating α. Again, using the arguments similar to those around [27, Eq. (5.5)], we can show that the Hessian can be given as follows:

$$\bar{D}^{**-1}\bar{Q}^{**}\sum_{t=1}^{n}\nabla^2_{\alpha\alpha'}l_t\bar{Q}^{**\prime}\bar{D}^{**-1}=\bar{D}^{**-1}\bar{Q}^{**}\sum_{t=1}^{n}R_t\bar{Q}^{**\prime}\bar{D}^{**-1}+o_p(1), \tag{A.8}$$

where $R_t=\mathrm{diag}(R_{1t},R_{2t})$, with R_{1t} and R_{2t} as defined in (4.7) and (4.8).

For any fixed positive constant K, let

$$\Xi_n\equiv\left\{(\tilde{\alpha},\tilde{\delta}):\|\bar{D}^{**}\bar{Q}^{**\prime-1}(\tilde{\alpha}-\alpha)\|\le K,\ \|\sqrt{n}(\tilde{\delta}-\delta)\|\le K\right\},$$

where $(\tilde{\alpha},\tilde{\delta})$ is a generic version of (α,δ). Using Assumptions 2.1-2.5 and a method similar to that in [31], it is easy to see that on Ξ_n,

$$\bar{D}^{**-1}\bar{Q}^{**}\sum_{t=1}^{n}\left(R_t|_{\tilde{\alpha},\tilde{\delta}}-R_t\right)\bar{Q}^{**\prime}\bar{D}^{**-1}=o_p(1), \tag{A.9}$$

$$\bar{D}^{**-1}\bar{Q}^{**}\sum_{t=1}^{n}\left(\nabla_\alpha l_t|_{\tilde{\alpha},\tilde{\delta}}-\nabla_\alpha l_t\right)=\bar{D}^{**-1}\bar{Q}^{**}\sum_{t=1}^{n}R_t(\tilde{\alpha}-\alpha)+o_p(1), \tag{A.10}$$

where R_t and $\nabla_\alpha l_t$ are evaluated at the true parameters (α,δ).

Denote

$$\begin{aligned}&\dot{Q}_1^{**}=(B_\perp\otimes I_r)\left(I_m\otimes(\dot{B}\bar{B}')'\right),\quad \dot{Q}_2^{**}=\mathrm{diag}\left((\dot{B}\bar{B}')^{-1}\otimes I_m,I_{(s-1)m^2}\right),\\ &\dot{\alpha}_1=vec\left((\dot{B}\bar{B}')^{-1}\hat{B}\right),\quad \dot{\alpha}_2=vec\left[\hat{A}(\dot{B}\bar{B}'),\hat{\Phi}_1^*,\dots,\hat{\Phi}_{s-1}^*\right],\quad \dot{\alpha}=\left[\dot{\alpha}_1',\dot{\alpha}_2'\right]'.\end{aligned}$$

It follows from the assertions (b), (c) and (d) of Lemma A.2 that $(\dot{\alpha},\hat{\delta})\in\Xi_n$. Thus, by (A.9) and the block-diagonality of R_t,

$$\begin{aligned}n^{-2}\sum_{t=1}^{n}\dot{Q}_1^{**}\left(R_{1t}|_{\dot{\alpha},\hat{\delta}}\right)\dot{Q}_1^{**\prime}&=n^{-2}\sum_{t=1}^{n}(B_\perp\otimes I_r)\left(R_{1t}|_{\dot{\alpha},\hat{\delta}}\right)(B_\perp'\otimes I_r)\\ &=n^{-2}\sum_{t=1}^{n}(B_\perp\otimes I_r)R_{1t}(B_\perp'\otimes I_r)+o_p(1),\end{aligned} \tag{A.11}$$

$$n^{-1}\sum_{t=1}^{n}\dot{Q}_2^{**}\left(R_{2t}|_{\dot{\alpha},\hat{\delta}}\right)\dot{Q}_2^{**\prime}=n^{-1}\sum_{t=1}^{n}\left(R_{2t}|_{\dot{\alpha},\hat{\delta}}\right)=n^{-1}\sum_{t=1}^{n}R_{2t}+o_p(1). \tag{A.12}$$

Refer to (A.8). By (A.9), (A.10) and the block-diagonality of R_t,

$$\begin{aligned}
n^{-1}\sum_{t=1}^{n}\dot{Q}_1^{**}\big(\nabla_{\alpha_1}l_t|_{\hat{\alpha},\hat{\delta}}\big) &= n^{-1}\sum_{t=1}^{n}(B_\perp\otimes I_r)\big(\nabla_{\alpha_1}l_t|_{\dot{\alpha},\hat{\delta}}\big)\\
&= n^{-1}\sum_{t=1}^{n}(B_\perp\otimes I_r)\nabla_{\alpha_1}l_t+\left[n^{-1}\sum_{t=1}^{n}(B_\perp\otimes I_r)R_{1t}(B'_\perp\otimes I_r)\right]\\
&\quad\times\big(\bar{B}_\perp\otimes I_r\big)(\dot{\alpha}_1-\alpha_1)+o_p(1),\\
n^{-\frac{1}{2}}\sum_{t=1}^{n}\dot{Q}_2^{**}\big(\nabla_{\alpha_2}l_t|_{\hat{\alpha},\hat{\delta}}\big) &= n^{-\frac{1}{2}}\sum_{t=1}^{n}\big(\nabla_{\alpha_2}l_t|_{\dot{\alpha},\hat{\delta}}\big)\\
&= n^{-\frac{1}{2}}\sum_{t=1}^{n}\nabla_{\alpha_2}l_t+\left(n^{-1}\sum_{t=1}^{n}R_{2t}\right)(\dot{\alpha}_2-\alpha_2)+o_p(1).
\end{aligned}\tag{A.13}$$

