Econometric Theory, **16**, 2000, 905–926. Printed in the United States of America.

VECTOR AUTOREGRESSIONS WITH UNKNOWN MIXTURES OF $I(0)$, $I(1)$, AND $I(2)$ COMPONENTS

YOOSOON CHANG
Rice University

This paper develops a new estimation method for nonstationary vector autoregressions (VAR's) with unknown mixtures of $I(0)$, $I(1)$, and $I(2)$ components. The method does not require prior knowledge on the exact number and location of unit roots in the system. It is, therefore, applicable for VAR's with any mixture of $I(0)$, $I(1)$, and $I(2)$ variables, which may be cointegrated in any form. The limit theory for the stationary component of our estimator is still normal, thereby preserving the usual VAR limit theory. Yet, the leading term of the nonstationary component of the estimator has mixed normal limit distribution and does not involve unit root distribution. Our method is an extension of the FM-VAR procedure by Phillips (1995, *Econometrica* 63, 1023–1078) and yields an estimator that is optimal in the sense of Phillips (1991, *Econometrica* 59, 283–306). Moreover, we show for a certain class of linear restrictions that the Wald tests based on the estimator are asymptotically distributed as a weighted sum of independent chi-square variates with weights between zero and one. For such restrictions, the limit distribution of the standard Wald test is nonstandard and nuisance parameter dependent. This has a direct application for Granger-causality testing in nonstationary VAR's.

1. INTRODUCTION

Nonstationary vector autoregressions (VAR's) with $I(1)$ processes have been investigated by many authors, and their statistical theory is now well established. The statistical theory for such VAR's is developed by Park and Phillips (1989) and Sims, Stock, and Watson (1990). The maximum likelihood estimation of those models in error correction model (ECM) or reduced rank form is proposed by Ahn and Reinsel (1988) and Johansen (1991). Toda and Phillips (1993, 1994) consider testing for causality in such nonstationary VAR's.

Phillips (1995) shows that the fully modified least squares (FM-OLS) regression by Phillips and Hansen (1990) provides an optimal inference for regressions with unknown mixtures of $I(0)$ and $I(1)$ regressors. Chang and Phillips

I am very grateful to Bruce E. Hansen for many helpful suggestions and encouragements. My thanks also go to two anonymous referees for their comments, which have greatly improved the content and exposition of the paper. I also thank Joon Y. Park for helpful discussions. Address correspondence to: Yoosoon Chang, Department of Economics, MS 22, Rice University, 6100 Main Street, Houston, TX 77005, USA.

 0266-4666/00 $9.50

(1995) extend the methodology to regressions including $I(2)$ regressors and propose the residual-based fully modified least squares (RBFM-OLS) procedure. The approach by Phillips (1995) and Chang and Phillips (1995) is in sharp contrast with other existing methods. All the existing optimal methods presume knowledge on the unit roots and cointegration in the model, which is in practice obtained through preliminary tests.

The theory for FM-OLS is valid also for VAR models with unknown mixtures of $I(0)$ and $I(1)$ components, as shown in Phillips (1995). However, the RBFM-OLS method by Chang and Phillips (1995) is not applicable to VAR's with unknown mixtures of $I(0)$, $I(1)$, and $I(2)$ components. The estimator is simply undefined in the context of VAR's. We propose in the paper a new method called *residual-based fully modified vector autoregression* (RBFM-VAR) procedure that is applicable to any VAR, as long as the individual variables are integrated of order not exceeding two. We allow for any unknown mixture of $I(0)$, $I(1)$, and $I(2)$ variables included in the VAR model. Moreover, the $I(1)$ and $I(2)$ variables may be cointegrated in any form among themselves.

The RBFM-VAR procedure is an extension of the FM-VAR methodology developed in Phillips (1995) and is optimal in the sense of Phillips (1991), though it does not require precise knowledge about the number of unit roots and double unit roots in individual series and the cointegrating relationships in the model. Naturally, our estimator has a limit distribution that is identical to that of the fully modified vector autoregression (FM-VAR) estimator by Phillips (1995) when the VAR includes only $I(0)$ and $I(1)$ components. For a certain class of linear restrictions, we show that the inference based on our estimator yields Wald tests that are asymptotically distributed as a weighted sum of independent chi-square variates with weights between zero and one.

The rest of the paper is organized as follows. Section 2 introduces the model with assumptions. Our RBFM-VAR estimator is proposed in Section 3, where we also investigate the asymptotic behavior of the estimator. Section 4 develops an asymptotic theory for the modified Wald tests based on the RBFM-VAR regression. The results from Monte Carlo simulations are reported in Section 5. Section 6 concludes the paper. Mathematical proofs are given in the Appendix.

The following terminology and notations are used in the paper. We denote by $\Omega = \sum_{k=-\infty}^{\infty} E(u_k u_0')$ the long run variance matrix of the stationary time series u_t and write $lr\,\mathrm{var}(u_t) = \Omega$. We use $\mathrm{BM}(\Omega)$ to denote a vector Brownian motion with covariance matrix Ω and write integrals with respect to Lebesque measure such as $\int_0^1 B(s)ds$ simply as $\int_0^1 B$. The notation $X_t \sim I(d)$ signifies that the time series $\{X_t\}$ is integrated of order d, so that $\triangle^d X_t \sim I(0)$, and this requires that $lr\,\mathrm{var}(\triangle^d X_t) > 0$. The inequality >0 denotes positive definite when applied to matrices. We use the symbols $\to_d$, $\to_p$, $\equiv$, and $:=$ to signify convergence in distribution, convergence in probability, equality in distribution, and notational definition, respectively. We also use $\mathrm{vec}(A)$ to stack the rows of a matrix A into a column vector and $[x]$ to denote the smallest integer $\leq x$. All the limits given in the paper are taken as the sample size $T \to \infty$.

2. THE MODEL AND PRELIMINARY RESULTS

Suppose we want to estimate a pth order VAR given by

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t = A(L)y_{t-1} + \varepsilon_t, \quad \textbf{(1)}$$

where $A(L) = \sum_{i=1}^{p} A_i L^{i-1}$. The system (1) is initialized at $t = -p+1,\ldots,0$. We let the initial values $\{y_{-p+1},\ldots, y_0\}$ be any random vectors including constants, because our asymptotics do not depend on them. To be more specific about the order of unit roots and cointegrating space, we write (1) as

$$\triangle^2 y_t = \Phi(L)\triangle^2 y_{t-1} + \Pi_1 \triangle y_{t-1} + \Pi_2 y_{t-1} + \varepsilon_t \quad \textbf{(2)}$$

in the ECM format used by Johansen (1995). The ranks of Π_1 and Π_2, and their ranges and null spaces, determine the nonstationary characteristics of the model. In what follows, we use the notations $\gamma_\perp$ and $\bar{\gamma}$, defined, respectively, by $\gamma_\perp' \gamma = 0$ and $\bar{\gamma} = \gamma(\gamma'\gamma)^{-1}$, for matrices γ $(n \times r)$ and $\gamma_\perp$ $(n \times (n-r))$ of full column rank.

Assumption 1. We assume

(a) ε_t is i.i.d. with zero mean, variance matrix $\Sigma_{\varepsilon\varepsilon} > 0$, and finite fourth order cumulants.
(b) The determinantal equation $|I - A(L)L| = 0$ has roots equal to one or outside the unit circle, i.e., $|L| \geq 1$.
(c) $\Pi_2 = \alpha\beta'$ has rank $r < n$, where α and β are $(n \times r)$ full rank matrices.
(d) $\bar{\alpha}_\perp' \Psi \bar{\beta}_\perp = \varphi\eta'$ has rank $s < n - r$, where $\Psi = \Pi_1 + \Pi_2$ and φ and η are $((n-r) \times s)$ matrices of full column rank.
(e) $\varphi_\perp' \bar{\alpha}_\perp' (\Psi \overline{\beta\alpha'} \Psi + I - \sum_{i=1}^{p-2} \Phi_i) \bar{\beta}_\perp \eta_\perp$ has full column rank $(n - r - s)$.

Remarks.

(a) When $r \neq 0$ and $s \neq 0$, it follows from Theorem 3 in Johansen (1995) that y_t is a mixture of $I(0)$, $I(1)$, and $I(2)$ processes under Assumption 1(a)–(e). Specifically, $\beta_2' \triangle^2 y_t$, $\beta_1' \triangle y_t$, and $\beta' y_t + \bar{\alpha}' \Psi \bar{\beta}_2 \beta_2' \triangle y_t$ are stationary processes, where $\beta_1 = \beta_\perp \eta$ and $\beta_2 = \bar{\beta}_\perp \eta_\perp$. Notice that the last stationary process listed involves cointegration of $I(2)$ process y_t with its own difference $\triangle y_t$, thereby establishing multicointegration or polynomial cointegration introduced in Engle and Yoo (1991). It follows that $(\beta, \beta_1)' y_t$, $\beta_2' \triangle y_t$ are $I(1)$ processes and $\beta_2' y_t$ is $I(2)$.
(b) In the case where $r = s = 0$, we have $\Pi_1 = \Pi_2 = 0$, and this implies that y_t is a noncointegrated $I(2)$ process.

Our estimation of $I(2)$-VAR (1) is based on the least squares regression

$$y_t = \Phi z_t + A w_t + \varepsilon_t, \quad \textbf{(3)}$$

where $z_t = (\triangle^2 y_{t-1}', \ldots, \triangle^2 y_{t-p+2}')'$, $w_t = (\triangle y_{t-1}', y_{t-1}')'$, and the coefficient matrices Φ and A are defined accordingly from A_i's in (1). We may recover the estimates for A_i's from those of Φ and A using the relationships

$$\Phi = (\Phi_1, \ldots, \Phi_{p-2}) \quad \text{with} \quad \Phi_i = \sum_{k=i}^{p} (k - i + 1) A_k$$

and

$$A = \left(-\sum_{k=2}^{p} (k-1) A_k, \sum_{k=1}^{p} A_k \right).$$

The regressors included in z_t earlier are the lagged second differences, and hence they are known to be stationary; however, those in w_t are the first differences and the levels of the data that are of unknown mixed order.

