Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Econometrics 135 (2006) 427–463
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Modified two-stage least-squares estimators for the estimation of a structural vector autoregressive integrated process

Cheng Hsiao[a,c], Siyan Wang[b,*]

[a]*Department of Economics, University of Southern California, Los Angeles, CA 90089-0253, USA*
[b]*Department of Economics, University of Delaware, Newark, DE 19716, USA*
[c]*The Wang Yanan Institute for Studies in Economics, Xiamen University, China*

Available online 24 August 2005

Abstract

We consider the estimation of a structural vector autoregressive model of nonstationary and possibly cointegrated variables without the prior knowledge of unit roots or rank of cointegration. We propose two modified two-stage least-squares estimators that are consistent and have limiting distributions that are either normal or mixed normal. Limited Monte Carlo studies are also conducted to evaluate their finite sample properties.

JEL classification: C32; C12; C13

Keywords: Structural vector autoregression; Unit root; Cointegration; Asymptotic properties; Hypothesis testing

1. Introduction

We consider the estimation of an equation in a structural vector autoregressive model (SVAR) involving integrated and possibly cointegrated variables without the prior knowledge of the location of unit roots or rank of cointegration. Although the

*Corresponding author. Tel.: +1 302 831 1924; fax: +1 302 831 6968.
E-mail address: wangs@udel.edu (S. Wang).

doi:10.1016/j.jeconom.2005.07.019

location of unit roots or rank of cointegration can provide information for identification and may improve the efficiency of the estimates, many econometric models are identified without prior information on this. For instance, the Klein–Goldberger (Klein et al., 1955) and the large-scale Wharton quarterly models (Klein and Evans, 1969) are identified through exclusion restrictions.

The SVAR we consider is different from the reduced-form VAR considered by Johansen (1988, 1991), Phillips (1995), or Sims et al. (1990) in that we allow more than one current variables to appear in each equation. The model is similar in spirit to the Cowles Commission structural equation specification in which each equation describes a behavioral or technological relation except that no strict exogeneity assumption has been imposed on some of the variables as in Hsiao (1997a, b). It is shown by Hsiao and Wang (2004) that an identified equation in such a system may be consistently estimated by the conventional two-stage or three-stage least-squares estimator (2SLS or 3SLS). However, their limiting distributions may be non-standard, hence a chi-square distribution may not approximate well the limiting distribution of a conventional Wald test statistic. In this paper we propose two modified estimators that are either asymptotically normally or mixed normally distributed, thus allow the construction of a Wald-type test statistic that is asymptotically chi-square distributed.

We set up the basic model in Section 2. We propose a modified two-stage least-squares estimator (M2SLS) in Section 3 and an alternatively modified two-stage least-squares estimator (A2SLS) in Section 4. Section 5 extends the discussion by adding an intercept term to the basic model. Section 6 provides some Monte Carlo studies comparing the performance of 2SLS, M2SLS, and A2SLS. Conclusions are in Section 7.

2. The model

Let $\underset{\sim t}{w}$ be an $m \times 1$ vector of random variables that can be represented by the following pth order autoregressive model:[1]

$$A(L)\underset{\sim t}{w} = \underset{\sim t}{\varepsilon}, \quad t = 1, \ldots, T, \tag{2.1}$$

where $A(L) = A_0 + A_1 L + \cdots + A_p L^p$ is a pth order matrix polynomial of the lag operator L. We assume that

A1 : A_0 is nonsingular.
A2 : The roots of $|A(L)| = 0$ are either 1 or outside the unit circle.
A3 : The $m \times 1$ error vector $\underset{\sim t}{\varepsilon}$ is independently, identically distributed (i.i.d.) with zero mean, nonsingular covariance matrix $\Sigma_{\varepsilon\varepsilon}$ and finite fourth cumulants.

[1]For ease of notations, we postulate (2.1) without the intercept term. The basic conclusions of this paper remain unchanged with the addition of intercept term in (2.1), see Section 5.

Since we are interested in the asymptotic properties of the estimators of (2.1), for ease of exposition, we shall also assume that the initial values, $\underset{\sim}{w}_0, \underset{\sim}{w}_{-1}, \ldots, \underset{\sim}{w}_{-p+1}$ are given.

Remark 2.1. Assumption A1 is needed to ensure that (2.1) contains m linearly independent behavioral equations. The purpose of A2 is to relax the stationary assumption implicitly assumed in the original Cowles Commission framework to allow for the presence of $I(1)$ variables. A3 is a standard assumption for VAR models. The existence of fourth moments is made to ensure that (functional) central limit theorem will hold in deriving the limiting distributions of the proposed estimators.