Recall that $\dot{Q}_1^{**\prime-1}\hat{\alpha}_1=(\bar{B}_\perp\otimes I_r)\dot{\alpha}_1$. By (4.5), (A.11) and (A.13),

$$\begin{aligned}
n\dot{Q}_1^{**\prime-1}\dot{\alpha}_1 &= n\dot{Q}_1^{**\prime-1}\hat{\alpha}_1-\left[n^{-2}\sum_{t=1}^{n}\dot{Q}_1^{**}\big(R_{1t}|_{\hat{\alpha},\hat{\delta}}\big)\dot{Q}_1^{**\prime}\right]^{-1}\left[n^{-1}\sum_{t=1}^{n}\dot{Q}_1^{**}\big(\nabla_{\alpha_1}l_t|_{\hat{\alpha},\hat{\delta}}\big)\right]\\
&= n(\bar{B}_\perp\otimes I_r)\dot{\alpha}_1-\left[n^{-2}\sum_{t=1}^{n}(B_\perp\otimes I_r)R_{1t}(B'_\perp\otimes I_r)\right]^{-1}\\
&\quad\times\left[n^{-1}\sum_{t=1}^{n}(B_\perp\otimes I_r)\nabla_{\alpha_1}l_t\right]-n(\bar{B}_\perp\otimes I_r)(\dot{\alpha}_1-\alpha_1)+o_p(1)\\
&= n(\bar{B}_\perp\otimes I_r)\alpha_1-\left[n^{-2}\sum_{t=1}^{n}(B_\perp\otimes I_r)R_{1t}(B'_\perp\otimes I_r)\right]^{-1}\\
&\quad\times\left[n^{-1}\sum_{t=1}^{n}(B_\perp\otimes I_r)\nabla_{\alpha_1}l_t\right]+o_p(1).
\end{aligned}\tag{A.14}$$

Note that

$$\dot{Q}_1^{**\prime-1}\dot{\alpha}_1-(\bar{B}_\perp\otimes I_r)\alpha_1=vec\big[\big((\dot{B}\bar{B}')^{-1}\dot{B}-B\big)\bar{B}'_\perp\big].$$

The item (a) now follows from (A.14) and the assertions (a), (b) of Lemma A.1.

On the other hand, by (4.6), (A.12) and (A.13),

$$\begin{aligned}
\sqrt{n}\dot{Q}_2^{**\prime-1}\dot{\alpha}_2 &= \sqrt{n}\dot{Q}_2^{**\prime-1}\hat{\alpha}_2-\left[n^{-1}\sum_{t=1}^{n}\dot{Q}_2^{**}\big(R_{2t}|_{\hat{\alpha},\hat{\delta}}\big)\dot{Q}_2^{**\prime}\right]^{-1}\left[n^{-\frac{1}{2}}\sum_{t=1}^{n}\dot{Q}_2^{**}\big(\nabla_{\alpha_2}l_t|_{\hat{\alpha},\hat{\delta}}\big)\right]\\
&= \sqrt{n}\dot{\alpha}_2-\left[n^{-1}\sum_{t=1}^{n}R_{2t}\right]^{-1}\left[n^{-\frac{1}{2}}\sum_{t=1}^{n}\nabla_{\alpha_2}l_t\right]-\sqrt{n}(\dot{\alpha}_2-\alpha_2)+o_p(1)
\end{aligned}$$

$$=\sqrt{n}\alpha_2-\left[n^{-1}\sum_{t=1}^{n}R_{2t}\right]^{-1}\left[n^{-\frac{1}{2}}\sum_{t=1}^{n}\nabla_{\alpha_2}l_t\right]+o_p(1). \tag{A.15}$$

The item (b) now follows from (A.15) and the assertions (a), (b) of Lemma A.1. This completes the proof. □

Proof of Lemma 5.1. From (5.4),

$$LR_G=(\dot{\varphi}-\ddot{\varphi})'\left(-\sum_{t=1}^{n}\dot{F}_t\right)(\dot{\varphi}-\ddot{\varphi})+o_p(1), \tag{A.16}$$

where the full-rank estimator

$$\begin{aligned}
&\dot{\varphi}=vec\left[\dot{C},\dot{\Phi}_1^*,\ldots,\dot{\Phi}_{s-1}^*\right], \quad \ddot{\varphi}\equiv vec\left[\dot{A}\dot{B},\dot{\Phi}_1^*,\ldots,\dot{\Phi}_{s-1}^*\right],\\
&\dot{F}_t=-\left(X_{t-1}X_{t-1}'\otimes\dot{V}_{t-1}^{-1}\right)-\frac{1}{4}\left(\nabla_{\varphi}\dot{h}_{t-1}\right)\dot{D}_{t-1}^{-2}\left(\dot{\Gamma}^{-1}\odot\dot{\Gamma}+I_m\right)\dot{D}_{t-1}^{-2}\left(\nabla_{\varphi}'\dot{h}_{t-1}\right).
\end{aligned}$$