We use a $(2n \times 2n)$ matrix H to separate out the $I(0)$, $I(1)$, and $I(2)$ components of the $(2n \times 1)$ regressor w_t of unknown mixed order. In the notation of Assumption 1, the matrix H is expressed as

$$H = (H_1', H_2', H_3')' = \left(\begin{array}{ccc|cc|c} \beta_2(\bar{\alpha}'\Psi\bar{\beta}_2)' & \beta_1 & \beta & \beta_2 & 0 & 0 \\ \beta & 0 & 0 & 0 & \beta_1 & \beta_2 \end{array} \right)' \tag{4}$$

and the corresponding inverse as

$$H^{-1} = (H^1, H^2, H^3) = \left(\begin{array}{ccc|cc|c} 0 & \bar{\beta}_1 & \bar{\beta} & \bar{\beta}_2 & 0 & 0 \\ \bar{\beta} & 0 & 0 & -\bar{\beta}\bar{\alpha}'\Psi\bar{\beta}_2 & \bar{\beta}_1 & \bar{\beta}_2 \end{array} \right).$$

The component matrices H_1, H_2, and H_3 are of ranks $m_1 = 2r + s$, $m_2 = n - r$, and $m_3 = n - r - s$, respectively. We then specify w_t as follows:

$$\begin{aligned} H_1 w_t &= w_{1t} = u_{1t}, \\ \triangle H_2 w_t &= \triangle w_{2t} = u_{2t}, \\ \triangle^2 H_3 w_t &= \triangle^2 w_{3t} = u_{3t}, \end{aligned} \tag{5}$$

where u_{1t}, u_{2t}, and u_{3t} generate, respectively, the $I(0)$, $I(1)$, and $I(2)$ components of w_t. The matrix H contains the information about the exact orders of integration of the individual components in the potentially nonstationary regressor w_t and the precise form of cointegration in the model (3). We emphasize that H is unknown and that the method proposed in the present paper assumes no such knowledge about H.

Define an $(np \times np)$ matrix G by

$$G = \begin{pmatrix} I_{n(p-2)} & 0 \\ 0 & H' \end{pmatrix}' = \begin{pmatrix} I_{n(p-2)} & 0 & 0 & 0 \\ 0 & H_1' & H_2' & H_3' \end{pmatrix}' =: (G_1', G_2', G_3')'$$

and its inverse by

$$G^{-1} = \begin{pmatrix} I_{n(p-2)} & 0 \\ 0 & H^{-1} \end{pmatrix} = \begin{pmatrix} I_{n(p-2)} & 0 & 0 & 0 \\ 0 & H^1 & H^2 & H^3 \end{pmatrix} =: (G^1, G^2, G^3).$$

The matrix G separates out the $I(0)$, $I(1)$, and $I(2)$ components of the entire regressor $x_t = (z_t', w_t')'$ in (3). We may now rewrite the model (3) as

$$\begin{aligned} y_t &= Fx_t + \varepsilon_t \\ &= F_1 x_{1t} + F_2 x_{2t} + F_3 x_{3t} + \varepsilon_t, \end{aligned} \tag{6}$$

where $F = (\Phi, A)$, $F_1 = FG^1 = (\Phi, A^1)$, $F_2 = FG^2 = A^2$, $F_3 = FG^3 = A^3$, with $A^i = AH^i$ for $i = 1,2,3$ and

$$\begin{aligned} x_{1t} &= G_1 x_t = (z_t', w_t' H_1')' = (z_t', w_{1t}')' \sim I(0), \\ x_{2t} &= G_2 x_t = H_2 w_t = w_{2t} \sim I(1), \\ x_{3t} &= G_3 x_t = H_3 w_t = w_{3t} \sim I(2) \end{aligned} \tag{7}$$

with $x_t = (x_{1t}', x_{2t}', x_{3t}')'$.

For the development of our asymptotic theory, we define $u_t = (\varepsilon_t', u_{2t}', u_{3t}')'$ to be an $(n + m_2 + m_3)$ vector stationary process. Because of Phillips and Solo (1992), the functional central limit theory (FCLT) for u_t holds; i.e., $T^{-1/2} \sum_{t=1}^{[T\cdot]} u_t \to_d B(\cdot) \equiv \mathrm{BM}(\Omega)$, where $B = (B_\varepsilon', B_2', B_3')'$ is a vector Brownian motion with covariance matrix $\Omega = \sum_{j=-\infty}^{\infty} \mathbf{E} u_j u_0'$.[1] We also define the contemporaneous covariance matrix Σ and the one-sided long run covariance matrix Δ of u_t by $\Sigma = \mathbf{E} u_0 u_0'$ and $\Delta = \sum_{j=0}^{\infty} \mathbf{E} u_j u_0'$. We partition Ω, Σ, and Δ conformably with the partition of u_t into cell submatrices, Ω_{ij}, Σ_{ij}, and Δ_{ij}, for $i, j = \varepsilon, 2, 3$.

Moreover, if we let $\varphi_t = \varepsilon_t \otimes x_{1t}$, then $\{\varphi_t\}$ is a martingale difference sequence (mds) with $\mathrm{var}(\varphi_t) = lr\,\mathrm{var}(\varphi_t) = \Sigma_{\varepsilon\varepsilon} \otimes \Sigma_{x11}$, because $\{\varepsilon_t\}$ is independent and identically distributed (i.i.d.) under Assumption 1. Therefore, we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varphi_t \to_d \mathbf{N}\left(0, \sum_{j=-\infty}^{\infty} \mathbf{E}(\varepsilon_t \varepsilon_{t+j}' \otimes x_{1t} x_{1t+j}')\right) \equiv \mathbf{N}(0, \Sigma_{\varepsilon\varepsilon} \otimes \Sigma_{x11}), \tag{8}$$

where $\Sigma_{x11} = \mathbf{E} x_{1t} x_{1t}'$.

3. THE RBFM-VAR ESTIMATOR AND ITS LIMIT THEORY

We now introduce a new method of estimating the $I(2)$-VAR model (1) that does not require prior knowledge about the number of unit roots and double unit roots in the system or pretesting to determine the dimension of the cointegration space. Our method is based on the regression formulated in (3), which can be viewed as a regression with an unknown mixture of $I(0)$, $I(1)$, and $I(2)$ processes. One may therefore consider directly applying the RBFM-OLS method of Chang and Phillips (1995) to estimate the model (3). Unfortunately, the method is not applicable here. The RBFM-OLS procedure corrects the endogeneity using the residual from the first order autoregression of the differenced nonstationary regressor $w_t = (\triangle y'_{t-1}, y'_{t-1})'$, which reduces in this case to $\hat{v}_t = (\hat{v}'_{1t}, \hat{v}'_{2t})'$ in

$$\begin{pmatrix} \triangle^2 y_{t-1} \\ \triangle y_{t-1} \end{pmatrix} = \hat{J} \begin{pmatrix} \triangle^2 y_{t-2} \\ \triangle y_{t-2} \end{pmatrix} + \begin{pmatrix} \hat{v}_{1t} \\ \hat{v}_{2t} \end{pmatrix}. \tag{9}$$

However, we have $\hat{v}_{1t} = \hat{v}_{2t}$, and this results in singularity in parameter estimates. To see this, write $\triangle^2 y_{t-1} = \triangle y_{t-1} - \triangle y_{t-2}$ and note that $\triangle y_{t-2}$ is included in the regressors $(\triangle^2 y'_{t-2}, \triangle y'_{t-2})'$ in (9). Therefore, the fitted residual $\hat{v}_{1t}$ from the regression for $\triangle^2 y_{t-1}$ becomes identical to the fitted residual from the regression for $\triangle y_{t-1}$, which is exactly $\hat{v}_{2t}$. The RBFM-OLS estimator is therefore not defined for the VAR models.

To introduce a new estimator, we write (3) and (6) in matrix format as

$$Y' = \Phi Z' + AW' + E' = FX' + E', \tag{10}$$

where $Y' = (y_1, \ldots, y_T)$, $Z' = (z_1, \ldots, z_T)$, $W' = (w_1, \ldots, w_T)$, $E' = (\varepsilon_1, \ldots, \varepsilon_T)$, $X = (Z, W)$, and $W = (\triangle Y'_{-1}, Y'_{-1})'$ with $Y_{-1} = (y_0, \ldots, y_{T-1})$. We use for the construction of our correction terms the preliminary ordinary least squares (OLS) residual $\hat{\varepsilon}_t$ and the process

$$\hat{v}_t = \begin{pmatrix} \hat{v}_{1t} \\ \hat{v}_{2t} \end{pmatrix} = \begin{pmatrix} \triangle^2 y_{t-1} \\ \triangle y_{t-1} - \hat{N} \triangle y_{t-2} \end{pmatrix}, \tag{11}$$

where $\hat{N}$ is the OLS coefficient estimate from the regression of $\triangle y_{t-1}$ on $\triangle y_{t-2}$. We also define $\hat{V} = (\triangle^2 Y_{-1}, \triangle Y_{-1} - \hat{N} \triangle Y_{-2})$.

Our estimator, which we call the RBFM-VAR estimator, is defined by

$$\hat{F}^+ = (\hat{\Phi}^+, \hat{A}^+) = (Y'Z, Y^{+'}W + T\hat{\Delta}^+)(X'X)^{-1} \tag{12}$$

with

$$Y^{+'} = Y' - \hat{\Omega}_{\hat{\varepsilon}\hat{v}} \hat{\Omega}_{\hat{v}\hat{v}}^{-1} \hat{V}' \quad \text{and} \quad \hat{\Delta}^+ = \hat{\Omega}_{\hat{\varepsilon}\hat{v}} \hat{\Omega}_{\hat{v}\hat{v}}^{-1} \hat{\Delta}_{\hat{v}\triangle w}, \tag{13}$$

where consistent estimates for various nuisance parameters are denoted by $\hat{}$, as we will explain subsequently. Note from the definition of $\hat{F}^+$ given in (12) that we leave the known to be stationary regressor Z intact and transform only the regressors of unknown mixed order W to correct for its potential endogeneity and serial correlation effects.