Let $A = [A_0, A_1, \ldots, A_p]$ and define a $(p+1)m$ dimensional nonsingular matrix $\tilde{M}$ as

$$\tilde{M} = \begin{bmatrix} I_m & I_m & \ldots & I_m \\ \underset{\sim}{0} & I_m & \ldots & I_m \\ \underset{\sim}{0} & \underset{\sim}{0} & \ldots & I_m \\ \ldots & \ldots & \ldots & \\ \underset{\sim}{0} & \ldots & \underset{\sim}{0} & I_m \end{bmatrix}. \tag{2.2}$$

Postmultiplying A by the matrix $\tilde{M}$, we obtain an error-correction representation of (2.1),

$$\sum_{j=0}^{p-1} A_j^* \nabla \underset{\sim}{w}_{t-j} + A_p^* \underset{\sim}{w}_{t-p} = \underset{\sim}{\varepsilon}_t, \tag{2.3}$$

where $\nabla = (1-L), A_j^* = \sum_{\ell=0}^{j} A_\ell, j = 0, 1, \ldots, p.$ Let $A^* = [A_0^*, \ldots, A_p^*] = [\tilde{A}_0^*, A_p^*]$, then $A^* = A\tilde{M}$. The coefficient matrices $\tilde{A}_1^*$ and A_p^* provide the implied short-run dynamics and long-run relations of the system (2.1) as defined in Hsiao (2001).

Model (2.1) is different from the conventional VAR model of Johansen (1988, 1991), Phillips (1995), Sims (1980), Sims et al. (1990), Tsay and Tiao (1990), etc. in that A_0 is not an m-rowed identity matrix I_m. In other words, more than one current variables can appear in an equation. It can be viewed as a Cowles Commission structural equation model without the strict exogeneity assumption on some elements of $\underset{\sim}{w}_t$ (e.g. Koopmans et al., 1950; Hsiao, 1997a). Multiplying A_0^{-1} to (2.1) yields the conventional VAR which may be viewed as a reduced-form representation of (2.1),

$$\underset{\sim}{w}_t = \Pi_1 \underset{\sim}{w}_{t-1} + \cdots + \Pi_p \underset{\sim}{w}_{t-p} + \underset{\sim}{v}_t, \tag{2.4}$$

where $\Pi_j = -A_0^{-1} A_j, \underset{\sim}{v}_t = A_0^{-1} \underset{\sim}{\varepsilon}_t$.

We shall assume that at least one root of $|A(L)| = 0$ is equal to 1. More specifically,

A4 :

(a) $A_p^* = \underset{\sim}{\alpha} \underset{\sim}{\beta}'$ where $\underset{\sim}{\alpha}$ and $\underset{\sim}{\beta}$ are $m \times r$ matrices of full column rank r, $0 \leqslant r \leqslant m-1$;

(b) $\underset{\sim\perp}{\alpha}' J \underset{\sim\perp}{\beta}$ is nonsingular, where $J = \sum_{j=0}^{p-1} A_j^*, \underset{\sim\perp}{\alpha}$ and $\underset{\sim\perp}{\beta}$ are $m \times (m-r)$ matrices of full column rank such that $\underset{\sim\perp}{\alpha}' \underset{\sim}{\alpha} = \underset{\sim}{0} = \underset{\sim\perp}{\beta}' \underset{\sim}{\beta}$. (If $r = 0$, then we take $\underset{\sim\perp}{\alpha} = I_m = \underset{\sim\perp}{\beta}$).

Under A1–A4, $\underset{\sim t}{w}$ has r cointegrating vectors (the columns of β) and $m - r$ unit roots. As shown by Johansen (1988, 1991) and Toda and Phillips (1993) A4 ensures that the Granger representation theorem (Engle and Granger, 1987) applies, so that $\nabla \underset{\sim t}{w}$ is stationary, $\underset{\sim}{\beta}' \underset{\sim t}{w}$ is stationary, and $\underset{\sim t}{w}$ is an $I(1)$ process when $r < m$.

Suppose that the gth equation of (2.1) satisfies the prior restrictions $\underset{\sim g}{a}' \Phi_g = \underset{\sim}{0}'$, where $\underset{\sim g}{a}'$ denotes the gth row of A and Φ_g denotes a $(p+1)m \times R_g$ matrix with known elements. Let $\Phi_g^* = \tilde{M}^{-1} \Phi_g$, the existence of prior restrictions $\underset{\sim g}{a}' \Phi_g = \underset{\sim}{0}'$ is equivalent to the existence of prior restrictions $\underset{\sim g}{a}^{*\prime} \Phi_g^* = \underset{\sim}{0}'$, where $\underset{\sim g}{a}^{*\prime}$ is the gth row of A^*. Hsiao (2001) proved the following lemma.

