Exploiting the GARCH Effects to Test for Cointegrating Rank[a]

Chor-yiu SIN
Wang Yanan Institute for Studies in Economics (WISE), Xiamen University, Xiamen, Fujian, China
Email: cysinhkbu@gmail.com

Shiqing LING[b]
Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong SAR, China
Email: maling@ust.hk

January 20, 2008

Abstract

This paper develops a test for conintegarting rank that exploits the possible conditional heteroskedasticity. We make use, and modify, the asymptotic distributions of the full rank and the reduced rank quasi-maximum likelihood (QMLE) estimators developed in Li, Ling and Wong (2001) and Seo (2007). The former distribution belongs to the class of locally asymptotically Brownian functional (LABF) while the latter belongs to the class of locally asymptotic mixed normal (LAMN). The modification is in virtue to (a) the widely used constant-correlation GARCH model adopted in this paper; (b) more importantly, no prior knowledge of the cointegrating structure. These two estimators are used to construct a likelihood ratio (LR) test for cointegrating rank, where the asymptotic distribution is in turn a functional of a standard Brownian motion and a standard normal vector, with d to-be-estimated nuisance parameters, where d is the difference between the number of variables and the cointegrating rank. To the best of our knowledge, tests of this sort do not exist in the literature. The critical values of the LR test can be simulated via the Monte Carlo method. The size and the power of this test are compared, via Monte Carlo experiments, with those of a test that ignores conditional heteroskedasticity. We also apply our test to an empirical example of three interest rates.

Keywords: Autoregressive model; Cointegrating structure; Cointegration; Full rank estimation; Multivariate GARCH process; Price discovery; Reduced rank estimation

JEL Classifications: C32, C51, G14

---

[a] This is a substantial revision of the paper, "Estimation and Testing for Partially Nonstationary Vector Autoregressive Models with GARCH". The current version of this paper was presented at the 2007 International Symposium on Financial Engineering and Risk Management (FERM2007) and the 2007 Singapore Econometrics Study Group (SESG) Meeting. We thank helpful comments from H. Peter Boswijk, Zongwu Cai, Ying Fang, Heejoon Han, Yongmiao Hong, Roberto Mariano, Jiazhu Pan, Peter Phillips, Yiu Kuen Tse, Qiwei Yao, Jun Yu and other participants at these conferences. The usual disclaimers apply.

[b] Send correspondence to: Shiqing LING, Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. Tel.: +852-2358-7459; Fax.: +852-2359-1016; Email: maling@ust.hk.

1 Introduction

First investigated in a series of papers such as Granger (1983) and Engle and Granger (1987), cointegration has been a leading topic in the literature of economics. Notable examples include consumption function, purchasing power parity, money demand function, hedging ratio of spot and futures exchange rates, and yield curves of different terms of maturities. The partially nonstationary multivariate AR model or cointegrated time series models without GARCH have been extensively discussed over the past twenty years. Other notable examples include Phillips and Durlauf (1986) and Stock and Watson (1993).

Economic time series related to financial markets often exhibit time-varying variances such as GARCH. Recently, a considerable number of papers, including Kroner and Sultan (1993), Brenner and Kroner (1995), Hasbrouck (1995, 2003), Alexakis and Apergis (1996), Gonzalo and Ng (2001), Li, Ling and Wong (2001) (henceforth LLW (2001)) and Seo (2007), investigate multivariate time series that exhibit both cointegration and time-varying variances. On the other hand, while Franses, Kofman and Moser (1994) and Lee and Tse (1996) perform Monte Carlo experiments on different tests for cointegration which ignore the presence of GARCH, the related asymptotic theory has yet been fully developed.

As a sequel of LLW (2001), this paper continues to generalize the univariate unit root process with time-varying variance to multivariate cases. We develop a test for conintegarting rank that exploits the possible conditional heteroskedasticity such as GARCH. We set off with the full rank estimation (estimation that does not impose the cointegrating restriction), which asymptotic theory resembles that in LLW (2001). Instead of the random coefficient AR model that excludes lags of conditional variances, we adopt the constant-correlation GARCH model first suggested by Bollerslev (1990), which is in turn contrast to that suggested in Seo (2007). Though the proof of the asymptotic theory is similar (see Lemma B.1 below), the gradient and the Hessian for estimation are somewhat different, and more involved than those in Seo (2007). More importantly, the difference is substantial in practice, and our model provides a much wider scope of applications. See, for instance, Tse (2000). We then consider the reduced rank estimation (estimation that imposes the cointegrating restriction). Contrast to LLW (2001) and Seo (2007), we do not assume we know a priori some components of the vector series are non-conitegrated. As we are testing for the cointegrating rank, it is unclear how we know that some components are non-cointegrated. Our approach is similar to that of Anderson (1951,2002) and Johansen (1988,1996). More precisely, as in Johansen (1988, 1996), we adopt the normalization factors $\left(\widetilde{\beta}'\bar{\beta}\right)^{-1}$ and $\left(\widetilde{\beta}'\bar{\beta}\right)$, where $\widetilde{\beta}'$ is the reduced-rank estimator for the cointegrating vectors. See Theorems 4.2 and 4.3. However, the Johansen's approach of computing eigenvectors is not directly applicable. All in all, our proof is somewhat different from those in Johansen (1988,1996), LLW (2001) and Seo (2007). In particular, we need, in the intermediate steps, to consider the asymptotic properties of $\widehat{\alpha}\left(\widetilde{\beta}'\bar{\beta}\right)$ and $\left(\widetilde{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}$, where $\left(\widehat{\alpha},\widehat{\beta}\right)$ is the initial estimator. See Lemma B.2 and the proofs of Theorems 4.2 and 4.3.

The main thrust of this paper is using these two estimators to construct a likelihood ratio (LR) test for cointegrating rank. To the best of our knowledge, tests of this sort do not exist in the literature. When the standardized error η_t is normally distributed, the null distribution is a functional of a standard Brownian motion and a standard normal vector with d to-be-estimated nuisance parameters, where d is the difference between the number of variables and the cointegrating rank. If η_t is not necessarily normal, the null distribution a functional of standard Brownian motions, with $d(d+1)$ nuisance parameters. Following the lines in Gourieroux and Monfort (1989), we also derive a Hausman-type test, which null distribution contains only d nuisance parameters. The critical value can be simulated via Monte Carlo method. In the univariate case, Seo (1999) and Ling, Li and McAleer (2003) construct a unit root test with the (Q)MLE of an AR-GARCH model. Their simulation results show that the unit root test based on the (Q)MLE is much more powerful than the Dickey Fuller test based on the least squares estimation (LSE). It is expected that the LR test based on the(Q)MLE of process (1.1)-(1.3) is more powerful than Johansen's test or Reinsel-Ahn's test which ignores GARCH.

Throughout this paper, we consider a $p-$dimensional autoregressive (AR) process $\{X_t\}$, which is generated by

$$X_t = \Pi_1 X_{t-1} + \cdots + \Pi_k X_{t-k} + \epsilon_t, \tag{1.1}$$

$$\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{pt})', \tag{1.2}$$

$$\epsilon_{it} = \eta_{it}\sqrt{h_{it-1}}, \;\; h_{it-1} = a_{i0} + \sum_{l=1}^{q} a_{il}\epsilon_{it-l}^2 + \sum_{l=1}^{s} b_{il}h_{it-1-l}, \tag{1.3}$$

where Π_j's are constant matrices. In (1.3), $\eta_t = (\eta_{1t}, \ldots, \eta_{pt})'$ is a sequence of independently and identically distributed (i.i.d.) random vectors with zero mean and $E(\eta_t\eta_t') = \Lambda \equiv (\lambda_{ij})_{p\times p}$, a positive definite matrix with $\lambda_{ii} = 1$ and $\lambda_{ij} = \lambda_{ji}$. It is easy to see that $E(\epsilon_t|\mathcal{F}_{t-1}) = 0$ and $E(\epsilon_t\epsilon_t'|\mathcal{F}_{t-1}) = V_{t-1} = D_{t-1}\Lambda D_{t-1}$, where $\mathcal{F}_{t-1} = \sigma\{\eta_\tau, \tau = t-1, t-2, \ldots\}$ and $D_{t-1} = diag(\sqrt{h_{1t-1}}, \ldots, \sqrt{h_{pt-1}})$. In other words, V_{t-1} is the conditional covariance matrix with constant correlation. We use the subscript "$t-1$" rather than "t", in order to signify that the conditional covariance has already realized in "$t-1$". The process ϵ_t in (1.2)-(1.3) is the multivariate generalized autoregressive conditional heteroskedasticity (GARCH) process of constant correlations. Assuming the ϵ_t's are i.i.d., under further conditions on Π_j's (see Assumptions 2.1 and 2.2 below), Johansen (1988, 1996), and Ahn and Reinsel (1990) show that, although some component series of $\{X_t\}$ exhibit nonstationary behavior, there are r linear combinations of $\{X_t\}$ that are stationary, that is, the components of $\{X_t\}$ are cointegrated.

This paper proceeds as follows. Section 2 discusses the structure of model (1.1)-(1.3). Section 3 and Section 4 derive the asymptotic distribution of the full rank estimator and the reduced rank estimator, respectively. Section 5 devises tests for cointegrating rank. Section 6 contains the practical details. It includes a sub-section on an illustrative case of $p = 2$, another on an illustrative case of $p = 3$, one on estimating the nuisance parameters, while the method of simulating critical values with some illustrative examples are also found. We report the Monte Carlo experiments and an empirical example of three interest rates in Sections 7 and 8 respectively. Conclusions can be found in Section 9. All the technical proofs are relegated to Appendix B. This paper focuses on the case in which there is not a constant term in the error-correction model (ECM). Discussions on a specific model with a constant term can be found in the end of Section 5. 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. L is denoted as the lag operator while $\Delta = 1 - L$.

2 Basic Properties of the Models

In this section, we discuss the basic properties of the models with some preliminary assumptions. Assumptions 2.1-2.2 are about the mean equation specified in Model (1.1); while Assumptions 2.3-2.5 are about the variance equation specified in Models (1.2)-(1.3). The first set is essentially adopted from Johansen (1988,1996) while the second set is essentially adopted from Ling and McAleer (2003a). Refer to (1.1) and define $\Pi(L) = I_p - \sum_{j=1}^{k} \Pi_j L^j$. We first make the following assumption.

Assumption 2.1. $\mid \Pi(z) \mid = 0$ implies that either $\mid z \mid > 1$ or $z = 1$. □

Re-parameterize process (1.1) as:

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \epsilon_t, \tag{2.1}$$

where $\Pi \equiv -\left(I_p - \sum_{l=1}^{k} \Pi_l\right)$, $\Gamma_j \equiv -\sum_{l=j+1}^{k} \Pi_l$. Following Johansen (1988,1996) and Ahn and Reinsel (1990), we decompose $\Pi = \alpha\beta'$, where α and β are two $p \times r$ matrices of rank r, $0 \leq r \leq p$. Define $d \equiv p - r$. Denote $\beta_\perp$ as a $p \times d$ matrix of full rank such that $\beta'\beta_\perp = 0_{r\times d}$. Define $\bar{\beta} = \beta(\beta'\beta)^{-1}$ and $\bar{\beta}_\perp = \beta_\perp(\beta_\perp'\beta_\perp)^{-1}$. Similarly, denote $\alpha_\perp$ as a $p \times d$ matrix of full rank such that $\alpha'\alpha_\perp = 0_{r\times d}$, and define $\bar{\alpha} = \alpha(\alpha'\alpha)^{-1}$ and $\bar{\alpha}_\perp = \alpha_\perp(\alpha_\perp'\alpha_\perp)^{-1}$.

Partition the true β' such that $\beta' = [\beta_1', \beta_2']$ where β_1 is $r \times r$ and β_2 is $d \times r$. In Ahn and Reinsel (1990) (see also Phillips and Durlauf, 1986, LLW, 2001 and Seo, 2007), it is assumed that we know a priori that the last d components of X_t are non-cointegrated. Given this piece of information, one can write $\Pi = \alpha\beta_1'[I_r, \beta_1'^{-1}\beta_2']$. All in all, in our notation, Ahn and Reinsel (1990) consistently estimate the parameters $\alpha\beta_1'$ and $\beta_1'^{-1}\beta_2'$. In contrast, we adopt the approach in Anderson (1951, 2002) and Johansen (1988, 1996) and do not assume this piece of information.

Next we impose the following condition:
Assumption 2.2. $0 \leq r < p$. $\mid \alpha_\perp' \Gamma \beta_\perp \mid \neq 0$, where $\Gamma \equiv \left(I_p - \sum_{j=1}^{k-1} \Gamma_j\right)$. □

Given Assumptions (2.1) and (2.2), by the proof of Theorem (4.2) in Johansen (1996), we have an

AR representation for the two linear combinations of X_t, $\beta' X_t$ and $\Delta\beta'_\perp X_t$:

$$\widetilde{A}(L)\left[\begin{array}{c}\beta' X_t\\ \Delta\beta'_\perp X_t\end{array}\right]=\left[\begin{array}{c}\bar{\alpha}'\\ \bar{\alpha}'_\perp\end{array}\right]\epsilon_t,\ \widetilde{A}(1)=\left[\begin{array}{cc}-I_r & \bar{\alpha}'\Gamma\bar{\beta}_\perp\\ 0_{d\times r} & \bar{\alpha}'_\perp\Gamma\bar{\beta}_\perp\end{array}\right]. \tag{2.2}$$

See (4.16) and (4.14) in Johansen (1996). Unlike Johansen (1988,1996) and Reinsel and Ahn (1990), we do not assume that ϵ_t's are i.i.d. Nevertheless, given Assumption 2.3 below, ϵ_t is stationary by the arguments in Ling and Li (1998). Thus, given Assumptions 2.1 and 2.2, so are $\beta' X_t$ and $\Delta\beta'_\perp X_t$.

Lastly we make the following assumptions on (1.2)-(1.3).

Assumption 2.3. For $i=1,\ldots,p$, $a_{i0}>0$, $a_{i1},\ldots,a_{iq};b_{i1},\ldots,b_{is}\geq 0$, and $\sum_{l=1}^{q}a_{il}+\sum_{l=1}^{s}b_{il}<1$. □

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

$$A_{it}=\left(\begin{array}{cccccc}a_{i1}\eta_{it}^2 & \ldots & a_{iq}\eta_{it}^2 & b_{i1}\eta_{it}^2 & \ldots & b_{is}\eta_{it}^2\\ & I_{q-1} & 0_{(q-1)\times 1} & & 0_{(q-1)\times s} & \\ a_{i1} & \ldots & a_{iq} & b_{i1} & \ldots & b_{is}\\ & 0_{(s-1)\times q} & & & I_{s-1} & 0_{(s-1)\times 1}\end{array}\right). \qquad \square$$

Assumption 2.5. η_t is symmetrically distributed. □

Assumptions 2.3-2.4 are the necessary and sufficient conditions for $E\left(vec\left[\epsilon_t\epsilon_t'\right]vec\left[\epsilon_t\epsilon_t'\right]'\right)<\infty$, see Ling and McAleer (2003a). Assumption 2.5 allows the mean parameters (see Sections 3 and 4) and the variance parameters (see Appendix A) to be estimated separately without altering the asymptotic distributions. The symmetry condition is also assumed in LLW (2001) and Seo (2007). We will have a discussion on that in Section 9.

3 Full Rank Estimation

Refer to Processes (2.2) and (1.3). In this section, we consider the full rank estimators for the mean parameters $\varphi\equiv vec[\Pi,\Gamma_1,\ldots,\Gamma_{k-1}]$ and the estimators for the variance parameters δ (see Appendix A for the details). Given $\{X_t: t=1,\cdots,T\}$, conditional on the initial values $X_t=0$ for $t\leq 0$, the log-likelihood function (LF) (with a constant ignored), as a function of the generic φ and δ, can be written as:

$$l_F(\varphi,\delta)=\sum_{t=1}^{T}l_t(\varphi,\delta)\ \text{ and }\ l_t(\varphi,\delta)\ =\ -\frac{1}{2}\epsilon_t' V_{t-1}^{-1}\epsilon_t-\frac{1}{2}\ln|V_{t-1}|, \tag{3.1}$$

where $\epsilon_t=\epsilon_t(\varphi)$ and $V_{t-1}=V_{t-1}(\varphi,\delta)$. Suppress the argument (φ,δ) when there is no ambiguity. Further denote the $p\times 1$- vector of variances as $h_{t-1}=(h_{1t-1},\ldots,h_{pt-1})'$. As argued in Appendix B, given the symmetry Assumption 2.5, φ and δ can be estimated separately without altering the asymptotic distributions. In the balance of this section, we confine our attention to estimating φ. The gradient and the (simplified) Hessian can be expressed as:

$$\nabla_\varphi l_t \ =\ (Y_{t-1}\otimes I_p)V_{t-1}^{-1}\epsilon_t-\frac{1}{2}\nabla_\varphi h_{t-1}w\left(D_{t-1}^{-2}\left(I_p-\epsilon_t\epsilon_t' V_{t-1}^{-1}\right)\right), \tag{3.2}$$

$$F_t \ =\ -\left(Y_{t-1}Y_{t-1}'\otimes V_{t-1}^{-1}\right)-\nabla_\varphi h_{t-1}\left(\Psi_p\left(\Lambda^{-1}D_{t-1}^{-2}\otimes D_{t-1}^{-2}\Lambda\right)\Psi_p'+D_{t-1}^{-4}\right)\nabla_\varphi' h_{t-1}/4, \tag{3.3}$$

where $Y_{t-1} = [X'_{t-1}, \Delta X'_{t-1}, \ldots, \Delta X'_{t-k+1}]'$, and for any square matrix χ, $w(\chi)$ is a vector containing the diagonal elements of χ, $\Psi'_p w(\chi) = vec(\chi)$. And $\nabla_\varphi h_{t-1}$ is recursively defined as:

$$\nabla_\varphi h_{t-1} = -2\sum_{l=1}^{q}(Y_{t-1-l} \otimes I_p) diag\left(a_{1l}\epsilon_{1t-l}, \ldots, a_{pl}\epsilon_{pt-l}\right) + \sum_{l=1}^{s}(\nabla_\varphi h_{t-1-l}) diag\left(b_{1l}, \ldots, b_{pl}\right).$$

See Section 6 for the details of $\nabla_\varphi l_t$ and F_t of the illustrative cases in which $p = 2$ and $p = 3$.

We first find an initial estimator $(\hat{\varphi}, \hat{\delta}) \in \Theta_T^{(F)}$, where $\Theta_T^{(F)}$ is defined in the proof of Theorem 3.2 below. For instance, we may adopt the LSE (least squares estimator) considered in LLW (2001) and use the residuals of the VAR for estimating the variance parameters. (See, for instance, Ling, Li and McAleer, 2003.) Given this initial estimator, we perform a one-step iteration:

$$\dot{\varphi} = \widehat{\varphi} - \left(\sum_{t=1}^{T} F_t|_{\widehat{\varphi},\widehat{\delta}}\right)^{-1} \left(\sum_{t=1}^{T} \nabla_\varphi l_t|_{\widehat{\varphi},\widehat{\delta}}\right), \tag{3.4}$$

Remark 3.1. The first terms in (3.2) and (3.3) are exactly the same as the counterparts in LLW (2001) (see 4.3 and the definition of M_t there). The only differences are the second terms in (3.2) and (3.3) due to a different model of conditional heteroskedasticity.

Remark 3.2. In practice, we may repeat the iterative procedure in (3.4), and get an estimator closer to the (quasi-) maximum likelihood estimator, (Q)MLE. The asymptotic distribution is not altered though.

Remark 3.3. Due to the different heteroskedasticity model, the algorithm of this one-step iteration is somewhat different from that in LLW (2001). However, the proofs of our Theorems 3.1 and 3.2 below are similar to that of Theorem 2 in LLW (2001). Thus we simply provide a sketchy proof in Appendix B.

Recall the full rank estimator $\dot{\varphi} = vec[\dot{\Pi}, \dot{\Gamma}_1, \ldots, \dot{\Gamma}_{k-1}]$. In Theorems 3.1 and 3.2 below, we state the asymptotic distribution of $T(\dot{\Pi} - \Pi)\bar{\beta}_\perp$ (the "nonstationary" mean parameters), and that of $\sqrt{T} vec[(\dot{\Pi} - \Pi)\bar{\beta}, (\dot{\Gamma}_1 - \Gamma_1), \ldots, (\dot{\Gamma}_{k-1} - \Gamma_{k-1})]$ (the "stationary" mean parameters). By the definition of $\bar{\beta}_\perp$, $\Pi\bar{\beta}_\perp = \alpha\beta'\beta_\perp(\beta'_\perp\beta_\perp)^{-1} = 0_{p\times d}$. On the other hand, in Theorem 3.1(a), $\Pi\bar{\beta} = \alpha\beta'\beta\,(\beta'\beta)^{-1} = \alpha$.

To facilitate the discussions, we first consider the heteroskedasticity model (1.2)-(1.3). For $i = 1, 2, \ldots, p$, ν_{il} is implicitly defined such that $\left(\sum_{l=1}^{q} a_{il}L^l\right)\left(1 - \sum_{l=1}^{s} b_{il}L^l\right)^{-1} = \sum_{l=0}^{\infty} \nu_{il}L^l$. We define:

$$e_{it-1} = \sum_{l=1}^{\infty} \nu_{il}\epsilon_{it-l},\ i = 1, \ldots, p, \text{and } e_{t-1} = diag\left(e_{1t-1}, \ldots, e_{pt-1}\right).$$

Refer to (3.2), we further define:

$$\epsilon_t^* = V_{t-1}^{-1}\epsilon_t + e_{t-1} w\left(D_{t-1}^2\left(I_p - \epsilon_t\epsilon_t' V_{t-1}^{-1}\right)\right).$$

Next we let $E\left[w\left(\eta_t\eta_t'\Lambda^{-1}\right)\left(w\left(\eta_t\eta_t'\Lambda^{-1}\right)\right)'\right] \equiv \left(\Delta_{ij}^*\right)_{p\times p}$. Corresponding to $\left\{\epsilon_t', \epsilon_t^{*'}\right\}'$, we define a $2p-$dimensional Brownian motion (BM) $(W_p'(u), W_p^{*\prime}(u))'$ with the covariance matrix:

$$u\begin{pmatrix} EV_{t-1} & I_p \\ I_p & \Omega_1^* \end{pmatrix}, \tag{3.5}$$

where $\Omega_1^* = E(V_{t-1}^{-1}) + \Sigma^*, \Sigma^* = \left(\Sigma_{ij}^*\right)_{p\times p}, \Sigma_{ij}^* = (\Delta_{ij}^* - 1)\sum_{l=1}^{\infty}\nu_{il}\nu_{jl}E\left(\frac{\epsilon_{it-l}\epsilon_{jt-l}}{h_{it-1}h_{jt-1}}\right)$. Define a $d-$dimensional standard BM,

$$B_d(u) \equiv \left(\alpha_\perp' EV_{t-1}\alpha_\perp\right)^{-1/2}\alpha_\perp' W_p(u). \tag{3.6}$$

Next we define a functional of the two correlated Brownian motions $W_p^*(u)$ and $B_d(u)$.

$$M^* = \left(\int_0^1 B_d(u)dW_p^*(u)'\right)'\left(\int_0^1 B_d(u)B_d(u)'du\right)^{-1}\left(\alpha_\perp' EV_{t-1}\alpha_\perp\right)^{-1/2}\left(\alpha_\perp'\Gamma\bar{\beta}_\perp\right). \tag{3.7}$$

In Theorem 3.1, we consider the asymptotic distribution of the MLE. That is, we assume $\eta_t \sim N(0,\Lambda)$. The case of QMLE can be found in Theorem 3.2.

Theorem 3.1. Suppose Assumptions 2.1-2.5 hold and $\eta_t \sim N(0,\Lambda)$. Then

$$\begin{aligned}(a)\qquad & T(\dot{\Pi}-\Pi)\bar{\beta}_\perp = T\dot{\Pi}\bar{\beta}_\perp \\ = \quad & \Omega_1^{*-1}\left(T^{-1}\sum_{t=1}^{T}\epsilon_t^* W_{t-1}'\right)\left(T^{-2}\sum_{t=1}^{T}W_{t-1}W_{t-1}'\right)^{-1} + o_P(1) \longrightarrow_{\mathcal{L}} \Omega_1^{*-1}M^*, \\ (b)\qquad & \sqrt{T}vec\left[\left(\dot{\Pi}\bar{\beta}-\alpha\right),\left(\dot{\Gamma}_1-\Gamma_1\right),\ldots,\left(\dot{\Gamma}_{k-1}-\Gamma_{k-1}\right)\right] \\ = \quad & -\left[T^{-1}\sum_{t=1}^{T}R_{2t}\right]^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\nabla_{\phi_2}l_t\right] + o_P(1) \longrightarrow_{\mathcal{L}} N(0,\Omega_2^{*-1}),\end{aligned}$$

where Ω_1^* and M^* are defined in (3.5) and (3.7) respectively. $W_{t-1} = \beta_\perp' X_{t-1}$, $\Omega_2^* = E(\nabla_{\phi_2}l_t\nabla_{\phi_2}'l_t)$, $\nabla_{\phi_2}l_t$ is defined in (4.6), and R_{2t} is defined in (4.8). □

Remark 3.4. As in Theorem 2(a) of LLW (2001), the asymptotic distribution of $T(\dot{\Pi}-\Pi)\bar{\beta}_\perp$ is a functional of two correlated high-dimensional Brownian motions.

Recall that $\Lambda = (\lambda_{ij})_{p\times p}$. Denote $\Lambda^{-1} = (\lambda^{ij})_{p\times p}$, and $\Delta = (\Delta_{ij})_{p\times p}$. For $i=j$, define $\Delta_{ii} = 1+\lambda^{ii}$; and for $i \neq j$, define $\Delta_{ij} = \lambda^{ij}\lambda_{ij}$.

$$\Omega_1 = E(V_{t-1}^{-1}) + \Sigma, \Sigma = (\Sigma_{ij})_{p\times p}, \Sigma_{ij} = \Delta_{ij}\sum_{l=1}^{\infty}\nu_{il}\nu_{jl}E\left(\frac{\epsilon_{it-l}\epsilon_{jt-l}}{h_{it-1}h_{jt-1}}\right). \tag{3.8}$$

Theorem 3.2. Suppose Assumptions 2.1-2.5 hold. Then

$$(a)\qquad T(\dot{\Pi}-\Pi)\bar{\beta}_\perp = T\dot{\Pi}\bar{\beta}_\perp$$

$$= \quad \Omega_1^{-1}\left(T^{-1}\sum_{t=1}^{T}\epsilon_t^* W_{t-1}'\right)\left(T^{-2}\sum_{t=1}^{T}W_{t-1}W_{t-1}'\right)^{-1} + o_P(1) \longrightarrow_{\mathcal{L}} \Omega_1^{-1}M^*,$$

$$(b) \qquad \sqrt{T}vec[(\dot{\Pi}\bar{\beta}-\alpha),(\dot{\Gamma}_1-\Gamma_1),\ldots,(\dot{\Gamma}_{k-1}-\Gamma_{k-1})]$$

$$= \quad -\left[T^{-1}\sum_{t=1}^{T}R_{2t}\right]^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\nabla_{\phi_2}l_t\right] + o_P(1) \longrightarrow_{\mathcal{L}} N(0,\Omega_2^{-1}\Omega_2^*\Omega_2^{-1}),$$

where Ω_1 and M^* are defined in (3.8) and (3.7) respectively. $\Omega_2 = -E(R_{2t})$, and the remaining variables are as defined in Theorem 3.1. □

Remark 3.5. The proof of Theorem 3.2(b) is standard while that of Theorem 3.2(a) follows the lines in the proof of Theorem 2(a) in LLW (2001), with different definitions of Ω_1^* and M^*.