In the formulae for the correction terms given in (13), $\hat{\Omega}_{\hat{\varepsilon}\hat{v}}$ and $\hat{\Omega}_{\hat{v}\hat{v}}$ are kernel estimates of the long run covariance matrices of $(\hat{\varepsilon}_t, \hat{v}_t)$ and $\hat{v}_t$, respectively. Similarly, $\hat{\Delta}_{\hat{v}\triangle w}$ is a kernel estimate of the one-sided long run covariance of $\hat{v}_t$ and $\triangle w_t$. These kernel estimates are defined in the general form, which can be found in Priestley (1981) or Hannan (1970). As in the analyses for the $I(1)$

cointegrated models in Phillips (1995) and for the $I(2)$ cointegrating regressions in Chang and Phillips (1995), the kernel estimation of both Ω and Δ continues to play an important role in developing the limit theory for our $I(2)$-VAR models. We use the same class of admissible kernels as in the aforementioned references.

We also employ the same expansion rate order symbol O_e defined in Phillips (1995) and Chang and Phillips (1995) to explicitly characterize rates of expansion of the lag truncation or the bandwidth $K = K(T)$ as $T \to \infty$. We use the definition $K = O_e(T^k)$ to impose some explicit conditions on how the bandwidth parameter K grows as $T \to \infty$. In particular, the bandwidth parameter expansion rate, k, is used in the kernel estimation of the long run covariance matrices appearing in the formulae for our correction terms given in (13).

We use a subscript coupling notation b by $b = 2,3$ to group the nonstationary regressors and their coefficient matrices in (6) as $x_{bt} = (x_{2t}', x_{3t}')'$ and $F_b = (F_2, F_3)$. We may then conveniently formulate the asymptotic theory in terms of the component submatrices F_1 and F_b that correspond to the stationary and nonstationary components of the regressors. Also define $D_T = \text{diag}(TI_{m_2}, T^2 I_{m_3})$ for normalization of the $I(1)$ and $I(2)$ components in our subsequent asymptotic analyses. We now present the limit theory for the RBFM-VAR estimator given in (12).

THEOREM 1. *Under Assumption* 1, *we have*

(a) $\sqrt{T}(\hat{F}^+ - F)G^1 \to_d \mathbf{N}(0, \Sigma_{\varepsilon\varepsilon} \otimes \Sigma_{x11}^{-1})$,
(b) $(\hat{F}^+ - F)G^b D_T \to_d \int_0^1 dB_{\varepsilon\cdot 2} \bar{B}_b' (\int_0^1 \bar{B}_b \bar{B}_b')^{-1} \equiv \mathbf{MN}(0, \Omega_{\varepsilon\varepsilon\cdot 2} \otimes (\int_0^1 \bar{B}_b \bar{B}_b')^{-1})$,

where $\bar{B}_b = (B_2', \bar{B}_3')'$ *with* $\bar{B}_3(r) = \int_0^r B_3(s)ds$ *and* $B_{\varepsilon\cdot 2} = B_\varepsilon - \Omega_{\varepsilon 2}\Omega_{22}^{-1} B_2 \equiv \text{BM}(\Omega_{\varepsilon\varepsilon\cdot 2})$ *with* $\Omega_{\varepsilon\varepsilon\cdot 2} = \Sigma_{\varepsilon\varepsilon} - \Omega_{\varepsilon 2}\Omega_{22}^{-1}\Omega_{2\varepsilon}$.

Part (a) *holds for the bandwidth parameter expansion rate* $K = O_e(T^k)$ *for* $k \in (\frac{1}{4}, \frac{1}{2})$. *Part* (b) *holds for* $k \in (0, \frac{1}{2})$. *The limit distributions in* (a) *and* (b) *are statistically independent.*

Remarks.

(a) The limit distribution of the RBFM-VAR estimator for the stationary component coefficient remains the same as the corresponding OLS estimator, which will be called OLS-VAR henceforth. Therefore our procedure does preserve the usual VAR limit theory for the stationary components in the absence of prior or pretest information on the cointegration space.
(b) The limit distribution of the RBFM-VAR estimator for the nonstationary coefficient is mixed normal. The mixed normality follows from the independence of the limit Brownian motions $B_{\varepsilon\cdot 2}$ and $\bar{B}_b$. The covariance matrix $\Omega_{\varepsilon\varepsilon\cdot 2}$ of $B_{\varepsilon\cdot 2}$ is singular along H_2 defined in (4). This implies in particular that the limiting distribution in part (b) is degenerate in the unit root direction. It is possible to analyze lower order asymptotics along this direction, but we do not pursue it any further in the paper.

(c) The statistical independence of the limit distributions in parts (a) and (b) in the preceding discussion is established by the i.i.d. property of ε_t. The form of the covariance matrix, $\Sigma_{\varepsilon\varepsilon} \otimes \Sigma_{x11}^{-1}$, in part (a) is also due to this property.

(d) The process $\hat{v}_t$ defined in (11) can be viewed as the residual from regression (9) with restrictions on the coefficient matrix $\hat{J}$, namely,

$$\begin{pmatrix} \triangle^2 y_{t-1} \\ \triangle y_{t-1} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \hat{N} \end{pmatrix} \begin{pmatrix} \triangle^2 y_{t-2} \\ \triangle y_{t-2} \end{pmatrix} + \begin{pmatrix} \hat{v}_{1t} \\ \hat{v}_{2t} \end{pmatrix}. \tag{14}$$

The zero restrictions on $\hat{J}$ remove the singularity problem that arises in the application of the RBFM-OLS procedure. To examine the preceding regression more explicitly, we further partition the rotation matrix H and its inverse as

$$H = \begin{pmatrix} H_{11}' & H_{21}' & H_{31}' \\ H_{12}' & H_{22}' & H_{32}' \end{pmatrix}' \quad \text{and} \quad H^{-1} = \begin{pmatrix} H^{11} & H^{21} & H^{31} \\ H^{12} & H^{22} & H^{32} \end{pmatrix}$$

and use these to respecify the model (14) as

$$H\triangle w_t = H \begin{pmatrix} 0 & 0 \\ 0 & \hat{N} \end{pmatrix} H^{-1} H \triangle w_{t-1} + H\hat{v}_t,$$

i.e.,

$$\triangle w_{1t} = \hat{J}_{11}\triangle w_{1t-1} + \hat{J}_{12}\triangle w_{2t-1} + \hat{J}_{13}\triangle w_{3t-1} + \hat{v}_{1t}, \tag{15}$$

$$\triangle w_{2t} = \hat{J}_{21}\triangle w_{1t-1} + \hat{J}_{22}\triangle w_{2t-1} + \hat{J}_{23}\triangle w_{3t-1} + \hat{v}_{2t}, \tag{16}$$

$$\triangle w_{3t} = \hat{J}_{31}\triangle w_{1t-1} + \hat{J}_{32}\triangle w_{2t-1} + \hat{J}_{33}\triangle w_{3t-1} + \hat{v}_{3t}, \tag{17}$$

where we use the notations $\hat{J}_{ij} = H_{i2}\hat{N}H^{j2}$, for $i,j = 1,2,3$. The probability limits of the coefficient matrices on the $I(1)$ regressor $\triangle w_{3t-1}$ in the regressions (15) and (16) are zero lest the regressions be spurious. However, $p\lim \hat{J}_{33} = I$, because the regression (17) is a full rank $I(1)$ regression. We may indeed show that $\hat{J}_{13} = \hat{J}_{23} = O_p(T^{-1})$ and $\hat{J}_{33} = I + O_p(T^{-1})$, because the OLS estimators for the coefficients of $I(1)$ variables are T consistent. The residual $\hat{v}_{ht} := (\hat{v}_{1t}', \hat{v}_{2t}', \hat{v}_{3t}')'$ can then be expressed as

$$\hat{v}_{ht} = \begin{pmatrix} \triangle u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix} - \begin{pmatrix} \hat{J}_{11} & \hat{J}_{12} \\ \hat{J}_{21} & \hat{J}_{22} \\ \hat{J}_{31} & \hat{J}_{32} \end{pmatrix} \begin{pmatrix} \triangle u_{1t-1} \\ u_{2t-1} \end{pmatrix} + O_p(T^{-1/2})$$

using the definitions of u_{1t}, u_{2t}, and u_{3t} given in (5).

(e) As can be seen clearly from the previous discussion, the process $\hat{v}_t$ extracts and locates the stationary processes u_{2t} and u_{3t} exactly where they are needed for the correction of the endogeneities in the $I(1)$ and $I(2)$ components. In the stationary direction, however, $\hat{v}_t$ contains $\triangle u_{1t}$, the difference of the stationary process u_{1t}, which has zero long run variance. The limit of the kernel estimate $\hat{\Omega}_{\hat{v}\hat{v}}$ of the long run variance of $\hat{v}_t$ will therefore be singular in the stationary direction. This is precisely why our correction terms constructed from $\hat{v}_t$ leave the usual VAR limit theory for stationary components intact, while successfully removing the endogeneity problem in the limit distribution of the nonstationary OLS-VAR estimates. To achieve this, we of course need to correct for the serial correlation effects induced by our correction terms, i.e., the one between v_t and w_t in the $I(1)$ direction. This is again done similarly by exploiting the asymptotic singularity of the kernel estimate $\hat{\Delta}_{\hat{v}\triangle w}$ of the one-sided long run covariance of $\hat{v}_t$ and $\triangle w_t$. See (A.4) and (A.5) in the Appendix.