Lemma 2.1. *Suppose that the gth equation of* (2.1) *is subject to the prior restrictions* $\underset{\sim g}{a}' \Phi_g = \underset{\sim}{0}'$. *A necessary and sufficient condition for the identification of the gth equation of* (2.1) *or* (2.2) *is that*

$$\operatorname{rank}(A\Phi_g) = m - 1, \tag{2.5}$$

or

$$\operatorname{rank}(A^* \Phi_g^*) = m - 1. \tag{2.6}$$

Remark 2.2. The identification condition (2.5) or (2.6) does not require the prior information about the existence or location of unit roots or rank of cointegration.

3. The modified two stage least-squares estimator

For ease of exposition, we assume that prior information is in the form of excluding certain variables, both current and lagged, from an equation. Let the gth equation of (2.1) be written as

$$\underset{\sim g}{w} = Z_g \underset{\sim g}{\delta} + \underset{\sim g}{\varepsilon}, \tag{3.1}$$

where $\underset{\sim g}{w}$ and $\underset{\sim g}{\varepsilon}$ denote the $T \times 1$ vectors of $(w_{g1}, \ldots, w_{gT})'$ and $(\varepsilon_{g1}, \ldots, \varepsilon_{gT})'$, respectively, and Z_g denotes the included current and lagged variables of $\underset{\sim t}{w}$. Let $X = (W_{-1}, W_{-2}, \ldots, W_{-p})$. The 2SLS estimator of $\underset{\sim g}{\delta}$ is given by

$$\underset{\sim g,\text{2SLS}}{\hat{\delta}} = [Z_g' X (X'X)^{-1} X' Z_g]^{-1} [Z_g' X (X'X)^{-1} X' \underset{\sim g}{w}]. \tag{3.2}$$

To derive the limiting distribution of 2SLS estimator, we let M_g be the nonsingular transformation matrix that transforms Z_g into $Z_g^* = Z_g M_g = (Z_{g1}^*, Z_{g2}^*)$, where Z_{g1}^* denotes the ℓ_g-dimensional linearly independent $I(0)$ variables and Z_{g2}^* denotes the T observations of b_g linearly independent $I(1)$ variables, then

$$
\begin{aligned}
\underset{\sim g}{w} &= Z_g M_g M_g^{-1} \underset{\sim g}{\delta} + \underset{\sim g}{\varepsilon} \\
&= Z_g^* \underset{\sim g}{\delta^*} + \underset{\sim g}{\varepsilon},
\end{aligned} \tag{3.3}
$$

where $\underset{\sim g}{\delta^*} = M_g^{-1} \underset{\sim g}{\delta} = (\underset{\sim g1}{\delta^{*\prime}}, \underset{\sim g2}{\delta^{*\prime}})'$ with $\underset{\sim g1}{\delta^*}$ and $\underset{\sim g2}{\delta^*}$ denoting the $\ell_g \times 1$ and $b_g \times 1$ vector, respectively. Such transformation always exists. For instance, if no cointegration relation exists among the g_Δ included variables, say $\underset{\sim gt}{\tilde{w}}$, then Z_{g1}^* consists of the first-differenced current and $p-1$ lagged included variables, Z_{g2}^* is simply the $T \times g_\Delta$ included $\underset{\sim gt}{\tilde{w}}$ lagged by p periods, $\underset{\sim g,t-p}{\tilde{w}}$. Suppose there exist $g_\Delta - b_g$ linearly independent cointegrating relations among the g_Δ included variables, $\underset{\sim gt}{\tilde{w}}$, then Z_{g1}^* consists of the current and $p-1$ lagged $\nabla \underset{\sim g}{\tilde{w}}$ and $\tilde{W}_{g1,-p} - \tilde{W}_{g2,-p} \underset{\sim g}{\pi}$, where $\tilde{W}_{g1,-p}$ is $T\times$ $(g_\Delta - b_g)$, $\tilde{W}_{g2,-p}$ is $T \times b_g$, $\underset{\sim g}{\pi}$ is $b_g \times (g_\Delta - b_g)$ of constants, and Z_{g2}^* consists of the T observed b_g linearly independent $I(1)$ variables $\tilde{W}_{g2,-p}$.