Denote $\ddot{A}=\dot{A}(\dot{B}\bar{B}')$ and $\ddot{B}=(\dot{B}\bar{B}')^{-1}\dot{B}$. Note $\dot{A}\dot{B}=\ddot{A}\ddot{B}$. Moreover,

$$\ddot{A}\ddot{B}-AB=(\ddot{A}-A)B+A(\ddot{B}-B)+(\ddot{A}-A)(\ddot{B}-B).$$

Recall that $B\bar{B}_\perp'=0_{rxd}$. By Theorem 4.2, $(\ddot{B}-B)\bar{B}_\perp'=O_p(n^{-1})$ and $(\ddot{A}-A)=O_p(n^{-1/2})$ under H_0. Hence,

$$\begin{aligned}
n(\ddot{A}\ddot{B}-AB)\bar{B}_\perp'&=n(\ddot{A}-A)B\bar{B}_\perp'+nA(\ddot{B}-B)\bar{B}_\perp'+(\ddot{A}-A)n(\ddot{B}-B)\bar{B}_\perp'\\
&=nA(\ddot{B}-B)\bar{B}_\perp'+O_p\left(n^{-\frac{1}{2}}\right).
\end{aligned} \tag{A.17}$$

On the other hand, by Theorem 4.1(a) and the arguments in [18, Lemma 13.2], $(\ddot{B}-B)=O_p(n^{-1})$. Therefore,

$$\begin{aligned}
\sqrt{n}(\ddot{A}\ddot{B}-AB)\bar{B}'&=\sqrt{n}(\ddot{A}-A)B\bar{B}'+\sqrt{n}\ddot{A}(\ddot{B}-B)\bar{B}'\\
&=\sqrt{n}(\ddot{A}-A)+O_p\left(n^{-\frac{1}{2}}\right).
\end{aligned} \tag{A.18}$$

But from the proofs of Theorem 4.2(b) and Theorem 3.1(b),

$$\sqrt{n}(\ddot{A}-A)-\sqrt{n}(\dot{C}-C)\bar{B}'=o_p(1). \tag{A.19}$$

All in all, by (A.17)-(A.19), (A.16) can be rewritten as

$$LR_G=vec\left[n(\dot{C}-A\ddot{B})\bar{B}_\perp'\right]'\left[n^{-2}\sum_{t=1}^{n}L_{1t}\right]vec\left[n(\dot{C}-A\ddot{B})\bar{B}_\perp'\right]+o_p(1), \tag{A.20}$$

where

$$L_{1t}=\left(Z_{1t-1}Z_{1t-1}'\otimes V_{t-1}^{-1}\right)+\sum_{l=1}^{t-1}\Big[Z_{1t-1-l}Z_{1t-1-l}'\otimes\left(\left(\Gamma^{-1}\odot\Gamma+I_m\right)\odot\nu_l\nu_l'\odot\Pi_{lt}\right)\Big],$$

$$Z_{1t-1}=B_\perp Y_{t-1}.$$

By Lemma A.1(b), Theorem 3.1(a) and Theorem 4.2(a),

$$\begin{aligned}
&n^{-2}\sum_{t=1}^{n}L_{1t}\ \longrightarrow_{\mathcal{L}}\ Z\otimes\Omega_1,\\
&n\dot{C}\bar{B}_\perp'\ \longrightarrow_{\mathcal{L}}\ \Omega_1^{-1}M^*,\\
&nA\ddot{B}\bar{B}_\perp'\ \longrightarrow_{\mathcal{L}}\ A\left(A'\Omega_1^{-1}A\right)^{-1}A'M^*,
\end{aligned}$$

where

$$Z\equiv\psi_{11}\Omega_{a_1}^{\frac{1}{2}}\left[\int_0^1 B_d(u)B_d(u)'du\right]\Omega_{a_1}^{\frac{1}{2}}\psi_{11}'$$

and M^* is as defined in Theorem 3.1. Therefore,

$$\begin{aligned}
LR_G\ \longrightarrow_{\mathcal{L}}\ &vec\Big[\left(\Omega_1^{-1}-A\left(A'\Omega_1A\right)^{-1}A'\right)M^*\Big]'(Z\otimes\Omega_1)\\
&\times vec\Big[\left(\Omega_1^{-1}-A\left(A'\Omega_1A\right)^{-1}A'\right)M^*\Big]\\
=&\operatorname{tr}\Big[\left(\Omega_1^{-1}-A\left(A'\Omega_1A\right)^{-1}A'\right)M^*ZM^{*'}\Big]. \qquad \text{(A.21)}
\end{aligned}$$