(f) When there are only $I(0)$ and $I(1)$ components in the system, the limit distribution given in part (b) becomes mixed normal with variance $\Omega_{\varepsilon\varepsilon\cdot 2} \otimes (\int_0^1 \bar{B}_2 \bar{B}_2')^{-1}$, which is identical to that of the FM-VAR estimator in Phillips (1995). Moreover, the conditional covariance matrix given in part (b) is identical to that of the maximum likelihood estimator (MLE) under Gaussian errors obtained by Kitamura (1995), because $\Omega_{\varepsilon\varepsilon\cdot 2} = \Sigma_{\varepsilon\varepsilon} - \Omega_{\varepsilon 2}\Omega_{22}^{-1}\Omega_{2\varepsilon} = \Sigma_{\varepsilon\varepsilon} - \Omega_{\varepsilon b}\Omega_{bb}^{+}\Omega_{b\varepsilon} = \Omega_{\varepsilon\varepsilon\cdot b}$. Our results in part (b) characterizing the asymptotic behaviors of our estimators correspond to those of the exact MLE under normality obtained by Johansen (1995, Theorem 5). However, it seems difficult to establish a direct comparison because the two estimators are based on different normalizations.[2]

4. HYPOTHESIS TESTING IN RBFM-VAR REGRESSION

We consider hypothesis testing in the VAR model (1) formulated as in (6). As usual, we write the general linear restrictions on the coefficient matrix F as

$$\mathcal{H}_0 : R\,\mathrm{vec}(F) = r, \qquad F = (\Phi, A), \qquad R\ (q \times n^2 p) \quad \text{of rank } q. \tag{18}$$

It is well known that the Wald test for the hypothesis (18) has chi-square limit distribution if the rank condition

$$\mathrm{rank}(R(\Sigma_{\varepsilon\varepsilon} \otimes G_1'\Sigma_{x11}^{-1}G_1)R') = q \tag{19}$$

holds. However, the rank condition (19) may fail. Importantly, such rank condition may fail in testing for Granger causality, as Toda and Phillips (1993, 1994) point out. They show that the limit theory of the causality test in nonstationary $I(1)$-VAR's may involve nuisance parameters and nonstandard distributions, if based on the OLS-VAR estimator. To alleviate such difficulty, we propose to use

$$W_F^{+} = T(R\,\mathrm{vec}\,\hat{F}^{+} - r)'(R(\hat{\Sigma}_{\hat{\varepsilon}\hat{\varepsilon}} \otimes T(X'X)^{-1})R')^{-1}(R\,\mathrm{vec}\,\hat{F}^{+} - r), \tag{20}$$

where $\hat{\Sigma}_{\hat{\varepsilon}\hat{\varepsilon}}$ is the usual covariance matrix estimate for the regression errors. It is a modified Wald test based on our RBFM-VAR estimator $\hat{F}^{+}$ defined in (12). Both W_F^{+} and the standard Wald test have the same χ_q^2 limiting distribution when the rank condition (19) is satisfied. However, they are expected to behave quite differently when the rank condition (19) fails.

To look more closely at the limit theory of W_F^{+} in the case of rank condition failure, we suppose that the restriction matrix R has the Kronecker product form $R = R_1 \otimes R_2'$, where R_1 $(q_1 \times n)$ and R_2 $(np \times q_2)$ are of rank q_1 and q_2, respectively, with $q_1 q_2 = q$. The causality restrictions may be formulated in this Kronecker product form that can be further restricted to (22), which follows. See the next section for an illustration. The rank condition (19) is then written accordingly as

$$\text{rank}(R_1 \Sigma_{\varepsilon\varepsilon} R_1' \otimes R_2' G_1' \Sigma_{x11}^{-1} G_1 R_2) = q_1 q_2, \tag{21}$$

which fails when $R_2' G_1'$ is of deficient row rank. This happens when the restriction R_2 isolates some of the nonstationary coefficients of A in $F = (\Phi, A)$. To effectively analyze such cases, we more specifically let $R_2 = \text{diag}(R_{2\Phi}, R_{2A})$ so that the restrictions on the potentially nonstationary coefficient A can be written out separately from those on the known to be stationary coefficient Φ as

$$\mathcal{H}_0' : R_1 \Phi R_{2\Phi} = R_{3\Phi} \quad \text{and} \quad R_1 A R_{2A} = R_{3A}, \tag{22}$$

where $\text{rank}(R_{2\Phi}) = q_\Phi$, $\text{rank}(R_{2A}) = q_A$, with $q_2 = q_\Phi + q_A$, and for some suitable matrices $R_{3\Phi}$ and R_{3A}. We may then write

$$R_{2A} = (R_{2A1}, R_{2Ab}) = (H^1 S_{A1}, H^b S_{Ab}), \tag{23}$$

where $\text{rank}(R_{2A1}) = q_{A1}$, and $\text{rank}(R_{2Ab}) = q_{Ab}$, with $q_A = q_{A1} + q_{Ab}$, and for some matrices S_{A1} and S_{Ab}. We assume without loss of generality that the matrix S_{A1} has full column rank.

When $q_{Ab} \neq 0$, i.e., when the restriction does relate to the nonstationary coefficients of A, the $R_2' G_1'$ becomes deficient in row rank, and consequently the rank condition (21) fails. The standard χ_q^2 limit theory therefore does not apply in this case. The following theorem provides the limit distribution of the modified Wald statistic W_F^+ in this case of the rank condition failure.

THEOREM 2. *Under Assumption* 1, *the modified Wald statistic* W_F^+ *for testing hypothesis* $\mathcal{H}_0 : R\,\text{vec}(F) = r$ *with* $R = R_1 \otimes R_2'$ *has a limit distribution that is a mixture of chi-square variates, for the bandwidth parameter expansion rate* $K = O_e(T^k)$ *for* $k \in (\frac{1}{4}, \frac{1}{2})$. *In particular, when* R_2 *has the form* $R_2 = \text{diag}(R_{2\Phi}, R_{2A})$, *where* R_{2A} *is given by* (23), *we have*

$$W_F^+ \to_d \chi^2_{q_1(q_\Phi + q_{A1})} + \sum_{i=1}^{q_1} d_i \chi^2_{q_{Ab}}(i),$$

where $\chi^2_{q_{Ab}}(i) \equiv$ *i.i.d.* $(\chi^2_{q_{Ab}})$, $i = 1, \ldots, q_1$ *and are independent of the* $\chi^2_{q_1(q_\Phi + q_{A1})}$ *member in the preceding equation. The coefficients* d_i, $i = 1, \ldots, q_1$ *are the eigenvalues of the matrix* $(R_1 \Omega_{\varepsilon\varepsilon\cdot 2} R_1')^{1/2} (R_1 \Sigma_{\varepsilon\varepsilon} R_1')^{-1} (R_1 \Omega_{\varepsilon\varepsilon\cdot 2} R_1')^{1/2}$. *The limit distribution of* W_F^+ *is bounded above by a* χ_q^2 *distribution.*

Remarks.

(a) From $\Omega_{\varepsilon\varepsilon\cdot 2} < \Sigma_{\varepsilon\varepsilon}$, it follows that $(R_1 \Omega_{\varepsilon\varepsilon\cdot 2} R_1')^{1/2} (R_1 \Sigma_{\varepsilon\varepsilon} R_1')^{-1} \times (R_1 \Omega_{\varepsilon\varepsilon\cdot 2} R_1')^{1/2} \leq I$, implying that the eigenvalues d_i, $i = 1, \ldots, q_1$ that appear in Theorem 2 satisfy $0 \leq d_i \leq 1$, $\forall i$. Consequently, the limit distribution of W_F^+ is bounded above by the variate $\chi^2_{q_1(q_\Phi + q_{A1})} + \sum_{i=1}^{q_1} \chi^2_{q_{Ab}}(i) \equiv \chi^2_{q_1(q_\Phi + q_A)}$. Thus, we can always construct an asymptotically conservative test for the hypothesis $\mathcal{H}_0'$ using a $\chi^2_{q_1(q_\Phi + q_A)} = \chi^2_{q_1 q_2} = \chi^2_q$ limit distribution. Thus, conventional critical values can be used to construct asymptotically valid, though conservative, tests in our RBFM-VAR regressions.

(b) The reason that the limit theory of W_F^+ is conservative is that W_F^+ uses the weighting metric $\hat{\Sigma}_{\hat{\varepsilon}\hat{\varepsilon}} \otimes (X'X)^{-1}$ for the entire coefficient matrix $F = (\Phi, A)$, irrespective of whether the associated variable is $I(0)$, $I(1)$, or $I(2)$. With our mds regression errors, this weighting matrix is proper for the stationary coefficient estimates; however, for the estimates of the nonstationary coefficients, it is heavier than it should be. (For a more detailed explanation, see Phillips, 1995, Remark 4.6(d).)

(c) The hypothesis formulated in (18) or (22) does not include the test of the rank of Π_2, except for the special case $\Pi_2 = 0$. The reader is referred to Johansen (1995) for the general rank test. On the other hand, Johansen also considers the hypothesis of the form $\Pi_2 R_2 = R_2$, with known restriction matrix R_2. This is just a linear hypothesis on Π_2, which is a special case of the restriction we consider here.

The limit theory presented in Theorem 2 establishes the extension of the results in Theorem 6.1 of Phillips (1995) to more general VAR models that allow for $I(2)$ processes and a wider range of cointegrations. Our theory includes causality tests and therefore offers an alternative to sequential test procedures such as the one in Toda and Phillips (1994) and to artificial model overfitting procedures such as the one introduced in Choi (1993).

5. A MONTE CARLO SIMULATION

To examine the finite sample behavior of the newly proposed RBFM-VAR estimator and test statistics, we perform a Monte Carlo simulation. For the simulation, we consider a VAR in $y_t = (y_{1t}, y_{2t})'$ generated by

$$\triangle y_{1t} = \rho_1 \triangle y_{1t-1} + \rho_2 (y_{1t-1} - \triangle y_{2t-1}) + \varepsilon_{1t}$$
$$\triangle^2 y_{2t} = \varepsilon_{2t}. \tag{24}$$

We set $\varepsilon_t \sim$ i.i.d. $\mathbf{N}(0, \Sigma)$ with $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ in the simulation.