Remark 3.6. It is not difficult to see that the asymptotic distribution of $T\dot{\Pi}\bar{\beta}_\perp$ in Theorem 3.1 or 3.2 belongs to the class of locally asymptotically Brownian functional (LABF). When the error η_t is normal, and h_{it-1}'s are not constant, $\dot{\Pi}$ is more efficient than the LSE of Π in Ahn and Reinsel (1990). See the discussions in Ling and McAleer (2003b).

4 Reduced Rank Estimation

We first rewrite (2.1) in a reduced rank form:

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{j=1}^{k-1}\Gamma_j \Delta X_{t-j} + \epsilon_t, \tag{4.1}$$

where α and β are as defined in Section 2.

In this section, we consider the reduced rank estimator for $\phi = [\phi_1', \phi_2']'$ with $\phi_1 = vec[\beta']$ and $\phi_2 = vec[\alpha, \Gamma_1, \ldots, \Gamma_{k-1}]$. Note the initial estimator for (α, β) suggested in Section 5 of Ahn and Reinsel (1990) are not applicable in our context, as we do not assume that the components of X_t are correctly arranged such that the last d components are non-cointegrated. Instead, we adopt Anderson (1951,2002)'s or Johansen (1988,1996)'s approach to obtain an initial estimator. The asymptotic properties are shown in Sub-section 4.1. In Sub-section 4.2, we use this initial estimator and suggest a reduced rank estimator that incorporates GARCH.

4.1 Initial Estimator for the Mean Parameters

This initial estimator is essentially the (Q)MLE 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 (1951,2002)'s or Johansen (1988,1996)'s estimator.

Denote this estimator as $\widehat{\phi} = [\widehat{\phi}_1', \widehat{\phi}_2']'$ with $\widehat{\phi}_1 = vec[\widehat{\beta}']$ and $\widehat{\phi}_2 = vec[\widehat{\alpha}, \widehat{\Gamma}_1, \ldots, \widehat{\Gamma}_{k-1}]$. Using the arguments in Lemma 13.2 of Johansen (1996) and Lemma B.1 below, we obtain the asymptotic distribution of the normalized $\widehat{\phi}_1$ and that of the normalized $\widehat{\phi}_2$. As in Johansen (1988,1996), $\widehat{\beta}'$ is pre-multiplied by $(\widehat{\beta}'\bar{\beta})^{-1}$ while $\widehat{\alpha}$ is post-multiplied by $\widehat{\beta}'\bar{\beta}$. A sketchy proof can be found in Appendix B.

Theorem 4.1. Suppose Assumptions 2.1-2.5 hold. Then

$$(a) \qquad T\left((\widehat{\beta}'\bar{\beta})^{-1}\widehat{\beta}' - \beta'\right)\bar{\beta}_\perp = T(\widehat{\beta}'\bar{\beta})^{-1}\widehat{\beta}'\bar{\beta}_\perp \longrightarrow_{\mathcal{L}} \left(\alpha'\left(EV_{t-1}\right)^{-1}\alpha\right)^{-1}\alpha'\left(EV_{t-1}\right)^{-1}M,$$

$$(b) \qquad \sqrt{T}vec\left[\left(\widehat{\alpha}\left(\widehat{\beta}'\bar{\beta}\right) - \alpha\right), \left(\widehat{\Gamma}_1 - \Gamma_1\right), \ldots, \left(\widehat{\Gamma}_{k-1} - \Gamma_{k-1}\right)\right] \longrightarrow_{\mathcal{L}} N\left(0, \Sigma_2^{-1}\Sigma_2^*\Sigma_2^{-1}\right),$$

where $M = (\int_0^1 B_d(u)dW_p(u)')'(\int_0^1 B_d(u)B_d(u)'du)^{-1}\left(\alpha_\perp' EV_{t-1}\alpha_\perp\right)^{-1/2}\left(\alpha_\perp'\Gamma\bar{\beta}_\perp\right)$, $\Sigma_2 = E(U_{t-1}U_{t-1}' \otimes I_p)$, $\Sigma_2^* = E(U_{t-1}U_{t-1}' \otimes V_{t-1})$, and the remaining variables are as defined in Theorem 3.1. □

Remark 4.1. The normalization factors $(\widehat{\beta}'\bar{\beta})^{-1}$ and $\widehat{\beta}'\bar{\beta}$ in (a) and (b) respectively have been found very useful in deriving a lot of hypothesis testing, when time-varying variances are not considered. Moreover, Theorem 4.1 is used in proving Lemma B.2, which in turn plays an important role in proving Theorem 4.3.

Remark 4.2. The asymptotic distribution in (a) is exactly the same as that in Lemma 13.2 in Johansen (1996), regardless of the presence of conditional heteroskedasticity (at least of that specified in (1.2)-(1.3)). In view of this result and other regularity conditions, the test for reduced rank considered in Johansen (1988,1996) can be shown to have the correct asymptotic size. See also Franses, Kofman and Moser (1994) and Lee and Tse (1996) for some results of Monte-Carlo experiments.

Remark 4.3. Suppose V_{t-1} is a constant matrix, say V. $\Sigma_2^{-1}\Sigma_2^*\Sigma_2^{-1} = ((EU_{t-1}U_{t-1}')^{-1} \otimes V)$ and thus the asymptotic distribution in (b) is exactly the same as that in Johansen (1996) (see, for instance, Lemma 13.2 there).

Remark 4.4. If, as in Ahn and Reinsel (1990) (see also LLW, 2001 and Seo, 2007), the components of X_t can be arranged so that the last d components are non-cointegrated, we can decompose $\widehat{\beta}' = [\widehat{\beta}_1', \widehat{\beta}_2']$, where $\widehat{\beta}_1$ is $r \times r$ and $\widehat{\beta}_2$ is $d \times r$. Provided that $\widehat{\beta}_1$ is invertible, it is easy to show that

$$T\left(\widehat{\beta}_1'^{-1}\widehat{\beta}_2' - \beta_1'^{-1}\beta_2'\right) \longrightarrow_{\mathcal{L}} \left(\alpha'\left(EV_{t-1}\right)^{-1}\alpha\right)^{-1}\alpha'\left(EV_{t-1}\right)^{-1}MP_{21}^{-1}, \tag{4.2}$$

$$\sqrt{T}vec[(\widehat{\alpha}\widehat{\beta}_1' - \alpha\beta_1'), (\widehat{\Gamma}_1 - \Gamma_1), \ldots, (\widehat{\Gamma}_{k-1} - \Gamma_{k-1})] \longrightarrow_{\mathcal{L}} N(0, \Sigma_2^{-1}\Sigma_2^*\Sigma_2^{-1}), \tag{4.3}$$

where P_{21} is a $d \times d$ matrix such that $\bar{\beta}_\perp \equiv [P_{11}', P_{21}']'$. The distribution in (4.2) is exactly the same as that in Ahn and Reinsel (1990), if their Jordan canonical form exists and $\alpha = \bar{\beta}$ up to an $r \times r$ invertible matrix.

4.2 Reduced Rank Estimation that Incorporates GARCH

This sub-section uses the initial estimator $\widehat{\phi}$ in Sub-section 4.1 and $\widehat{\delta}$ suggested in Li, Ling and McAleer (2003) to obtain a new reduced rank estimation that incorporates GARCH. The likelihood function (LF) based on the error-correction form is similar to that in (3.1), but now it is a function of the generic ϕ and δ instead. The log-likelihood function can be written as:

$$l_F(\phi,\delta)=\sum_{t=1}^{T}l_t(\phi,\delta) \text{ and } l_t(\phi,\delta) = -\frac{1}{2}\epsilon_t' V_{t-1}^{-1}\epsilon_t-\frac{1}{2}\ln|V_{t-1}|, \tag{4.4}$$

where $\epsilon_t=\epsilon_t(\phi)$ and $V_{t-1}=V_{t-1}(\phi,\delta)$. Suppress the argument (ϕ,δ) when there is no ambiguity. As argued in Appendix B, given the symmetry Assumption 2.5, ϕ and δ can be estimated separately without altering the asymptotic distributions. In the balance of this section, we confine our attention to estimating ϕ. The estimation for δ can be found in Appendix A. The gradient (with respect to ϕ) can be expressed as:

$$\nabla_{\phi_1}l_t = (X_{t-1}\otimes\alpha')V_{t-1}^{-1}\epsilon_t-\frac{1}{2}\nabla_{\phi_1}h_{t-1}w\left(D_{t-1}^{-2}\left(I_p-\epsilon_t\epsilon_t'V_{t-1}^{-1}\right)\right), \tag{4.5}$$

$$\nabla_{\phi_2}l_t = (U_{t-1}\otimes I_p)V_{t-1}^{-1}\epsilon_t-\frac{1}{2}\nabla_{\phi_2}h_{t-1}w\left(D_{t-1}^{-2}\left(I_p-\epsilon_t\epsilon_t'V_{t-1}^{-1}\right)\right), \tag{4.6}$$

where $U_{t-1}=[X_{t-1}'\beta,\Delta X_{t-1}',\ldots,\Delta X_{t-k+1}']'$. The (simplified) Hessians can be expressed as:

$$R_{1t} = -\left(X_{t-1}X_{t-1}'\otimes\alpha' V_{t-1}^{-1}\alpha\right)-\nabla_{\phi_1}h_{t-1}\left(\Psi_p\left(\Lambda^{-1}D_{t-1}^{-2}\otimes D_{t-1}^{-2}\Lambda\right)\Psi_p'+D_{t-1}^{-4}\right)\nabla_{\phi_1}'h_{t-1}/4; \tag{4.7}$$

$$R_{2t} = -\left(U_{t-1}U_{t-1}'\otimes V_{t-1}^{-1}\right)-\nabla_{\phi_2}h_{t-1}\left(\Psi_p\left(\Lambda^{-1}D_{t-1}^{-2}\otimes D_{t-1}^{-2}\Lambda\right)\Psi_p'+D_{t-1}^{-4}\right)\nabla_{\phi_2}'h_{t-1}/4, \tag{4.8}$$

where $\nabla_{\phi_1}h_{t-1}$ and $\nabla_{\phi_2}h_{t-1}$ are recursively defined as:

$$\nabla_{\phi_1}h_{t-1} = -2\sum_{l=1}^{q}(X_{t-1-l}\otimes\alpha')diag\left(a_{1l}\epsilon_{1t-l},\ldots,a_{pl}\epsilon_{pt-l}\right)+\sum_{l=1}^{s}(\nabla_{\phi_1}h_{t-1-l})diag\left(b_{1l},\ldots,b_{pl}\right);$$

$$\nabla_{\phi_2}h_{t-1} = -2\sum_{l=1}^{q}(U_{t-1-l}\otimes I_p)diag\left(a_{1l}\epsilon_{1t-l},\ldots,a_{pl}\epsilon_{pt-l}\right)+\sum_{l=1}^{s}(\nabla_{\phi_2}h_{t-1-l})diag\left(b_{1l},\ldots,b_{pl}\right).$$

See Section 6 for the details of $\nabla_{\phi_1}l_t$, $\nabla_{\phi_2}l_t$, R_{1t} and R_{2t} of the illustrative cases in which $p=2$ and $p=3$.

Given an initial estimator $(\widehat{\phi},\widehat{\delta})\in\Theta_T^{(R)}$ ($\Theta_T^{(R)}$ is defined in the proof of Theorem 4.3), we perform a one-step iteration:

$$\widetilde{\phi}_1 = \widehat{\phi}_1-\left(\sum_{t=1}^{T}R_{1t}|_{\widehat{\phi},\widehat{\delta}}\right)^{-1}\left(\sum_{t=1}^{T}\nabla_{\phi_1}l_t|_{\widehat{\phi},\widehat{\delta}}\right), \tag{4.9}$$

$$\widetilde{\phi}_2 = \widehat{\phi}_2-\left(\sum_{t=1}^{T}R_{2t}|_{\widehat{\phi},\widehat{\delta}}\right)^{-1}\left(\sum_{t=1}^{T}\nabla_{\phi_2}l_t|_{\widehat{\phi},\widehat{\delta}}\right), \tag{4.10}$$

Remark 4.5. As for the full rank estimation, in practice, we may repeat the iterative procedure in (4.9)-(4.10), and get an estimator closer to the (Q)MLE.

The asymptotic distribution of the normalized $\widetilde{\phi}_1$ and that of the normalized $\widetilde{\phi}_2$ are given in Theorems 4.2 and 4.3 below. Theorem 4.2 deals with the MLE in which we assume $\eta_t\sim N(0,\Lambda)$. The general case of QMLE can be found in Theorem 4.3. Similar to Johansen (1988,1996), in Theorems 4.2 and 4.3, $\widetilde{\beta}'$ is pre-multiplied by $(\widetilde{\beta}'\bar{\beta})^{-1}$ while $\widetilde{\alpha}$ is post-multiplied by $\widetilde{\beta}'\bar{\beta}$.

Theorem 4.2. Suppose Assumptions 2.1-2.5 hold and $\eta_t \sim N(0, \Lambda)$. Then

$$
\begin{aligned}
(a) \qquad & T\left((\widetilde{\beta}'\bar{\beta})^{-1}\widetilde{\beta}' - \beta'\right)\bar{\beta}_{\perp} = T(\widetilde{\beta}'\bar{\beta})^{-1}\widetilde{\beta}'\bar{\beta}_{\perp} \\
= \quad & \left(\alpha'\Omega_1^*\alpha\right)^{-1}\alpha'\left(T^{-1}\sum_{t=1}^{T}\epsilon_t^* W_{t-1}'\right)\left(T^{-2}\sum_{t=1}^{T}W_{t-1}W_{t-1}'\right)^{-1} + o_P(1) \longrightarrow_{\mathcal{L}} \left(\alpha'\Omega_1^*\alpha\right)^{-1}\alpha' M^*, \\
(b) \qquad & \sqrt{T} vec\left[\left(\widetilde{\alpha}\left(\widetilde{\beta}'\bar{\beta}\right) - \alpha\right), \left(\widetilde{\Gamma}_1 - \Gamma_1\right), \ldots, \left(\widetilde{\Gamma}_{k-1} - \Gamma_{k-1}\right)\right] \\
= \quad & -\left[T^{-1}\sum_{t=1}^{T} R_{2t}\right]^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\nabla_{\phi_2} l_t\right] + o_P(1) \longrightarrow_{\mathcal{L}} N\left(0, \Omega_2^{*-1}\right),
\end{aligned}
$$

where Ω_1^* and M^* are defined in (3.5) and (3.7) respectively. The remaining variables are as defined in Theorem 3.1. □

Theorem 4.3. Suppose Assumptions 2.1-2.5 hold. Then

$$
\begin{aligned}
(a) \qquad & T\left((\widetilde{\beta}'\bar{\beta})^{-1}\widetilde{\beta}' - \beta'\right)\bar{\beta}_{\perp} = T(\widetilde{\beta}'\bar{\beta})^{-1}\widetilde{\beta}'\bar{\beta}_{\perp} \\
= \quad & \left(\alpha'\Omega_1\alpha\right)^{-1}\alpha'\left(T^{-1}\sum_{t=1}^{T}\epsilon_t^* W_{t-1}'\right)\left(T^{-2}\sum_{t=1}^{T}W_{t-1}W_{t-1}'\right)^{-1} + o_P(1) \longrightarrow_{\mathcal{L}} \left(\alpha'\Omega_1\alpha\right)^{-1}\alpha' M^*, \\
(b) \qquad & \sqrt{T} vec\left[\left(\widetilde{\alpha}\left(\widetilde{\beta}'\bar{\beta}\right) - \alpha\right), \left(\widetilde{\Gamma}_1 - \Gamma_1\right), \ldots, \left(\widetilde{\Gamma}_{k-1} - \Gamma_{k-1}\right)\right] \\
= \quad & -\left[T^{-1}\sum_{t=1}^{T} R_{2t}\right]^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\nabla_{\phi_2} l_t\right] + o_P(1) \\
= \quad & -\left[T^{-1}\sum_{t=1}^{T} R_{2t}\right]^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\nabla_{\phi_2} l_t\right] + o_P(1) \longrightarrow_{\mathcal{L}} N\left(0, \Omega_2^{-1}\Omega_2^*\Omega_2^{-1}\right),
\end{aligned}
$$

where Ω_1 and M^* are defined in (3.8) and (3.7) respectively. The remaining variables are as defined in Theorem 3.2. □

Remark 4.6. The normalization factors $(\widetilde{\beta}'\bar{\beta})^{-1}$ and $\widetilde{\beta}'\bar{\beta}$ in (a) and (b) are adopted such that $(\widetilde{\beta}'\bar{\beta})^{-1}\widetilde{\beta}' - \beta' = O_P(T^{-1})$. See, for instance, the discussion in Section 13.2 of Johansen (1996). As one can see in Appendix B, this property plays an important role in proving Theorem 5.1, Lemma 5.1 and Theorem 5.2. See (B.18) below.

Remark 4.7. The asymptotic distributions in 4.2(b) and 4.3(b) are exactly the same as those in Theorem 3.1(b) and Theorem 3.2(b) respectively. In other words, the asymptotic distribution of the "stationary" mean parameters is unaltered, regardless of we imposing the reduced rank. In contrast, the asymptotic distributions in 4.2(a) and 4.3(a) (those of the "nonstationary" mean parameters) are different from those in Theorem 3.1(a) and Theorem 3.2(a). This difference forms the basis of the likelihood ratio (LR) test or the Hausman-type test for cointegrating rank. See Section 5.

Remark 4.8. Decompose $\widetilde{\beta}' = [\widetilde{\beta}_1', \widetilde{\beta}_2']$, where $\widetilde{\beta}_1$ is $r \times r$ and $\widetilde{\beta}_2$ is $d \times r$. If the components of X_t can be arranged as in Ahn and Reinsel (1990) such that the last d components are non-cointegrated, and $\widetilde{\beta}_1'$ is invertible, it is easy to show that

$$
T\left(\widetilde{\beta}_1'^{-1}\widetilde{\beta}_2' - \beta_1'^{-1}\beta_2'\right) \longrightarrow_{\mathcal{L}} \left(\alpha'\Omega_1\alpha\right)^{-1}\alpha' M^* P_{21}^{-1}, \qquad (4.11)
$$

$$\sqrt{T}vec\left[\left(\widetilde{\alpha}\widetilde{\beta}_1'-\alpha\beta_1'\right),\left(\widetilde{\Gamma}_1-\Gamma_1\right),\ldots,\left(\widetilde{\Gamma}_{k-1}-\Gamma_{k-1}\right)\right]\longrightarrow_{\mathcal{L}} N\left(0,\Omega_2^{-1}\Omega_2^{*}\Omega_2^{-1}\right), \tag{4.12}$$

where P_{21} is defined around (4.3). The distribution in (4.11) is similar to that in LLW (2001), with different definitions of Ω_1 and M^* due to the different heteroskedasticity model.

Remark 4.9. It is not difficult to see that the asymptotic distribution of $T(\widetilde{\beta}'\bar{\beta})^{-1}\widetilde{\beta}'\bar{\beta}_{\perp}$ in Theorem 4.2 or 4.3 belongs to the class of locally asymptotically mixed normal (LAMN). When the error η_t is normal, and h_{it-1}'s are not constant, $(\widetilde{\beta}'\bar{\beta})^{-1}\widetilde{\beta}'$ is more efficient than $(\widehat{\beta}'\bar{\beta})^{-1}\widehat{\beta}'$ in Subsection 4.1. See the discussions in Ling and McAleer (2003b).

5 Testing for Cointegrating Rank

This section applies the asymptotic distributions in Theorems 3.1, 3.2, 4.2 and 4.3 to construct tests for reduced rank. The null and the alternative hypotheses are:

$$H_0: rank(\Pi)=r<p \text{ vs } H_a: rank(\Pi)=p. \tag{5.1}$$

We first consider the likelihood ratio (LR) test:

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

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 $\widetilde{\phi}$ are respectively the full rank (see Section 3) and the reduced rank (see Section 4) estimators for φ and ϕ.

Recall the full-rank estimator $\dot{\varphi}=vec[\dot{\Pi},\dot{\Gamma}_1,\ldots,\dot{\Gamma}_{k-1}]$ (see (3.4)). Define another estimator $\widetilde{\varphi}\equiv vec[\widetilde{\alpha}\widetilde{\beta}',\widetilde{\Gamma}_1,\ldots,\widetilde{\Gamma}_{k-1}]$, where we recall that $\widetilde{\phi}_1=vec[\widetilde{\beta}']$ is obtained from (4.9) and $\widetilde{\phi}_2\equiv vec[\widetilde{\alpha},\widetilde{\Gamma}_1,\ldots,\widetilde{\Gamma}_{k-1}]$ is obtained from (4.10). Expansions which are similar to those standardly used in the likelihood theory give (see, for instance, Section 5.6 of Gourieroux and Monfort, 1989):

$$LR_G = (\dot{\varphi}-\widetilde{\varphi})'\left(-\sum_{t=1}^{T}\widetilde{F}_t\right)(\dot{\varphi}-\widetilde{\varphi})+o_P(1), \tag{5.3}$$

where we refer to the (simplified) Hessian in (3.3) and define

$$\widetilde{F}_t=-\left(Y_{t-1}Y_{t-1}'\otimes\widetilde{V}_{t-1}^{-1}\right)-\nabla_{\varphi}\widetilde{h}_{t-1}\left(\Psi_p\left(\widetilde{\Lambda}^{-1}\widetilde{D}_{t-1}^{-2}\otimes\widetilde{D}_{t-1}^{-2}\widetilde{\Lambda}\right)\Psi_p'+\widetilde{D}_{t-1}^{-4}\right)\nabla_{\varphi}'\widetilde{h}_{t-1}/4. \tag{5.4}$$

Remark 5.1. The approximation in (5.3) hinges on the fact that $\sqrt{T}(\dot{\delta}-\widetilde{\delta})=o_P(1)$ under the null. In fact for practical convenience, we may replace both $\dot{\delta}$ and $\widetilde{\delta}$ by $\widehat{\delta}$ which is suggested in Li, Ling and McAleer (2003).

Remark 5.2. In defining $\widetilde{\varphi}$, for $j = 1, \ldots, k-1$, one may use $\dot{\Gamma}_j$ (the full rank estimator) instead of $\widetilde{\Gamma}_j$ (the reduced rank estimator), since $\sqrt{T}(\dot{\Gamma}_j - \widetilde{\Gamma}_j) = o_P(1)$ under the null. In fact, due to this fact and other asymptotic approximations, one may have some variations of LR_G. They include:

$$(\dot{\varphi} - \widetilde{\varphi})' \left(\sum_{t=1}^{T} Y_{t-1} Y_{t-1}' \otimes \widetilde{\Omega}_1 \right) (\dot{\varphi} - \widetilde{\varphi}),$$

$$\left(vec\left(\dot{\Pi} - \widetilde{\alpha}\widetilde{\beta}' \right) \right)' \left(\sum_{t=1}^{T} X_{t-1} X_{t-1}' \otimes \widetilde{\Omega}_1 \right) vec\left(\dot{\Pi} - \widetilde{\alpha}\widetilde{\beta}' \right).$$

See the proof of Lemma 5.1 below. In fact, as one can see from the proof of Lemma 5.1, $\widetilde{\Omega}_1$ above may be replaced by $\widetilde{\alpha}_\perp \left(\widetilde{\alpha}_\perp' \widetilde{\Omega}_1^{-1} \widetilde{\alpha}_\perp \right)^{-1} \widetilde{\alpha}_\perp'$. See (B.23).

Remark 5.3. The first term on the RHS of (5.3) (and its variations) gives an asymptotically equivalent form of the LR_G, which is non-negative and computationally easier. We use this form in the Monte Carlo experiments as well as in the empirical example in Sections 8 and 9 below. On the other hand, this form suggests a Hausman-type test statistic, which is also non-negative. See (5.6) below.

Theorem 5.1. Suppose Assumptions 2.1-2.5 hold and $\eta_t \sim N(0, \Lambda)$. Then under the null H_0, the LR test for cointegrating rank,

$$LR_G \longrightarrow_{\mathcal{L}} tr \left\{ \left[\zeta \left(I_d - \Xi \right)^{1/2} + \Phi \Xi^{1/2} \right]' \left[\zeta \left(I_d - \Xi \right)^{1/2} + \Phi \Xi^{1/2} \right] \right\}, \tag{5.5}$$

where Ξ is a diagonal matrix which diagonal elements are the d eigenvalues of $I_d - \Upsilon\Upsilon'$, $\Upsilon = \left(\alpha_\perp' \Omega_1^{-1} \alpha_\perp \right)^{1/2} (\alpha_\perp' (EV_{t-1}) \alpha_\perp)^{-1/2}$. $vec(\Phi) \sim N(0, I_{d^2})$, which is independent of $\zeta =$ $\left[\int_0^1 B_d(u) B_d(u)' du \right]^{-1/2} \int_0^1 B_d(u) dB_d(u)'$, $B_d(u)$ is a $d-$dimensional standard Brownian motion. □

Remark 5.4. If $\Upsilon = I_d$, which is the case when ϵ_t is conditional homoskedastic and $\Omega_1 = (EV_{t-1})^{-1}$ the distribution in Theorem 5.1 boils down to $tr\{\zeta'\zeta\}$, the distribution in Reinsel and Ahn (1992) and that of a special case in Johansen (1996).

In general we may not want to assume $\eta_t \sim N(0, \Lambda)$. The asymptotic distribution in this general case is reported in Lemma 5.1.