Following the lines on [38, p. 359], we can rewrite $\Omega_1^{-1}-A(A'\Omega_1A)^{-1}A'$ as

$$\Omega_1^{-1}\left(\Omega_1-\Omega_1A\left(A'\Omega_1A\right)^{-1}A'\Omega_1\right)\Omega_1^{-1}=\Omega_1^{-1}A_\perp\left(A_\perp'\Omega_1^{-1}A_\perp\right)^{-1}A_\perp'\Omega_1^{-1}.$$

Therefore, the asymptotic distribution in (A.21) can be rewritten as

$$\operatorname{tr}\left[\left(\int_0^1 B_d(u)dV_d^*(u)'\right)'\left(\int_0^1 B_d(u)B_d(u)'du\right)^{-1}\left(\int_0^1 B_d(u)dV_d^*(u)'\right)\right], \qquad \text{(A.22)}$$

where

$$V_d^*(u)\equiv\left(A_\perp'\Omega_1^{-1}A_\perp\right)^{-\frac{1}{2}}A_\perp'\Omega_1^{-1}W_m^*(u).$$

By the definitions around (3.7), we can write $B_d(u)$ as

$$\left(\bar{A}_\perp'(EV_{t-1})\bar{A}_\perp\right)^{-\frac{1}{2}}\bar{A}_\perp'W_m(u)=\left(A_\perp'(EV_{t-1})A_\perp\right)^{-\frac{1}{2}}A_\perp'W_m(u).$$

Therefore,

$$
\begin{aligned}
E\left[B_d(u)V_d^*(u)'\right]=&\left(A_\perp'(EV_{t-1})A_\perp\right)^{-\frac{1}{2}}A_\perp' E\left[W_m(u)W_m^*(u)'\right]\Omega_1^{-1}A_\perp \\
&\times\left(A_\perp'\Omega_1^{-1}A_\perp\right)^{-\frac{1}{2}} \\
=&\,u\left(A_\perp'(EV_{t-1})A_\perp\right)^{-\frac{1}{2}}\left(A_\perp'\Omega_1^{-1}A_\perp\right)^{\frac{1}{2}}=u\Upsilon'.
\end{aligned}
$$

Thus, we can rewrite $V_d^*(u)$ as a linear combination of two independent d-dimensional standard BMs

$$
\begin{aligned}
\Upsilon B_d(u)+\Big[&\left(A_\perp'\Omega_1^{-1}A_\perp\right)^{-\frac{1}{2}}A_\perp'\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}A_\perp \\
&\times\left(A_\perp'\Omega_1^{-1}A_\perp\right)^{-\frac{1}{2}}-\Upsilon\Upsilon'\Big]^{\frac{1}{2}}V_d(u).
\end{aligned}\tag{A.23}
$$

The proof is complete. $\square$

Proof of Theorem 5.1. Similar to (A.20) in the proof of Lemma 5.1, we have

$$
H_G=vec\left[n(\dot{C}-A\ddot{B})\bar{B}_\perp'\right]'\left[n^{-2}\sum_{t=1}^{n}L_{1t}^H\right]vec\left[n(\dot{C}-A\ddot{B})\bar{B}_\perp'\right]+o_p(1),\tag{A.24}
$$

where by the construction of F_t^H in (5.6),

$$
L_{1t}^H=\left(Z_{1t-1}Z_{1t-1}'\otimes A_\perp\left(A_\perp'\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}A_\perp\right)^{-1}A_\perp'\right).
$$

By Lemma A.1(b),

$$
n^{-2}\sum_{t=1}^{n}L_{1t}^H \ \longrightarrow_{\mathcal{L}}\ Z\otimes A_\perp\left(A_\perp'\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}A_\perp\right)^{-1}A_\perp',
$$

where Z is as defined in the proof of Lemma 5.1 and M^* is as defined in Theorem 3.1. Therefore, similar to (A.21),

$$
\begin{aligned}
H_G\ \longrightarrow_{\mathcal{L}}\ & vec\left[\left(\Omega_1^{-1}-A(A'\Omega_1A)^{-1}A'\right)M^*\right]'(Z\otimes A_\perp\Sigma A_\perp') \\
&\times vec\left[\left(\Omega_1^{-1}-A(A'\Omega_1A)^{-1}A'\right)M^*\right],
\end{aligned}
$$

where $\Sigma\equiv(A_\perp'\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}A_\perp)^{-1}$. However, as argued in the proof of Lemma 5.1,

$$
\Omega_1^{-1}-A\left(A'\Omega_1A\right)^{-1}A'=\Omega_1^{-1}A_\perp\left(A_\perp'\Omega_1^{-1}A_\perp\right)^{-1}A_\perp'\Omega_1^{-1}.
$$

Therefore,

$$H_G \;\longrightarrow_{\mathcal{L}}\; \mathrm{tr}[\Omega M^* Z M^{*'}],$$

where

$$\begin{aligned}\Omega \equiv & \left(\Omega_1^{-1}A_\perp(A_\perp'\Omega_1^{-1}A_\perp)^{-1}A_\perp'\Omega_1^{-1}\right)(A_\perp\Sigma A_\perp')\\ &\times\left(\Omega_1^{-1}A_\perp(A_\perp'\Omega_1^{-1}A_\perp)^{-1}A_\perp'\Omega_1^{-1}\right)\\ =&\,\Omega_1^{-1}A_\perp\Sigma A_\perp'\Omega_1^{-1}.\end{aligned}$$