Let M_x be a nonsingular transformation matrix such that $XM_x = (X_1^*, X_2^*)$, where X_1^* consists of the linearly independent $I(0)$ variables and X_2^* consists of the linearly independent $I(1)$ variables, say dimension b. It is shown by Hsiao and Wang (2004) that

Lemma 3.1. *The* 2*SLS estimate of* $\underset{\sim g}{\delta^*}$ *is consistent and*

$$
\sqrt{T}(\underset{\sim g1,2\text{SLS}}{\hat{\delta}^*} - \underset{\sim g1}{\delta^*}) \Longrightarrow \mathrm{N}(\underset{\sim}{0}, \sigma_g^2 (M_{z_{g1}x_1}^* M_{x_1x_1}^{*-1} M_{x_1z_{g1}}^*)^{-1}), \tag{3.4}
$$

$$
\begin{aligned}
T(\underset{\sim g2,2\text{SLS}}{\hat{\delta}^*} - \underset{\sim g2}{\delta^*}) \Longrightarrow & \left\{ \int B_{z_{g2}^*} B'_{x_2^*} \,\mathrm{d}r \left(\int B_{x_2^*} B'_{x_2^*} \,\mathrm{d}r \right)^{-1} \int B_{x_2^*} B'_{z_{g2}^*} \,\mathrm{d}r \right\}^{-1} \\
& \times \left\{ \int B_{z_{g2}^*} B'_{x_2^*} \,\mathrm{d}r \left(\int B_{x_2^*} B'_{x_2^*} \,\mathrm{d}r \right)^{-1} \left[\int B_{x_2^*} \,\mathrm{d}B_{\varepsilon_g} \right] \right\},
\end{aligned} \tag{3.5}
$$

where $\Longrightarrow$ *denotes convergence in distribution of the associated probability measures,*

$$
M_{z_{g1},x_1}^* = \operatorname{plim} \frac{1}{T} Z_{g1}^{*\prime} X_1^*, \quad M_{x_1x_1}^* = \operatorname{plim} \frac{1}{T} X_1^{*\prime} X_1^*, \tag{3.6}
$$

B_{ε_g} *denotes the Brownian motion of* ε_{gt} *with variance* $\sigma_g^2, B_{x_2^*}$ *denotes a* $b \times 1$ *vector Brownian motion of* $\nabla \underset{\sim 2t}{x^*}$ *with covariance matrix* $\Omega_{\nabla x_2^* \nabla x_2^*}$ *where* $\Omega_{\nabla x_2^* \nabla x_2^*}$ *is the long-run covariance matrix of* $\nabla \underset{\sim 2t}{x^*}$, *and* $B_{z_{g2}^*}$ *denotes a* $b_g \times 1$ *vector Brownian motion of* $\nabla \underset{\sim g2,t}{z^*}$

which appears in the gth equation. Moreover $\sqrt{T}(\hat{\underset{\sim}{\delta}}^*_{g1,2SLS} - \underset{\sim}{\delta}^*_{g1})$ *and* $T(\hat{\underset{\sim}{\delta}}^*_{g2,2SLS} - \underset{\sim}{\delta}^*_{g2})$ *are asymptotically independent.*

The limiting distribution of (3.5) is nonstandard. It involves a matrix unit root distribution that arises from using lagged $\underset{\sim}{w}$ as instruments when $\underset{\sim}{w}_t$ is $I(1)$ and is contemporaneously correlated with $\underset{\sim}{\varepsilon}_t$. The long-run "endogeneities" of the nonstationary instruments X_2^* leads to a miscentering and skewness of the limiting distribution of (3.5). However, since $\hat{\underset{\sim}{\delta}}_{g,2SLS} = M_g \hat{\underset{\sim}{\delta}}^*_{g,2SLS}$, the limiting distribution of $\hat{\underset{\sim}{\delta}}_{g,2SLS}$ is given by the components of $\hat{\underset{\sim}{\delta}}^*_{g,2SLS}$ that have slower rate of convergence. Therefore, if $p > 1$ and interest is in testing a particular coefficient, say $\delta_{gk} = c_k$, then the conventional test statistic, $(\hat{\delta}_{gk,2SLS} - c_k)/\mathrm{Sd}(\hat{\delta}_{gk,2SLS})$ is asymptotically t-distributed. However, inference about the null hypothesis $P\underset{\sim}{\delta}_g = \underset{\sim}{c}$ can be tricky, where P and $\underset{\sim}{c}$ are known matrix and vector of proper dimensions, respectively.If $\sqrt{T}P(\hat{\underset{\sim}{\delta}}_{g,2SLS} - \underset{\sim}{\delta}_g)$ has a singular convariance matrix, it means that there exists a nonsingular matrix L such that