The preceding data generating process for y_t can be written in the ECM form as in (2) with $\Phi(L) = 0$,

$$\Pi_1 = \begin{pmatrix} {}_1\pi_{11} & {}_1\pi_{12} \\ {}_1\pi_{21} & {}_1\pi_{22} \end{pmatrix} = \begin{pmatrix} \rho_1 - 1 & -\rho_2 \\ 0 & 0 \end{pmatrix},$$
$$\Pi_2 = \begin{pmatrix} {}_2\pi_{11} & {}_2\pi_{12} \\ {}_2\pi_{21} & {}_2\pi_{22} \end{pmatrix} = \begin{pmatrix} \rho_2 & 0 \\ 0 & 0 \end{pmatrix},$$

where the parameters ρ_1 and ρ_2 are required to be $|\rho_1| < 1$ and $-2(1 + \rho_1) < \rho_2 \leq 0$ under condition (b) of Assumption 1. Note that when $\rho_2 = 0$ and $\rho_1 = 1$, we have $\Pi_1 = \Pi_2 = 0$. This is exactly the case discussed in Remark (b) following Assumption 1. Here, the model (24) becomes $\triangle^2 y_t = \varepsilon_t$, which means that both y_{1t} and y_{2t} are $I(2)$ with no cointegration. If $\rho_2 = 0$ and $|\rho_1| < 1$, then (24) is written explicitly as $\triangle y_{1t} = \rho_1 \triangle y_{1t-1} + \varepsilon_{1t}$ and $\triangle^2 y_{2t} = \varepsilon_{2t}$. This implies that y_{2t} is still $I(2)$ but y_{1t} becomes $I(1)$ for all $|\rho_1| < 1$.

When $\rho_2 \neq 0$, the implications from Theorem 3 of Johansen (1995) directly apply. In the notations used in Assumption 1, we have from the reduced rank restriction $\Pi_2 = \alpha\beta'$ that $\alpha = (\rho_2, 0)'$ and $\beta = (1,0)'$. It is straightforward to see that $r = 1$ and $s = 0$ in conditions (c) and (d) of Assumption 1. Condition (e) in Assumption 1 is also trivially satisfied. This then implies that y_t is composed of $I(0)$, $I(1)$, and $I(2)$ processes. More specifically,

$$\begin{aligned}\beta' y_t &= y_{1t} \sim I(1),\\ \beta_2' y_t &= y_{2t} \sim I(2),\\ \beta' y_t + \bar{\alpha}' \Psi \bar{\beta}_2 \beta_2' \triangle y_t &= y_{1t} - \triangle y_{2t} \sim I(0),\end{aligned}$$

as discussed in Remark (a) following Assumption 1.

We look at the following three cases, each of which is defined by the values of the parameters ρ_1 and ρ_2:

Case A $(\rho_1, \rho_2) = (1,0)$. In this case, both y_1 and y_2 are $I(2)$ processes with no cointegrating relationship. Furthermore, none of y_1 and y_2 Granger-causes the other.

Case B $(\rho_1, \rho_2) = (0.5,0)$. One can easily see that y_1 reduces to $I(1)$ process under this specification, because $|\rho_1| < 1$. The other variable y_2 remains to be $I(2)$ process. As in Case A, no Granger causality exists in either direction.

Case C $(\rho_1, \rho_2) = (-0.3, -0.15)$. As in Case B, y_1 and y_2 are $I(1)$ and $I(2)$, respectively. However, y_2 in this case Granger-causes y_1.

We test whether y_{1t} is caused by y_{2t}. Then the null hypothesis of noncausality can be formulated as

$$\mathcal{H}_0 : {}_1\pi_{12} = 0 \quad \text{and} \quad {}_2\pi_{12} = 0, \tag{25}$$

which can also be expressed as $R\,\text{vec}(\Pi_1, \Pi_2) = r$ as in (18) with

$$R = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad r = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

where R and r can be written, respectively, as $R = R_1 \otimes R_2'$ and $r = \text{vec}(R_3)$ with

$$R_1 = (1,0), \qquad R_2 = \begin{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{pmatrix}, \quad \text{and} \quad R_3 = (0,0).$$

The null hypothesis is tested via Wald tests constructed from the OLS-VAR and the RBFM-VAR estimators for the coefficient matrices Π_1 and Π_2.

For each set of simulations, samples of sizes 150 and 500 are drawn 10,000 times to compare the finite sample performances of the OLS-VAR and the

RBFM-VAR estimators. Also the Wald tests based on the OLS-VAR and the RBFM-VAR estimators are compared in terms of their finite sample sizes and power properties. We explore how close the finite sample sizes of these Wald tests are in relation to the nominal sizes of the bounding variate χ_2^2.

Table 1 reports the finite sample biases and standard deviations (s.d.) for the OLS-VAR and RBFM-VAR estimators of Π_1 and Π_2 for Cases A–C when $T = 150$. The results from the simulations with $T = 500$ are similar to those from the simulations with $T = 150$ and thus are not reported. Figures 1–3 present the density estimates for the OLS-VAR and the RBFM-VAR estimates. Here we only report the results for Π_1 and for the sample size $T = 150$. The results for Π_2 and for the simulations with $T = 500$ do not provide much additional information. Each figure has a set of four density estimates for the individual coefficients of $\Pi_1 = ({}_1\pi_{ij})$, $i, j = 1,2$. Table 2 reports for Cases A–C the finite sample sizes and rejection probabilities of the standard Wald test W_F constructed from the OLS-VAR estimates and the modified Wald test W_F^+ based on the RBFM-VAR estimators defined in (20).

As one can see from Table 1 and Figures 1–3, the RBFM-VAR estimators generally perform better in finite samples than the OLS-VAR estimators in terms of both biases and variances. The former have smaller biases and variances than the latter in most cases. This, however, is not so for every case. There are a few cases where the OLS-VAR estimators outperform the RBFM-VAR counterparts. This is indeed expected from our theory. There are stationary components in the model, for which no correction is needed. For the coefficients of the stationary components, the OLS-VAR estimators are efficient, and our method introduces unnecessary correction terms. The unnecessary correction would incur additional finite sample biases and variations. Though we do not report the details to save space, these additional biases and variations disappear as the sample size increases.

The finite sample sizes of the modified Wald test W_F^+ constructed from the RBFM-VAR estimator are relatively much closer to the nominal sizes. As can be seen from Table 2, the standard Wald test W_F based on the OLS-VAR estimator has serious size distortions for both Cases A and B. Worse, this problem appears to persist even for large samples. The size distortions of the standard Wald test are enormous even when the sample size is as large as 500. The null of noncausality would therefore be overrejected significantly if based on the standard Wald test. For Case C, the reported numbers are the rejection probabilities for the modified and standard Wald tests. They are smaller for the modified Wald test, compared to the standard Wald test. As a result, the rejection of the null hypothesis is more likely if one uses the modified Wald test.

6. CONCLUSION

The RBFM-VAR procedure we proposed in the paper can be used to statistically analyze VAR models without specifying nonstationary characteristics of the model. In particular, we allow for the presence of $I(2)$ variables and coin-

Table 1. Finite sample biases and standard deviations

$T = 150$	Case	${}_1\pi_{11}$	${}_1\pi_{12}$	${}_1\pi_{21}$	${}_1\pi_{22}$	${}_2\pi_{11}$	${}_2\pi_{12}$	${}_2\pi_{21}$	${}_2\pi_{22}$
				OLS-VAR estimators					
Bias	A	−.72279	.00045	−.00360	−.71256	−.00648	.00017	−.00020	−.00659
	B	−.26780	.04559	−.03925	−.43318	−.26286	−.00007	−.00258	−.00032
	C	.07918	.14560	−.04801	−.35663	−.33216	.00078	.09794	.00033
s.d.	A	.61219	.58508	.59164	.60788	.02488	.02550	.02577	.02475
	B	.90883	.48336	.91076	.48995	.26825	.00922	.26363	.00887
	C	.92990	.61381	.94687	.65698	.66624	.00462	.71439	.00582
				RBFM-VAR estimators					
Bias	A	−.46002	−.00253	−.00402	−.45717	−.00230	.00003	−.00009	−.00243
	B	−.22124	−.03669	−.06553	−.26108	−.08025	−.00007	.00206	−.00014
	C	−.06287	−.43333	−.01431	−.21680	.41416	.00039	.03557	.00008
s.d.	A	.49459	.47492	.47206	.49179	.01283	.01262	.01336	.01253
	B	.99153	.34266	1.0031	.36156	.14827	.00416	.14247	.00403
	C	.96448	.65763	.98883	.62593	.68866	.00284	.63427	.00228

Note: The actual numbers reported are scaled by $\sqrt{T}$ for both biases and standard deviations.

Table 2. Finite sample sizes and rejection probabilities

Case	$T = 150$	1% test	5% test	10% test	$T = 500$	1% test	5% test	10% test
A	W_F	0.195	0.404	0.529	W_F	0.197	0.403	0.518
	W_F^+	0.031	0.090	0.150	W_F^+	0.031	0.084	0.130
B	W_F	0.105	0.274	0.395	W_F	0.090	0.255	0.378
	W_F^+	0.011	0.044	0.080	W_F^+	0.015	0.045	0.077
C	W_F	0.761	0.902	0.947	W_F	1.000	1.000	1.000
	W_F^+	0.317	0.524	0.628	W_F^+	0.979	0.994	0.998