As a result, we can rewrite the asymptotic distribution in (A.25) as

$$\mathrm{tr}\left[\left(\int_0^1 B_d(u)dV_d^H(u)'\right)'\left(\int_0^1 B_d(u)B_d(u)'du\right)^{-1}\left(\int_0^1 B_d(u)dV_d^H(u)'\right)\right], \quad \text{(A.25)}$$

where $V_d^H(u)\equiv\Sigma^{-1/2}A_\perp'\Omega_1^{-1}W_m^*(u)$. Therefore, contrast to $E[B_d(u)V_d^*(u)']$ in Lemma 5.1,

$$E\left[B_d(u)V_d^H(u)'\right]=u\left(A_\perp'(EV_{t-1})A_\perp\right)^{-\frac{1}{2}}(A_\perp'\Omega_1^{-1}A_\perp)\Sigma^{\frac{1}{2}}=u\Upsilon^{H'}.$$

Thus, we can rewrite $V_d^H(u)$ as a linear combination of two independent d-dimensional standard BMs

$$\begin{aligned}&\Upsilon^H B_d(u)+\left[\Sigma^{\frac{1}{2}}A_\perp'\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}A_\perp\Sigma^{\frac{1}{2}}-\Upsilon^H\Upsilon^{H'}\right]^{\frac{1}{2}}V_d(u)\\ =&\,\Upsilon^H B_d(u)+\left[I_d-\Upsilon^H\Upsilon^{H'}\right]^{\frac{1}{2}}V_d(u).\end{aligned} \quad \text{(A.26)}$$

Thus, the asymptotic distribution can be simplified as

$$\begin{aligned}\mathrm{tr}\Bigg\{&\left[\int_0^1\Upsilon^H B_d(u)dB_d(u)'\Upsilon^{H'}+\int_0^1\Upsilon^H B_d(u)dV_d(u)'\left(I_d-\Upsilon^H\Upsilon^{H'}\right)^{\frac{1}{2}}\right]'\\ &\times\left[\int_0^1\Upsilon^H B_d(u)B_d(u)'\Upsilon^{H'}du\right]^{-1}\\ &\times\left[\int_0^1\Upsilon^H B_d(u)dB_d(u)'\Upsilon^{H'}+\int_0^1\Upsilon^H B_d(u)dV_d(u)'(I_d-\Upsilon^H\Upsilon^{H'})^{\frac{1}{2}}\right]\Bigg\}.\end{aligned}$$

However, $\Upsilon^H B_d(u)\sim N(0,\Upsilon^H\Upsilon^{H'})$. Abusing the notation, we write $\Upsilon^H B_d(u)$ as $(\Upsilon^H\Upsilon^{H'})^{1/2}\ B_d(u)$, where $B_d(u)$ is (another) d-dimensional standard BM independent of $V_d(u)$. Therefore, cancelling some of the $(\Upsilon^H\Upsilon^{H'})^{1/2}$ terms, the asymptotic distribution can be expressed as

$$\begin{aligned}\operatorname{tr}\Bigg\{&\left[\int_0^1 B_d(u)dB_d(u)'\big(\Upsilon^H\Upsilon^{H'}\big)^{\frac{1}{2}}+\int_0^1 B_d(u)dV_d(u)'\big(I_d-\Upsilon^H\Upsilon^{H'}\big)^{\frac{1}{2}}\right]'\\&\times\left[\int_0^1 B_d(u)B_d(u)'du\right]^{-1}\\&\times\left[\int_0^1 B_d(u)dB_d(u)'\big(\Upsilon^H\Upsilon^{H'}\big)^{\frac{1}{2}}+\int_0^1 B_d(u)dV_d(u)'\big(I_d-\Upsilon^H\Upsilon^{H'}\big)^{\frac{1}{2}}\right]\Bigg\}.\end{aligned}$$

Since $(I_d-\Upsilon^H\Upsilon^{H'})$ is a real symmetric matrix, we can decompose it as $\Theta\Lambda_d^H\Theta'$, where Θ is an orthogonal matrix such that $\Theta'\Theta=I_d$. In view of $(\Upsilon^H\Upsilon^{H'})^{1/2}=\Theta(I_d-\Lambda_d^H)^{1/2}\Theta'$ and $(I_d-\Upsilon^H\Upsilon^{H'})^{1/2}=\Theta\Lambda_d^{H1/2}\Theta'$ and due to the orthogonality of Θ, we can write the asymptotic distribution as

$$\begin{aligned}\operatorname{tr}\Bigg\{&\left[\int_0^1 \Theta'B_d(u)dB_d(u)'\Theta\big(I_d-\Lambda_d^H\big)^{\frac{1}{2}}\Theta'+\int_0^1 \Theta'B_d(u)dV_d(u)'\Theta\Lambda_d^{H\frac{1}{2}}\Theta'\right]'\\&\times\left[\int_0^1 \Theta'B_d(u)B_d(u)'du\Theta\right]^{-1}\\&\times\left[\int_0^1 \Theta'B_d(u)dB_d(u)'\Theta\big(I_d-\Lambda_d^H\big)^{\frac{1}{2}}\Theta'+\int_0^1 \Theta'B_d(u)dV_d(u)'\Theta\Lambda_d^{H\frac{1}{2}}\Theta'\right]\Bigg\}.\end{aligned}$$