$$LP\underset{\sim}{\delta}_g = LP^*\underset{\sim}{\delta}^*_g = \begin{bmatrix} \tilde{P}_{11} & \tilde{P}_{12} \\ \underset{\sim}{0} & \tilde{P}_{22} \end{bmatrix} \begin{bmatrix} \underset{\sim}{\delta}^*_{g1} \\ \underset{\sim}{\delta}^*_{g2} \end{bmatrix} \tag{3.7}$$

with nonzero $\tilde{P}_{22}$. Then

$$\begin{aligned}
&(P\hat{\underset{\sim}{\delta}}_{g,2SLS} - \underset{\sim}{c})' \,\mathrm{Cov}\,(P\hat{\underset{\sim}{\delta}}_{g,2SLS})^{-1}(P\hat{\underset{\sim}{\delta}}_{g,2SLS} - \underset{\sim}{c}) \\
&\quad = \left\{ \begin{bmatrix} \tilde{P}_{11} & \tilde{P}_{12} \\ \underset{\sim}{0} & \tilde{P}_{22} \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{\delta}}^*_{g1,2SLS} \\ \hat{\underset{\sim}{\delta}}^*_{g2,2SLS} \end{bmatrix} - L\underset{\sim}{c} \right\}' \mathrm{Cov}\,(LP\hat{\underset{\sim}{\delta}}_{g,2SLS})^{-1} \\
&\qquad \times \left\{ \begin{bmatrix} \tilde{P}_{11} & \tilde{P}_{12} \\ \underset{\sim}{0} & P_{22} \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{\delta}}^*_{g1,2SLS} \\ \hat{\underset{\sim}{\delta}}^*_{g2,2SLS} \end{bmatrix} - L\underset{\sim}{c} \right\} \\
&\quad \Longrightarrow T\left(\tilde{P}_{11}\hat{\underset{\sim}{\delta}}^*_{g1,2SLS} + \tilde{P}_{12}\hat{\underset{\sim}{\delta}}^*_{g2,2SLS} - \underset{\sim}{\tilde{c}}_1\right)' \mathrm{Cov}\left(\sqrt{T}\tilde{P}_{11}\hat{\underset{\sim}{\delta}}^*_{g1,2SLS}\right)^{-1} \\
&\qquad \times \left(\tilde{P}_{11}\hat{\underset{\sim}{\delta}}^*_{g1,2SLS} + \tilde{P}_{12}\hat{\underset{\sim}{\delta}}^*_{g2,2SLS} - \underset{\sim}{\tilde{c}}_1\right) + T^2\left(\tilde{P}_{22}\hat{\underset{\sim}{\delta}}^*_{g2,2SLS} - \underset{\sim}{\tilde{c}}_2\right)' \\
&\qquad \times \mathrm{Cov}\left(T\tilde{P}_{22}\hat{\underset{\sim}{\delta}}^*_{g2,2SLS}\right)^{-1}\left(\tilde{P}_{22}\hat{\underset{\sim}{\delta}}^*_{g2,2SLS} - \underset{\sim}{\tilde{c}}_2\right),
\end{aligned} \tag{3.8}$$

where $L\underset{\sim}{c} = (\tilde{\underset{\sim}{c}}_1', \tilde{\underset{\sim}{c}}_2')'$. The first term on the right-hand side of (3.8) is asymptotically chi-square distributed. The second term, according to Lemma 3.1 has a nonstandard distribution. Hence (3.8) is not asymptotically chi-square distributed.

Remark 3.1. Our interest lies in the statistical properties of the estimators of $\underset{\sim}{\delta}_g$, not $\underset{\sim}{\delta}_g^*$ (or $\underset{\sim}{\delta}_g^{**}$ to be introduced in Section 4). The matrices Z_g^* and X^* and the corresponding parameter vector $\underset{\sim}{\delta}_g^*$ are introduced for the ease of deriving the limiting distributions of 2SLS of $\underset{\sim}{\delta}_g$ and the corresponding Wald test statistic. The transformed matrices Z_g^* or X^* is not used in actual estimation or in constructing Wald test statistics. Therefore, it is sufficient to know that transformation of Z_g or X to Z_g^* or X^* (or Z_g^{**} or X^{**} in later section) exists. For instance, consider a three equation model of the form

$$A_0\underset{\sim}{w}_t + A_1\underset{\sim}{w}_{t-1} + A_2\underset{\sim}{w}_{t-2} = \underset{\sim}{\varepsilon}_t, \tag{3.9}$$