Lemma 5.1. Suppose Assumptions 2.1-2.5 hold. Then under the null H_0, the LR test for cointegrating rank,

$$LR_G \longrightarrow_{\mathcal{L}} tr \left[\left(\int_0^1 B_d(u) dV_d(u)' \right)' \left(\int_0^1 B_d(u) B_d(u)' du \right)^{-1} \left(\int_0^1 B_d(u) dV_d(u)' \right) \right],$$

where $V_d(u) = (\Upsilon\Upsilon')^{1/2} B_d(u) + \left[\left(\alpha_\perp' \Omega_1^{-1} \alpha_\perp \right)^{-1/2} \alpha_\perp' \Omega_1^{-1} \Omega_1^* \Omega_1^{-1} \alpha_\perp \left(\alpha_\perp' \Omega_1^{-1} \alpha_\perp \right)^{-1/2} - \Upsilon\Upsilon' \right]^{1/2} B_d^*(u)$, Υ is defined as in Theorem 5.1, and $(B_d'(u), B_d^{*\prime}(u))'$ is a $2d-$dimensional standard Brownian motion. □

Remark 5.5. When ϵ_t is conditional homoscedastic, $\Omega_1^* = \Omega_1 = (EV_{t-1})^{-1}$, $\Upsilon = I_d$ and $V_d(u) = B_d(u)$. The distribution of LR_G is $tr\{\zeta'\zeta\}$, which is exactly the same as that in Reinsel and Ahn (1992) and that of a special case in Johansen (1996).

Remark 5.6. In principle, the critical value of the asymptotic distribution in Lemma 5.1 can be simulated via Monte Carlo method. However, the number of nuisance parameters equals $d(d+1)$, contrast to only d parameters in Theorem 5.1. As one can see from the proof of Lemma 5.1, the number of nuisance parameters is reduced if we force $E[V_d(u)V_d(u)']$ to be I_d, regardless of η_t being normal or not. The construction of the Hausman-type test is an attempt to do so.

Modifying upon (5.3), we propose a Hausman-type test statistic:

$$H_G \equiv (\dot{\varphi} - \widetilde{\varphi})' \left(\sum_{t=1}^{T} Y_{t-1}Y_{t-1}' \otimes \widetilde{\alpha}_\perp \left(\widetilde{\alpha}_\perp' \widetilde{\Omega}_1^{-1} \widetilde{\Omega}_1^* \widetilde{\Omega}_1^{-1} \widetilde{\alpha}_\perp \right)^{-1} \widetilde{\alpha}_\perp' \right) (\dot{\varphi} - \widetilde{\varphi}) . \tag{5.6}$$

Remark 5.7. Similar to Remark 5.2, as $\sqrt{T}(\dot{\Gamma}_j - \widetilde{\Gamma}_j) = o_P(1)$ under the null, H_G can alternatively defined as:

$$\left(vec\left(\dot{\Pi} - \widetilde{\alpha}\widetilde{\beta}' \right) \right)' \left(\sum_{t=1}^{T} X_{t-1}X_{t-1}' \otimes \widetilde{\alpha}_\perp \left(\widetilde{\alpha}_\perp' \widetilde{\Omega}_1^{-1} \widetilde{\Omega}_1^* \widetilde{\Omega}_1^{-1} \widetilde{\alpha}_\perp \right)^{-1} \widetilde{\alpha}_\perp' \right) vec\left(\dot{\Pi} - \widetilde{\alpha}\widetilde{\beta}' \right) .$$

The following theorem gives the asymptotic distribution of H_G.

Theorem 5.2. Suppose the assumptions in Lemma 5.1 hold. Then under the null H_0, the Hausman-type test statistic,

$$H_G \longrightarrow_{\mathcal{L}} tr\left\{ \left[\zeta \left(I_d - \Xi^H \right)^{1/2} + \Phi \Xi^{H1/2} \right]' \left[\zeta \left(I_d - \Xi^H \right)^{1/2} + \Phi \Xi^{H1/2} \right] \right\}, \tag{5.7}$$

where Ξ^H is a diagonal matrix which diagonal elements are the d eigenvalues of $I_d - \Upsilon^H \Upsilon^{H\prime}$, $\Upsilon^H = \left(\alpha_\perp' \Omega_1^{-1} \Omega_1^* \Omega_1^{-1} \alpha_\perp \right)^{-1/2} \left(\alpha_\perp' \Omega_1^{-1} \alpha_\perp \right) \left(\alpha_\perp' (EV_{t-1}) \alpha_\perp \right)^{-1/2}$. Φ and ζ are as defined in Theorem 5.1. □

Remark 5.8. Similar to the arguments in Remark 5.5, when ϵ_t is conditional homoskedastic, $\Omega_1^* = \Omega_1 = (EV_{t-1})^{-1}$, $\Upsilon^H = I_d$. The asymptotic distribution of H_G is $tr\{\zeta'\zeta\}$, is exactly the same as that in Reinsel and Ahn (1992) and that of a special case in Johansen (1996).

Remark 5.9. From Theorems 5.1 and 5.2, when $\Omega_1^* = \Omega_1 \neq (EV_{t-1})^{-1}$ in general, $\Upsilon^H = \Upsilon$, $\Xi^H = \Xi$ and the asymptotic distributions in both theorems are the same.

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),

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \epsilon_t + \mu, \tag{5.8}$$

where μ is unknown but we assume that $\alpha_\perp'\mu = 0$. Alternatively, $\mu = \alpha\rho_0$, where ρ_0 is $r \times 1$. This is a model used in many empirical examples, including that in Reinsel and Ahn (1992) and one particular case in Johansen (1996). See (4.6), p.49 in Johansen (1996). We thus consider a counterpart of (4.1):

$$\Delta X_t - \overline{\Delta X} = \alpha\beta'(X_{t-1} - \bar{X}) + \sum_{j=1}^{k-1} \Gamma_j(\Delta X_{t-j} - \overline{\Delta X}) + (\epsilon_t - \bar{\epsilon}) . \tag{5.9}$$

And there is not a drift in the nonstationary part $\beta_\perp' X_t$. Denote the corresponding LR test and the Hausman-type test as $LR_{G\mu}$ and $H_{G\mu}$ respectively. The following two corollaries can be obtained straightforwardly from Theorem 5.1 and Theorem 5.2 respectively. The proofs are thus omitted.

Corollary 5.1. Suppose Assumptions 2.1-2.5 hold for Equation (5.6) and $\eta_t \sim N(0, \Lambda)$. Then under the null H_0, the LR test for cointegrating rank,

$$LR_{G\mu} \longrightarrow_{\mathcal{L}} tr\left\{\left[\bar{\zeta}\left(I_d - \Xi\right)^{1/2} + \Phi\Xi^{1/2}\right]'\left[\bar{\zeta}\left(I_d - \Xi\right)^{1/2} + \Phi\Xi^{1/2}\right]\right\}, \tag{5.10}$$

where Ξ is as defined in Theorem 5.1. $vec(\Phi) \sim N\left(0, I_{d^2}\right)$, which is independent of $\bar{\zeta} = [\int_0^1 \bar{B}_d(u)\bar{B}_d(u)'du]^{-1/2} \int_0^1 \bar{B}_d(u)dB_d(u)'$, $\bar{B} \equiv [B_d(u) - \int_0^1 B_d(u)du]$, $B_d(u)$ is a $d-$dimensional standard Brownian motion. □

Corollary 5.2. Suppose Assumptions 2.1-2.5 hold for Equation (5.6). Then under the null H_0, the Hausman-type test statistic,

$$H_{G\mu} \longrightarrow_{\mathcal{L}} tr\left\{\left[\bar{\zeta}\left(I_d - \Xi^H\right)^{1/2} + \Phi\Xi^{H1/2}\right]'\left[\bar{\zeta}\left(I_d - \Xi^H\right)^{1/2} + \Phi\Xi^{H1/2}\right]\right\}, \tag{5.11}$$

where Ξ^H is as defined in Theorem 5.2. Φ and $\bar{\zeta}$ are as defined in Corollary 5.1. □

6 Practical Details

6.1 An Illustrative Case of $p = 2$

In this sub-section, we illustrative how the full rank estimation and reduced rank estimation are proceeded, with a VAR(1)-GARCH(1,1), when $p = 2$ and $r = 1$. We first express the two common terms in the one-step iterations (see (3.4), (4.9) and (4.10)). Let $V_{t-1}^{-1} = \left(v_{t-1}^{ij}\right)_{2\times 2}$.

$$w\left(D_{t-1}^{-2}\left(I_2 - \epsilon_t\epsilon_t'V_{t-1}^{-1}\right)\right) = \begin{pmatrix} \left(1 - \epsilon_{1t}^2 v_{t-1}^{11} - \epsilon_{1t}\epsilon_{2t}v_{t-1}^{12}\right)/h_{1t-1} \\ \left(1 - \epsilon_{2t}\epsilon_{1t}v_{t-1}^{12} - \epsilon_{2t}^2 v_{t-1}^{22}\right)/h_{2t-1} \end{pmatrix}. \tag{6.1}$$

Recall that $\Lambda = (\lambda_{ij})_{2\times 2}$. Let $\Lambda^{-1} = (\lambda^{ij})_{2\times 2}$.

$$\left(\Psi_2\left(\Lambda^{-1}D_{t-1}^{-2} \otimes D_{t-1}^{-2}\Lambda\right)\Psi_2' + D_{t-1}^{-4}\right) = \begin{pmatrix} \left(1 + \lambda^{11}\right)/h_{1t-1}^2 & \lambda^{12}\lambda_{12}/h_{1t-1}h_{2t-1} \\ \lambda^{12}\lambda_{12}/h_{2t-1}h_{1t-1} & \left(1 + \lambda^{22}\right)/h_{2t-1}^2 \end{pmatrix}. \tag{6.2}$$

The gradient and the (simplified) Hessian of the full rank estimation:

$$\nabla_\varphi l_t = \begin{pmatrix} X_{1t-1}\left(v_{t-1}^{11}\epsilon_{1t} + v_{t-1}^{12}\epsilon_{2t}\right) \\ X_{1t-1}\left(v_{t-1}^{12}\epsilon_{1t} + v_{t-1}^{22}\epsilon_{2t}\right) \\ X_{2t-1}\left(v_{t-1}^{11}\epsilon_{1t} + v_{t-1}^{12}\epsilon_{2t}\right) \\ X_{2t-1}\left(v_{t-1}^{12}\epsilon_{1t} + v_{t-1}^{22}\epsilon_{2t}\right) \end{pmatrix} - \frac{1}{2}\nabla_\varphi h_{t-1} w\left(D_{t-1}^{-2}\left(I_2 - \epsilon_t\epsilon_t' V_{t-1}^{-1}\right)\right),$$

$$F_t = -\begin{pmatrix} X_{1t-1}^2 v_{t-1}^{11} & X_{1t-1}^2 v_{t-1}^{12} & X_{1t-1}X_{2t-1}v_{t-1}^{11} & X_{1t-1}X_{2t-1}v_{t-1}^{12} \\ X_{1t-1}^2 v_{t-1}^{12} & X_{1t-1}^2 v_{t-1}^{22} & X_{1t-1}X_{2t-1}v_{t-1}^{12} & X_{1t-1}X_{2t-1}v_{t-1}^{22} \\ X_{2t-1}X_{1t-1}v_{t-1}^{11} & X_{2t-1}X_{1t-1}v_{t-1}^{12} & X_{2t-1}^2 v_{t-1}^{11} & X_{2t-1}^2 v_{t-1}^{12} \\ X_{2t-1}X_{1t-1}v_{t-1}^{12} & X_{2t-1}X_{1t-1}v_{t-1}^{22} & X_{2t-1}^2 v_{t-1}^{12} & X_{2t-1}^2 v_{t-1}^{22} \end{pmatrix}$$
$$-\frac{1}{4}\nabla_\varphi h_{t-1}\left(\Psi_2\left(\Lambda^{-1}D_{t-1}^{-2}\otimes D_{t-1}^{-2}\Lambda\right)\Psi_2' + D_{t-1}^{-4}\right)\nabla_\varphi' h_{t-1},$$

where the 4×2- vector $\nabla_\varphi h_{t-1}$ is recursively defined as:

$$\nabla_\varphi h_{t-1} = -2\begin{pmatrix} a_{11}X_{1t-2}\epsilon_{1t-1} & 0 \\ 0 & a_{21}X_{1t-2}\epsilon_{2t-1} \\ a_{11}X_{2t-2}\epsilon_{1t-1} & 0 \\ 0 & a_{21}X_{2t-2}\epsilon_{2t-1} \end{pmatrix} + \nabla_\varphi h_{t-2}\begin{pmatrix} b_{11} & 0 \\ 0 & b_{21} \end{pmatrix}.$$

And, with $\alpha \equiv (\alpha_1, \alpha_2)'$, the gradients and the (simplified) Hessians of the reduced rank estimation:

$$\nabla_{\phi_1} l_t = \begin{pmatrix} \alpha_1 X_{1t-1}\left(v_{t-1}^{11}\epsilon_{1t} + v_{t-1}^{12}\epsilon_{2t}\right) + \alpha_2 X_{1t-1}\left(v_{t-1}^{12}\epsilon_{1t} + v_{t-1}^{22}\epsilon_{2t}\right) \\ \alpha_1 X_{2t-1}\left(v_{t-1}^{11}\epsilon_{1t} + v_{t-1}^{12}\epsilon_{2t}\right) + \alpha_2 X_{2t-1}\left(v_{t-1}^{12}\epsilon_{1t} + v_{t-1}^{22}\epsilon_{2t}\right) \end{pmatrix}$$
$$-\frac{1}{2}\nabla_{\phi_1} h_{t-1} w\left(D_{t-1}^{-2}\left(I_2 - \epsilon_t\epsilon_t' V_{t-1}^{-1}\right)\right),$$

$$\nabla_{\phi_2} l_t = \begin{pmatrix} \beta' X_{t-1}\left(v_{t-1}^{11}\epsilon_{1t} + v_{t-1}^{12}\epsilon_{2t}\right) \\ \beta' X_{t-1}\left(v_{t-1}^{12}\epsilon_{1t} + v_{t-1}^{22}\epsilon_{2t}\right) \end{pmatrix} - \frac{1}{2}\nabla_{\phi_2} h_{t-1} w\left(D_{t-1}^{-2}\left(I_2 - \epsilon_t\epsilon_t' V_{t-1}^{-1}\right)\right);$$

$$R_{1t} = -\begin{pmatrix} X_{1t-1}^2 \alpha' V_{t-1}^{-1}\alpha & X_{1t-1}X_{2t-1}\alpha' V_{t-1}^{-1}\alpha \\ X_{2t-1}X_{1t-1}\alpha' V_{t-1}^{-1}\alpha & X_{2t-1}^2\alpha' V_{t-1}^{-1}\alpha \end{pmatrix}$$
$$-\frac{1}{4}\nabla_{\phi_1} h_{t-1}\left(\Psi_2\left(\Lambda^{-1}D_{t-1}^{-2}\otimes D_{t-1}^{-2}\Lambda\right)\Psi_2' + D_{t-1}^{-4}\right)\nabla_{\phi_1}' h_{t-1},$$

$$R_{2t} = -\begin{pmatrix} \beta' X_{t-1}X_{t-1}'\beta v_{t-1}^{11} & \beta' X_{t-1}X_{t-1}'\beta v_{t-1}^{12} \\ \beta' X_{t-1}X_{t-1}'\beta v_{t-1}^{12} & \beta' X_{t-1}X_{t-1}'\beta v_{t-1}^{22} \end{pmatrix}$$
$$-\frac{1}{4}\nabla_{\phi_2} h_{t-1}\left(\Psi_2\left(\Lambda^{-1}D_{t-1}^{-2}\otimes D_{t-1}^{-2}\Lambda\right)\Psi_2' + D_{t-1}^{-4}\right)\nabla_{\phi_2}' h_{t-1},$$

where the 2×2- vectors $\nabla_{\phi_1} h_{t-1}$ and $\nabla_{\phi_2} h_{t-1}$ are recursively defined as:

$$\nabla_{\phi_1} h_{t-1} = -2\begin{pmatrix} \alpha_1 a_{11} X_{1t-2}\epsilon_{1t-1} & \alpha_2 a_{21} X_{1t-2}\epsilon_{2t-1} \\ \alpha_1 a_{11} X_{2t-2}\epsilon_{1t-1} & \alpha_2 a_{21} X_{2t-2}\epsilon_{2t-1} \end{pmatrix} + \nabla_{\phi_1} h_{t-2}\begin{pmatrix} b_{11} & 0 \\ 0 & b_{21} \end{pmatrix},$$

$$\nabla_{\phi_2} h_{t-1} = -2\begin{pmatrix} a_{11}\beta' X_{t-1}\epsilon_{1t-1} & 0 \\ 0 & a_{21}\beta' X_{t-1}\epsilon_{2t-1} \end{pmatrix} + \nabla_{\phi_2} h_{t-2}\begin{pmatrix} b_{11} & 0 \\ 0 & b_{21} \end{pmatrix}.$$

6.2 An Illustrative Case of $p = 3$

In this sub-section, we illustrative how the full rank estimation and reduced rank estimation are proceeded, with a VAR(1)-GARCH(1,1), when $p = 3$ and $r = 1$ or 2. We first express the two common terms in the one-step iterations (see (3.4), (4.9) and (4.10)). Let $V_{t-1}^{-1} = \left(v_{t-1}^{ij}\right)_{3\times 3}$.

$$w\left(D_{t-1}^{-2}\left(I_3 - \epsilon_t\epsilon_t' V_{t-1}^{-1}\right)\right) = \begin{pmatrix} \left(1-\epsilon_{1t}^2 v_{t-1}^{11} - \epsilon_{1t}\epsilon_{2t}v_{t-1}^{12} - \epsilon_{1t}\epsilon_{3t}v_{t-1}^{13}\right)/h_{1t-1} \\ \left(1-\epsilon_{2t}\epsilon_{1t} v_{t-1}^{12} - \epsilon_{2t}^2 v_{t-1}^{22} - \epsilon_{2t}\epsilon_{3t}v_{t-1}^{23}\right)/h_{2t-1} \\ \left(1-\epsilon_{3t}\epsilon_{1t} v_{t-1}^{13} - \epsilon_{3t}\epsilon_{2t}v_{t-1}^{23} - \epsilon_{3t}^2 v_{t-1}^{33}\right)/h_{3t-1} \end{pmatrix}. \tag{6.3}$$

Recall that $\Lambda = (\lambda_{ij})_{3\times 3}$. Let $\Lambda^{-1} = (\lambda^{ij})_{3\times 3}$.

$$\begin{aligned} &\left(\Psi_3\left(\Lambda^{-1}D_{t-1}^{-2}\otimes D_{t-1}^{-2}\Lambda\right)\Psi_3' + D_{t-1}^{-4}\right) \\ &= \begin{pmatrix} \left(1+\lambda^{11}\right)/h_{1t-1}^2 & \lambda^{12}\lambda_{12}/h_{1t-1}h_{2t-1} & \lambda^{13}\lambda_{13}/h_{1t-1}h_{3t-1} \\ \lambda^{12}\lambda_{12}/h_{2t-1}h_{1t-1} & \left(1+\lambda^{22}\right)/h_{2t-1}^2 & \lambda^{23}\lambda_{23}/h_{2t-1}h_{3t-1} \\ \lambda^{13}\lambda_{13}/h_{3t-1}h_{1t-1} & \lambda^{23}\lambda_{23}/h_{3t-1}h_{2t-1} & \left(1+\lambda^{33}\right)/h_{3t-1}^2 \end{pmatrix}. \end{aligned} \tag{6.4}$$

The gradient and the (simplified) Hessian of the full rank estimation:

$$\nabla_\varphi l_t = \begin{pmatrix} X_{1t-1}\left(v_{t-1}^{11}\epsilon_{1t} + v_{t-1}^{12}\epsilon_{2t} + v_{t-1}^{13}\epsilon_{3t}\right) \\ X_{1t-1}\left(v_{t-1}^{12}\epsilon_{1t} + v_{t-1}^{22}\epsilon_{2t} + v_{t-1}^{23}\epsilon_{3t}\right) \\ X_{1t-1}\left(v_{t-1}^{13}\epsilon_{1t} + v_{t-1}^{23}\epsilon_{2t} + v_{t-1}^{33}\epsilon_{3t}\right) \\ X_{2t-1}\left(v_{t-1}^{11}\epsilon_{1t} + v_{t-1}^{12}\epsilon_{2t} + v_{t-1}^{13}\epsilon_{3t}\right) \\ X_{2t-1}\left(v_{t-1}^{12}\epsilon_{1t} + v_{t-1}^{22}\epsilon_{2t} + v_{t-1}^{23}\epsilon_{3t}\right) \\ X_{2t-1}\left(v_{t-1}^{13}\epsilon_{1t} + v_{t-1}^{23}\epsilon_{2t} + v_{t-1}^{33}\epsilon_{3t}\right) \\ X_{3t-1}\left(v_{t-1}^{11}\epsilon_{1t} + v_{t-1}^{12}\epsilon_{2t} + v_{t-1}^{13}\epsilon_{3t}\right) \\ X_{3t-1}\left(v_{t-1}^{12}\epsilon_{1t} + v_{t-1}^{22}\epsilon_{2t} + v_{t-1}^{23}\epsilon_{3t}\right) \\ X_{3t-1}\left(v_{t-1}^{13}\epsilon_{1t} + v_{t-1}^{23}\epsilon_{2t} + v_{t-1}^{33}\epsilon_{3t}\right) \end{pmatrix} - \frac{1}{2}\nabla_\varphi h_{t-1} w\left(D_{t-1}^{-2}\left(I_3 - \epsilon_t\epsilon_t' V_{t-1}^{-1}\right)\right);$$

$$F_t = \begin{pmatrix} f_{11t} & f_{12t} & f_{13t} \\ f_{12t} & f_{22t} & f_{23t} \\ f_{13t} & f_{23t} & f_{33t} \end{pmatrix} - \frac{1}{4}\nabla_\varphi h_{t-1}\left(\Psi_3\left(\Lambda^{-1}D_{t-1}^{-2}\otimes D_{t-1}^{-2}\Lambda\right)\Psi_3' + D_{t-1}^{-4}\right)\nabla_\varphi' h_{t-1},$$

where

$$f_{11t} = -\begin{pmatrix} X_{1t-1}^2 v_{t-1}^{11} & X_{1t-1}^2 v_{t-1}^{12} & X_{1t-1}^2 v_{t-1}^{13} \\ X_{1t-1}^2 v_{t-1}^{12} & X_{1t-1}^2 v_{t-1}^{22} & X_{1t-1}^2 v_{t-1}^{23} \\ X_{1t-1}^2 v_{t-1}^{13} & X_{1t-1}^2 v_{t-1}^{23} & X_{1t-1}^2 v_{t-1}^{33} \end{pmatrix},$$

$$f_{12t} = -\begin{pmatrix} X_{1t-1}X_{2t-1} v_{t-1}^{11} & X_{1t-1}X_{2t-1} v_{t-1}^{12} & X_{1t-1}X_{2t-1} v_{t-1}^{13} \\ X_{1t-1}X_{2t-1} v_{t-1}^{12} & X_{1t-1}X_{2t-1} v_{t-1}^{22} & X_{1t-1}X_{2t-1} v_{t-1}^{23} \\ X_{1t-1}X_{2t-1} v_{t-1}^{13} & X_{1t-1}X_{2t-1} v_{t-1}^{23} & X_{1t-1}X_{2t-1} v_{t-1}^{33} \end{pmatrix},$$

$$f_{13t} = -\begin{pmatrix} X_{1t-1}X_{3t-1} v_{t-1}^{11} & X_{1t-1}X_{3t-1} v_{t-1}^{12} & X_{1t-1}X_{3t-1} v_{t-1}^{13} \\ X_{1t-1}X_{3t-1} v_{t-1}^{12} & X_{1t-1}X_{3t-1} v_{t-1}^{22} & X_{1t-1}X_{3t-1} v_{t-1}^{23} \\ X_{1t-1}X_{3t-1} v_{t-1}^{13} & X_{1t-1}X_{3t-1} v_{t-1}^{23} & X_{1t-1}X_{3t-1} v_{t-1}^{33} \end{pmatrix},$$

$$f_{22t} = -\begin{pmatrix} X_{2t-1}^2 v_{t-1}^{11} & X_{2t-1}^2 v_{t-1}^{12} & X_{2t-1}^2 v_{t-1}^{13} \\ X_{2t-1}^2 v_{t-1}^{12} & X_{2t-1}^2 v_{t-1}^{22} & X_{2t-1}^2 v_{t-1}^{23} \\ X_{2t-1}^2 v_{t-1}^{13} & X_{2t-1}^2 v_{t-1}^{23} & X_{2t-1}^2 v_{t-1}^{33} \end{pmatrix},$$

$$f_{23t} = -\begin{pmatrix} X_{2t-1}X_{3t-1} v_{t-1}^{11} & X_{2t-1}X_{3t-1} v_{t-1}^{12} & X_{2t-1}X_{3t-1} v_{t-1}^{13} \\ X_{2t-1}X_{3t-1} v_{t-1}^{12} & X_{2t-1}X_{3t-1} v_{t-1}^{22} & X_{2t-1}X_{3t-1} v_{t-1}^{23} \\ X_{2t-1}X_{3t-1} v_{t-1}^{13} & X_{2t-1}X_{3t-1} v_{t-1}^{23} & X_{2t-1}X_{3t-1} v_{t-1}^{33} \end{pmatrix},$$

$$f_{33t} = -\begin{pmatrix} X_{3t-1}^2 v_{t-1}^{11} & X_{3t-1}^2 v_{t-1}^{12} & X_{3t-1}^2 v_{t-1}^{13} \\ X_{3t-1}^2 v_{t-1}^{12} & X_{3t-1}^2 v_{t-1}^{22} & X_{3t-1}^2 v_{t-1}^{23} \\ X_{3t-1}^2 v_{t-1}^{13} & X_{3t-1}^2 v_{t-1}^{23} & X_{3t-1}^2 v_{t-1}^{33} \end{pmatrix},$$

where the 9×3- vector $\nabla_{\varphi} h_{t-1}$ is recursively defined as:

$$\nabla_{\varphi} h_{t-1} = -2 \begin{pmatrix} a_{11}X_{1t-2}\epsilon_{1t-1} & 0 & 0 \\ 0 & a_{21}X_{1t-2}\epsilon_{2t-1} & 0 \\ 0 & 0 & a_{31}X_{1t-2}\epsilon_{3t-1} \\ a_{11}X_{2t-2}\epsilon_{1t-1} & 0 & 0 \\ 0 & a_{21}X_{2t-2}\epsilon_{2t-1} & 0 \\ 0 & 0 & a_{31}X_{2t-2}\epsilon_{3t-1} \\ a_{11}X_{3t-2}\epsilon_{1t-1} & 0 & 0 \\ 0 & a_{21}X_{3t-2}\epsilon_{2t-1} & 0 \\ 0 & 0 & a_{31}X_{3t-2}\epsilon_{3t-1} \end{pmatrix} + \nabla_{\varphi} h_{t-2} \begin{pmatrix} b_{11} & 0 & 0 \\ 0 & b_{21} & 0 \\ 0 & 0 & b_{31} \end{pmatrix}.$$