Since $\Theta'B_d(u)\sim N(0,\Theta'\Theta)=N(0,I_d)$, similar to the previous arguments, and abusing the notation, we can write $\Theta'B_d(u)$ and $\Theta'V_d(u)$ as two independent standard BMs $B_d(u)$ and $V_d(u)$ respectively. Cancelling the orthogonal Θ, we have

$$\begin{aligned}&\operatorname{tr}\Bigg\{\left[\int_0^1 B_d(u)dB_d(u)'\big(I_d-\Lambda_d^H\big)^{\frac{1}{2}}+\int_0^1 B_d(u)dV_d(u)'\Lambda_d^{H\frac{1}{2}}\right]'\\&\qquad\times\left[\int_0^1 B_d(u)B_d(u)'du\right]^{-1}\\&\qquad\times\left[\int_0^1 B_d(u)dB_d(u)'\big(I_d-\Lambda_d^H\big)^{\frac{1}{2}}+\int_0^1 B_d(u)dV_d(u)'\Lambda_d^{H\frac{1}{2}}\right]\Bigg\}\\&=\operatorname{tr}\left\{\left[\zeta\big(I_d-\Lambda_d^H\big)^{\frac{1}{2}}+\Phi\Lambda_d^{H\frac{1}{2}}\right]'\left[\zeta\big(I_d-\Lambda_d^H\big)^{\frac{1}{2}}+\Phi\Lambda_d^{H\frac{1}{2}}\right]\right\}.\end{aligned}\tag{A.27}$$

The proof is complete. □

Proof of Theorem 5.2. When $\Omega_1^*=\Omega_1$, $\Upsilon^H=\Upsilon$ and $\Lambda_d^H=\Lambda_d$, then by Theorem 5.1,

$$\begin{aligned}LR_G=H_G+o_p(1)\ \longrightarrow_{\mathcal{L}}\ &\mathrm{tr}\left\{\left[\zeta\big(I_d-\Lambda_d^H\big)^{\frac{1}{2}}+\Phi\Lambda_d^{H\frac{1}{2}}\right]'\left[\zeta\big(I_d-\Lambda_d^H\big)^{\frac{1}{2}}+\Phi\Lambda_d^{H\frac{1}{2}}\right]\right\}\\ =&\,\mathrm{tr}\left\{\left[\zeta(I_d-\Lambda_d)^{\frac{1}{2}}+\Phi\Lambda_d^{\frac{1}{2}}\right]'\left[\zeta(I_d-\Lambda_d)^{\frac{1}{2}}+\Phi\Lambda_d^{\frac{1}{2}}\right]\right\}.\end{aligned}$$

This completes the proof. □

Appendix B. Additional formulae

We first let $\vec{\varepsilon}_t=(\varepsilon_{1t}^2,\ldots,\varepsilon_{mt}^2)'$. Further let $\gamma_l=(\gamma_{1l},\ldots,\gamma_{ml})'$, for each $i=1,\ldots,m$, γ_{il} is implicitly defined such that

$$\left(1-\sum_{l=1}^{p}b_{il}L^l\right)^{-1}=\sum_{l=0}^{\infty}\gamma_{il}L^l.$$

Refer to the discussion right before (3.3)-(3.4)

$$\begin{aligned}\nabla_\varphi h_{t-1}&=-2\sum_{l=1}^{q}(X_{t-1-l}\otimes I_m)\mathrm{diag}(a_l\odot\varepsilon_{t-l})+\sum_{l=1}^{p}(\nabla_\varphi h_{t-1-l})\mathrm{diag}(b_l)\\ &=-2\sum_{l=1}^{t-1}(X_{t-1-l}\otimes I_m)\mathrm{diag}(\nu_l\odot\varepsilon_{t-l}),\end{aligned}\tag{B.1}$$

$$\nabla_{\delta_1}h_{t-1}=\big(\nabla'_{a_0}h_{t-1};\nabla'_{a_1}h_{t-1},\ldots,\nabla'_{a_q}h_{t-1};\nabla'_{b_1}h_{t-1},\ldots,\nabla'_{b_p}h_{t-1}\big)',\tag{B.2}$$

where

$$\begin{aligned}\nabla_{a_0}h_{t-1}&=I_m+\sum_{l=1}^{p}(\nabla_{a_0}h_{t-1-l})\mathrm{diag}(b_l)=\sum_{l=0}^{\infty}\mathrm{diag}(\gamma_l),\\ \nabla_{a_j}h_{t-1}&=\mathrm{diag}(\vec{\varepsilon}_{t-j})+\sum_{l=1}^{p}(\nabla_{a_j}h_{t-1-l})\mathrm{diag}(b_l)\\ &=\sum_{l=0}^{\infty}\mathrm{diag}(\gamma_l\odot\vec{\varepsilon}_{t-l-j}),\qquad j=1,\ldots,q,\\ \nabla_{b_j}h_{t-1}&=\mathrm{diag}(h_{t-1-j})+\sum_{l=1}^{p}(\nabla_{b_j}h_{t-1-l})\mathrm{diag}(b_l)\end{aligned}$$

$$=\sum_{l=0}^{\infty}\mathrm{diag}(\gamma_l\odot h_{t-1-l-j}),\quad j=1,\ldots,p.$$