where

$$A_0 = \begin{pmatrix} 1 & a_{0,12} & 0 \\ 0 & 1 & a_{0,23} \\ a_{0,31} & 0 & 1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} a_{1,11} & a_{1,12} & 0 \\ 0 & a_{1,22} & a_{1,23} \\ a_{1,31} & 0 & a_{1,33} \end{pmatrix},$$

$$A_2 = \begin{pmatrix} a_{2,11} & a_{2,12} & 0 \\ 0 & a_{2,22} & a_{2,23} \\ a_{2,31} & 0 & a_{2,33} \end{pmatrix},$$

and all three equations satisfy the rank condition for identification (2.5). Consider the first equation ($g = 1$) of (3.9). We can rewrite it in the form of (3.1),

$$\underset{\sim}{w}_1 = Z_1\underset{\sim}{\delta}_1 + \underset{\sim}{\varepsilon}_1, \tag{3.10}$$

where $Z_1 = (\underset{\sim}{w}_2, \underset{\sim}{w}_{1,-1}, \underset{\sim}{w}_{2,-1}, \underset{\sim}{w}_{1,-2}, \underset{\sim}{w}_{2,-2})$, and $\underset{\sim}{\delta}_1 = -(a_{0,12}, a_{1,11}, a_{1,12}, a_{2,11}, a_{2,12})'$.

Suppose that A_2 takes the form

$$A_2 = A_0 - A_1 + \underset{\sim}{\alpha}'\underset{\sim}{\beta},$$

where $\underset{\sim}{\alpha}$ and $\underset{\sim}{\beta}$ are $3 \times r$ matrices, $0 \leqslant r < 3$. When $r = 0$, there is no cointegration among w_{1t}, w_{2t} and w_{3t}. Then $Z_1^* = Z_1 M_1 = (Z_{11}^*, Z_{12}^*)$, where

$$M_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{pmatrix},$$

$Z_{11}^* = (\bigtriangledown \underset{\sim 2}{w}, \bigtriangledown \underset{\sim 1,-1}{w}, \bigtriangledown \underset{\sim 2,-1}{w})$, and $Z_{12}^{*\prime} = (\underset{\sim 1,-2}{w}, \underset{\sim 2,-2}{w})$, and $\underset{\sim 1}{\delta^*} = M_1^{-1} \underset{\sim 1}{\delta} = (\underset{\sim 11}{\delta^{*\prime}}, \underset{\sim 12}{\delta^{*\prime}})', \underset{\sim 11}{\delta^{*\prime}} = -(a_{0,12}, a_{1,11}, a_{0,12} + a_{1,12}), \underset{\sim 12}{\delta^{*\prime}} = -(a_{1,11} + a_{2,11}, a_{0,12} + a_{1,12} + a_{2,12})$. The instruments $X = (W_{-1}, W_{-2})$ and $X^* = XM_x = (X_1^*, X_2^*)$, where

$$M_x = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{pmatrix},$$

$X_1^* = (\bigtriangledown W_{-1})$, and $X_2^* = (W_{-2})$.

Suppose that

$$\underset{\sim}{\beta}' = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \quad \text{and} \quad \alpha' = \begin{pmatrix} \alpha_{11} & 0 & \alpha_{31} \\ 0 & \alpha_{22} & \alpha_{32} \end{pmatrix},$$

then model (3.9) is in the spirit of King et al. (1991) three-equation model in which there are two cointegrating relations ($w_{1t} - w_{2t}$ (money and income), and $w_{2t} - w_{3t}$ (income and interest rate)). The corresponding transformation of Z_1^* and X^* then becomes $Z_1^* = Z_1 M_1 = (Z_{11}^*, Z_{12}^*)$ with

$$M_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & -1 & 1 \end{pmatrix},$$

$Z_{11}^* = (\bigtriangledown \underset{\sim 2}{w}, \bigtriangledown \underset{\sim 1,-1}{w}, \bigtriangledown \underset{\sim 2,-1}{w}, \underset{\sim 1,-2}{w} - \underset{\sim 2,-2}{w})$ and $Z_{12}^* = (\underset{\sim 2,-2}{w})$, and $X^* = XM_x = (X_1^*, X_2^*)$,

$$M_x = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 1 \end{pmatrix},$$

$X_1^* = (\bigtriangledown W_{-1}, \underset{\sim 1,-2}{w} - \underset{\sim 2,-2}{w}, \underset{\sim 2,-2}{w} - \underset{\sim 3,-2}{w}), X_2^* = (\underset{\sim 3,-2}{w})$. (The parameter vector $\underset{\sim 1}{\delta^*} = (\underset{\sim 11}{\delta^{*\prime}}, \underset{\sim 12}{\delta^{*\prime}})'$ now has the form, $\underset{\sim 11}{\delta^{*\prime}} = -(a_{0,12}, a_{1,11}, a_{0,12} + a_{1,12}, a_{1,11} + a_{2,11})$ and $\underset{\sim 12}{\delta^{*\prime}} = -(a_{0,12} + a_{1,12} + a_{2,12} + a_{1,11} + a_{2,11})$.)