Next we consider the reduced rank estimation of the sub-case $r = 1$. With $\alpha \equiv (\alpha_{11}, \alpha_{21}, \alpha_{31})'$, the gradients and the (simplified) Hessians of the reduced rank estimation are:

$$\nabla_{\phi_1} l_t =$$
$$\begin{pmatrix} \alpha_{11}X_{1t-1}\left(v^{11}_{t-1}\epsilon_{1t} + v^{12}_{t-1}\epsilon_{2t} + v^{13}_{t-1}\epsilon_{3t}\right) + \alpha_{21}X_{1t-1}\left(v^{12}_{t-1}\epsilon_{1t} + v^{22}_{t-1}\epsilon_{2t} + v^{23}_{t-1}\epsilon_{3t}\right) + \alpha_{31}X_{1t-1}\left(v^{13}_{t-1}\epsilon_{1t} + v^{23}_{t-1}\epsilon_{2t} + v^{33}_{t-1}\epsilon_{3t}\right) \\ \alpha_{11}X_{2t-1}\left(v^{11}_{t-1}\epsilon_{1t} + v^{12}_{t-1}\epsilon_{2t} + v^{13}_{t-1}\epsilon_{3t}\right) + \alpha_{21}X_{2t-1}\left(v^{12}_{t-1}\epsilon_{1t} + v^{22}_{t-1}\epsilon_{2t} + v^{23}_{t-1}\epsilon_{3t}\right) + \alpha_{31}X_{2t-1}\left(v^{13}_{t-1}\epsilon_{1t} + v^{23}_{t-1}\epsilon_{2t} + v^{33}_{t-1}\epsilon_{3t}\right) \\ \alpha_{11}X_{3t-1}\left(v^{11}_{t-1}\epsilon_{1t} + v^{12}_{t-1}\epsilon_{2t} + v^{13}_{t-1}\epsilon_{3t}\right) + \alpha_{21}X_{3t-1}\left(v^{12}_{t-1}\epsilon_{1t} + v^{22}_{t-1}\epsilon_{2t} + v^{23}_{t-1}\epsilon_{3t}\right) + \alpha_{31}X_{3t-1}\left(v^{13}_{t-1}\epsilon_{1t} + v^{23}_{t-1}\epsilon_{2t} + v^{33}_{t-1}\epsilon_{3t}\right) \end{pmatrix}$$
$$-\frac{1}{2}\nabla_{\phi_1} h_{t-1} w\left(D^{-2}_{t-1}\left(I_3 - \epsilon_t\epsilon'_t V^{-1}_{t-1}\right)\right),$$

$$\begin{aligned} \nabla_{\phi_2} l_t &= \begin{pmatrix} \beta' X_{t-1}\left(v^{11}_{t-1}\epsilon_{1t} + v^{12}_{t-1}\epsilon_{2t} + v^{13}_{t-1}\epsilon_{3t}\right) \\ \beta' X_{t-1}\left(v^{12}_{t-1}\epsilon_{1t} + v^{22}_{t-1}\epsilon_{2t} + v^{23}_{t-1}\epsilon_{3t}\right) \\ \beta' X_{t-1}\left(v^{13}_{t-1}\epsilon_{1t} + v^{23}_{t-1}\epsilon_{2t} + v^{33}_{t-1}\epsilon_{3t}\right) \end{pmatrix} - \frac{1}{2}\nabla_{\phi_2} h_{t-1} w\left(D^{-2}_{t-1}\left(I_3 - \epsilon_t\epsilon'_t V^{-1}_{t-1}\right)\right); \\ R_{1t} &= -\begin{pmatrix} X^2_{1t-1}\alpha' V^{-1}_{t-1}\alpha & X_{1t-1}X_{2t-1}\alpha' V^{-1}_{t-1}\alpha & X_{1t-1}X_{3t-1}\alpha' V^{-1}_{t-1}\alpha \\ X_{2t-1}X_{1t-1}\alpha' V^{-1}_{t-1}\alpha & X^2_{2t-1}\alpha' V^{-1}_{t-1}\alpha & X_{2t-1}X_{3t-1}\alpha' V^{-1}_{t-1}\alpha \\ X_{3t-1}X_{1t-1}\alpha' V^{-1}_{t-1}\alpha & X_{3t-1}X_{2t-1}\alpha' V^{-1}_{t-1}\alpha & X^2_{3t-1}\alpha' V^{-1}_{t-1}\alpha \end{pmatrix} \\ &\quad -\frac{1}{4}\nabla_{\phi_1} h_{t-1}\left(\Psi_3\left(\Lambda^{-1}D^{-2}_{t-1} \otimes D^{-2}_{t-1}\Lambda\right)\Psi'_3 + D^{-4}_{t-1}\right)\nabla'_{\phi_1} h_{t-1}, \\ R_{2t} &= -\begin{pmatrix} \beta' X_{t-1}X'_{t-1}\beta v^{11}_{t-1} & \beta' X_{t-1}X'_{t-1}\beta v^{12}_{t-1} & \beta' X_{t-1}X'_{t-1}\beta v^{13}_{t-1} \\ \beta' X_{t-1}X'_{t-1}\beta v^{12}_{t-1} & \beta' X_{t-1}X'_{t-1}\beta v^{22}_{t-1} & \beta' X_{t-1}X'_{t-1}\beta v^{23}_{t-1} \\ \beta' X_{t-1}X'_{t-1}\beta v^{13}_{t-1} & \beta' X_{t-1}X'_{t-1}\beta v^{23}_{t-1} & \beta' X_{t-1}X'_{t-1}\beta v^{33}_{t-1} \end{pmatrix} \\ &\quad -\frac{1}{4}\nabla_{\phi_2} h_{t-1}\left(\Psi_3\left(\Lambda^{-1}D^{-2}_{t-1} \otimes D^{-2}_{t-1}\Lambda\right)\Psi'_3 + D^{-4}_{t-1}\right)\nabla'_{\phi_2} h_{t-1}, \end{aligned}$$

where the 3×3- vectors $\nabla_{\phi_1} h_{t-1}$ and $\nabla_{\phi_2} h_{t-1}$ are recursively defined as:

$$\begin{aligned} \nabla_{\phi_1} h_{t-1} &= -2\begin{pmatrix} \alpha_{11}a_{11}X_{1t-2}\epsilon_{1t-1} & \alpha_{21}a_{21}X_{1t-2}\epsilon_{2t-1} & \alpha_{31}a_{31}X_{1t-2}\epsilon_{3t-1} \\ \alpha_{11}a_{11}X_{2t-2}\epsilon_{1t-1} & \alpha_{21}a_{21}X_{2t-2}\epsilon_{2t-1} & \alpha_{31}a_{31}X_{2t-2}\epsilon_{3t-1} \\ \alpha_{11}a_{11}X_{3t-2}\epsilon_{1t-1} & \alpha_{21}a_{21}X_{3t-2}\epsilon_{2t-1} & \alpha_{31}a_{31}X_{3t-2}\epsilon_{3t-1} \end{pmatrix} + \nabla_{\phi_1} h_{t-2}\begin{pmatrix} b_{11} & 0 & 0 \\ 0 & b_{21} & 0 \\ 0 & 0 & b_{31} \end{pmatrix}, \\ \nabla_{\phi_2} h_{t-1} &= -2\begin{pmatrix} a_{11}\beta' X_{t-1}\epsilon_{1t-1} & 0 & 0 \\ 0 & a_{21}\beta' X_{t-1}\epsilon_{2t-1} & 0 \\ 0 & 0 & a_{31}\beta' X_{t-1}\epsilon_{3t-1} \end{pmatrix} + \nabla_{\phi_2} h_{t-2}\begin{pmatrix} b_{11} & 0 & 0 \\ 0 & b_{21} & 0 \\ 0 & 0 & b_{31} \end{pmatrix}. \end{aligned}$$

Lastly we consider the reduced rank estimation of the sub-case $r = 2$. $\alpha \equiv (\alpha_1, \alpha_2)$, $\alpha_1 \equiv (\alpha_{11}, \alpha_{21}, \alpha_{31})'$ and $\alpha_2 \equiv (\alpha_{12}, \alpha_{22}, \alpha_{32})'$. $\beta = (\beta_1, \beta_2)$, where both β_1 and β_2 are 3×1- vectors. The gradients and the (simplified) Hessians of the reduced rank estimation are:

$$\nabla_{\phi_1} l_t =$$
$$\begin{pmatrix} \alpha_{11}X_{1t-1}\left(v^{11}_{t-1}\epsilon_{1t} + v^{12}_{t-1}\epsilon_{2t} + v^{13}_{t-1}\epsilon_{3t}\right) + \alpha_{21}X_{1t-1}\left(v^{12}_{t-1}\epsilon_{1t} + v^{22}_{t-1}\epsilon_{2t} + v^{23}_{t-1}\epsilon_{3t}\right) + \alpha_{31}X_{1t-1}\left(v^{13}_{t-1}\epsilon_{1t} + v^{23}_{t-1}\epsilon_{2t} + v^{33}_{t-1}\epsilon_{3t}\right) \\ \alpha_{12}X_{1t-1}\left(v^{11}_{t-1}\epsilon_{1t} + v^{12}_{t-1}\epsilon_{2t} + v^{13}_{t-1}\epsilon_{3t}\right) + \alpha_{22}X_{1t-1}\left(v^{12}_{t-1}\epsilon_{1t} + v^{22}_{t-1}\epsilon_{2t} + v^{23}_{t-1}\epsilon_{3t}\right) + \alpha_{32}X_{1t-1}\left(v^{13}_{t-1}\epsilon_{1t} + v^{23}_{t-1}\epsilon_{2t} + v^{33}_{t-1}\epsilon_{3t}\right) \\ \alpha_{11}X_{2t-1}\left(v^{11}_{t-1}\epsilon_{1t} + v^{12}_{t-1}\epsilon_{2t} + v^{13}_{t-1}\epsilon_{3t}\right) + \alpha_{21}X_{2t-1}\left(v^{12}_{t-1}\epsilon_{1t} + v^{22}_{t-1}\epsilon_{2t} + v^{23}_{t-1}\epsilon_{3t}\right) + \alpha_{31}X_{2t-1}\left(v^{13}_{t-1}\epsilon_{1t} + v^{23}_{t-1}\epsilon_{2t} + v^{33}_{t-1}\epsilon_{3t}\right) \\ \alpha_{12}X_{2t-1}\left(v^{11}_{t-1}\epsilon_{1t} + v^{12}_{t-1}\epsilon_{2t} + v^{13}_{t-1}\epsilon_{3t}\right) + \alpha_{22}X_{2t-1}\left(v^{12}_{t-1}\epsilon_{1t} + v^{22}_{t-1}\epsilon_{2t} + v^{23}_{t-1}\epsilon_{3t}\right) + \alpha_{32}X_{2t-1}\left(v^{13}_{t-1}\epsilon_{1t} + v^{23}_{t-1}\epsilon_{2t} + v^{33}_{t-1}\epsilon_{3t}\right) \\ \alpha_{11}X_{3t-1}\left(v^{11}_{t-1}\epsilon_{1t} + v^{12}_{t-1}\epsilon_{2t} + v^{13}_{t-1}\epsilon_{3t}\right) + \alpha_{21}X_{3t-1}\left(v^{12}_{t-1}\epsilon_{1t} + v^{22}_{t-1}\epsilon_{2t} + v^{23}_{t-1}\epsilon_{3t}\right) + \alpha_{31}X_{3t-1}\left(v^{13}_{t-1}\epsilon_{1t} + v^{23}_{t-1}\epsilon_{2t} + v^{33}_{t-1}\epsilon_{3t}\right) \\ \alpha_{12}X_{3t-1}\left(v^{11}_{t-1}\epsilon_{1t} + v^{12}_{t-1}\epsilon_{2t} + v^{13}_{t-1}\epsilon_{3t}\right) + \alpha_{22}X_{3t-1}\left(v^{12}_{t-1}\epsilon_{1t} + v^{22}_{t-1}\epsilon_{2t} + v^{23}_{t-1}\epsilon_{3t}\right) + \alpha_{32}X_{3t-1}\left(v^{13}_{t-1}\epsilon_{1t} + v^{23}_{t-1}\epsilon_{2t} + v^{33}_{t-1}\epsilon_{3t}\right) \end{pmatrix}$$
$$-\frac{1}{2}\nabla_{\phi_1} h_{t-1} w\left(D^{-2}_{t-1}\left(I_3 - \epsilon_t\epsilon'_t V^{-1}_{t-1}\right)\right),$$

$$\nabla_{\phi_2} l_t = \begin{pmatrix} \beta_1' X_{t-1}\left(v_{t-1}^{11}\epsilon_{1t} + v_{t-1}^{12}\epsilon_{2t} + v_{t-1}^{13}\epsilon_{3t}\right) \\ \beta_1' X_{t-1}\left(v_{t-1}^{12}\epsilon_{1t} + v_{t-1}^{22}\epsilon_{2t} + v_{t-1}^{23}\epsilon_{3t}\right) \\ \beta_1' X_{t-1}\left(v_{t-1}^{13}\epsilon_{1t} + v_{t-1}^{23}\epsilon_{2t} + v_{t-1}^{33}\epsilon_{3t}\right) \\ \beta_2' X_{t-1}\left(v_{t-1}^{11}\epsilon_{1t} + v_{t-1}^{12}\epsilon_{2t} + v_{t-1}^{13}\epsilon_{3t}\right) \\ \beta_2' X_{t-1}\left(v_{t-1}^{12}\epsilon_{1t} + v_{t-1}^{22}\epsilon_{2t} + v_{t-1}^{23}\epsilon_{3t}\right) \\ \beta_2' X_{t-1}\left(v_{t-1}^{13}\epsilon_{1t} + v_{t-1}^{23}\epsilon_{2t} + v_{t-1}^{33}\epsilon_{3t}\right) \end{pmatrix} - \frac{1}{2}\nabla_{\phi_2} h_{t-1} w\left(D_{t-1}^{-2}\left(I_3 - \epsilon_t\epsilon_t' V_{t-1}^{-1}\right)\right);$$

$$R_{1t} = -\begin{pmatrix} X_1^2\alpha_1'V^{-1}\alpha_1 & X_1^2\alpha_1'V^{-1}\alpha_2 & X_1X_2\alpha_1'V^{-1}\alpha_1 & X_1X_2\alpha_1'V^{-1}\alpha_2 & X_1X_3\alpha_1'V^{-1}\alpha_1 & X_1X_3\alpha_1'V^{-1}\alpha_2 \\ X_1^2\alpha_2'V^{-1}\alpha_1 & X_1^2\alpha_2'V^{-1}\alpha_2 & X_1X_2\alpha_2'V^{-1}\alpha_1 & X_1X_2\alpha_2'V^{-1}\alpha_2 & X_1X_3\alpha_2'V^{-1}\alpha_1 & X_1X_3\alpha_2'V^{-1}\alpha_2 \\ X_2X_1\alpha_1'V^{-1}\alpha_1 & X_2X_1\alpha_1'V^{-1}\alpha_2 & X_2^2\alpha_1'V^{-1}\alpha_1 & X_2^2\alpha_1'V^{-1}\alpha_2 & X_2X_3\alpha_1'V^{-1}\alpha_1 & X_2X_3\alpha_1'V^{-1}\alpha_2 \\ X_2X_1\alpha_2'V^{-1}\alpha_1 & X_2X_1\alpha_2'V^{-1}\alpha_2 & X_2^2\alpha_2'V^{-1}\alpha_1 & X_2^2\alpha_2'V^{-1}\alpha_2 & X_2X_3\alpha_2'V^{-1}\alpha_1 & X_2X_3\alpha_2'V^{-1}\alpha_2 \\ X_3X_1\alpha_1'V^{-1}\alpha_1 & X_3X_1\alpha_1'V^{-1}\alpha_2 & X_3X_2\alpha_1'V^{-1}\alpha_1 & X_3X_2\alpha_1'V^{-1}\alpha_2 & X_3^2\alpha_1'V^{-1}\alpha_1 & X_3^2\alpha_1'V^{-1}\alpha_2 \\ X_3X_1\alpha_2'V^{-1}\alpha_1 & X_3X_1\alpha_2'V^{-1}\alpha_2 & X_3X_2\alpha_2'V^{-1}\alpha_1 & X_3X_2\alpha_2'V^{-1}\alpha_2 & X_3^2\alpha_2'V^{-1}\alpha_1 & X_3^2\alpha_2'V^{-1}\alpha_2 \end{pmatrix}$$
$$-\frac{1}{4}\nabla_{\phi_1} h_{t-1}\left(\Psi_3\left(\Lambda^{-1}D_{t-1}^{-2}\otimes D_{t-1}^{-2}\Lambda\right)\Psi_3' + D_{t-1}^{-4}\right)\nabla_{\phi_1}' h_{t-1},$$

$$R_{2t} = -\begin{pmatrix} \beta_1'XX'\beta_1v_{t-1}^{11} & \beta_1'XX'\beta_1v_{t-1}^{12} & \beta_1'XX'\beta_1v_{t-1}^{13} & \beta_1'XX'\beta_2v_{t-1}^{11} & \beta_1'XX'\beta_2v_{t-1}^{12} & \beta_1'XX'\beta_2v_{t-1}^{13} \\ \beta_1'XX'\beta_1v_{t-1}^{12} & \beta_1'XX'\beta_1v_{t-1}^{22} & \beta_1'XX'\beta_1v_{t-1}^{23} & \beta_1'XX'\beta_2v_{t-1}^{12} & \beta_1'XX'\beta_2v_{t-1}^{22} & \beta_1'XX'\beta_2v_{t-1}^{23} \\ \beta_1'XX'\beta_1v_{t-1}^{13} & \beta_1'XX'\beta_1v_{t-1}^{23} & \beta_1'XX'\beta_1v_{t-1}^{33} & \beta_1'XX'\beta_2v_{t-1}^{13} & \beta_1'XX'\beta_2v_{t-1}^{23} & \beta_1'XX'\beta_2v_{t-1}^{33} \\ \beta_2'XX'\beta_1v_{t-1}^{11} & \beta_2'XX'\beta_1v_{t-1}^{12} & \beta_2'XX'\beta_1v_{t-1}^{13} & \beta_2'XX'\beta_2v_{t-1}^{11} & \beta_2'XX'\beta_2v_{t-1}^{12} & \beta_2'XX'\beta_2v_{t-1}^{13} \\ \beta_2'XX'\beta_1v_{t-1}^{12} & \beta_2'XX'\beta_1v_{t-1}^{22} & \beta_2'XX'\beta_1v_{t-1}^{23} & \beta_2'XX'\beta_2v_{t-1}^{12} & \beta_2'XX'\beta_2v_{t-1}^{22} & \beta_2'XX'\beta_2v_{t-1}^{23} \\ \beta_2'XX'\beta_1v_{t-1}^{13} & \beta_2'XX'\beta_1v_{t-1}^{23} & \beta_2'XX'\beta_1v_{t-1}^{33} & \beta_2'XX'\beta_2v_{t-1}^{13} & \beta_2'XX'\beta_2v_{t-1}^{23} & \beta_2'XX'\beta_2v_{t-1}^{33} \end{pmatrix}$$
$$-\frac{1}{4}\nabla_{\phi_2} h_{t-1}\left(\Psi_3\left(\Lambda^{-1}D_{t-1}^{-2}\otimes D_{t-1}^{-2}\Lambda\right)\Psi_3' + D_{t-1}^{-4}\right)\nabla_{\phi_2}' h_{t-1},$$

where in some places, we suppress the subscript $t-1$, and the 6×3- vectors $\nabla_{\phi_1} h_{t-1}$ and $\nabla_{\phi_2} h_{t-1}$ are recursively defined as:

$$\nabla_{\phi_1} h_{t-1} = -2\begin{pmatrix} \alpha_{11}a_{11}X_{1t-2}\epsilon_{1t-1} & \alpha_{21}a_{21}X_{1t-2}\epsilon_{2t-1} & \alpha_{31}a_{31}X_{1t-2}\epsilon_{3t-1} \\ \alpha_{12}a_{11}X_{1t-2}\epsilon_{1t-1} & \alpha_{22}a_{21}X_{1t-2}\epsilon_{2t-1} & \alpha_{32}a_{31}X_{1t-2}\epsilon_{3t-1} \\ \alpha_{11}a_{11}X_{2t-2}\epsilon_{1t-1} & \alpha_{21}a_{21}X_{2t-2}\epsilon_{2t-1} & \alpha_{31}a_{31}X_{2t-2}\epsilon_{3t-1} \\ \alpha_{12}a_{11}X_{2t-2}\epsilon_{1t-1} & \alpha_{22}a_{21}X_{2t-2}\epsilon_{2t-1} & \alpha_{32}a_{31}X_{2t-2}\epsilon_{3t-1} \\ \alpha_{11}a_{11}X_{3t-2}\epsilon_{1t-1} & \alpha_{21}a_{21}X_{3t-2}\epsilon_{2t-1} & \alpha_{31}a_{31}X_{3t-2}\epsilon_{3t-1} \\ \alpha_{12}a_{11}X_{3t-2}\epsilon_{1t-1} & \alpha_{22}a_{21}X_{3t-2}\epsilon_{2t-1} & \alpha_{32}a_{31}X_{3t-2}\epsilon_{3t-1} \end{pmatrix} + \nabla_{\phi_1} h_{t-2}\begin{pmatrix} b_{11} & 0 & 0 \\ 0 & b_{21} & 0 \\ 0 & 0 & b_{31} \end{pmatrix},$$

$$\nabla_{\phi_2} h_{t-1} = -2\begin{pmatrix} a_{11}\beta_1'X_{t-1}\epsilon_{1t-1} & 0 & 0 \\ 0 & a_{21}\beta_1'X_{t-1}\epsilon_{2t-1} & 0 \\ 0 & 0 & a_{31}\beta_1'X_{t-1}\epsilon_{3t-1} \\ a_{11}\beta_2'X_{t-1}\epsilon_{1t-1} & 0 & 0 \\ 0 & a_{21}\beta_2'X_{t-1}\epsilon_{2t-1} & 0 \\ 0 & 0 & a_{31}\beta_2'X_{t-1}\epsilon_{3t-1} \end{pmatrix} + \nabla_{\phi_2} h_{t-2}\begin{pmatrix} b_{11} & 0 & 0 \\ 0 & b_{21} & 0 \\ 0 & 0 & b_{31} \end{pmatrix}.$$

6.3 Estimating the Nuisance Parameters

When we apply Theorem 5.1 and Corollary 5.1, we need to consistently estimate the d eigenvalues of $I_d - \Upsilon\Upsilon' = I_d - \left(\alpha_\perp'\Omega_1^{-1}\alpha_\perp\right)^{1/2}(\alpha_\perp'(EV_{t-1})\alpha_\perp)^{-1}\left(\alpha_\perp'\Omega_1^{-1}\alpha_\perp\right)^{1/2}$. On the other hand, when we apply Theorem 5.2 and Corollary 5.2, we need to consistently estimate the d eigenvalues of $I_d - \Upsilon^H\Upsilon^{H\prime} = I_d - \left(\alpha_\perp'\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp\right)^{-1/2}\left(\alpha_\perp'\Omega_1^{-1}\alpha_\perp\right)(\alpha_\perp'(EV_{t-1})\alpha_\perp)^{-1}\left(\alpha_\perp'\Omega_1^{-1}\alpha_\perp\right)\left(\alpha_\perp'\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp\right)^{-1/2}$. All in all, we need to consistently estimate EV_{t-1}, $\alpha_\perp$, Ω_1 and Ω_1^*.

$$\widetilde{EV_{t-1}} = T^{-1}\sum_{t=1}^{T}\widetilde{V}_{t-1}. \quad (6.5)$$

On the other hand, by the definition of $\alpha_\perp$ (see around (2.2) above),

$$\widetilde{\alpha}_\perp = (I_p - cc'\widetilde{\alpha}(\widetilde{\alpha}'cc'\widetilde{\alpha})^{-1}\widetilde{\alpha}')c_\perp, \quad (6.6)$$

where $c = (I_r, 0_{r\times d})'$ and $c_\perp = (0_{d\times r}, I_d)'$. See p.48 of Johansen (1996) for a similar estimator.

Next, we refer to the definitions of Σ^*_{ij} in Ω^*_1 and Σ_{ij} in Ω_1 (see (3.5) and (3.8) respectively). Due to the form of h_{t-1}'s and the symmetric distribution of η_t,

$$\sum_{l=1}^{\infty}\nu_{il}\nu_{jl}E\left(\frac{\epsilon_{it-l}\epsilon_{jt-l}}{h_{it-1}h_{jt-1}}\right) = E\left[\left(\sum_{l=1}^{\infty}\frac{\nu_{il}\epsilon_{it-l}}{h_{it-1}}\right)\left(\sum_{l=1}^{\infty}\frac{\nu_{jl}\epsilon_{jt-l}}{h_{jt-1}}\right)\right] \equiv E\left[\left(\frac{e_{it-1}}{h_{it-1}}\right)\left(\frac{e_{jt-1}}{h_{jt-1}}\right)\right].$$

For $i = 1,\ldots,p$, set the initial values $\widetilde{e}_{i,-1} = \cdots = \widetilde{e}_{i,-s+1} = 0$. For $t = 1,\ldots,T$, recursively compute:

$$\widetilde{e}_{it-1} = \sum_{l=1}^{q}\widetilde{a}_{il}\widetilde{\epsilon}_{it-l} + \sum_{l=1}^{s}\widetilde{b}_{il}\widetilde{e}_{it-1-l}. \tag{6.7}$$

$\widetilde{\Delta}^* = \left(\widetilde{\Delta}^*_{ij}\right)_{p\times p} \equiv \frac{1}{T}\sum_{t=1}^{T} w\left(\widetilde{\eta}_t\widetilde{\eta}_t'\widetilde{\Lambda}^{-1}\right)\left(w\left(\widetilde{\eta}_t\widetilde{\eta}_t'\widetilde{\Lambda}^{-1}\right)\right)'$. Denote $\widetilde{\Lambda}^{-1} = \left(\widetilde{\lambda}^{ij}\right)_{p\times p}$, where for $i = j$, $\widetilde{\Delta}_{ii} \equiv 1 + \widetilde{\lambda}^{ii}$ and for $i \neq j$, $\widetilde{\Delta}_{ij} \equiv \widetilde{\lambda}^{ij}\widetilde{\lambda}_{ij}$. The estimates for Ω^*_1 and Ω_1 are:

$$\widetilde{\Omega}^*_1 = \frac{1}{T}\sum_{t=1}^{T}\widetilde{V}_{t-1}^{-1} + \widetilde{\Sigma}^*, \widetilde{\Sigma}^* = \left(\widetilde{\Sigma}^*_{ij}\right)_{p\times p}, \text{and } \widetilde{\Sigma}^*_{ij} = \left(\widetilde{\Delta}^*_{ij} - 1\right)\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{e}_{it-1}\widetilde{e}_{jt-1}}{\widetilde{h}_{it-1}\widetilde{h}_{jt-1}}; \tag{6.8}$$

$$\widetilde{\Omega}_1 = \frac{1}{T}\sum_{t=1}^{T}\widetilde{V}_{t-1}^{-1} + \widetilde{\Sigma}, \widetilde{\Sigma} = \left(\widetilde{\Sigma}_{ij}\right)_{p\times p}, \text{and } \widetilde{\Sigma}_{ij} = \widetilde{\Delta}_{ij}\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{e}_{it-1}\widetilde{e}_{jt-1}}{\widetilde{h}_{it-1}\widetilde{h}_{jt-1}}. \tag{6.9}$$

6.4 Simulating the Critical Value

In Section 5, we propose two types of tests for cointegrating rank. One is the LR (likelihood ratio) test while the other is the Hausman-type test. In this section, as an illustration, we simulate and tabulate the relevant critical values for the cases $d = 1$ and $d = 2$. Cases of higher dimensions can be done similarly.