Next refer to the discussion right before (4.5)-(4.6),

$$\begin{aligned}\nabla_{\alpha_1}h_{t-1}&=-2\sum_{l=1}^{q}\big(Y_{t-1-l}\otimes A'\big)\mathrm{diag}(a_l\odot\varepsilon_{t-l})+\sum_{l=1}^{p}(\nabla_{\alpha_1}h_{t-1-l})\mathrm{diag}(b_l)\\&=-2\sum_{l=1}^{t-1}\big(Y_{t-1-l}\otimes A'\big)\mathrm{diag}(\nu_l\odot\varepsilon_{t-l}),\end{aligned}\tag{B.3}$$

$$\begin{aligned}\nabla_{\alpha_2}h_{t-1}&=-2\sum_{l=1}^{q}(U_{t-1-l}\otimes I_m)\mathrm{diag}(a_l\odot\varepsilon_{t-l})+\sum_{l=1}^{p}(\nabla_{\alpha_2}h_{t-1-l})\mathrm{diag}(b_l)\\&=-2\sum_{l=1}^{t-1}(U_{t-1-l}\otimes I_m)\mathrm{diag}(\nu_l\odot\varepsilon_{t-l}).\end{aligned}\tag{B.4}$$

Refer to the discussion around (3.4). When $m=2$,

$$\Psi_2=\begin{pmatrix}1&0&0&0\\0&0&0&1\end{pmatrix},\tag{B.5}$$

$$N_2=\begin{pmatrix}1&0&0&0\\0&\frac{1}{2}&\frac{1}{2}&0\\0&\frac{1}{2}&\frac{1}{2}&0\\0&0&0&1\end{pmatrix},\quad \tilde{L_2}=\begin{pmatrix}0&1&0&0\end{pmatrix}.\tag{B.6}$$

See respectively [36, pp.109, 48-49, 96-97] for details.

Refer to the discussion around (5.8)

$$\begin{aligned}\nabla_{\alpha_1}h^H_{t-1}&=-2\sum_{l=1}^{q}\big(Y_{t-1-l}\otimes A^{H'}\big)\mathrm{diag}(a_l\odot\varepsilon_{t-l})+\sum_{l=1}^{p}(\nabla_{\alpha_1}h_{t-1-l})\mathrm{diag}(b_l)\\&=-2\sum_{l=1}^{t-1}\big(Y_{t-1-l}\otimes A^{H'}\big)\mathrm{diag}(\nu_l\odot\varepsilon_{t-l}).\end{aligned}\tag{B.7}$$

References

[1] S. K. Ahn and G. C. Reinsel, *Estimation for partially nonstationary multivariate models*, J. Amer. Statist. Assoc. 85 (1990), 813–823.

[2] T. W. Anderson, *Estimating linear restrictions on regression coefficients for multivariate normal distributions*, Ann. Math. Statist. 22(3) (1951), 327–351.

[3] T. W. Anderson, *Reduced rank regression in cointegrated models*, J. Econometrics 106(2) (2002), 203–216.

[4] T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*, J. Econometrics, 31(3) (1986), 307–327.

[5] T. Bollerslev, *Modelling the coherence in the short-run nominal exchange rates: A multivariate generalized ARCH approach*, Rev. Econ. Stat. 72(3) (1990), 498–505.

[6] T. Bollerslev and J. M. Wooldridge, *Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances*, Econom. Rev. 11(2) (1992), 143–172.

[7] B. Q. Cai, J. T. Gao, and D. Tjøstheim, *A new class of bivariate threshold cointegration models*, J. Bus. Econ. Stat. 35(2) (2017), 288–305.

[8] G. Cavaliere, L. De Angelis, A. Rahbek, and A. M. R. Taylor, *Determining the cointegration rank in heteroskedastic var models of unknown order*, Econometric Theory, 34(2) (2018), 349–382.

[9] G. Cavaliere, H. B. Nielsen, and A. Rahbek, *Bootstrap testing of hypotheses on cointegration relations in vector autoregressive models*, Econometrica 83 (2015), 813–831.

[10] N. H. Chan and R. M. Zhang, *Inference for unit-root models with infinite variance GARCH errors*, Statist. Sinica 20(4) (2010), 1363–1393.

[11] R. F. Engle, *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*, Econometrica 50(4) (1982), 987–1007.

[12] R. Engle, *Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models*, J. Bus. Econ. Stat. 20(3) (2002), 339–350.

[13] R. F. Engle and C. W. J. Granger, *Co-integration and error correction: Representation, estimation and testing*, Econometrica 55(2) (1987), 251–276.

[14] C. Francq and J. M. Zakoian, *GARCH Models: Structure, Statistical Inference and Financial Applications*, John Wiley & Sons, 2019.

[15] P. H. Franses, P. Kofman, and J. Moser, *Garch effects on a test for cointegration*, Rev. Quant. Finance Account. 4(1) (1994), 19–26.

[16] C. Gourieroux and A. Monfort, *A general framework for testing a null hypothesis in a "mixed" form*, J. Econometrics 5(1) (1989), 63–82.