We note that the application of 2SLS does not provide asymptotically normal or mixed normal estimator because of the long-run endogeneities between lagged $I(1)$ instruments and the (current) shocks of the system (Hsiao and Wang, 2004). But if we can condition on the innovations driving the common trends it will allow us to establish the independence between Brownian motion of the errors of the conditional system involving the cointegrating relations and the innovations driving the common trends. The idea of the modified 2SLS estimator is to apply the 2SLS method to the equation conditional on the innovations driving the common trends. Unfortunately, the direction of nonstationarity is generally unknown. Neither does the identification condition given by Lemma 2.1 require such knowledge. In the event that such knowledge is unavailable, we propose to modify Phillips (1995) fully modified VAR estimator that is used to estimate the reduced-form VAR of the form (2.4) with desirable properties.

Rewrite (3.1) as

$$\begin{aligned}\underset{\sim g}{w} &= Z_g \tilde{M}_g \tilde{M}_g^{-1} \underset{\sim g}{\delta} + \underset{\sim g}{\varepsilon} \\ &= (Z_{g1}^{**} \quad Z_{g2}^{**}) \begin{pmatrix} \underset{\sim g1}{\delta^{**}} \\ \underset{\sim g2}{\delta^{**}} \end{pmatrix} + \underset{\sim g}{\varepsilon} \\ &= Z_g^{**} \underset{\sim g}{\delta^{**}} + \underset{\sim g}{\varepsilon}, \end{aligned} \tag{3.11}$$

where $Z_g^{**} = Z_g \tilde{M}_g = (Z_{g1}^{**}, Z_{g2}^{**}), Z_{g1}^{**} = (\triangledown W_g, \triangledown \tilde{W}_{g,-1}, \ldots, \triangledown \tilde{W}_{g,-p+1}), Z_{g2}^{**} = \tilde{W}_{g,-p}, \underset{\sim g}{\delta^{**}} = \tilde{M}_g^{-1} \underset{\sim g}{\delta}, \triangledown \tilde{W}_{g,-j}$ denoting the $T \times g_\Delta$ stacked first difference of the included variable $\triangledown \underset{\sim g,t-j}{\tilde{w}}$ and $\triangledown W_g$ denoting the $T \times (g_\Delta - 1)$ first difference of the included variables $\triangledown \underset{\sim gt}{\tilde{w}}$ excluding $\triangledown w_{gt}$. The decomposition $(Z_{g1}^{**}, Z_{g2}^{**})$ and $\underset{\sim g}{\delta^{**}} = (\underset{\sim g1}{\delta^{**\prime}}, \underset{\sim g2}{\delta^{**\prime}})'$ are identical to (Z_{g1}^{*}, Z_{g2}^{*}) if there is no cointegrating relations among $\underset{\sim gt}{\tilde{w}}$, $\underset{\sim g}{\pi} = \underset{\sim}{0}$. Unlike $(Z_{g1}^{*}, Z_{g2}^{*}), (Z_{g1}^{**}, Z_{g2}^{**})$ are well defined and observable. When $Z_{g1}^{*} \neq Z_{g1}^{**}$, there exists a nonsingular transformation matrix D_g such that $(Z_{g1}^{**}, Z_{g2}^{**}) D_g = (Z_{g1}^{*}, Z_{g2}^{*})$. Then

$$\underset{\sim g}{\delta^{*}} = D_g^{-1} \underset{\sim g}{\delta^{**}}. \tag{3.12}$$

Remark 3.2. Using the example (3.9), $Z_1^{**} = Z_1 \tilde{M}_1 = (Z_{11}^{**}, Z_{12}^{**})$, where

$$\tilde{M}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{pmatrix},$$

with $Z_{11}^{**} = (\nabla \underset{\sim 2}{w}, \nabla \underset{\sim 1,-1}{w}, \nabla \underset{\sim 2,-1}{w})$ and $Z_{12}^{**} = (\underset{\sim 1,-2}{w}, \underset{\sim 2,-2}{w})$, and $\underset{\sim 11}{\delta}^{**\prime} = -(a_{0,12}, a_{1,11}, a_{0,12} + a_{1,12})$, $\underset{\sim 12}{\delta}^{**\prime} = -(a_{1,11} + a_{2,11}, a_{0,12} + a_{1,12} + a_{2,12})$ irrespective of the cointegration rank in the system.