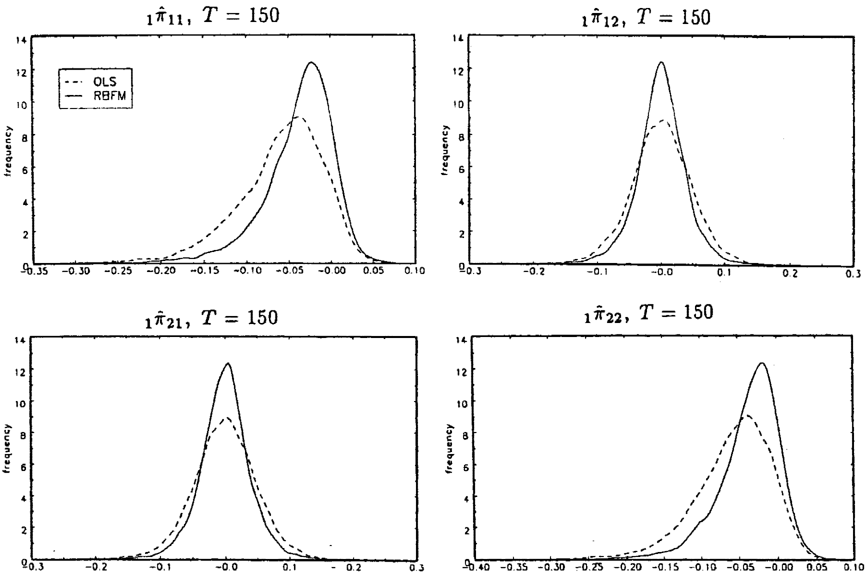

FIGURE 1. Densities of OLS-VAR and RBFM-VAR estimates for Case A

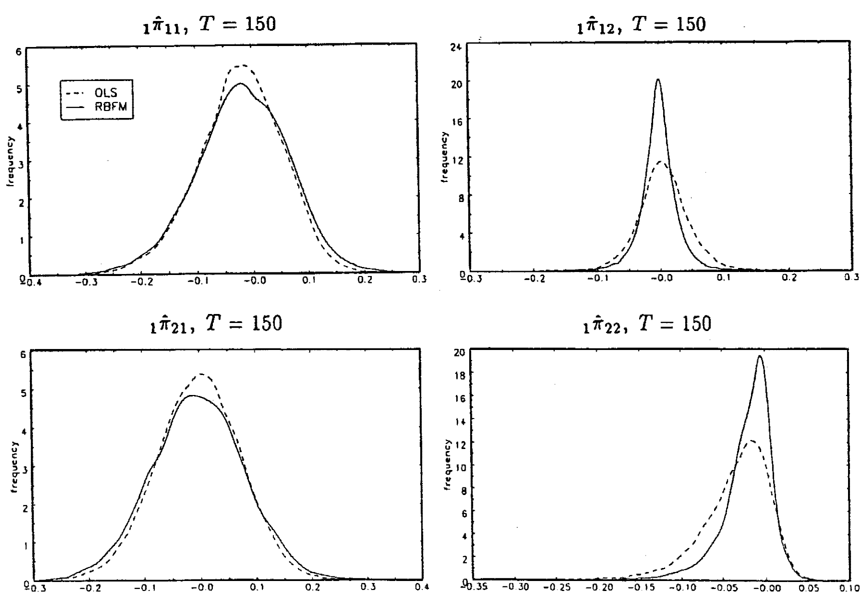

FIGURE 2. Densities of OLS-VAR and RBFM-VAR estimates for Case B

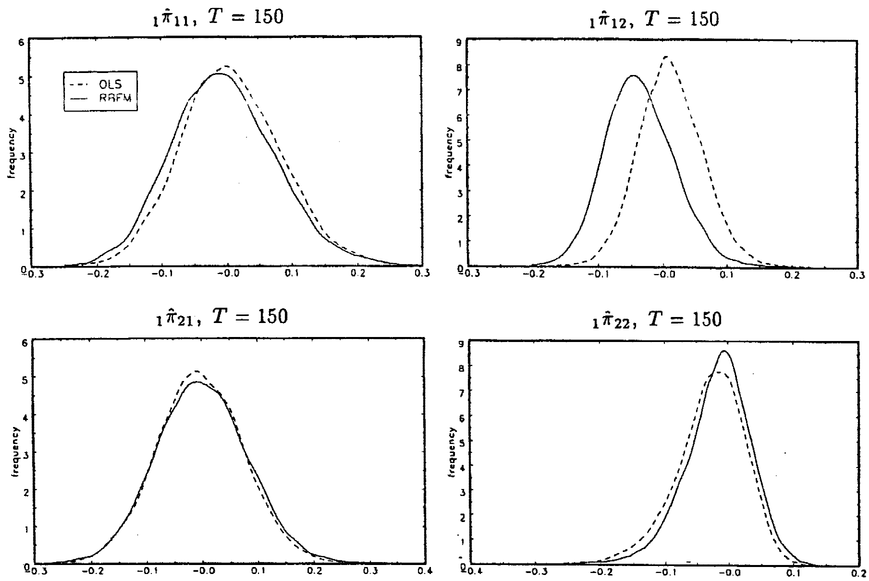

FIGURE 3. Densities of OLS-VAR and RBFM-VAR estimates for Case C

tegrations of the form $CI(1,1)$, $CI(2,2)$, and $CI(2,1)$ and for multicointegration in the VAR systems. The asymptotic theory established in the paper shows, however, that the RBFM-VAR estimator is consistent and that its leading term has mixed normal limit distribution. This is achieved without the specification of the nonstationary characteristics of the regressors and the precise configurations of cointegration space.

The mixed normal limit distributions of the RBFM-VAR estimates simplify statistical inference in cointegrated $I(2)$-VAR's. Wald tests that are based on the RBFM-VAR estimator are shown to have a limit theory that involves a linear combination of independent chi-square variates. The limit distribution is bounded above by the usual chi-square distribution with degrees of freedom equal to the number of restrictions being tested. Thus, the conventional critical values can be used to construct asymptotically valid, but conservative, tests in our RBFM-VAR regressions. This has a direct application for Granger-causality tests in nonstationary VAR models.

NOTES

1. Note that the long run variance Ω is singular. To see this, we may write, as in Johansen (1995), $u_{2t} = C(L)\varepsilon_t$ for an $(m_2 \times n)$ infinite matrix lag polynomial $C(L)$. Then it follows that $B_2 = C(1)B_\varepsilon$. Also note that $B_3 = (I_{m_3}, 0)B_2$, which can be seen easily from the definitions given in (4) and (5).

2. Our results here are not comparable to the limit theories in Johansen (1997) because he uses a different parameterization there.

REFERENCES

Ahn, S.K. & G.C. Reinsel (1988) Nested reduced rank autoregressive models for multiple time series. *Journal of the American Statistical Association* 83, 849–856.

Chang, Y. & P.C.B. Phillips (1995) Time series regression with mixtures of integrated processes. *Econometric Theory* 11, 1033–1094.

Choi, I. (1993) Asymptotic normality of least squares estimators for higher order autoregressive integrated processes with some applications. *Econometric Theory* 9, 236–282.

Engle, R.F. & B.S. Yoo (1991) Cointegrated economic time series: A survey with new results. In: W.J. Granger & R.F. Engle (eds.), *Long-Run Economic Relations: Readings in Cointegration*, pp. 237–266. Oxford: Oxford University Press.

Hannan, E.J. (1970) *Multiple Time Series*. New York: Wiley.

Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.

Johansen, S. (1995) A statistical analysis of cointegration for $I(2)$ variables. *Econometric Theory* 11, 25–59.

Johansen, S. (1997) Likelihood analysis of the $I(2)$ model. *Scandinavian Journal of Statistics* 24, 433–462.

Kitamura, Y. (1995) Estimation of cointegrated systems with $I(2)$ processes. *Econometric Theory* 11, 1–24.

Park, J.Y. & P.C.B. Phillips (1989) Statistical inference in regressions with integrated processes, part 2. *Econometric Theory* 5, 95–131.

Phillips, P.C.B. (1991) Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.

Phillips, P.C.B. (1995) Fully modified least squares and vector autoregression. *Econometrica* 63, 1023–1078.

Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regressions with $I(1)$ processes. *Review of Economic Studies* 57, 99–125.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

Priestley, M.B. (1981) *Spectral Analysis and Time Series*, vols I. and II. New York: Academic Press.

Sims, C.A., J.H. Stock, & M.W. Watson (1990) Inference in linear time series models with some unit roots. *Econometrica* 58, 113–144.

Toda, H. & P.C.B. Phillips (1993) Vector autoregressions and causality. *Econometrica* 61, 1367–1393.

Toda, H. & P.C.B. Phillips (1994) Vector autoregressions and causality: A theoretical overview and simulation study. *Econometric Reviews* 13, 259–285.

APPENDIX: PROOFS

Proof of Theorem 1. As shown in Remark (d) following Theorem 1, the process $\hat{v}_t$ defined in (11) that we use as the basis for our correction terms has the same representation as the residual employed in the RBFM-OLS procedure. (See Chang and Phillips, 1995, equation (13), p. 1044). Moreover, we admit the same class of kernel functions $\omega(\cdot)$ used in forming long run covariance matrix estimates and characterize rates of expansion of the bandwidth parameter $K = K(T)$ by using the expansion rate symbol O_e in the same manner as in Chang and Phillips (1995). Therefore, the asymptotic results established in that reference are directly applicable to our present analysis for explicitly characterizing the limit behaviors of the kernel estimates of the long run and one-sided long run covariance matrices used in the construction of the RBFM-VAR correction terms given in (13).

We begin by rewriting the RBFM-VAR estimator $\hat{F}^+$ defined in (12) as

$$\begin{aligned}\hat{F}^+ &= (Y'Z, Y'W - \hat{\Omega}_{\hat{\varepsilon}\hat{v}}\hat{\Omega}_{\hat{v}\hat{v}}^{-1}(\hat{V}'W - T\hat{\Delta}_{\hat{v}\triangle w}))(X'X)^{-1}\\ &= (FX'X + E'(Z,W) - (0, \hat{\Omega}_{\hat{\varepsilon}\hat{v}}\hat{\Omega}_{\hat{v}\hat{v}}^{-1}(\hat{V}'W - T\hat{\Delta}_{\hat{v}\triangle w})))(X'X)^{-1}\end{aligned}$$

because $X = (Z,W)$. Then the estimation error in $\hat{F}^+$ follows as

$$\hat{F}^+ - F = (E'Z, E'W - \hat{\Omega}_{\hat{\varepsilon}\hat{v}}\hat{\Omega}_{\hat{v}\hat{v}}^{-1}(\hat{V}'W - T\hat{\Delta}_{\hat{v}\triangle w}))(X'X)^{-1},$$

which can be written more explicitly as

$$\begin{aligned}&(E'Z, E'W - \hat{\Omega}_{\hat{\varepsilon}\hat{v}}H'(H\hat{\Omega}_{\hat{v}\hat{v}}H')^{-1}H(\hat{V}'W - T\hat{\Delta}_{\hat{v}\triangle w}))G'(GX'XG')^{-1}G\\ &\quad = (E'Z, E'WH' - \hat{\Omega}_{\hat{\varepsilon}\hat{v}_h}\hat{\Omega}_{\hat{v}_h\hat{v}_h}^{-1}(\hat{V}_h'W - T\hat{\Delta}_{\hat{v}_h\triangle w})H')(GX'XG')^{-1}G\\ &\quad = [E'X_1 - (0, \hat{\Omega}_{\hat{\varepsilon}\hat{v}_h}\hat{\Omega}_{\hat{v}_h\hat{v}_h}^{-1}(\hat{V}_h'U_1 - T\hat{\Delta}_{\hat{v}_h\triangle u_1}))|\\ &\qquad \times E'X_b - \hat{\Omega}_{\hat{\varepsilon}\hat{v}_h}\hat{\Omega}_{\hat{v}_h\hat{v}_h}^{-1}(\hat{V}_h'X_b - T\hat{\Delta}_{\hat{v}_h\triangle x_b})](GX'XG')^{-1}G \qquad \textbf{(A.1)}\end{aligned}$$