[17] C. W. J. Granger, *Co-Integrated Variables and Error Correction Models*, Discussion Paper, University of California at San Diego, 1983.

[18] S. Johansen, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, 1995.

[19] S. Johansen, *Times Series: Cointegration*, Centre for Research in Econometric Analysis of Time Series, Aarhus University, 2014.

[20] S. Johansen and M. Nielsen, *Nonstationary cointegration in the fractionally cointegrated VAR model*, J. Time Series Anal. 40(4) (2019), 519–543.

[21] K. Juselius, *Domestic and foreign effects on prices in an open economy: The case of Denmark*, J. Policy Model. 14(4) (1992), 401–428.

[22] D. Kristensen and A. Rahbek, *Testing and inference in nonlinear cointegrating vector error correction models*, Econometric Theory 29(6) (2013), 1238–1288.

[23] T. Lange, *Tail behavior and OLS estimation in AR-GARCH models*, Statist. Sinica 21 (2011), 1191–1200.

[24] T. Lange, A. Rahbek, and S. T. Jensen, *Estimation and asymptotic inference in the AR-ARCH model*, Econom. Rev. 30(2) (2011), 129–153.

[25] T. H. Lee and Y. Tse, *Cointegration tests with conditional heteroskedasticity*, J. Econometrics 73(2) (1996), 401–410.

[26] G. Li and W. K. Li, *Least absolute deviation estimation for unit root processes with GARCH errors*, Econometric Theory 25(5) (2009), 1208–1227.

[27] W. K. Li, S. Ling, and H. Wong, *Estimation for partially nonstationary multivariate autoregressive models with conditional heteroskedasticity*, Biometrika 88(4) (2001), 1135–1152.

[28] Z. Liang, Z. Lin, and C. Hsiao, *Local linear estimation of a nonparametric cointegration model*, Econometric Rev. 34(6-10) (2015), 881–905.

[29] Y. Lin, Y. Tu, and Q. Yao, *Estimation for double-nonlinear cointegration*, J. Econometrics 216(1) (2020), 175–191.

[30] S. Ling and W. K. Li, *Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors*, J. Time Series Anal. 18(5) (1997), 447–464.

[31] S. Ling and W. K. Li, *Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors*, Ann. Statist. 26(1) (1998), 84–125.

[32] S. Ling and W. K. Li, *Asymptotic inference for unit root with GARCH(1,1) errors*, Econometric Theory 19(4) (2003), 541–564.

[33] S. Ling, W. K. Li, and M. McAleer, *Estimation and testing for unit root processes with GARCH(1,1) errors: Theory and Monte Carlo study*, Econometric Rev. 22(2) (2003), 179–202.

[34] S. Ling and M. McAleer, *Asymptotic theory for a vector ARMA-GARCH model*, Econometric Theory 19(2) (2003), 280–310.

[35] S. Ling and M. McAleer, *On adaptive estimation in nonstationary ARMA models with GARCH errors*, Ann. Statist. 31(2) (2003), 642–674.

[36] J. R. Magnus, *Linear Structures*, Griffin, 1988.

[37] P. C. B. Phillips and S. N. Durlauf, *Multiple time series regression with integrated processes*, Rev. Econ. Stud. 53(4) (1986), 473–495.

[38] G. C. Reinsel and S. K. Ahn, *Vector autoregressive models with unit root and reduced rank structure: Estimation likelihood ratio test, and forecasting*, J. Time Series Anal. 13(4) (1992), 353–375.

[39] B. Seo, *Distribution theory for unit root tests with conditional heteroskedasticity*, J. Econometrics 91(1) (1999), 113–144.

[40] R. She and S. Ling, *Inference in heavy-tailed vector error correction models*, J. Econometrics 214(2) (2020), 433–450.

[41] K. Shinki and Z. J. Zhang, *Asymptotic theory for fractionally integrated asymmetric power ARCH models*, J. Statist. Plann. Inference 142(11) (2012), 2871–2890.

[42] J. H. Stock and M. W. Watson, *Testing for common trends*, J. Amer. Statist. Assoc. 83(404) (1988), 1097–1107.

[43] J. H. Stock and M. W. Watson, *A simple estimator of cointegrating vectors in higher order integrated system*, Econometrica 61(4) (1993), 783–820.

[44] Y. K. Tse, *A test for constant correlations in a multivariate GARCH model*, J. Econometrics 98(1) (2000), 107–127.

[45] Q. Y. Wang, *Martingale limit theorem revisited and nonlinear cointegrating regression*, Econometric Theory 30(3) (2014), 509–535.

[46] Q. Y. Wang and P. C. B. Phillips, *A specification test for nonlinear nonstationary models*, Ann. Statist. 40(2) (2012), 727–758.

[47] R. M. Zhang, C. Li, and L. Peng, *Inference for the tail index of a GARCH(1,1) model and an AR(1) model with ARCH(1) errors*, Econometric Rev. 38(2) (2019), 151–169.

[48] R. M. Zhang and S. Ling, *Asymptotic inference for AR models with heavy-tailed G-GARCH noises*, Econometric Theory 31(4) (2015), 880–890.