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

$$tr\left\{\left[\zeta\left(I_d - \Xi\right)^{1/2} + \Phi\Xi^{1/2}\right]'\left[\zeta\left(I_d - \Xi\right)^{1/2} + \Phi\Xi^{1/2}\right]\right\} \text{ and} \tag{6.10}$$

$$tr\left\{\left[\zeta\left(I_d - \Xi^H\right)^{1/2} + \Phi\Xi^{H1/2}\right]'\left[\zeta\left(I_d - \Xi^H\right)^{1/2} + \Phi\Xi^{H1/2}\right]\right\} \text{ respectively.} \tag{6.11}$$

See Theorems 5.1 and 5.2. Theorem 5.1 depends on the fact that $\Omega^*_1 = \Omega_1$ while Theorem 5.2 does not. Further, if $\Xi = \Xi^H$, the distribution in (6.10) is exactly the same as that in (6.11).

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

$$tr\left\{\left[\bar{\zeta}\left(I_d - \Xi\right)^{1/2} + \Phi\Xi^{1/2}\right]'\left[\bar{\zeta}\left(I_d - \Xi\right)^{1/2} + \Phi\Xi^{1/2}\right]\right\} \text{ and} \tag{6.12}$$

$$tr\left\{\left[\bar{\zeta}\left(I_d - \Xi^H\right)^{1/2} + \Phi\Xi^{H1/2}\right]'\left[\bar{\zeta}\left(I_d - \Xi^H\right)^{1/2} + \Phi\Xi^{H1/2}\right]\right\} \text{ respectively.} \tag{6.13}$$

8 See Corollaries 5.1 and 5.2. Similarly, Corollary 5.1 depends on the fact that $\Omega^*_1 = \Omega_1$ while Corollary 5.2 does not. Again if $\Xi = \Xi^H$, the distribution in (6.12) is exactly the same as that in

(6.13).

The critical values of the distribution in (6.10)-(6.13) can be simulated via Monte Carlo method. For $d = 1$, we denote $\Xi = \xi_1$ or $\Xi^H = \xi_1$; while for $d = 2$, we denote $\Xi_2 = diag\,(\xi_1, \xi_2)$ or $\Xi^H = diag\,(\xi_1, \xi_2)$. For each independent replication, $vec\,(\Phi)$ is generated from a d^2-dimensional standard normal distribution, while the T (the sample size) ϵ_s's are generated from T i.i.d. $d-$dimensional standard normal distribution, which is also independent of that of Φ. $W_t \equiv \sum_{\tau=1}^{t} \epsilon_\tau$. When one considers (6.10)-(6.11), $\zeta \equiv \left(T^{-2} \sum_{t=1}^{T} W_{t-1} W_{t-1}'\right)^{-1/2} \left(T^{-1} \sum_{t=1}^{T} W_{t-1} \epsilon_t'\right)$. When one considers (6.12)-(6.13), $\bar{\zeta} \equiv \left(T^{-2} \sum_{t=1}^{T} \left(W_{t-1} - \bar{W}\right) \left(W_{t-1} - \bar{W}\right)'\right)^{-1/2} \left(T^{-1} \sum_{t=1}^{T} \left(W_{t-1} - \bar{W}\right) \epsilon_t'\right)$, $\bar{W} \equiv T^{-1} \sum_{t=1}^{T} W_{t-1}$.

We simulate the critical values with $d = 1$ and $d = 2$ and (ξ_1, ξ_2) ranging from 0.0 to 1.0. For intermediate values of (ξ_1, ξ_2), the critical values could be obtained by interpolation. The simulated critical values with 100,000 replications and $T = 2,000$ are tabulated in Tables 6.1-6.3.

TABLE 6.1: 90% Simulated Critical Values

H_G or LR_G	(ξ_1, ξ_2)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
d=1		2.995	2.978	2.964	2.941	2.914	2.883	2.845	2.811	2.782	2.746	2.720
	0.0	10.479	10.386	10.312	10.234	10.119	10.003	9.906	9.796	9.680	9.551	9.397
	0.1		10.295	10.217	10.125	10.018	9.919	9.808	9.679	9.561	9.455	9.305
	0.2			10.116	10.028	9.916	9.819	9.691	9.579	9.453	9.322	9.152
	0.3				9.931	9.816	9.693	9.565	9.442	9.310	9.163	9.004
d=2	0.4					9.707	9.576	9.440	9.313	9.176	9.018	8.866
	0.5						9.444	9.310	9.177	9.030	8.866	8.698
	0.6							9.153	9.015	8.857	8.698	8.527
	0.7								8.847	8.688	8.520	8.349
	0.8									8.526	8.345	8.160
	0.9										8.166	7.966
	1.0											7.750
$H_{G\mu}$ or $LR_{G\mu}$	(ξ_1, ξ_2)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
d=1		6.588	6.327	6.051	5.767	5.457	5.113	4.715	4.272	3.794	3.270	2.720
	0.0	15.842	15.578	15.270	14.955	14.630	14.277	13.913	13.537	13.132	12.712	12.237
	0.1		15.264	14.977	14.650	14.326	13.984	13.616	13.219	12.808	12.393	11.914
	0.2			14.661	14.341	13.985	13.630	13.259	12.861	12.457	12.034	11.583
	0.3				14.000	13.658	13.310	12.919	12.515	12.087	11.666	11.196
d=2	0.4					13.306	12.961	12.572	12.152	11.709	11.264	10.793
	0.5						12.595	12.191	11.774	11.309	10.850	10.352
	0.6							11.780	11.356	10.897	10.412	9.890
	0.7								10.921	10.456	9.956	9.409
	0.8									9.963	9.466	8.899
	0.9										8.942	8.340
	1.0											7.750

TABLE 6.2: 95% Simulated Critical Values

H_G or LR_G	(ξ_1, ξ_2)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
d=1		4.153	4.140	4.138	4.108	4.083	4.043	4.013	3.963	3.920	3.867	3.829
	0.0	12.286	12.237	12.158	12.073	11.987	11.887	11.789	11.676	11.559	11.446	11.251
	0.1		12.140	12.071	11.987	11.902	11.818	11.692	11.578	11.434	11.284	11.137
	0.2			11.973	11.879	11.791	11.691	11.566	11.433	11.293	11.141	11.001
	0.3				11.752	11.669	11.570	11.432	11.296	11.158	11.010	10.838
d=2	0.4					11.557	11.438	11.310	11.171	11.024	10.847	10.685
	0.5						11.322	11.176	11.049	10.854	10.693	10.519
	0.6							11.035	10.894	10.713	10.529	10.333
	0.7								10.719	10.555	10.353	10.150
	0.8									10.342	10.144	9.961
	0.9										9.932	9.711
	1.0											9.455
$H_{G\mu}$ or $LR_{G\mu}$	(ξ_1, ξ_2)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
d=1		8.167	7.932	7.656	7.373	7.049	6.679	6.251	5.763	5.220	4.576	3.829
	0.0	18.064	17.783	17.508	17.201	16.877	16.516	16.137	15.739	15.332	14.908	14.314
	0.1		17.530	17.212	16.897	16.564	16.206	15.833	15.439	15.004	14.561	13.983
	0.2			16.917	16.583	16.246	15.894	15.495	15.094	14.650	14.178	13.609
	0.3				16.253	15.906	15.539	15.147	14.736	14.277	13.806	13.228
d=2	0.4					15.556	15.153	14.747	14.345	13.888	13.372	12.834
	0.5						14.741	14.352	13.919	13.470	12.958	12.385
	0.6							13.924	13.484	13.018	12.479	11.911
	0.7								13.031	12.553	11.975	11.372
	0.8									12.045	11.429	10.809
	0.9										10.861	10.183
	1.0											9.455

TABLE 6.3: 99% Simulated Critical Values

H_G or LR_G	(ξ_1, ξ_2)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
d=1		7.018	6.941	6.939	6.931	6.929	6.895	6.842	6.839	6.774	6.718	6.657
	0.0	16.278	16.144	16.041	15.986	15.895	15.802	15.716	15.623	15.530	15.435	15.330
	0.1		16.105	15.991	15.920	15.806	15.643	15.552	15.482	15.337	15.247	15.037
	0.2			15.898	15.812	15.647	15.556	15.405	15.319	15.191	15.023	14.921
	0.3				15.702	15.609	15.471	15.318	15.202	15.021	14.870	14.802
d=2	0.4					15.510	15.374	15.231	15.087	14.928	14.747	14.660
	0.5						15.298	15.115	14.954	14.820	14.612	14.463
	0.6							14.993	14.809	14.622	14.480	14.279
	0.7								14.668	14.435	14.259	14.040
	0.8									14.255	14.064	13.817
	0.9										13.770	13.570
	1.0											13.405
$H_{G\mu}$ or $LR_{G\mu}$	(ξ_1, ξ_2)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
d=1		11.690	11.469	11.291	11.027	10.727	10.293	9.817	9.316	8.655	7.789	6.657
	0.0	22.745	22.477	22.192	21.901	21.541	21.236	20.842	20.390	19.987	19.498	18.761
	0.1		22.311	22.013	21.670	21.346	20.978	20.528	20.137	19.640	19.157	18.321
	0.2			21.739	21.389	21.052	20.695	20.250	19.815	19.281	18.757	17.991
	0.3				21.093	20.787	20.360	19.921	19.441	18.904	18.329	17.595
d=2	0.4					20.385	20.004	19.548	19.072	18.549	17.907	17.105
	0.5						19.545	19.122	18.639	18.047	17.429	16.611
	0.6							18.673	18.199	17.607	16.943	16.061
	0.7								17.617	17.093	16.376	15.493
	0.8									16.478	15.794	14.899
	0.9										15.178	14.140
	1.0											13.405

7 Monte Carlo Experiments

This section examines the performance of the likelihood ratio (LR) test and the Hausman-type test, both proposed in Section 5, in finite samples through Monte Carlo experiments. The experiments are based on the sample size, $T = 200, 400, 800$ with $1,000$ replications.

Throughout, we use the following VAR(1) (without or with a constant μ) with GARCH(1,1) error:

$$\Delta X_t = \Pi Y_{t-1} + \epsilon_t + \mu; \tag{7.1}$$

$$\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{pt})', \;\; \epsilon_{it} = \eta_{it}\sqrt{h_{it-1}},$$

$$h_{it-1} = a_{i0} + a_{i1}\epsilon_{it-1}^2 + b_{i1}h_{it-2}, \;\; i = 1, \ldots, p, a_{i0} = 1 - a_{i1} - b_{i1}; \tag{7.2}$$

$$\eta_t = (\eta_{1t}, \ldots, \eta_{pt})', \;\; E[\eta_t\eta_t'] = \Lambda = \begin{bmatrix} 1 & \cdots & \lambda \\ \vdots & \ddots & \vdots \\ \lambda & \cdots & 1 \end{bmatrix}. \tag{7.3}$$

Two sets of DGPs are generated. For the first set, $p = 2$, and for the second set, $p = 3$. Details are as follows.

DGP Set (1)

DGP(1*a*) $\Pi = \alpha\beta' = \begin{bmatrix} -0.4 \\ 0.12 \end{bmatrix} \begin{bmatrix} 1 & -2.5 \end{bmatrix};$

$\mu = 0_{2\times1}, \alpha;$

$(a_{i1}, b_{i1}) = (0.3, 0.65), (0.05, 0.90), (0.10, 0.85), (0.3, 0.699), (0.05, 0.949), (0.10, 0.899);$

$\lambda = 0.0, 0.5, 0.9.$

DGP(1*b*) $\Pi = \varsigma I_2, \varsigma = 0.01, 0.05, 0.1;$

$\mu = 0_{2\times1}, \begin{bmatrix} 1 \\ 1 \end{bmatrix};$

$(a_{i1}, b_{i1}) = (0.3, 0.65);$

$\lambda = 0.0.$

In DGP(1a), $p = 2$, $r = 1$ and our focus is the size of testing $H_0 : r = 1$. The first choice of μ corresponds to a DGP without a constant while the second one corresponds to that with a constant. See around (5.7). The different choices of (a_{i1}, b_{i1}) and λ allow one to see how the tests are sensitive to the parameters in the GARCH(1,1) model. Moreover, apart from normality of η_t, we also try T-distribution with degree of freedom equal to 8 and 5. They are denoted by t_8 and t_5 respectively. These two DGPs aim to see if the tests are sensitive to the leptokurtic standardized innovations. See the discussions in Lee and Tse (1996).

In DGP(1b), $p = r = 2$ and our focus is the power of testing $H_0 : r = 1$. As in DGP(1a), the first choice of μ corresponds to a DGP without a constant while the second one corresponds to that with a constant. Here we assume normality with $(a_{i1}, b_{i1}) = (0.30, 0.65)$ and $\lambda = 0.0$.

DGP Set (2)

$$DGP(2a) \qquad \Pi = \alpha\beta' = \begin{bmatrix} -0.4 & 0.4 \\ 0.12 & -0.34 \\ 0.1 & 0.166 \end{bmatrix} \begin{bmatrix} 1 & 0 & -0.8 \\ 0 & 1 & -0.48 \end{bmatrix};$$

$$\mu = 0_{3\times1}, \alpha;$$

$$(a_{i1}, b_{i1}) = (0.3, 0.65);$$

$$\lambda = 0.0.$$

$$DGP(2b) \qquad \Pi = \varsigma I_3, \varsigma = 0.01, 0.05, 0.1;$$

$$\mu = 0_{3\times1}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix};$$

$$(a_{i1}, b_{i1}) = (0.3, 0.65);$$

$$\lambda = 0.0.$$

In DGP(2a), $p = 3$, $r = 2$ and our focus is the size of testing $H_0 : r = 2$. In DGP(2b), $p = r = 3$ and our focus is the power of testing $H_0 : r = 2$. In both DGPs, the first choice of μ corresponds to a DGP without a constant while the second one corresponds to that with a constant. See around (5.7). In both DGP(2a) and DGP(2b), we assume normality with $(a_{i1}, b_{i1}) = (0.30, 0.65)$ and $\lambda = 0.0$.

For comparison, apart from the two tests that incorporates GARCH, namely the LR test (denoted by LR_G or $LR_{G\mu}$) and the Hausman-type test (denoted by H_G or $H_{G\mu}$), we also consider the test that ignores GARCH (denoted by LR_{NG} or $LR_{NG\mu}$). See Johansen (1996) and Reinsel and Ahn (1992). For the bi-variate DGP (Set 1), the rejection percentages of DGP(1a) are reported in Table 7.1 - Table 7.4, while the corresponding summary statistics are reported in Table 7.10 - Table 7.13. On the other hand, the rejection percentages of DGP(1b) can be found in Table 7.5 and Table 7.6. For the tri-variate DGPs (Set 2), the rejection percentages of DGP(2a) are reported in Table 7.7, while the corresponding summary statistics are reported in Table 7.14 - Table 7.15. On the other hand, the rejection percentages of DGP(2b) can be found in Table 7.8 and Table 7.9. For notational simplicity, there is not a constant if not otherwise stated.

As one can see from Tables 7.1 - 7.4 and Table 7.7, all three tests are of the reasonably correct finite-sample size, even when the number of observations is as small as 200. The tests slightly over-reject when the sample size is 200 or 400, and the over-rejections are comparable, except possibly for our LR test when the distribution of η_t is t_8 or t_5. Moreover, as one can see from Tables 7.5, 7.6, 7.8 and 7.9, it is clear that our Hausman-type test is by and large more powerful than the test that ignores GARCH.

As one can see from Table 7.10 - Table 7.15, the biases of the estimators with GARCH are comparable to those that do not take into account GARCH innovations, if not smaller than. More precisely, if GARCH is incorporated, the biases of the stationary mean parameters α_{ij}'s are smaller, while those of the nonstationary parameters B_{ij}'s are very close to the true parameters. Further, the standard deviations and the mean squared errors are definitely smaller, even when the sample size is as small as 200. That said, one should be careful interpreting the standard deviations of the estimators for nonstationary parameters B_{12}, B_{13} or B_{23}, since their distributions have fat tails.

All in all, due to the higher power and higher precision, the tests and the estimators suggested in this paper are recommended, except possibly when the sample size is as small as 200, or there are many mean parameters to be estimated (such as the case when $p = 3$ and $r = 2$ and there is a constant in the VAR model). Further, a test of multivariate heteroskedasticity may be performed before we decide to use the methodologies suggested in this paper. This approach is adopted in the empirical example reported in the next section.

TABLE 7.1 Rejection Percentage of DGP(1a)
p = 2 and r = 1, normality with $\lambda = 0.0$. $H_0 : r = 1$

T		(a_{i1}, b_{i1})=(0.30,0.65)			(a_{i1}, b_{i1})=(0.05,0.90)			(a_{i1}, b_{i1})=(0.10,0.85)		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
200	LR_{NG}	0.100	0.059	0.010	0.084	0.049	0.013	0.091	0.050	0.010
	LR_G	0.128	0.076	0.024	0.090	0.054	0.018	0.102	0.058	0.013
	H_G	0.121	0.075	0.021	0.090	0.054	0.018	0.102	0.057	0.013
400	LR_{NG}	0.111	0.066	0.023	0.106	0.053	0.008	0.102	0.053	0.007
	LR_G	0.131	0.074	0.028	0.112	0.051	0.010	0.106	0.052	0.018
	H_G	0.126	0.070	0.027	0.112	0.051	0.010	0.105	0.051	0.018
800	LR_{NG}	0.114	0.067	0.014	0.099	0.045	0.008	0.095	0.050	0.009
	LR_G	0.118	0.067	0.022	0.103	0.050	0.009	0.102	0.051	0.014
	H_G	0.115	0.065	0.022	0.103	0.050	0.009	0.102	0.051	0.014

TABLE 7.2: Rejection Percentage of DGP(1a)
p = 2 and r = 1, normality with $\lambda = 0.0$. $H_0 : r = 1$

T		(a_{i1}, b_{i1})=(0.30,0.699)			(a_{i1}, b_{i1})=(0.05,0.949)			(a_{i1}, b_{i1})=(0.10,0.899)		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
200	LR_{NG}	0.100	0.060	0.024	0.097	0.053	0.008	0.101	0.049	0.012
	LR_G	0.195	0.143	0.086	0.108	0.057	0.015	0.122	0.080	0.023
	H_G	0.176	0.121	0.069	0.108	0.057	0.015	0.122	0.080	0.023
400	LR_{NG}	0.123	0.083	0.028	0.100	0.055	0.004	0.111	0.057	0.014
	LR_G	0.192	0.150	0.080	0.126	0.070	0.021	0.153	0.103	0.043
	H_G	0.174	0.125	0.070	0.125	0.070	0.021	0.150	0.100	0.043
800	LR_{NG}	0.116	0.076	0.034	0.094	0.050	0.010	0.097	0.057	0.015
	LR_G	0.158	0.104	0.056	0.122	0.067	0.021	0.145	0.087	0.034
	H_G	0.136	0.092	0.042	0.122	0.067	0.021	0.144	0.087	0.033

TABLE 7.3: Rejection Percentage of DGP(1a)
p = 2 and r = 1, $(a_{i1}, b_{i1}) = (0.30, 0.65)$, $\lambda = 0.0$. $H_0 : r = 1$

T		t_8			t_5		
		10%	5%	1%	10%	5%	1%
200	LR_{NG}	0.137	0.081	0.024	0.110	0.058	0.018
	LR_G	0.159	0.111	0.054	0.173	0.123	0.053
	H_G	0.151	0.097	0.039	0.146	0.105	0.038
400	LR_{NG}	0.113	0.065	0.020	0.101	0.055	0.012
	LR_G	0.132	0.072	0.022	0.148	0.098	0.035
	H_G	0.112	0.061	0.016	0.123	0.074	0.020
800	LR_{NG}	0.122	0.078	0.024	0.108	0.062	0.024
	LR_G	0.122	0.076	0.022	0.147	0.091	0.038
	H_G	0.113	0.064	0.016	0.111	0.069	0.023

TABLE 7.4: Rejection Percentage of DGP(1a)
p = 2 and r = 1, normality with $(a_{i1}, b_{i1}) = (0.30, 0.65)$. $H_0 : r = 1$

T		$\lambda = 0.5$			$\lambda = 0.9$			$\lambda = 0.0$, *w/ a constant*		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
200	LR_{NG}	0.102	0.058	0.011	0.113	0.063	0.018	0.141	0.077	0.027
	LR_G	0.142	0.082	0.027	0.162	0.101	0.054	0.216	0.153	0.082
	H_G	0.131	0.078	0.024	0.149	0.092	0.045	0.214	0.145	0.080
400	LR_{NG}	0.107	0.061	0.014	0.132	0.067	0.010	0.141	0.086	0.037
	LR_G	0.135	0.077	0.022	0.156	0.109	0.042	0.191	0.126	0.064
	H_G	0.126	0.069	0.021	0.135	0.089	0.032	0.188	0.121	0.059
800	LR_{NG}	0.112	0.051	0.020	0.101	0.055	0.020	0.125	0.080	0.022
	LR_G	0.115	0.070	0.010	0.133	0.082	0.030	0.144	0.091	0.030
	H_G	0.108	0.063	0.007	0.113	0.061	0.017	0.141	0.089	0.024

TABLE 7.5: Rejection Percentage of DGP(1b)
p = 2 and r = 2, normality with $(a_{i1}, b_{i1}) = (0.30, 0.65)$, $\lambda = 0.0$. $H_0 : r = 1$

T		$\varsigma = -0.01$			$\varsigma = -0.05$			$\varsigma = -0.1$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
200	LR_{NG}	0.073	0.027	0.000	0.754	0.500	0.112	0.984	0.958	0.699
	LR_G	0.279	0.174	0.054	0.934	0.881	0.637	0.998	0.996	0.963
	H_G	0.275	0.171	0.053	0.935	0.880	0.635	0.998	0.996	0.961
400	LR_{NG}	0.152	0.067	0.005	0.984	0.953	0.681	1.000	0.997	0.990
	LR_G	0.557	0.408	0.190	0.995	0.993	0.979	0.998	0.998	0.998
	H_G	0.552	0.400	0.178	0.995	0.993	0.978	0.998	0.998	0.998
800	LR_{NG}	0.540	0.301	0.044	0.997	0.996	0.991	1.000	0.999	0.999
	LR_G	0.902	0.808	0.557	0.999	0.999	0.998	1.000	1.000	0.999
	H_G	0.896	0.809	0.547	0.999	0.999	0.998	1.000	1.000	0.999

TABLE 7.6: Rejection Percentage of DGP(1b)
p = 2 and r = 2, normality with $(a_{i1}, b_{i1}) = (0.30, 0.65)$, $\lambda = 0.0$, w/ a constant. $H_0 : r = 1$

T		$\varsigma = -0.01$			$\varsigma = -0.05$			$\varsigma = -0.1$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
200	LR_{NG}	0.323	0.196	0.077	0.604	0.431	0.173	0.931	0.828	0.503
	LR_G	0.487	0.365	0.174	0.879	0.795	0.558	0.992	0.982	0.919
	H_G	0.480	0.357	0.167	0.876	0.792	0.543	0.991	0.980	0.921
400	LR_{NG}	0.390	0.275	0.106	0.932	0.865	0.555	0.997	0.993	0.970
	LR_G	0.562	0.438	0.223	0.991	0.983	0.927	0.998	0.998	0.997
	H_G	0.555	0.432	0.216	0.991	0.983	0.926	0.998	0.998	0.996
800	LR_{NG}	0.526	0.366	0.162	0.995	0.994	0.967	0.999	0.999	0.998
	LR_G	0.819	0.683	0.442	0.999	0.999	0.999	1.000	1.000	1.000
	H_G	0.815	0.680	0.430	0.999	0.999	0.998	1.000	1.000	1.000

TABLE 7.7: Rejection Percentage of DGP(2a)
p = 3 and r = 2, normality with $(a_{i1}, b_{i1}) = (0.30, 0.65)$, $\lambda = 0.0$.
$H_0 : r = 2$

T		*w/o a constant*			*w/ a constant*		
		10%	5%	1%	10%	5%	1%
200	LR_{NG}	0.107	0.061	0.015	0.119	0.067	0.025
	LR_G	0.145	0.076	0.028	0.194	0.124	0.050
	H_G	0.139	0.072	0.025	0.190	0.118	0.047
400	LR_{NG}	0.119	0.069	0.019	0.128	0.081	0.024
	LR_G	0.126	0.080	0.030	0.169	0.116	0.040
	H_G	0.123	0.079	0.029	0.166	0.112	0.038
800	LR_{NG}	0.124	0.069	0.019	0.120	0.058	0.017
	LR_G	0.128	0.063	0.015	0.134	0.072	0.022
	H_G	0.120	0.056	0.013	0.130	0.068	0.020

TABLE 7.8: Rejection Percentage of DGP(2b)
p = 3 and r = 3, normality with $(a_{i1}, b_{i1}) = (0.30, 0.65)$, $\lambda = 0.0$. $H_0 : r = 2$

T		$\varsigma = -0.01$			$\varsigma = -0.05$			$\varsigma = -0.1$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
200	LR_{NG}	0.020	0.004	0.000	0.631	0.306	0.035	0.979	0.939	0.575
	LR_G	0.225	0.121	0.042	0.915	0.828	0.567	0.998	0.995	0.954
	H_G	0.220	0.116	0.038	0.915	0.827	0.558	0.998	0.995	0.957
400	LR_{NG}	0.069	0.021	0.000	0.974	0.927	0.547	1.000	0.997	0.984
	LR_G	0.508	0.384	0.165	0.999	0.996	0.976	0.999	0.999	0.999
	H_G	0.500	0.372	0.154	0.998	0.996	0.976	0.999	0.999	0.999
800	LR_{NG}	0.421	0.164	0.011	0.997	0.994	0.989	0.999	0.998	0.997
	LR_G	0.889	0.789	0.519	0.999	0.999	0.998	1.000	1.000	1.000
	H_G	0.884	0.782	0.509	0.999	0.999	0.997	1.000	1.000	1.000

TABLE 7.9: Rejection Percentage of DGP(2b)
p = 3 and r = 3, normality with $(a_{i1}, b_{i1}) = (0.30, 0.65)$, $\lambda = 0.0$, w/ a constant. $H_0 : r = 2$

T		$\varsigma = -0.01$			$\varsigma = -0.05$			$\varsigma = -0.1$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
200	LR_{NG}	0.134	0.052	0.009	0.386	0.193	0.034	0.886	0.688	0.266
	LR_G	0.354	0.241	0.084	0.783	0.681	0.435	0.993	0.971	0.847
	H_G	0.346	0.237	0.078	0.783	0.677	0.425	0.992	0.971	0.841
400	LR_{NG}	0.165	0.065	0.018	0.856	0.718	0.304	0.995	0.989	0.937
	LR_G	0.439	0.321	0.121	0.995	0.983	0.910	0.999	0.998	0.998
	H_G	0.428	0.311	0.116	0.993	0.982	0.907	0.998	0.998	0.998
800	LR_{NG}	0.283	0.156	0.023	0.993	0.989	0.946	0.998	0.998	0.996
	LR_G	0.743	0.614	0.366	0.999	0.998	0.997	1.000	1.000	1.000
	H_G	0.740	0.605	0.350	0.998	0.998	0.997	1.000	1.000	1.000

TABLE 7.10: Empirical Mean (and Empircal Standard Deviation) of DGP(1a)
p = 2 and r = 1, normality with $\lambda = 0.0$.