because $W_1 = U_1$, $(Z,W_1) = X_1$, and $W_b = X_b$ by the definitions (5) and (7). We use $\hat{v}_h$ and $\hat{V}_h$ to denote the H-transformed $\hat{v}$ and $\hat{V}$. We have $XG' = (X_1, X_b)$ and

$$\begin{aligned}(GX'XG')^{-1} &= \begin{pmatrix} X_1'X_1' & X_1'X_b \\ X_b'X_1 & X_b'X_b\end{pmatrix}^{-1}\\ &= \begin{pmatrix} (X_1'Q_bX_1)^{-1} & -(X_1'X_1)^{-1}X_1'X_b(X_b'Q_1X_b)^{-1} \\ -(X_b'X_b)^{-1}X_b'X_1(X_1'Q_bX_1)^{-1} & (X_b'Q_1X_b)^{-1}\end{pmatrix}\end{aligned} \qquad \textbf{(A.2)}$$

with $Q_i = I - X_i(X_i'X_i)^{-1}X_i'$ for $i = 1, b$.

It follows from Lemma 3(d) and (e) of Chang and Phillips (1995) that

$$\hat{\Omega}_{\hat{\varepsilon}\hat{v}_h} = \hat{\Omega}_{\varepsilon v_h} + o_p(1) \quad \text{and} \quad \hat{\Omega}_{\hat{v}_h\hat{v}_h} = \hat{\Omega}_{v_hv_h} + o_p(1) \qquad \textbf{(A.3)}$$

for the bandwidth parameter expansion rate $k \in (0, \frac{1}{2})$, and, thus, we can use $\hat{\Omega}_{\varepsilon v_h}$ and $\hat{\Omega}_{v_hv_h}$ in lieu of $\hat{\Omega}_{\hat{\varepsilon}\hat{v}_h}$ and $\hat{\Omega}_{\hat{v}_h\hat{v}_h}$ without affecting our later asymptotic analyses. We also have from Lemma 6(c) and (d) of Chang and Phillips (1995) that

$$T^{-1/2}\hat{\Omega}_{\varepsilon v_h}\hat{\Omega}_{v_hv_h}^{-1}(\hat{V}_h'U_1 - T\hat{\Delta}_{\hat{v}_h\triangle u_1}) = O_p(T^{1/2}K^{-2}) + O_p(KT^{-1/2}) + O_p(K^{-1}) \qquad \textbf{(A.4)}$$

and

$$\begin{aligned}&\hat{\Omega}_{\varepsilon v_h}\hat{\Omega}_{v_hv_h}^{-1}(\hat{V}_h'X_b - T\hat{\Delta}_{\hat{v}_h\triangle x_b})D_T^{-1}\\ &\quad = \Omega_{\varepsilon b}\Omega_{bb}^{+}\overline{N}_{bbT} + O_p(K^{5/2}T^{-3/2}) + O_p(K^{3/2}T^{-1}), \qquad \textbf{(A.5)}\end{aligned}$$

where $\bar{N}_{bbT} \to_d \int_0^1 dB_b \bar{B}_b'$. Notice that we use the Moore–Penrose inverse in the preceding discussion, because Ω_{bb} is singular in general as a result of the correlation between B_2 and B_3 discussed following (7).

Part (a). Notice that

$$GG^1 = \begin{pmatrix} I & 0 \\ 0 & H \end{pmatrix}\begin{pmatrix} I & 0 \\ 0 & H^1 \end{pmatrix} = \begin{pmatrix} I_{n(p-2)} & 0 \\ 0 & HH^1 \end{pmatrix} = \begin{pmatrix} I_{n(p-2)+m_1} \\ 0 \end{pmatrix}$$

because $HH^1 = (I_{m_1}, 0)'$, and thus GG^1 picks up the first column of $(GX'XG')^{-1}$, which is given in (A.2). Then it follows from (A.1)–(A.5) that

$$\begin{aligned}
&\sqrt{T}(\hat{F}^+ - F)G^1 \\
&\quad = \sqrt{T}(E'Z, E'W - \hat{\Omega}_{\hat{\varepsilon}\hat{v}_h}\hat{\Omega}_{\hat{v}_h\hat{v}_h}^{-1}(\hat{V}_h'W - T\hat{\Delta}_{\hat{v}_h\triangle w}))G'(GX'XG')^{-1}GG^1 \\
&\quad = \sqrt{T}(E'X_1 - (0, \hat{\Omega}_{\varepsilon v_h}\hat{\Omega}_{v_h v_h}^{-1}(\hat{V}_h'U_1 - \hat{\Delta}_{\hat{v}_h\triangle u_1})))(X_1'Q_bX_1)^{-1} \\
&\qquad - \sqrt{T}(E'X_b - \hat{\Omega}_{\varepsilon v_h}\hat{\Omega}_{v_h v_h}^{-1}(\hat{V}_h'X_b - T\hat{\Delta}_{\hat{v}_h\triangle x_b}))(X_b'X_b)^{-1}X_b'X_1(X_1'Q_bX_1)^{-1} \\
&\quad = \sqrt{T}(T^{-1}E'X_1 - (0, \hat{\Omega}_{\varepsilon v_h}\hat{\Omega}_{v_h v_h}^{-1}(T^{-1}\hat{V}_h'U_1 - \hat{\Delta}_{\hat{v}_h\triangle u_1})))(T^{-1}X_1'Q_bX_1)^{-1} \\
&\qquad - T^{-1/2}(E'X_bD_T^{-1} - \hat{\Omega}_{\varepsilon v_h}\hat{\Omega}_{v_h v_h}^{-1}(\hat{V}_h'X_b - T\hat{\Delta}_{\hat{v}_h\triangle x_b})D_T^{-1}) \\
&\qquad \times (D_T(X_b'X_b)^{-1}D_T)(D_T^{-1}X_b'X_1)(T^{-1}X_1'Q_bX_1)^{-1} \\
&\quad = T^{-1/2}E'X_1(T^{-1}X_1'X_1)^{-1} + O_p(T^{1/2}K^{-2}) + O_p(KT^{-1/2}) + O_p(K^{-1}).
\end{aligned}$$

The error terms appearing in the preceding expression are of order $o_p(1)$ for a bandwidth expansion rate $k \in (\frac{1}{4}, \frac{1}{2})$. Then it follows immediately from (8) that

$$\sqrt{T}(\hat{F}^+ - F)G^1 \to_d \mathbf{N}(0, \Sigma_{\varepsilon\varepsilon} \otimes \Sigma_{x11}^{-1}),$$

where $\Sigma_{x11} = \mathbf{E}x_{1t}x_{1t}'$, which is shown to be positive definite in Lemma 1(iii) of Toda and Phillips (1993), and this completes the proof.

Part (b). Similarly, it follows from $GG^b = (0, I_{m_b})'$ and (A.1)–(A.5) that

$$\begin{aligned}
&(\hat{F}^+ - F)G^bD_T \\
&\quad = (E'Z, E'W_h - \hat{\Omega}_{\hat{\varepsilon}\hat{v}_h}\hat{\Omega}_{\hat{v}_h\hat{v}_h}^{-1}(\hat{V}_h'W_h - T\hat{\Delta}_{\hat{v}_h\triangle w_h}))(G'X'XG)^{-1}G'G^bD_T \\
&\quad = -(T^{-1}E'X_1 - (0, \hat{\Omega}_{\varepsilon v_h}\hat{\Omega}_{v_h v_h}^{-1}(T^{-1}\hat{V}_h'U_1 - \hat{\Delta}_{\hat{v}_h\triangle u_1})))(T^{-1}X_1'X_1)^{-1} \\
&\qquad \times X_1'X_bD_T^{-1}D_T(X_b'Q_1X_b)^{-1}D_T \\
&\qquad + (E'X_b - \hat{\Omega}_{\varepsilon v_h}\hat{\Omega}_{v_h v_h}^{-1}(\hat{V}_h'X_b - T\hat{\Delta}_{\hat{v}_h\triangle x_b}))D_T^{-1}D_T(X_b'Q_1X_b)^{-1}D_T \\
&\quad = (\bar{N}_{\varepsilon bT} - \Omega_{\varepsilon b}\Omega_{bb}^+\bar{N}_{bbT})D_T(X_b'Q_1X_b)^{-1}D_T + O_p(K^{5/2}T^{-3/2}) + O_p(K^{3/2}T^{-1}),
\end{aligned}$$

where $\bar{N}_{\varepsilon bT} - \Omega_{\varepsilon b}\Omega_{bb}^+\bar{N}_{bbT} \to_d \int_0^1 d(B_\varepsilon - \Omega_{\varepsilon b}\Omega_{bb}^+B_b)\bar{B}_b' = \int_0^1 d(B_\varepsilon - \Omega_{\varepsilon 2}\Omega_{22}^{-1}B_2)\bar{B}_b'$. All the error terms in the equation are $o_p(1)$ for $k \in (0, \frac{3}{5})$, and therefore the stated follows for $k \in (0, \frac{1}{2}) \cap (0, \frac{3}{5}) = (0, \frac{1}{2})$.