T		$(0.30, 0.65)$			$(0.05, 0.90)$			$(0.10, 0.85)$		
		$\beta_{21} =$ -2.50	$\alpha_{11} =$ -0.40	$\alpha_{21} =$ 0.12	$\beta_{21} =$ -2.50	$\alpha_{11} =$ -0.40	$\alpha_{21} =$ 0.12	$\beta_{21} =$ -2.50	$\alpha_{11} =$ -0.40	$\alpha_{21} =$ 0.12
200	ng	-2.4977	-0.4027	0.1247	-2.4975	-0.4021	0.1227	-2.4972	-0.4023	0.1230
		(0.0508)	(0.0356)	(0.0377)	(0.0462)	(0.0264)	(0.0272)	(0.0467)	(0.0274)	(0.0290)
	g	-2.4994	-0.3991	0.1220	-2.4978	-0.4015	0.1222	-2.4979	-0.4009	0.1219
		(0.0404)	(0.0307)	(0.0275)	(0.0466)	(0.0267)	(0.0268)	(0.0455)	(0.0274)	(0.0270)
400	ng	-2.4972	-0.4018	0.1219	-2.4972	-0.4012	0.1199	-2.4974	-0.4013	0.1203
		(0.0240)	(0.0265)	(0.0311)	(0.0228)	(0.0183)	(0.0184)	(0.0229)	(0.0188)	(0.0205)
	g	-2.4971	-0.3985	0.1196	-2.4971	-0.4012	0.1196	-2.4971	-0.4009	0.1194
		(0.0186)	(0.0236)	(0.0293)	(0.0226)	(0.0186)	(0.0182)	(0.0216)	(0.0188)	(0.0185)
800	ng	-2.4977	-0.4012	0.1206	-2.4976	-0.4001	0.1208	-2.4976	-0.4003	0.1208
		(0.0131)	(0.0249)	(0.0253)	(0.0124)	(0.0127)	(0.0139)	(0.0125)	(0.0132)	(0.0153)
	g	-2.4978	-0.3980	0.1206	-2.4975	-0.3996	0.1209	-2.4975	-0.3992	0.1207
		(0.0095)	(0.0213)	(0.0149)	(0.0121)	(0.0128)	(0.0135)	(0.0115)	(0.0133)	(0.0136)

Empirical stadnard deviation is in bracket. ng: no garch, g: garch.

TABLE 7.11: Empirical Mean (and Empircal Standard Deviation) of DGP(1a)
p = 2 and r = 1, normality with $\lambda = 0.0$.

T		(0.30, 0.699)			(0.05, 0.949)			(0.10, 0.899)		
		$\beta_{21} = -2.50$	$\alpha_{11} = -0.40$	$\alpha_{21} = 0.12$	$\beta_{21} = -2.50$	$\alpha_{11} = -0.40$	$\alpha_{21} = 0.12$	$\beta_{21} = -2.50$	$\alpha_{11} = -0.40$	$\alpha_{21} = 0.12$
200	ng	-2.4991	-0.4043	0.1271	-2.4963	-0.4024	0.1226	-2.4957	-0.4027	0.1229
		(0.0538)	(0.0565)	(0.0469)	(0.0487)	(0.0287)	(0.0269)	(0.0503)	(0.0347)	(0.0294)
	g	-2.5012	-0.3955	0.1216	-2.4966	-0.4016	0.1220	-2.4970	-0.4007	0.1215
		(0.0369)	(0.0420)	(0.0289)	(0.0480)	(0.0286)	(0.0259)	(0.0443)	(0.0326)	(0.0252)
400	ng	-2.4952	-0.4035	0.1214	-2.4967	-0.4015	0.1198	-2.4969	-0.4020	0.1197
		(0.0298)	(0.0437)	(0.0401)	(0.0245)	(0.0204)	(0.0189)	(0.0271)	(0.0253)	(0.0228)
	g	-2.4973	-0.3982	0.1200	-2.4964	-0.4013	0.1195	-2.4972	-0.4004	0.1190
		(0.0166)	(0.0312)	(0.0280)	(0.0232)	(0.0201)	(0.0178)	(0.0220)	(0.0230)	(0.0181)
800	ng	-2.4974	-0.4015	0.1219	-2.4975	-0.4001	0.1210	-2.4976	-0.4004	0.1212
		(0.0139)	(0.0406)	(0.0334)	(0.0137)	(0.0155)	(0.0145)	(0.0136)	(0.0210)	(0.0179)
	g	-2.4981	-0.3987	0.1206	-2.4975	-0.3996	0.1207	-2.4976	-0.3990	0.1205
		(0.0081)	(0.0316)	(0.0174)	(0.0123)	(0.0155)	(0.0130)	(0.0097)	(0.0168)	(0.0128)

Empirical stadnard deviation is in bracket. ng: no garch, g: garch.

TABLE 7.12: Empirical Mean (and Empircal Standard Deviation)
of DGP(1a). p = 2 and r = 1, $(a_{i1}, b_{i1}) = (0.30, 0.65)$, $\lambda = 0.0$.

T		t_8			t_5		
		$\beta_{21} = -2.50$	$\alpha_{11} = -0.40$	$\alpha_{21} = 0.12$	$\beta_{21} = -2.50$	$\alpha_{11} = -0.40$	$\alpha_{21} = 0.12$
200	no garch	-2.4969	-0.4016	0.1243	-2.4993	-0.4004	0.1251
		(0.0563)	(0.0358)	(0.0405)	(0.0535)	(0.0362)	(0.0414)
	garch	-2.4975	-0.3990	0.1220	-2.4982	-0.3984	0.1220
		(0.0426)	(0.0316)	(0.0308)	(0.0468)	(0.0372)	(0.0340)
400	no garch	-2.4985	-0.4022	0.1239	-2.4973	-0.3997	0.1221
		(0.0274)	(0.0282)	(0.0332)	(0.0268)	(0.0291)	(0.0317)
	garch	-2.4980	-0.3977	0.1209	-2.4975	-0.3988	0.1206
		(0.0199)	(0.0250)	(0.0231)	(0.0209)	(0.0284)	(0.0248)
800	no garch	-2.4979	-0.4007	0.1200	-2.4977	-0.3999	0.1223
		(0.0130)	(0.0234)	(0.0258)	(0.0141)	(0.0212)	(0.0289)
	garch	-2.4975	-0.3993	0.1197	-2.4981	-0.3997	0.1205
		(0.0091)	(0.0202)	(0.0190)	(0.0104)	(0.0234)	(0.0235)

Empirical stadnard deviation is in bracket.

TABLE 7.13: Empirical Mean (and Empircal Standard Deviation) of DGP(1a)
p = 2 and r = 1, normality with $(a_{i1}, b_{i1}) = (0.30, 0.65)$.

T		$\lambda = 0.50$			$\lambda = 0.90$			$\lambda = 0.0$, *w/ a constant*		
		$\beta_{21} = -2.50$	$\alpha_{11} = -0.40$	$\alpha_{21} = 0.12$	$\beta_{21} = -2.50$	$\alpha_{11} = -0.40$	$\alpha_{21} = 0.12$	$\beta_{21} = -2.50$	$\alpha_{11} = -0.40$	$\alpha_{21} = 0.12$
200	*ng*	-2.4975	-0.4020	0.1254	-2.4973	-0.3980	0.1255	-2.4971	-0.4040	0.1268
		(0.0379)	(0.0397)	(0.0447)	(0.0176)	(0.0565)	(0.0598)	(0.0748)	(0.0361)	(0.0378)
	g	-2.4995	-0.3981	0.1215	-2.4979	-0.3981	0.1215	-2.5021	-0.3977	0.1211
		(0.0308)	(0.0331)	(0.0322)	(0.0137)	(0.0415)	(0.0422)	(0.0654)	(0.0310)	(0.0301)
400	*ng*	-2.4975	-0.4017	0.1211	-2.4980	-0.3992	0.1207	-2.4985	-0.4022	0.1230
		(0.0182)	(0.0309)	(0.0369)	(0.0093)	(0.0434)	(0.0484)	(0.0355)	(0.0266)	(0.0311)
	g	-2.4977	-0.3981	0.1195	-2.4979	-0.3994	0.1195	-2.4993	-0.3979	0.1192
		(0.0140)	(0.0234)	(0.0268)	(0.0078)	(0.0329)	(0.0345)	(0.0289)	(0.0236)	(0.0276)
800	*ng*	-2.4977	-0.4014	0.1206	-2.4977	-0.4002	0.1206	-2.4981	-0.4015	0.1212
		(0.0100)	(0.0289)	(0.0299)	(0.0052)	(0.0386)	(0.0396)	(0.0183)	(0.0250)	(0.0252)
	g	-2.4980	-0.3977	0.1203	-2.4977	-0.3972	0.1205	-2.4987	-0.3975	0.1202
		(0.0071)	(0.0229)	(0.0176)	(0.0038)	(0.0317)	(0.0248)	(0.0134)	(0.0206)	(0.0152)

Empirical stadnard deviation is in bracket. ng: no garch, g: garch.

TABLE 7.14: Empirical Mean (and Empircal Standard Deviation) of DGP(2a)
p = 3 and r = 2, normality with $(a_{i1}, b_{i1}) = (0.30, 0.65)$, $\lambda = 0.0$.

T		$\beta_{31} = -0.80$	$\beta_{32} = -0.48$	$\alpha_{11} = -0.40$	$\alpha_{12} = 0.40$	$\alpha_{21} = 0.12$	$\alpha_{22} = -0.34$	$\alpha_{31} = 0.10$	$\alpha_{32} = 0.166$
200	no garch	-0.8019	-0.4812	-0.4117	0.4072	0.1233	-0.3600	0.1085	0.1730
		(0.0521)	(0.0394)	(0.0667)	(0.0749)	(0.0683)	(0.0884)	(0.0596)	(0.0676)
	garch	-0.7993	-0.4791	-0.3997	0.4030	0.1232	-0.3465	0.1000	0.1666
		(0.0454)	(0.0346)	(0.0578)	(0.0664)	(0.0491)	(0.0733)	(0.0517)	(0.0571)
400	no garch	-0.7987	-0.4804	-0.4086	0.4031	0.1205	-0.3504	0.1044	0.1698
		(0.0268)	(0.0461)	(0.0560)	(0.0545)	(0.0459)	(0.0708)	(0.0465)	(0.0465)
	garch	-0.7981	-0.4839	-0.3994	0.4015	0.1180	-0.3400	0.0980	0.1642
		(0.0254)	(0.1535)	(0.0741)	(0.0477)	(0.0931)	(0.0576)	(0.0400)	(0.0407)
800	no garch	-0.7999	-0.4800	-0.4050	0.4033	0.1193	-0.3446	0.1002	0.1687
		(0.0142)	(0.0108)	(0.0458)	(0.0388)	(0.0346)	(0.0570)	(0.0334)	(0.0345)
	garch	-0.7995	-0.4798	-0.3975	0.3982	0.1212	-0.3399	0.0980	0.1640
		(0.0108)	(0.0081)	(0.0419)	(0.0370)	(0.0259)	(0.0392)	(0.0292)	(0.0310)

Empirical stadnard deviation is in bracket.

TABLE 7.15: Empirical Mean (and Empircal Standard Deviation) of DGP(2a)
p = 3 and r = 2, normality with $(a_{i1}, b_{i1}) = (0.30, 0.65)$, $\lambda = 0.0$, w/ a constant.

T		$\beta_{31} = -0.80$	$\beta_{32} = -0.48$	$\alpha_{11} = -0.40$	$\alpha_{12} = 0.40$	$\alpha_{21} = 0.12$	$\alpha_{22} = -0.34$	$\alpha_{31} = 0.10$	$\alpha_{32} = 0.166$
200	no garch	-0.8027	-0.4814	-0.4167	0.4088	0.1224	-0.3670	0.1119	0.1750
		(0.0836)	(0.0614)	(0.0676)	(0.0761)	(0.0691)	(0.0902)	(0.0602)	(0.0687)
	garch	-0.8001	-0.4787	-0.3974	0.4020	0.1239	-0.3469	0.0979	0.1651
		(0.0684)	(0.0519)	(0.0578)	(0.0655)	(0.0490)	(0.0705)	(0.0503)	(0.0565)
400	no garch	-0.7989	-0.4797	-0.4110	0.4043	0.1204	-0.3541	0.1064	0.1705
		(0.0385)	(0.0525)	(0.0566)	(0.0546)	(0.0459)	(0.0715)	(0.0467)	(0.0466)
	garch	-0.7989	-0.4844	-0.3974	0.4004	0.1170	-0.3393	0.0974	0.1630
		(0.0351)	(0.1732)	(0.0773)	(0.0487)	(0.1091)	(0.0603)	(0.0405)	(0.0396)
800	no garch	-0.8006	-0.4801	-0.4060	0.4038	0.1194	-0.3463	0.1011	0.1692
		(0.0192)	(0.0149)	(0.0459)	(0.0388)	(0.0347)	(0.0570)	(0.0334)	(0.0345)
	garch	-0.7996	-0.4795	-0.3962	0.3974	0.1213	-0.3393	0.0970	0.1634
		(0.0146)	(0.0109)	(0.0413)	(0.0363)	(0.0258)	(0.0396)	(0.0289)	(0.0306)

Empirical stadnard deviation is in bracket.

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 2004 which amounts to 540 observations. We first estimate a $VAR(k)$ with a constant with different order k, where $k = 1, \ldots, 10$. $VAR(6)$ attains the lowest AIC. For comparison with the results in Reinsel and Ahn (1992), we also use the series from January 1960 to December 1979 which amounts to 240 observations. $VAR(4)$ instead attains the lowest AIC. The residuals from the full-sample and this first sub-sample are both applied to a test for multivariate heteroscedasticity, along the lines in Ling and Li (1997) (see also Sin, 2005). The χ^2 test statistics with different numbers of terms are reported in Table 8.1.

Table 8.1 clearly shows that the hypothesis of homoscedasticity is rejected and suggests that ε_t is not i.i.d. In view of this, apart from Johansen's estimation, we also perform the full-rank as well as the reduced-rank estimation elucidated in Sections 3 and 4, which incorporate a GARCH(1,1) model. Results of the $LR_{NG\mu}$ test (the LR test that ignores GARCH) and the $H_{G\mu}$ test (the Hausman-type test), which has similar performance as the $LR_{G\mu}$ (the LR test that incorporates GARCH) in our Monte-Carlo experiments reported in Section 7, are summarized in Table 8.2(a). While we confine our discussion to the case that $k = 6$, only for completeness we report the results with different k. Table 8.2(a) shows the hypothesis that $r = 0$ or that $r = 1$ is rejected by both tests. Both tests do not reject the hypothesis that $r = 2$. All in all, when the full-sample is used, we conclude that the interest rates are nonstationary and there are two cointegrating vectors.

We also report in Table 8.2(b) the test for reduced rank with the first sub-sample. While we confine our discussion to the case that $k = 4$, only for completeness we report the results with different k. As one can see in the table, similar to the results in Reinsel and Ahn (1992), $LR_{NG\mu}$ can hardly reject or only marginally rejects (at 5%) the null of $r = 1$. Despite the possible slight over-rejection, as suggested in Section 7, the $H_{G\mu}$ rejects it (at around 1%), except possibly when $k = 3$ when it is arguably the case of insufficient number of lags.

TABLE 8.1: Test Statistics for Multivariate Heteroskedasticity

no. of terms		1	2	3	4	5	6	7	8	9	10
full-sample	$k = 6$	90.87	128.8	163.2	172.9	198.0	232.3	250.5	269.2	292.8	299.4
1st sub-sample	$k = 4$	23.08	34.91	43.87	50.20	68.43	89.00	101.9	149.9	155.9	158.0

TABLE 8.2(a): Testing for Cointegrating Rank: full-sample

	$H_0 : r = 0$		$H_0 : r = 1$		$H_0 : r = 2$	
s	$LR_{NG\mu}$	$H_{G\mu}$	$LR_{NG\mu}$	$H_{G\mu}$	$LR_{NG\mu}$	$H_{G\mu}$
1	140.3 (.000)	521.3 (.000)	47.49 (.000)	135.0 (.000)	1.009 (.758)	2.443 (.130)
2	124.7 (.000)	441.0 (.000)	57.98 (.000)	161.3 (.000)	0.8001 (.795)	1.828 (.187)
3	90.94 (.000)	321.6 (.000)	42.58 (.000)	192.7 (.000)	0.8683 (.783)	1.816 (.188)
4	99.43 (.000)	355.4 (.000)	45.66 (.000)	208.5 (.000)	0.7930 (.796)	1.673 (.207)
5	88.46 (.000)	333.5 (.000)	38.92 (.000)	206.9 (.000)	1.067 (.748)	2.470 (.126)
6	86.73 (.000)	364.7 (.000)	33.15 (.000)	207.7 (.000)	0.7503 (.804)	1.823 (.188)

p-value is in bracket.

TABLE 8.2(b): Testing for Cointegrating Rank: 1st sub-sample

	$H_0: r=0$		$H_0: r=1$		$H_0: r=2$	
k	$LR_{NG\mu}$	$H_{G\mu}$	$LR_{NG\mu}$	$H_{G\mu}$	$LR_{NG\mu}$	$H_{G\mu}$
1	59.85 (.000)	165.6 (.000)	13.04 (.219)	16.00 (.006)	0.0035 (.990)	0.3024 (.584)
2	48.47 (.000)	90.61 (.000)	14.07 (.167)	15.04 (.008)	0.0114 (.981)	0.0023 (.962)
3	37.85 (.010)	63.62 (.000)	12.89 (.228)	5.995 (.234)	0.0015 (.993)	0.0017 (.967)
4	48.45 (.000)	101.1 (.000)	18.23 (.047)	16.02 (.005)	0.0000 (.999)	0.0072 (.932)
5	48.08 (.000)	130.7 (.000)	14.45 (.150)	16.08 (.005)	0.0001 (.998)	0.0070 (.933)
6	44.03 (.000)	102.8 (.000)	15.39 (.114)	15.12 (.007)	0.0024 (.992)	0.0113 (.916)

p-value is in bracket.

We close this section with the estimation of the parameters. As one can see in that follows, the estimators of the "nonstationary mean" parameters B are rather stable, regardless one ignores or incorporates GARCH. This is consistent with the Monte-Carlo findings in Section 7. In contrast, the estimators of the "stationary mean" parameters α (as well as other "stationary mean" parameters, which are not reported for brevity) are not. Moreover, it is noted that in the GARCH(1,1) estimation, the sum of a_{i1} and b_{i1} is close to 1 but does not exceed 0.97, while the coefficients of the cross-correlation lie between 0.77 and 0.93.

Estimation ignoring GARCH: full-sample

$$\widehat{\Pi} = \begin{bmatrix} -0.1985 & 0.1201 & 0.0968 \\ -0.0025 & -0.2356 & 0.2503 \\ 0.0033 & -0.1060 & 0.1033 \end{bmatrix}, \ \widehat{\alpha}\widehat{\beta}' = \begin{bmatrix} -0.1964 & 0.1204 \\ -0.0031 & -0.2357 \\ 0.0060 & -0.1056 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1.1356 \\ 0 & 1 & -1.0464 \end{bmatrix}. \tag{8.1}$$

Estimation incorporating GARCH: full-sample

$$\dot{\Pi} = \begin{bmatrix} -0.1435 & -0.0066 & 0.1588 \\ 0.0715 & -0.3549 & 0.2752 \\ 0.0906 & -0.2704 & 0.1599 \end{bmatrix}, \ \widetilde{\alpha}\widetilde{\beta}' = \begin{bmatrix} -0.0930 & -0.0854 \\ 0.1072 & -0.4052 \\ 0.0995 & -0.2675 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1.1258 \\ 0 & 1 & -1.0328 \end{bmatrix}; \tag{8.2}$$

$$\dot{h}_{t-1} = \begin{bmatrix} \dot{h}_{1t-1} \\ \dot{h}_{2t-1} \\ \dot{h}_{3t-1} \end{bmatrix} = \begin{bmatrix} 0.0003 + 0.3707\dot{\epsilon}^2_{1t-1} + 0.5955\dot{h}_{1t-2} \\ 0.0003 + 0.1716\dot{\epsilon}^2_{2t-1} + 0.7941\dot{h}_{2t-2} \\ 0.0006 + 0.2159\dot{\epsilon}^2_{3t-1} + 0.7089\dot{h}_{3t-2} \end{bmatrix}, \ \dot{\Lambda} = \begin{bmatrix} 1.0000 & 0.8058 & 0.7736 \\ 0.8058 & 1.0000 & 0.9283 \\ 0.7736 & 0.9283 & 1.0000 \end{bmatrix}. \tag{8.3}$$

9 Conclusions

In this paper, we extend Li, Ling and Wong (2001)'s partially nonstationary model from RCAR to VAR-GARCH, which is more appealing in practice. The constant correlation multivariate GARCH suggested by Bollerslev (1990) is adopted. More importantly, we adopt Anderson (1951) and Johansen (1988)'s approach and do not assume that we do not know a priori some variables of interest are non-cointegrated. We make use, and modify, the asymptotic distributions of the full rank and the reduced rank quasi-maximum likelihood (QMLE) estimators developed in Li, Ling and Wong (2001) and Seo (2007). The main thrust of this paper is using these two estimators to construct a likelihood ratio (LR) test for cointegrating rank, where the asymptotic distribution is in turn a functional of a standard Brownian motion and a standard normal vector, with d to-be-estimated nuisance parameters, where d is the difference between the number of variables and the cointegrating rank. To the best of our knowledge, tests of this sort do not exist in the literature. The null distributions of these tests are non-standard but the critical value can be easily simulated via Monte Carlo method. On the other hand, Monte-Carlo experiments show that our tests are more powerful, with size comparable to the conventional test. Applying our test to three U.S. interest rates, we find evidence that there are two cointegrating vectors, with fewer observations than the conventional test.

Throughout the paper, we confine our attention to cases that (1) the correlations in the multivariate GARCH model are time-invariant; (2) the univariate GARCH model is symmetric in the errors ϵ_{it}'s; and (3) the standardized errors η_{it}'s are symmetrically distributed. In view of the growing literature of time-varying correlations (see, for instance, Tse, 2000), (1) seems to be quite restrictive. That said, extending the probability theories developed in Chan and Wei (1988) and Ling and Li (1998) is not straightforward. (2) and (3) are also restrictive, in view of the extended GARCH models such as GJR (Glosten, Jagannathan and Runkle, 1993) and the typical skewness of financial data. To accommodate these variations, either the distributions derived in this paper needs to be modified, or we may need to modify the estimation methods and/or the testing procedures. We leave this challenging task to future research.

A Appendix: Estimating the Variance Parameters

In this appendix, we consider the full rank estimators for the variance parameters $\delta \equiv [\delta_1', \delta_2']'$, demoted as $\dot{\delta}$. Those for the reduced rank estimators, denoted as $\tilde{\delta}$, are similar and the asymptotic distributions are exactly the same. The details are thus omitted.

Refer to Process (1.2)-(1.3). The variance parameters consist of two sub-vectors δ_1 and δ_2, where $\delta_1 \equiv [a_0', a_1', \ldots, a_q', b_1', \ldots, b_s']'$, $a_j \equiv [a_{1j}, ..., a_{pj}]'$, $b_l \equiv [b_{1l}, ..., \ b_{pl}]', j = 0, 1, ..., q, l = 1, ..., s$, and $\delta_2 \equiv \widetilde{\nu}(\Lambda)$, which is obtained from $vec(\Lambda)$ by eliminating the supradiagonal and the diagonal elements of Λ (see Magnus, 1988, p.27). Further let $\mu_l = (\mu_{1l}, \ldots, \mu_{pl})'$, for each $i = 1, \ldots, p$, μ_{il} is implicitly defined such that:

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

The gradient and the (simplified) Hessian, with respect to δ, can be expressed as:

$$\nabla_\delta l_t = \begin{pmatrix} -\frac{1}{2}\nabla_{\delta_1} h_{t-1} w(D_{t-1}^{-2}(I_p - \epsilon_t \epsilon_t' V_{t-1}^{-1})) \\ -\widetilde{\nu}(\Lambda^{-1} - \Lambda^{-1} D_{t-1}^{-1} \epsilon_t \epsilon_t' D_{t-1}^{-1} \Lambda^{-1}) \end{pmatrix}, \tag{A. 1}$$

$$S_t = (S_{ijt})_{2\times 2}, \tag{A. 2}$$

with

$$
\begin{aligned}
S_{11t} &= -\nabla_{\delta_1} h_{t-1} \left(\Psi_p \left(\Lambda^{-1} D_{t-1}^{-2} \otimes D_{t-1}^{-2} \Lambda \right) \Psi_p' + D_{t-1}^{-4} \right) \nabla_{\delta_1}' h_{t-1}/4, \\
S_{12t} &= -\nabla_{\delta_1} h_{t-1} D_{t-1}^{-2} \Psi_p (I_p \otimes \Lambda^{-1}) N_p \widetilde{L}_p', \\
S_{21t} &= S_{12t}', \\
S_{22t} &= -2 \widetilde{L}_p N_p \left(\Lambda^{-1} \otimes \Lambda^{-1} \right) N_p \widetilde{L}_p',
\end{aligned}
$$

where for a $p \times p$ matrix χ, $\Psi_p' w(\chi) = vec(\chi)$, $N_p vec(\chi) = \frac{1}{2}(\chi + \chi')$ and $\widetilde{L}_p \widetilde{\nu}(\chi) = vec(\chi)$ (provided χ is strictly lower triangular). See respectively p.109, pp.48-49 and pp.96-97 of Magnus (1988) for details. In particular, when $p = 2$,

$$
\Psi_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{A. 3}
$$

$$
N_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{A. 4}
$$

$$
\tilde{L}_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix}. \tag{A. 5}
$$

See Magnus (1988) for the case of a general p; in addition, and $\nabla_{\delta_1} h_{t-1}$ is recursively defined as:

$$
\nabla_{\delta_1} h_{t-1} = (\nabla_{a_0}' h_{t-1}; \nabla_{a_1}' h_{t-1}, \ldots, \nabla_{a_q}' h_{t-1}; \nabla_{b_1}' h_{t-1}, \ldots, \nabla_{b_s}' h_{t-1})', \tag{A. 6}
$$

where

$$
\begin{aligned}
\nabla_{a_0} h_{t-1} &= I_p + \sum_{l=1}^{s} (\nabla_{a_0} h_{t-1-l})\, diag(b_l) = \sum_{l=0}^{\infty} diag(\mu_l); \\
\nabla_{a_j} h_{t-1} &= diag\left(\epsilon_{1t-j}^2, \ldots, \epsilon_{pt-j}^2\right) + \sum_{l=1}^{s} \left(\nabla_{a_j} h_{t-1-l} \right) diag(b_l) \\
&= \sum_{l=0}^{\infty} diag\left(\mu_{1l} \epsilon_{1t-l-j}^2, \ldots, \mu_{pl} \epsilon_{pt-l-j}^2 \right), j = 1, \ldots, q; \\
\nabla_{b_j} h_{t-1} &= diag\left(h_{t-1-j}\right) + \sum_{l=1}^{s} \left(\nabla_{b_j} h_{t-1-l} \right) diag(b_l) \\
&= \sum_{l=0}^{\infty} diag\left(\mu_{1l} h_{1,t-1-l-j}, \ldots, \mu_{pl} h_{p,t-1-l-j} \right), j = 1, \ldots, s.
\end{aligned}
$$

We first find an initial estimator $(\hat{\varphi}, \hat{\delta}) \in \Theta_T^{(F)}$, where $\Theta_T^{(F)}$ is defined in the proof of Theorem 3.2 below. Given this initial estimator, we perform a one-step iteration:

$$
\dot{\delta} = \hat{\delta} - \left(\sum_{t=1}^{T} S_t|_{\hat{\varphi}, \hat{\delta}} \right)^{-1} \left(\sum_{t=1}^{T} \nabla_\delta l_t|_{\hat{\varphi}, \hat{\delta}} \right). \tag{A. 7}
$$

The following lemmas give the asymptotic distribution of $\sqrt{T}\left(\dot{\delta} - \delta\right)$. The proofs are similar to that of Lemma 1 in LLW (2001) and thus it is omitted.

Lemma A.1. Suppose Assumptions 2.1-2.5 hold and $\eta_t \sim N(0,\Gamma)$. Then

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

where $\Omega_{\delta}^{*} = E(\nabla_{\delta} l_t \nabla'_{\delta} l_t)$. □

Lemma A.2. Suppose Assumptions 2.1-2.5 hold. Then

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

where $\Omega_{\delta} = -E(S_t)$ and Ω_{δ}^{*} is as defined in Lemma A.1. □

B Appendix: Technical Proofs

We first consider Theorem 3.1. Refer to the one-step iteration on estimating φ (the full rank estimator for the mean parameters) in (3.4). Denote $Q_{\varphi} = diag(Q \otimes I_p, I_{(k-1)p^2})$ and $D_{\varphi} = diag(TI_{dp}, \sqrt{T}I_{rp+(k-1)p^2})$, where $Q' = [\beta_{\perp}, \beta]$. With a bit different definitions of $W_p^{*}(u)$, Ω_1^{*}, Ω_1, Ω_2^{*} and Ω_2 as defined in Section 3, the proof of the following lemma is similar to that of Lemma 1 in LLW (2001). The proof is thus omitted.

Lemma B.1. Suppose Assumptions 2.1-2.5 hold. Then

$$(a) \qquad \sum_{t=1}^{T} D_{\varphi}^{-1} Q_{\varphi} \nabla_{\varphi} l_t \longrightarrow_{\mathcal{L}} \left\{ vec\left[\left(\int_0^1 B_d(u) dW_p^{*}(u)'\right)' \left(\alpha'_{\perp} EV_{t-1}\alpha_{\perp}\right)^{1/2} \left(\alpha'_{\perp}\Gamma\bar{\beta}_{\perp}\right)^{-1'}\right]', \left[N\left(0,\Omega_2^{*}\right)\right]'\right\}',$$

$$(b) \qquad -\sum_{t=1}^{T} D_{\varphi}^{-1} Q_{\varphi} F_t Q'_{\varphi} D_{\varphi}^{-1} \longrightarrow_{\mathcal{L}}$$

$$diag\left\{\left[\left(\alpha'_{\perp}\Gamma\bar{\beta}_{\perp}\right)^{-1}\left(\alpha'_{\perp} EV_{t-1}\alpha_{\perp}\right)^{1/2}\int_0^1 B_d(u)B_d(u)'du \left(\alpha'_{\perp} EV_{t-1}\alpha_{\perp}\right)^{1/2}\left(\alpha'_{\perp}\Gamma\bar{\beta}_{\perp}\right)^{-1'} \otimes \Omega_1\right], \Omega_2\right\},$$

where all the variables are as defined in Theorem 3.1. □

Proof of Theorem 3.1. Directly comes from Theorem 3.2, with $\Omega_1 = \Omega_1^{*}$ and $\Omega_2 = \Omega_2^{*}$. □

Proof of Theorem 3.2. Using Assumptions 2.1-2.5 and the arguments around (4.3) in LLW (2001), we can show that:

$$T^{-1/2} D_{\varphi}^{-1} Q_{\varphi}\left(\sum_{t=1}^{T}\nabla^2_{\varphi\delta'} l_t\right) = o_P(1).$$

Thus, φ and δ can be estimated separately without altering the asymptotic distributions.

For any fixed positive constant K, let $\Theta_T^{(F)} \equiv \left\{\left(\check{\varphi},\check{\delta}\right) : \|D_{\varphi} Q_{\varphi}^{'-1}\left(\check{\varphi}-\varphi\right)\| \leq K \ \ and \ \ \|\sqrt{T}\left(\check{\delta}-\delta\right)\| \leq K\right\}$, where $(\check{\varphi},\check{\delta})$ is a generic version of (φ,δ). Using Assumptions 2.1-2.5 and a method similar to that in Ling and Li (1998), it is easy to see that provided that the initial estimator, $(\hat{\varphi}, \hat{\delta}) \in \Theta_T^{(F)}$, the one-step iteration (3.4) renders the full rank estimator $\acute{\varphi}$ an asymptotic expansion such that:

$$D_\varphi Q_\varphi'^{-1}(\acute{\varphi} - \varphi) = -\left(\sum_{t=1}^T D_\varphi^{-1} Q_\varphi F_t Q_\varphi' D_\varphi^{-1}\right)^{-1} \left(\sum_{t=1}^T D_\varphi^{-1} Q_\varphi \nabla_\varphi l_t\right) + o_P(1), \qquad \text{(B. 1)}$$

Theorem 3.2 then follows from (B.1) and Lemma B.1. □

Proof of Theorem 4.1. From the proof of Lemma 13.2 in Johansen (1996), in our notation,

$$\begin{aligned} & T\left(\widehat{\beta}'\bar{\beta}\right)^{-1} \widehat{\beta}'\bar{\beta}_\perp \\ = & \left(\alpha'(EV_{t-1})^{-1}\alpha\right)^{-1} \alpha'(EV_{t-1})^{-1} \left(T^{-1}\sum_{t=1}^T \epsilon_t W_{t-1}'\right)\left(T^{-2}\sum_{t=1}^T W_{t-1}W_{t-1}'\right)^{-1} + o_P(1), \end{aligned}$$

where we recall that $W_{t-1} = \beta_\perp' X_{t-1}$. By arguments similar to those for proving Lemma B.1(a),

$$T^{-1}\sum_{t=1}^T \epsilon_t W_{t-1}' \longrightarrow_{\mathcal{L}} \left[\int_0^1 B_d(u) dW_p(u)'\right]' (\alpha_\perp' EV_{t-1}\alpha_\perp)^{1/2} (\alpha_\perp' \Gamma \bar{\beta}_\perp)^{-1'}$$

On the other hand, by arguments similar to those for proving Lemma B.1(b),

$$T^{-2}\sum_{t=1}^T W_{t-1}W_{t-1}' \longrightarrow_{\mathcal{L}} (\alpha_\perp'\Gamma\bar{\beta}_\perp)^{-1} (\alpha_\perp' EV_{t-1}\alpha_\perp)^{1/2} \left[\int_0^1 B_d(u)B_d(u)'du\right] (\alpha_\perp' EV_{t-1}\alpha_\perp)^{1/2} (\alpha_\perp'\Gamma\bar{\beta}_\perp)^{-1'}.$$

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

Proof of Theorem 4.2. Directly comes from Theorem 4.3, with $\Omega_1 = \Omega_1^*$ and $\Omega_2 = \Omega_2^*$. □

The following lemma is useful for proving Theorem 4.3.
Lemma B.2. Under the assumptions in Theorem 4.3, it follows that

(a) $\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\left(\widetilde{\beta}' - \widehat{\beta}'\right) = O_P\left(T^{-1/2}\right)$,

(b) $\widehat{\alpha}\left(\widetilde{\beta}'\bar{\beta}\right) = \widehat{\alpha}\left(\widehat{\beta}'\bar{\beta}\right) + O_P\left(T^{-1/2}\right) = \alpha + O_P\left(T^{-1/2}\right)$,

(c) $\left(\widetilde{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}'\bar{\beta}_\perp = \left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}'\bar{\beta}_\perp + O_P\left(T^{-3/2}\right) = \beta'\bar{\beta}_\perp + O_P\left(T^{-1}\right) = O_P\left(T^{-1}\right)$,

(d) $\left(\widetilde{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}'\bar{\beta} = \left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}'\bar{\beta} + O_P\left(T^{-1/2}\right) = \beta'\bar{\beta} + O_P\left(T^{-1/2}\right) = I_r + O_P\left(T^{-1/2}\right)$. □

Proof. (a). We first denote $D_{\phi_1} = diag\left(TI_{rd}, \sqrt{T}I_{r^2}\right)$ and $\widehat{Q}_{\phi_1} = \mathcal{Q}\left(I_p \otimes \left(\widehat{\beta}'\bar{\beta}\right)'\right)$, with $\mathcal{Q} = (Q \otimes I_r)$, where we recall that $Q' = [\beta_\perp, \beta]$. Also denote $\widehat{\phi}_1 = vec\left(\widehat{\beta}'\right)$, $\acute{\phi}_1 = vec\left(\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}'\right)$ and $\widetilde{\phi}_1 = vec\left(\widetilde{\beta}'\right)$. $\widehat{\phi}_2$, $\acute{\phi}_2$ and $\widetilde{\phi}_2$ are defined accordingly. $\widehat{\phi}$, $\acute{\phi}$ and $\widetilde{\phi}$ are also defined accordingly. Denote $P \equiv Q^{-1} = (\bar{\beta}_\perp, \bar{\beta})$. Since $\widehat{Q}_{\phi_1}'^{-1} = (P' \otimes I_r)\left(I_p \otimes \left(\widehat{\beta}'\bar{\beta}\right)^{-1}\right)$, we have:

$$\begin{aligned} \left(I_p \otimes \left(\widehat{\beta}'\bar{\beta}\right)^{-1}\right)\left(\widetilde{\phi}_1 - \widehat{\phi}_1\right) &= \mathcal{Q}' D_{\phi_1}^{-1} D_{\phi_1} (P' \otimes I_r)\left(I_p \otimes \left(\widehat{\beta}'\bar{\beta}\right)^{-1}\right)\left(\widetilde{\phi}_1 - \widehat{\phi}_1\right) \\ &= \mathcal{Q}' D_{\phi_1}^{-1}\left[D_{\phi_1}\widehat{Q}_{\phi_1}'^{-1}\left(\widetilde{\phi}_1 - \widehat{\phi}_1\right)\right]. \end{aligned}$$

As $\mathcal{Q}' D_{\phi_1}^{-1} = O(T^{-1/2})$, it suffices to show $D_{\phi_1}\widehat{Q}_{\phi_1}'^{-1}\left(\widetilde{\phi}_1 - \widehat{\phi}_1\right) = O_P(1)$. By (4.9),

$$
\begin{aligned}
& D_{\phi_1}\widehat{Q}_{\phi_1}'^{-1}\left(\widetilde{\phi}_1 - \widehat{\phi}_1\right) \\
&= -\left[\sum_{t=1}^{T} D_{\phi_1}^{-1}\widehat{Q}_{\phi_1}\left(R_{1t}|_{\widehat{\phi},\widehat{\delta}}\right)\widehat{Q}_{\phi_1}' D_{\phi_1}^{-1}\right]^{-1}\left[\sum_{t=1}^{T} D_{\phi_1}^{-1}\widehat{Q}_{\phi_1}\left(\nabla_{\phi_1} l_t|_{\widehat{\phi},\widehat{\delta}}\right)\right] \\
&= -\left[\sum_{t=1}^{T} D_{\phi_1}^{-1}\mathcal{Q}\left(R_{1t}|_{\acute{\alpha},\widehat{\delta}}\right)\mathcal{Q}' D_{\phi_1}^{-1}\right]^{-1}\left[\sum_{t=1}^{T} D_{\phi_1}^{-1}\mathcal{Q}\left(\nabla_{\phi_1} l_t|_{\acute{\phi},\widehat{\delta}}\right)\right]. \qquad \text{(B. 2)}
\end{aligned}
$$

By Theorem 4.1, $T\left(\acute{\phi}_1 - \phi_1\right) = O_P(1)$, $\sqrt{T}\left(\acute{\phi}_2 - \phi_2\right) = O_P(1)$, which are in addition to $\sqrt{T}\left(\widehat{\delta} - \delta\right) = O_P(1)$. It is not difficult to see that:

$$
\sum_{t=1}^{T} D_{\phi_1}^{-1}\mathcal{Q}\left(R_{1t}|_{\acute{\phi},\widehat{\delta}}\right)\mathcal{Q}' D_{\phi_1}^{-1} = \sum_{t=1}^{T} D_{\phi_1}^{-1}\mathcal{Q}R_{1t}\mathcal{Q}' D_{\phi_1}^{-1} + o_P(1). \qquad \text{(B. 3)}
$$

On the other hand, by a Taylor's expansion and (B.3), with R_{1t}^* and l_t^* being evaluated at a mid-point of $(\acute{\phi}, \widehat{\delta})$ and (ϕ, δ),

$$
\begin{aligned}
& \sum_{t=1}^{T} D_{\phi_1}^{-1}\mathcal{Q}\left(\nabla_{\phi_1} l_t|_{\acute{\phi},\widehat{\delta}}\right) \\
&= \sum_{t=1}^{T} D_{\phi_1}^{-1}\mathcal{Q}\nabla_{\phi_1} l_t + \sum_{t=1}^{T} D_{\phi_1}^{-1}\mathcal{Q}R_{1t}^*\left(\acute{\phi}_1 - \phi_1\right) + \sum_{t=1}^{T} D_{\phi_1}^{-1}\mathcal{Q}\left(\nabla_{\phi_1\phi_1'} l_t^*\right)\left(\acute{\phi}_2 - \phi_2\right) \\
&= \sum_{t=1}^{T} D_{\phi_1}^{-1}\mathcal{Q}\nabla_{\phi_1} l_t + \left[\sum_{t=1}^{T} D_{\phi_1}^{-1}\mathcal{Q}R_{1t}\mathcal{Q}' D_{\phi_1}^{-1} + o_P(1)\right] T^{-1} D_{\phi_1}\left(P' \otimes I_r\right)\left[T\left(\acute{\phi}_1 - \phi_1\right)\right] \\
&\quad + \left[T^{-1/2}\sum_{t=1}^{T} D_{\phi_1}^{-1}\mathcal{Q}\left(\nabla_{\phi_1\phi_2'} l_t^*\right)\right]\sqrt{T}\left(\acute{\phi}_2 - \phi_2\right). \qquad \text{(B. 4)}
\end{aligned}
$$

It is not difficult to see that $T^{-1/2}\sum_{t=1}^{T} D_{\phi_1}^{-1}\mathcal{Q}\left(\nabla_{\phi_1\phi_2'} l_t^*\right)$ is $O_P(1)$. So is the RHS of (B.4). By Lemmas A.1(a)-(b), (B.2), (B.3) and (B.4), $D_{\phi_1}\widehat{Q}_{\phi_1}'^{-1}\left(\widetilde{\phi}_1 - \widehat{\phi}_1\right) = O_P(1)$. Thus (a) holds.

(b). By the $\sqrt{T}$-consistency of $\widehat{\alpha}\left(\widehat{\beta}'\bar{\beta}\right)$ for α (Theorem 4.1), and (a) of this lemma,

$$
\widehat{\alpha}\left(\widetilde{\beta}'\bar{\beta}\right) = \widehat{\alpha}\left(\widehat{\beta}'\bar{\beta}\right) + \widehat{\alpha}\left(\widehat{\beta}'\bar{\beta}\right)\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\left(\widetilde{\beta}' - \widehat{\beta}'\right)\bar{\beta} = \widehat{\alpha}\left(\widehat{\beta}'\bar{\beta}\right) + O_P(1)O_P(T^{-1/2}).
$$

Thus, (b) holds.

(c) and (d). Denote $\acute{\beta}' = \left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}'$.

$$
\left(\widetilde{\beta}'\bar{\beta}\right)^{-1}\widetilde{\beta}' = \left[\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widetilde{\beta}'\bar{\beta}\right]^{-1}\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widetilde{\beta}' = \left[\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widetilde{\beta}'\bar{\beta}\right]^{-1}\acute{\beta}'. \qquad \text{(B. 5)}
$$

Using the formula $dF^{-1} = -F^{-1}(dF)F^{-1}$ for the $r \times r$ matrix F with $F(x) = \left[x\bar{\beta}'\right]^{-1}$, and applying a Taylor's expansion to $\left[\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widetilde{\beta}'\bar{\beta}\right]^{-1}$ around $\acute{\beta}'\bar{\beta}$, we have

$$
\left[\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widetilde{\beta}'\bar{\beta}\right]^{-1} = \left[\acute{\beta}'\bar{\beta}\right]^{-1} - \left[\beta^{*\prime}\bar{\beta}\right]^{-1}\left[\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widetilde{\beta}' - \acute{\beta}'\right]\bar{\beta}\left[\beta^{*\prime}\bar{\beta}\right]^{-1},
$$

where β^* lies between $\widetilde{\beta}\left(\bar{\beta}'\widehat{\beta}\right)^{-1}$ and $\acute{\beta}$. Therefore, the RHS of (B.5) equals:

$$\left[\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}'\bar{\beta}\right]^{-1}\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}' - \left[\beta^{*\prime}\bar{\beta}\right]^{-1}\left[\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widetilde{\beta}' - \acute{\beta}'\right]\bar{\beta}\left[\beta^{*\prime}\bar{\beta}\right]^{-1}\acute{\beta}'$$
$$= \left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}' - \left[\beta^{*\prime}\bar{\beta}\right]^{-1}\left[\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widetilde{\beta}' - \acute{\beta}'\right]\bar{\beta}\left[\beta^{*\prime}\bar{\beta}\right]^{-1}\acute{\beta}'. \qquad \text{(B. 6)}$$

By (a) of this lemma, $\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widetilde{\beta}' - \acute{\beta}' = O_P(T^{-1/2})$. By this, we can show that $\left[\beta^{*\prime}\bar{\beta}\right]^{-1} = O_P(1)$. $\bar{\beta}$ and $\acute{\beta}$ are also $O_P(1)$. By (B.6), (d) holds.

By Theorem 4.1, $\acute{\beta}'\bar{\beta}_\perp = O_P(T^{-1})$. By (B.6),

$$\left[\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}'\bar{\beta}\right]^{-1}\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}'\bar{\beta}_\perp - \left[\beta^{*\prime}\bar{\beta}\right]^{-1}\left[\left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widetilde{\beta}' - \acute{\beta}'\right]\bar{\beta}\left[\beta^{*\prime}\bar{\beta}\right]^{-1}\acute{\beta}'\bar{\beta}_\perp$$
$$= \left(\widehat{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}'\bar{\beta}_\perp + O_P(T^{-3/2}).$$

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

Proof of Theorem 4.3. Denote $D_\phi = diag(TI_{rd}, \sqrt{T}I_{rp+(k-1)p^2})$ and $Q_\phi = diag\left(\left(\beta_\perp' \otimes I_r\right), I_{rp+(k-1)p^2}\right)$. Using Assumptions 2.1-2.5 and the arguments around (5.3) in LLW (2001), we can show that:

$$T^{-1/2}D_\phi^{-1}Q_\phi\left(\sum_{t=1}^T \nabla^2_{\phi\delta'}l_t\right) = o_P(1).$$

Thus, ϕ and δ can be estimated separately without altering the asymptotic distributions. Again, using the arguments similar to those around (5.5) in LLW (2001), we can show that the Hessian can be given as follows.

$$D_\phi^{-1}Q_\phi\sum_{t=1}^T \nabla^2_{\phi\phi'}l_t Q_\phi' D_\phi^{-1} = D_\phi^{-1}Q_\phi\sum_{t=1}^T R_t Q_\phi' D_\phi^{-1} + o_P(1), \qquad \text{(B. 7)}$$

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

For any fixed positive constant K, let $\Theta_T^{(R)} \equiv \left\{\left(\check{\phi}, \check{\delta}\right) : \|D_\phi Q_\phi'^{-1}\left(\check{\phi} - \phi\right)\| \leq K \ \ and \ \ \|\sqrt{T}\left(\check{\delta} - \delta\right)\| \leq K\right\}$, where $(\check{\phi}, \check{\delta})$ is a generic version of (ϕ, δ). Using Assumptions 2.1-2.5 and a method similar to that in Ling and Li (1998), it is easy to see that on $\Theta_T^{(R)}$,

$$D_\phi^{-1}Q_\phi\sum_{t=1}^T\left(R_t|_{\check{\phi},\check{\delta}} - R_t\right)Q_\phi' D_\phi^{-1} = o_P(1), \qquad \text{(B. 8)}$$
$$D_\phi^{-1}Q_\phi\sum_{t=1}^T\left(\nabla_\phi l_t|_{\check{\phi},\check{\delta}} - \nabla_\phi l_t\right) = D_\phi^{-1}Q_\phi\sum_{t=1}^T R_t\left(\check{\phi} - \phi\right) + o_P(1), \qquad \text{(B. 9)}$$

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

Denote $\widetilde{Q}_{\phi_1} = (\beta_\perp' \otimes I_r)\left(I_p \otimes \left(\widetilde{\beta}'\bar{\beta}\right)'\right)$, $\widetilde{Q}_{\phi_2} = diag\left(\left(\widetilde{\beta}'\bar{\beta}\right)^{-1} \otimes I_p, I_{(k-1)p^2}\right)$, $\grave{\phi}_1 = vec\left(\left(\widetilde{\beta}'\bar{\beta}\right)^{-1}\widehat{\beta}'\right)$, $\grave{\phi}_2 = vec\left[\widehat{\alpha}\left(\widetilde{\beta}'\bar{\beta}\right), \widehat{\Pi}_1, \ldots, \widehat{\Pi}_{k-1}\right]$, and $\grave{\phi} = \left[\grave{\phi}_1', \grave{\phi}_2'\right]'$. By Lemma A.2 and Lemmas B.2(b)-(d),

$(\grave{\phi}, \widehat{\delta}) \in \Theta_T^{(R)}$. Thus by (B.8) and the block-diagonality of R_t,

$$\begin{aligned} T^{-2}\sum_{t=1}^{T}\widetilde{Q}_{\phi_1}\left(R_{1t}|_{\widehat{\phi},\widehat{\delta}}\right)\widetilde{Q}'_{\phi_1} &= T^{-2}\sum_{t=1}^{T}\left(\beta'_\perp\otimes I_r\right)\left(R_{1t}|_{\grave{\phi},\widehat{\delta}}\right)\left(\beta_\perp\otimes I_r\right) \\ &= T^{-2}\sum_{t=1}^{T}\left(\beta'_\perp\otimes I_r\right)R_{1t}\left(\beta_\perp\otimes I_r\right)+o_P(1), \qquad \text{(B. 10)} \\ T^{-1}\sum_{t=1}^{T}\widetilde{Q}_{\phi_2}\left(R_{2t}|_{\widehat{\phi},\widehat{\delta}}\right)\widetilde{Q}'_{\phi_2} &= T^{-1}\sum_{t=1}^{T}\left(R_{2t}|_{\grave{\phi},\widehat{\delta}}\right) = T^{-1}\sum_{t=1}^{T}R_{2t}+o_P(1). \qquad \text{(B. 11)} \end{aligned}$$

Refer to (B.7). By (B.8), (B.9) and the block-diagonality of R_t, it is easy to see that:

$$\begin{aligned} & T^{-1}\sum_{t=1}^{T}\widetilde{Q}_{\phi_1}\left(\nabla_{\phi_1}l_t|_{\widehat{\phi},\widehat{\delta}}\right) \\ = \; & T^{-1}\sum_{t=1}^{T}\left(\beta'_\perp\otimes I_r\right)\left(\nabla_{\phi_1}l_t|_{\grave{\phi},\widehat{\delta}}\right) \\ = \; & T^{-1}\sum_{t=1}^{T}\left(\beta'_\perp\otimes I_r\right)\nabla_{\phi_1}l_t+\left[T^{-1}\sum_{t=1}^{T}\left(\beta'_\perp\otimes I_r\right)R_{1t}\left(\beta_\perp\otimes I_r\right)\right]\left(\bar{\beta}'_\perp\otimes I_r\right)\left(\grave{\phi}_1-\phi_1\right)+o_P(1), \text{(B. 12)} \\ & T^{-1/2}\sum_{t=1}^{T}\widetilde{Q}_{\phi_2}\left(\nabla_{\phi_2}l_t|_{\widehat{\phi},\widehat{\delta}}\right) \\ = \; & T^{-1/2}\sum_{t=1}^{T}\left(\nabla_{\phi_2}l_t|_{\grave{\phi},\widehat{\delta}}\right) \\ = \; & T^{-1/2}\sum_{t=1}^{T}\nabla_{\phi_2}l_t+\left(T^{-1}\sum_{t=1}^{T}R_{2t}\right)\left(\grave{\phi}_2-\phi_2\right)+o_P(1). \qquad \text{(B. 13)} \end{aligned}$$

Recall that $\widetilde{Q}'^{-1}_{\phi_1}\widehat{\phi}_1 = \left(\bar{\beta}'_\perp\otimes I_r\right)\grave{\phi}_1$. By (4.5), (B.10) and (B.12), we have the following asymptotic expansion:

$$\begin{aligned} T\widetilde{Q}'^{-1}_{\phi_1}\widetilde{\phi}_1 &= T\widetilde{Q}'^{-1}_{\phi_1}\widehat{\phi}_1-\left[T^{-2}\sum_{t=1}^{T}\widetilde{Q}_{\phi_1}\left(R_{1t}|_{\widehat{\phi},\widehat{\delta}}\right)\widetilde{Q}'_{\phi_1}\right]^{-1}\left[T^{-1}\sum_{t=1}^{T}\widetilde{Q}_{\phi_1}\left(\nabla_{\phi_1}l_t|_{\widehat{\phi},\widehat{\delta}}\right)\right] \\ &= T\left(\bar{\beta}'_\perp\otimes I_r\right)\grave{\phi}_1-\left[T^{-2}\sum_{t=1}^{T}\left(\beta'_\perp\otimes I_r\right)R_{1t}\left(\beta_\perp\otimes I_r\right)\right]^{-1}\left[T^{-1}\sum_{t=1}^{T}\left(\beta'_\perp\otimes I_r\right)\nabla_{\phi_1}l_t\right] \\ &\quad -T\left(\bar{\beta}'_\perp\otimes I_r\right)\left(\grave{\phi}_1-\phi_1\right)+o_P(1) \\ &= T\left(\bar{\beta}'_\perp\otimes I_r\right)\phi_1-\left[T^{-2}\sum_{t=1}^{T}\left(\beta'_\perp\otimes I_r\right)R_{1t}\left(\beta_\perp\otimes I_r\right)\right]^{-1}\left[T^{-1}\sum_{t=1}^{T}\left(\beta'_\perp\otimes I_r\right)\nabla_{\phi_1}l_t\right] \\ &\quad +o_P(1). \qquad \text{(B. 14)} \end{aligned}$$

Note that $\widetilde{Q}'^{-1}_{\phi_1}\widetilde{\phi}_1 = vec\left[\left(\left(\widetilde{\beta}'\bar{\beta}\right)^{-1}\widetilde{\beta}'-\beta'\right)\bar{\beta}_\perp\right]$. By (B.14) and Lemma B.1(a)-(b), (a) is proved.