Proof of Theorem 2. From

$$GR_2 = \begin{pmatrix} I & 0 \\ 0 & H \end{pmatrix} \left(\begin{array}{c|c} R_{2\Phi} & 0 \\ 0 & (H^1 S_{A1}, H^b S_{Ab}) \end{array} \right) = \begin{pmatrix} R_{2\Phi} & 0 & 0 \\ 0 & S_{A1} & 0 \\ 0 & 0 & S_{Ab} \end{pmatrix} \tag{A.6}$$

we have

$$\begin{aligned} R_1(\hat{F}^+ - F)R_2 &= R_1(\hat{F}^+ - F)G^{-1}GR_2 \\ &= \left(R_1(\hat{\Phi}^+ - \Phi, \hat{A}_1^+ - A_1) \begin{pmatrix} R_{2\Phi} & 0 \\ 0 & S_{A1} \end{pmatrix} \middle| R_1(\hat{A}_b^+ - A_b)S_{Ab} \right). \end{aligned} \tag{A.7}$$

Define $\bar{D}_T = \operatorname{diag}(I_{n(p-2)+m_1}, \sqrt{T} I_{m_2}, T^{3/2} I_{m_3})$. Then it follows from (A.6) that

$$\begin{aligned} &\bar{D}_T R_2' T(X'X)^{-1} R_2 \bar{D}_T \\ &\quad = \bar{D}_T R_2' G'(G')^{-1} T(X'X)^{-1} G^{-1} G R_2 \bar{D}_T \\ &\quad = R_2' G' \bar{D}_T \sqrt{T} (GX'XG')^{-1} \sqrt{T} \bar{D}_T G R_2 \\ &\quad = R_2' G' \begin{pmatrix} \sqrt{T} & 0 \\ 0 & D_T \end{pmatrix} \begin{pmatrix} X_1' X_1' & X_1' X_b \\ X_b' X_1 & X_b' X_b \end{pmatrix}^{-1} \begin{pmatrix} \sqrt{T} & 0 \\ 0 & D_T \end{pmatrix} G R_2 \\ &\quad \to_d \begin{pmatrix} \begin{matrix} R_{2\Phi}' & 0 \\ 0 & S_{A1}' \end{matrix} & 0 \\ 0 & S_{Ab}' \end{pmatrix} \begin{pmatrix} \Sigma_{x11} & 0 \\ 0 & \int_0^1 \bar{B}_b \bar{B}_b' \end{pmatrix}^{-1} \begin{pmatrix} \begin{matrix} R_{2\Phi} & 0 \\ 0 & S_{A1} \end{matrix} & 0 \\ 0 & S_{Ab} \end{pmatrix} \\ &\quad = \begin{pmatrix} \begin{pmatrix} R_{2\Phi}' & 0 \\ 0 & S_{A1}' \end{pmatrix} \Sigma_{x11}^{-1} \begin{pmatrix} R_{2\Phi} & 0 \\ 0 & S_{A1} \end{pmatrix} & 0 \\ 0 & S_{Ab}' \left(\int_0^1 \bar{B}_b \bar{B}_b' \right)^{-1} S_{Ab} \end{pmatrix} =: \begin{pmatrix} \mathcal{Q}_1 & 0 \\ 0 & \mathcal{Q}_2 \end{pmatrix}. \end{aligned} \tag{A.8}$$

It follows also from (A.7) and the results in Theorem 1 that

$$\begin{aligned} &\operatorname{vec}(\sqrt{T} R_1 (\hat{F}^+ - F) R_2 \bar{D}_T) \\ &\quad = \operatorname{vec}\left(\sqrt{T} R_1 (\hat{F}_1^+ - F_1) \begin{pmatrix} R_{2\Phi} & 0 \\ 0 & S_{A1} \end{pmatrix}, R_1 (\hat{F}_b^+ - F_b) D_T S_{Ab} \right) \to_d (\mathcal{Z}_1, \mathcal{Z}_b), \end{aligned}$$

where

$$\mathcal{Z}_1 = \mathbf{N}(0, R_1 \Sigma_{\varepsilon\varepsilon} R_1' \otimes \mathcal{Q}_1) \quad \text{and} \quad \mathcal{Z}_b = \mathbf{MN}(0, R_1 \Omega_{\varepsilon\varepsilon\cdot 2} R_1' \otimes \mathcal{Q}_b) \tag{A.9}$$

using the notations $\mathcal{Q}_1$ and $\mathcal{Q}_b$ defined in (A.8).

We now consider the asymptotics for the Wald statistic when the restriction matrix R has the form $R = R_1 \otimes R_2'$. The Wald statistic W_F^+ in this case is obtained from the

Wald statistic given in (20) simply by replacing R and r by $R_1 \otimes R_2'$ and $\mathrm{vec}(R_1FR_2)$, respectively, namely,

$$T(\mathrm{vec}\, R_1(\hat{F}^+ - F)R_2)'((R_1\hat{\Sigma}_{\varepsilon\varepsilon}R_1')^{-1} \otimes (R_2'T(X'X)^{-1}R_2)^{-1})\mathrm{vec}\, R_1(\hat{F}^+ - F)R_2$$
$$= \mathrm{tr}((R_1\hat{\Sigma}_{\varepsilon\varepsilon}R_1')^{-1}\sqrt{T}R_1(\hat{F}^+ - F)R_2\bar{D}_T$$
$$\times (\bar{D}_T R_2'T(X'X)^{-1}R_2\bar{D}_T)^{-1}(\sqrt{T}R_1(\hat{F}^+ - F)R_2\bar{D}_T)').$$

Then it follows directly from (A.8) and (A.9) that

$$W_F^+ \to_d \mathrm{tr}(R_1\Sigma_{\varepsilon\varepsilon}R_1')^{-1}\mathcal{Z}_1\mathcal{Q}_1^{-1}\mathcal{Z}_1' + \mathrm{tr}(R_1\Sigma_{\varepsilon\varepsilon}R_1')^{-1}\mathcal{Z}_b\mathcal{Q}_b^{-1}\mathcal{Z}_b' =: (\mathcal{W}_1 + \mathcal{W}_b). \quad \textbf{(A.10)}$$

Define $\mathcal{Z}_1^* = (R_1\Sigma_{\varepsilon\varepsilon}R_1')^{-1/2}\mathcal{Z}_1\mathcal{Q}_1^{-1/2}$, and use this to write $\mathcal{W}_1$ as

$$\mathcal{W}_1 = \mathrm{tr}(\mathcal{Z}_1^{*\prime}\mathcal{Z}_1^*) = (\mathrm{vec}(\mathcal{Z}_1^*))'\mathrm{vec}(\mathcal{Z}_1^*) \equiv \chi^2_{q_1(q_\Phi+q_{A1})} \quad \textbf{(A.11)}$$

because $\mathrm{vec}(\mathcal{Z}_1^*) = ((R_1\Sigma_{\varepsilon\varepsilon}R_1')^{-1/2} \otimes \mathcal{Q}_1^{-1/2})\mathrm{vec}(\mathcal{Z}_1) \equiv \mathbf{N}(0, I_{q_1(q_\Phi+q_{A1})})$.

To analyze the second term $\mathcal{W}_b$ of (A.10), we need to deal with the potential singularity of the variance matrix of $\mathcal{Z}_b$ defined in (A.9) that arises from the singularity of $\Omega_{\varepsilon\varepsilon\cdot 2}$. To do so, we use a $(q_1 \times q_1)$ orthogonal matrix $K = (K_1, K_2)$ whose component matrices K_1 and K_2 are of ranks q_1^* and $q_1 - q_1^*$, respectively. We may write $R_1\Omega_{\varepsilon\varepsilon\cdot 2}R_1' = K_1\Lambda K_1'$, where Λ is a q_1^*-dimensional diagonal matrix. Define $M = (K_1\Lambda^{1/2}, K_2)$ and use its inverse $M^{-1} = (K_1\Lambda^{-1/2}, K_2)'$ to transform $\mathcal{Z}_b$ as $\mathcal{Z}_b^* = M^{-1}\mathcal{Z}_b\mathcal{Q}_b^{-1/2}$, where

$$\mathrm{vec}(\mathcal{Z}_b^*) = (M^{-1} \otimes \mathcal{Q}_b^{-1/2})\mathrm{vec}(\mathcal{Z}_b) \equiv \mathbf{MN}\left(0, \begin{pmatrix} I_{q_1^*} & 0 \\ 0 & 0 \end{pmatrix} \otimes I_{q_{A_b}}\right).$$

Define $M_* = (K_1\Lambda^{1/2}, 0)$. We may then write $\mathcal{W}_b$ as

$$\mathcal{W}_b = \mathrm{tr}(M'(R_1\Sigma_{\varepsilon\varepsilon}R_1')^{-1}M\mathcal{Z}_b^*\mathcal{Z}_b^{*\prime}) = \mathrm{tr}(M_*'(R_1\Sigma_{\varepsilon\varepsilon}R_1')^{-1}M_*\mathcal{Z}_b^*\mathcal{Z}_b^{*\prime})$$

and let C be a q_1-dimensional orthogonal matrix such that $C'C = I_{q_1}$, for which

$$C'M_*'(R_1\Sigma_{\varepsilon\varepsilon}R_1')^{-1}M_*C = D = \mathrm{diag}(d_1, \ldots, d_{q_1}),$$

where $(q_1 - q_1^*)$ number of d_i's are zero. Note that the d_i's are eigenvalues of $(R_1\Omega_{\varepsilon\varepsilon\cdot 2}R_1')^{1/2}(R_1\Sigma_{\varepsilon\varepsilon}R_1')^{-1}(R_1\Omega_{\varepsilon\varepsilon\cdot 2}R_1')^{1/2}$, because $(R_1\Omega_{\varepsilon\varepsilon\cdot 2}R_1')^{1/2} = M_*K'$. We finally define $\bar{\mathcal{Z}}_b^* = C'\mathcal{Z}_b^*$. Then it follows that

$$\mathcal{W}_b = \mathrm{tr}(D\bar{\mathcal{Z}}_b^*\bar{\mathcal{Z}}_b^{*\prime}) = \sum_{i=1}^{q_1} d_i \sum_{j=1}^{q_{Ab}} (\chi_1^2)_i = \sum_{i=1}^{q_1} d_i\,\chi^2_{q_{Ab}}(i), \quad \textbf{(A.12)}$$

where $\chi^2_{q_{Ab}}(i) \sim$ i.i.d. $(\chi^2_{q_{Ab}})$, for $i = 1, \ldots, q_1$, because $\mathrm{vec}(\bar{\mathcal{Z}}_b^*) = (C' \otimes I)\mathrm{vec}(\mathcal{Z}_b^*)$. The stated result now follows immediately from (A.10)–(A.12). ∎