On the other hand, by (4.6), (B.11) and (B.13), we have the following asymptotic expansion:

$$\sqrt{T}\widetilde{Q}'^{-1}_{\phi_2}\widetilde{\phi}_2 = \sqrt{T}\widetilde{Q}'^{-1}_{\phi_2}\widehat{\phi}_2-\left[T^{-1}\sum_{t=1}^{T}\widetilde{Q}_{\phi_2}\left(R_{2t}|_{\widehat{\phi},\widehat{\delta}}\right)\widetilde{Q}'_{\phi_2}\right]^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\widetilde{Q}_{\phi_2}\left(\nabla_{\phi_2}l_t|_{\widehat{\phi},\widehat{\delta}}\right)\right]$$

$$= \sqrt{T}\grave{\phi}_2 - \left[T^{-1}\sum_{t=1}^{T} R_{2t}\right]^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\nabla_{\phi_2} l_t\right] - \sqrt{T}\left(\grave{\phi}_2 - \phi_2\right) + o_P(1)$$

$$= \sqrt{T}\phi_2 - \left[T^{-1}\sum_{t=1}^{T} R_{2t}\right]^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\nabla_{\phi_2} l_t\right] + o_P(1). \quad \text{(B. 15)}$$

By (B.15) and Lemma B.1(a)-(b), (b) is also proved. This completes the proof. □

Proof of Theorem 5.1. Refer to (B.25) in the Proof of Lemma 5.1. When $\Omega_1 = \Omega_1^*$,

$$V_d(u) = (\Upsilon\Upsilon')^{1/2} B_d(u) + (I_d - \Upsilon\Upsilon')^{1/2} B_d^*(u).$$

Moreover, $\Upsilon = \Upsilon^H$ and thus $V_d(u) = V_d^H(u)$. Thus the asymptotic distribution is identical to that in (B.29) and the result follows from the proof of Theorem 5.2. □

Proof of Lemma 5.1. From (5.3), it is not difficult to see that

$$LR_G = (\dot{\varphi} - \widetilde{\varphi})'\left(\sum_{t=1}^{T} Y_{t-1}Y_{t-1}' \otimes \Omega_1\right)(\dot{\varphi} - \widetilde{\varphi}) + o_P(1), \quad \text{(B. 16)}$$

where we recall that $\dot{\varphi} = vec[\dot{\Pi}, \dot{\Gamma}_1, \ldots, \dot{\Gamma}_{k-1}]$ and $\widetilde{\varphi} \equiv vec[\widetilde{\alpha}\widetilde{\beta}', \widetilde{\Gamma}_1, \ldots, \widetilde{\Gamma}_{k-1}]$.

Denote $\ddot{\alpha} = \widetilde{\alpha}\left(\widetilde{\beta}'\bar{\beta}\right)$ and $\ddot{\beta}' = \left(\widetilde{\beta}'\bar{\beta}\right)^{-1}\widetilde{\beta}'$. Note $\widetilde{\alpha}\widetilde{\beta}' = \ddot{\alpha}\ddot{\beta}'$. Moreover,

$$\ddot{\alpha}\ddot{\beta}' - \alpha\beta' = (\ddot{\alpha} - \alpha)\beta' + \alpha\left(\ddot{\beta}' - \beta'\right) + (\ddot{\alpha} - \alpha)\left(\ddot{\beta}' - \beta'\right).$$

Recall that $\beta'\bar{\beta}_\perp = 0_{r\times d}$. By Theorem 4.3, $\left(\ddot{\beta}' - \beta'\right)\bar{\beta}_\perp = O_P(T^{-1})$ and $(\ddot{\alpha} - \alpha) = O_P(T^{-1/2})$. Hence,

$$T\left(\ddot{\alpha}\ddot{\beta}' - \alpha\beta'\right)\bar{\beta}_\perp = T(\ddot{\alpha} - \alpha)\beta'\bar{\beta}_\perp + T\alpha\left(\ddot{\beta}' - \beta'\right)\bar{\beta}_\perp + (\ddot{\alpha} - \alpha)T\left(\ddot{\beta}' - \beta'\right)\bar{\beta}_\perp$$

$$= T\alpha\left(\ddot{\beta}' - \beta'\right)\bar{\beta}_\perp + O_P(T^{-1/2}). \quad \text{(B. 17)}$$

On the other hand, by Theorem 4.3(a) and the arguments in Lemma 13.2 of Johansen (1996), $\left(\ddot{\beta} - \beta\right) = O_P(T^{-1})$. Therefore,

$$\sqrt{T}\left(\ddot{\alpha}\ddot{\beta} - \alpha\beta'\right)\bar{\beta} = \sqrt{T}(\ddot{\alpha} - \alpha)\beta'\bar{\beta} + \sqrt{T}\ddot{\alpha}\left(\ddot{\beta}' - \beta'\right)\bar{\beta}$$

$$= \sqrt{T}(\ddot{\alpha} - \alpha) + O_P(T^{-1/2}). \quad \text{(B. 18)}$$

But from the proofs of Theorem 4.3(b) and Theorem 3.2(b),

$$\sqrt{T}(\ddot{\alpha} - \alpha) - \sqrt{T}\left(\dot{\Pi} - \Pi\right)\bar{\beta} = o_P(1), \quad \text{(B. 19)}$$

$$\sqrt{T}\left(\widetilde{\Pi}_j - \Pi_j\right) - \sqrt{T}\left(\dot{\Pi}_j - \Pi_j\right) = o_P(1), j = 1, \ldots, k-1. \quad \text{(B. 20)}$$

All in all, by (B.17)-(B.20) and Lemma B.1(b), (B.16) can be re-written as:

$$LR_G = \left[vec\left[T\left(\dot{\Pi} - \alpha\ddot{\beta}'\right)\bar{\beta}_\perp\right]\right]'\left[T^{-2}\sum_{t=1}^{T} W_{t-1}W_{t-1}' \otimes \Omega_1\right] vec\left[T\left(\dot{\Pi} - \alpha\ddot{\beta}'\right)\bar{\beta}_\perp\right] + o_P(1), \text{(B. 21)}$$

where we recall that $W_{t-1} = \beta_\perp' X_{t-1}$. By Lemma B.1(b), Theorem 3.2(a) and Theorem 4.3(a),

$$\begin{aligned} T^{-2}\sum_{t=1}^{T} W_{t-1}W_{t-1}' \otimes \Omega_1 &\longrightarrow_{\mathcal{L}} W \otimes \Omega_1; \\ T\dot{\Pi}\bar{\beta}_\perp &\longrightarrow_{\mathcal{L}} \Omega_1^{-1}M^*; \\ T\alpha\ddot{\beta}'\bar{\beta}_\perp &\longrightarrow_{\mathcal{L}} \alpha\left(\alpha'\Omega_1\alpha\right)^{-1}\alpha' M^*, \end{aligned}$$

where W is defined as $\left(\alpha_\perp'\Gamma\bar{\beta}_\perp\right)^{-1}\left(\alpha_\perp'\left(EV_{t-1}\right)\alpha_\perp\right)\left(\int_0^1 B_d(u)B_d(u)'du\right)\left(\alpha_\perp'\left(EV_{t-1}\right)\alpha_\perp\right)\left(\alpha_\perp'\Gamma\bar{\beta}_\perp\right)^{-1'}$ and we recall from (3.7) that $M^* =$ $\left(\int_0^1 B_d(u)dW_p^*(u)'\right)'\left(\int_0^1 B_d(u)B_d(u)'du\right)^{-1}\left(\alpha_\perp'\left(EV_{t-1}\right)\alpha_\perp\right)\left(\alpha_\perp'\Gamma\bar{\beta}_\perp\right)^{-1}$. Following the lines on p.359 of Reinsel and Ahn (1992), we can re-write $\Omega_1^{-1} - \alpha(\alpha'\Omega_1\alpha)^{-1}\alpha'$ as:

$$\Omega_1^{-1}(\Omega_1 - \Omega_1\alpha(\alpha'\Omega_1\alpha)^{-1}\alpha'\Omega_1)\Omega_1^{-1} = \Omega_1^{-1}\alpha_\perp(\alpha_\perp'\Omega_1^{-1}\alpha_\perp)^{-1}\alpha_\perp'\Omega_1^{-1}.$$

Therefore,

$$\begin{aligned} LR_G &\longrightarrow_{\mathcal{L}} \left[vec\left[\left(\Omega_1^{-1} - \alpha(\alpha'\Omega_1\alpha)^{-1}\alpha'\right)M^*\right]\right]'\left[W\otimes\Omega_1\right]vec\left[\left(\Omega_1^{-1} - \alpha(\alpha'\Omega_1\alpha)^{-1}\alpha'\right)M^*\right] \quad \text{(B. 22)} \\ &= tr\left\{\Omega_1\left(\Omega_1^{-1}\alpha_\perp(\alpha_\perp'\Omega_1^{-1}\alpha_\perp)^{-1}\alpha_\perp'\Omega_1^{-1}\right)M^*WM^{*'}\left(\Omega_1^{-1}\alpha_\perp(\alpha_\perp'\Omega_1^{-1}\alpha_\perp)^{-1}\alpha_\perp'\Omega_1^{-1}\right)\right\} \\ &= tr\left\{\left(\alpha_\perp'\Omega_1^{-1}\alpha_\perp\right)^{-1/2}\alpha_\perp'\Omega_1^{-1}M^*WM^{*'}\Omega_1^{-1}\alpha_\perp\left(\alpha_\perp'\Omega_1^{-1}\alpha_\perp\right)^{-1/2}\right\}, \end{aligned}$$

since

$$\Omega_1^{-1}\alpha_\perp(\alpha_\perp'\Omega_1^{-1}\alpha_\perp)^{-1}\alpha_\perp'\Omega_1^{-1}\Omega_1\left(\Omega_1^{-1}\alpha_\perp(\alpha_\perp'\Omega_1^{-1}\alpha_\perp)^{-1}\alpha_\perp'\Omega_1^{-1}\right) = \Omega_1^{-1}\alpha_\perp(\alpha_\perp'\Omega_1^{-1}\alpha_\perp)^{-1}\alpha_\perp'\Omega_1^{-1}. \quad \text{(B. 23)}$$

Therefore, the asymptotic distribution can be re-written as:

$$tr\left\{\left(\int_0^1 B_d(u)dV_d(u)'\right)'\left(\int_0^1 B_d(u)B_d(u)'du\right)^{-1}\left(\int_0^1 B_d(u)dV_d(u)'\right)\right\}, \quad \text{(B. 24)}$$

where $V_d(u)$ is defined as $\left(\alpha_\perp'\Omega_1^{-1}\alpha_\perp\right)^{-1/2}\alpha_\perp'\Omega_1^{-1}W_p^*(u)$. By (3.6), $E[B_d(u)V_d(u)'] =$ $(\alpha_\perp'(EV_{t-1})\alpha_\perp)^{-1/2}\alpha_\perp' E[W_p(u)W_p^*(u)']\Omega_1^{-1}\alpha_\perp(\alpha_\perp'\Omega_1^{-1}\alpha_\perp)^{-1/2} = u(\alpha_\perp'(EV_{t-1})\alpha_\perp)^{-1/2}(\alpha_\perp'\Omega_1^{-1}\alpha_\perp)^{1/2}$ $= u\Upsilon'$. All in all, we can re-write $V_d(u)$ as a linear combination of two independent $d-$dimensional standard BMs:

$$\Upsilon B_d(u) + [(\alpha_\perp'\Omega_1^{-1}\alpha_\perp)^{-1/2}\alpha_\perp'\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp(\alpha_\perp'\Omega_1^{-1}\alpha_\perp)^{-1/2} - \Upsilon\Upsilon']^{1/2}B_d^*(u).$$

But $\Upsilon B_d(u) \sim N(0, \Upsilon\Upsilon')$. Abusing the notation, we write $\Upsilon B_d(u)$ as $(\Upsilon\Upsilon')^{1/2}B_d(u)$, where $B_d(u)$ is (another) $d-$dimensional standard BM independent of $B_d^*(u)$. Canceling some of the $\Upsilon^{-1}(\Upsilon\Upsilon')^{1/2}$ terms, the asymptotic distribution is the same as that in (B.24), with V_d defined as:

$$(\Upsilon\Upsilon')^{1/2}B_d(u) + [(\alpha_\perp'\Omega_1^{-1}\alpha_\perp)^{-1/2}\alpha_\perp'\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp(\alpha_\perp'\Omega_1^{-1}\alpha_\perp)^{-1/2} - \Upsilon\Upsilon']^{1/2}B_d^*(u). \quad \text{(B. 25)}$$

The proof is complete. □

Proof of Theorem 5.2. Similar to (B.21) in the proof of Lemma 5.1, we have:

$$H_G = \left[vec\left[T\left(\dot{\Pi} - \alpha\ddot{\beta}'\right)\bar{\beta}_\perp\right]\right]'\left[T^{-2}\sum_{t=1}^{T}W_{t-1}W_{t-1}'\otimes\Omega_1^H\right]vec\left[T\left(\dot{\Pi} - \alpha\ddot{\beta}'\right)\bar{\beta}_\perp\right] + o_P(1), \text{(B. 26)}$$

where we recall that $W_{t-1} = \beta'_\perp X_{t-1}$, $\Omega_1^H = \alpha_\perp(\alpha'_\perp\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp)^{-1}\alpha'_\perp$. Therefore, similar to (B.22),

$$\begin{aligned} H_G \quad &\longrightarrow_{\mathcal{L}} \quad \left[vec\left[\left(\Omega_1^{-1} - \alpha(\alpha'\Omega_1\alpha)^{-1}\alpha'\right)M^*\right]\right]'\left[W\otimes\Omega_1^H\right]vec\left[\left(\Omega_1^{-1} - \alpha(\alpha'\Omega_1\alpha)^{-1}\alpha'\right)M^*\right] \quad \text{(B. 27)} \\ &= \quad tr\left\{\Omega_1^H\left(\Omega_1^{-1}\alpha_\perp(\alpha'_\perp\Omega_1^{-1}\alpha_\perp)^{-1}\alpha'_\perp\Omega_1^{-1}\right)M^*WM^{*'}\left(\Omega_1^{-1}\alpha_\perp(\alpha'_\perp\Omega_1^{-1}\alpha_\perp)^{-1}\alpha'_\perp\Omega_1^{-1}\right)\right\} \\ &= \quad tr\left\{\left(\alpha'_\perp\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp\right)^{-1/2}\alpha'_\perp\Omega_1^{-1}M^*WM^{*'}\Omega_1^{-1}\alpha_\perp\left(\alpha'_\perp\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp\right)^{-1/2}\right\}, \end{aligned}$$

since

$$\begin{aligned} &\Omega_1^{-1}\alpha_\perp(\alpha'_\perp\Omega_1^{-1}\alpha_\perp)^{-1}\alpha'_\perp\Omega_1^{-1}\Omega_1^H\left(\Omega_1^{-1}\alpha_\perp(\alpha'_\perp\Omega_1^{-1}\alpha_\perp)^{-1}\alpha'_\perp\Omega_1^{-1}\right) \\ = \; &\Omega_1^{-1}\alpha_\perp(\alpha'_\perp\Omega_1^{-1}\alpha_\perp)^{-1}\alpha'_\perp\Omega_1^{-1}\alpha_\perp(\alpha'_\perp\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp)^{-1}\alpha'_\perp\left(\Omega_1^{-1}\alpha_\perp(\alpha'_\perp\Omega_1^{-1}\alpha_\perp)^{-1}\alpha'_\perp\Omega_1^{-1}\right) \\ = \; &\Omega_1^{-1}\alpha_\perp(\alpha'_\perp\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp)^{-1}\alpha'_\perp\Omega_1^{-1}. \qquad \text{(B. 28)} \end{aligned}$$

Therefore, the asymptotic distribution can be re-written as:

$$tr\left\{\left(\int_0^1 B_d(u)dV_d^H(u)'\right)'\left(\int_0^1 B_d(u)B_d(u)'du\right)^{-1}\left(\int_0^1 B_d(u)dV_d^H(u)'\right)\right\}, \qquad \text{(B. 29)}$$

where $V_d^H(u)$ is defined as $\left(\alpha'_\perp\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp\right)^{-1/2}\alpha'_\perp\Omega_1^{-1}W_p^*(u)$. Therefore, contrast to that in Lemma 5.1, $E[B_d(u)V_d^H(u)'] = (\alpha'_\perp(EV_{t-1})\alpha_\perp)^{-1/2}\alpha'_\perp E[W_p(u)W_p^*(u)']\Omega_1^{-1}\alpha_\perp(\alpha'_\perp\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp)^{-1/2}$ $= u(\alpha'_\perp(EV_{t-1})\alpha_\perp)^{-1/2}\alpha'_\perp\Omega_1^{-1}\alpha_\perp(\alpha'_\perp\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp)^{-1/2} = u\Upsilon^{H'}$. All in all, we can re-write $V_d^H(u)$ as a linear combination of two independent $d-$dimensional standard BMs:

$$\begin{aligned} &\Upsilon^H B_d(u) + \left[(\alpha'_\perp\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp)^{-1/2}\alpha'_\perp\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp(\alpha'_\perp\Omega_1^{-1}\Omega_1^*\Omega_1^{-1}\alpha_\perp)^{-1/2} - \Upsilon^H\Upsilon^{H'}\right]^{1/2}B_d^*(u) \\ &= \Upsilon^H B_d(u) + \left[I_d - \Upsilon^H\Upsilon^{H'}\right]^{1/2}B_d^*(u). \qquad \text{(B. 30)} \end{aligned}$$

Thus the asymptotic distribution in (B.29) can be re-written as:

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

But $\Upsilon^H B_d(u) \sim N(0, \Upsilon^H\Upsilon^{H'})$. Abusing the notation, we write $\Upsilon^H B_d(u)$ as $\left(\Upsilon^H\Upsilon^{H'}\right)^{1/2}B_d(u)$, where $B_d(u)$ is (another) $d-$dimensional standard BM independent of $B_d^*(u)$. Canceling some of the $\left(\Upsilon^H\right)^{-1}\left(\Upsilon^H\Upsilon^{H'}\right)^{1/2}$ terms, we can re-write the asymptotic distribution in (B.29):

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

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

$$tr\left\{\left[\int_0^1 \Theta' B_d(u) dB_d(u)'\Theta\left(I_d-\Xi^H\right)^{1/2}\Theta' + \int_0^1 \Theta' B_d(u) dB_d^*(u)'\Theta\Xi^{H1/2}\Theta'\right]'\left[\int_0^1 \Theta' B_d(u)B_d(u)'du\Theta\right]^{-1} \cdot \left[\int_0^1 \Theta' B_d(u) dB_d(u)'\Theta\left(I_d-\Xi^H\right)^{1/2}\Theta' + \int_0^1 \Theta' B_d(u) dB_d^*(u)'\Theta\Xi^{H1/2}\Theta'\right]\right\}.$$

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 $\Theta' B_d^*(u)$ as two independent standard BMs $B_d(u)$ and $B_d^*(u)$ respectively. As $\Theta'\Theta = I_d$, we have:

$$\begin{aligned} & tr\left\{\left[\int_0^1 B_d(u) dB_d(u)'\left(I_d-\Xi^H\right)^{1/2} + \int_0^1 B_d(u) dB_d^*(u)'\Xi^{H1/2}\right]'\left[\int_0^1 B_d(u)B_d(u)'du\right]^{-1} \cdot \left[\int_0^1 B_d(u) dB_d(u)'\left(I_d-\Xi^H\right)^{1/2} + \int_0^1 B_d(u) dB_d^*(u)'\Xi^{H1/2}\right]\right\} \\ & = tr\left\{\left[\zeta\left(I_d-\Xi^H\right)^{1/2} + \Phi\Xi^{H1/2}\right]'\left[\zeta\left(I_d-\Xi^H\right)^{1/2} + \Phi\Xi^{H1/2}\right]\right\}. \end{aligned}$$

The proof is complete. □

REFERENCES

Ahn, S. K. and G.C. Reinsel, 1990, Estimation for partially nonstationary multivariate models, Journal of the American Statistical Association 85, 813-823.

Alexakis, P. and N. Apergis, 1996, ARCH effects and cointegration: is the foreign exchange market efficient?, Journal of Banking and Finance 20, 687-697.

Anderson, T.W., 1951, Estimating linear restrictions on regression coefficients for multivariate normal distributions, Annals of Mathematical Statistics 22, 327-351. [Correction, Annals of Statistics 8, 1980, p.1400.]

Anderson, T.W., 2002, Reduced rank regression in cointegrated models, Journal of Econometrics 106, 203-216.

Bollerslev, T., 1990, Modelling the coherence in the short-run nominal exchange rates: a multivariate generalized ARCH approach, Review of Economics and Statistics 72, 498-505.

Brenner, R.J. and K.F. Kroner, 1995, Arbitrage, cointegration, and testing the unbiasedness hypothesis in financial markets, Journal of Financial and Quantitative Analysis 30, 23-42.

Chan, N.H. and C.Z. Wei, 1988, Limiting distributions of least squares estimates of unstable autoregressive processes, Annals of Statistics 16, 367-401.

Engle, R.F. and C.W.J. Granger, 1987, Cointegration and error correction: representation, estimation and testing, Econometrica 55, 251-276.

Franses, P.H., P. Kofman, and J. Moser, 1994, Garch effects on a test for cointegration, Review of Quantitative Finance and Accounting 4, 19-26.

Glosten, L.R., R. Jagannathan, and D.E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, Journal of Finance 48, 1779-1802.

Gonzalo, J., Ng, S., 2001, A systematic framework for analyzing the dynamic effects of permanent and transitory shocks, Journal of Economic Dynamics and Control 25, 1527-1546.

Gourieroux, C. and A. Monfort, 1989, A general framework for testing a null hypothesis in a "mixed" form, Journal of Econometrics 5, 63-82.

Granger, C.W.J., 1983, Cointegrated variables and error correction models, Discussion Paper, Department of Economics, University of California at San Diego.

Hasbrouck, J., 1995, One security, many markets: determining the contributions to price discovery, Journal of Finance 50, 1175-1199.

Hasbrouck, J., 2003, Intraday price formation in U.S. equity index markets, Journal of Finance 58, 2375-2400.

Johansen, S., 1988, Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control 12, 231-254.

Johansen, S., 1996, Likelihood-based inference in cointegrated vector autoregressive models. New York: Oxford University Press.

Kroner, K.F. and J. Sultan, 1993, Time-varying distributions and dynamic hedging with foreign currency futures, Journal of Financial and Quantitative Analysis 28, 535-551.

Lee, T.H. and Y. Tse, 1996, Cointegration tests with conditional heteroskedasticity, Journal of Econometrics 73, 401-410.

Li, W.K., S. Ling, and H. Wong, 2001, Estimation for partially nonstationary multivariate autoregressive models with conditional heteroskedasticity, Biometrika 88, 1135-1152.

Ling, S. and W.K. Li, 1998, Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors, Annals of Statistics 26, 84-125.

Ling, S., W.K. Li, and M. McAleer, 2003, Estimation and testing for unit root processes with GARCH(1,1) errors: theory and Monte Carlo study, Econometric Reviews 22, 179-202.

Ling, S. and M. McAleer, 2003a, Asymptotic theory for a vector ARMA-GARCH model, Econometric Theory 19, 280-310.

Ling, S. and M. McAleer, 2003b, On adaptive estimation in nonstationary ARMA models with GARCH errors, Annals of Statistics 31, 642-674.

Magnus, J.R., 1988, Linear Structures. New York: Oxford University Press.

Phillips, P.C.B. and S.N. Durlauf, 1986, Multiple time series regression with integrated processes, Review of Economic Studies 53, 473-495.

Reinsel, G.C. and S.K. Ahn, 1992, Vector AR models with unit root and reduced rank structure: estimation, likelihood ratio test, and forecasting, Journal of Time Series Analysis 13, 353-375.

Seo, B., 1999, Distribution theory for unit root tests with conditional heteroskedasticity, Journal of Econometrics 91, 113-144.

Seo, B., 2007, Asymptotic distribution of the cointegrating vector estimator in error correction models with conditional heteroskedasticity, Journal of Econometrics 137, 68-111.

Stock, J.H. and M.W. Watson, 1993, A simple estimator of cointegrating vectors in higher order integrated system, Econometrica 61, 783-820.

Tse, Y.K., 2000, A test for constant correlations in a multivariate GARCH model, Journal of Econometrics 98, 107-127.