Let

$$C_g = (W'_{-p} \nabla W_{-p} - T\Delta_{\nabla w \nabla w})\Omega^{-}_{\nabla w \nabla w}\Omega_{\nabla w \varepsilon_g}, \tag{3.13}$$

where Ω_{uv} and Δ_{uv} denote the long-run covariance and the one-sided long-run covariance matrix of two sets of $I(0)$ variables, $(\underset{\sim t}{u}, \underset{\sim t}{v})$,

$$\Omega_{uv} = \sum_{j=-\infty}^{\infty} \Gamma_{uv}(j), \tag{3.14}$$

and

$$\Delta_{uv} = \sum_{j=0}^{\infty} \Gamma_{uv}(j), \tag{3.15}$$

where $\Gamma_{uv}(j) = \mathrm{E}\underset{\sim t}{u}\underset{\sim t-j}{v}'$. Let

$$\hat{C}_g = (W'_{-p} \nabla W_{-p} - T\hat{\Delta}_{\nabla w \nabla w})\hat{\Omega}^{-1}_{\nabla w \nabla w}\hat{\Omega}_{\nabla w \varepsilon_g}, \tag{3.16}$$

where $\hat{\Omega}_{uv}$ and $\hat{\Delta}_{uv}$ are the kernel estimates of Ω_{uv} and Δ_{uv}, which, following Phillips (1995), takes the form

$$\hat{\Omega}_{uv} = \sum_{j=-T+1}^{T-1} K(j/k)\hat{\Gamma}_{uv}(j), \tag{3.17}$$

and

$$\hat{\Delta}_{uv} = \sum_{j=0}^{T-1} K(j/k)\hat{\Gamma}_{uv}(j), \tag{3.18}$$

where $K(\cdot)$ is a kernel function and k is a truncation or bandwidth parameter, and $\hat{\Gamma}_{uv}(j)$ is the sample covariance function of $(\underset{\sim t}{u}, \underset{\sim t-j}{v})$,

$$\hat{\Gamma}_{uv}(j) = \frac{1}{T}\sum_{t=j+1}^{T} \underset{\sim t}{\hat{u}}\underset{\sim t-j}{\hat{v}}'. \tag{3.19}$$

A modified 2SLS estimator following Phillips (1995) fully modified VAR estimator can be defined as

$$\underset{\sim g,\mathrm{m2SLS}}{\hat{\delta}^{**}} = \{Z_g^{**\prime}X^{**}(X^{**\prime}X^{**})^{-1}X^{**\prime}Z_g^{**}\}^{-1} \times \left\{ Z_g^{**\prime}X^{**}(X^{**\prime}X^{**})^{-1}\begin{pmatrix} X_1^{**\prime}\underset{\sim g}{w} \\ X_2^{**\prime}\underset{\sim g}{w} - \hat{C}_g \end{pmatrix}\right\}, \tag{3.20}$$

where $X^{**} = X\tilde{M}_x = (X_1^{**}, X_2^{**}), X_1^{**} = (\triangledown W_{-1}, \ldots, \triangledown W_{-p+1})$, and $X_2^{**} = W_{-p}$. Just like $(Z_{g1}^{**}, Z_{g2}^{**}), (X_1^{**}, X_2^{**})$ are well defined and observable.

Following Phillips (1995), we assume that

Assumption KL. The Kernel function $K(\cdot) : R \to [0, 1]$ in (3.17) and (3.18) is a twice continuously differentiable even function with:

(a) $K(0) = 1, K'(0) = 0, K''(0) \neq 0$; and either
(b) $K(x) = 0, |x| \geqslant 1$, with $\lim_{|x| \to 1}[K(x)/(1-|x|)^2] = \text{constant}$, or
(c) $K(x) = O(x^{-2})$ as $|x| \to \infty$.

Assumption BW. The bandwidth parameter k in (3.17) and (3.18) has an expansion rate of the form:

$k = O_e(T^q)$ for some $q \in (1/4,\ 2/3)$, where the symbol O_e is the expansion rate symbol such that

$$k = O_e(T^q) \text{ if } k \sim c_T T^q \text{ as } T \to \infty$$

for some c_T which is slowly varying at infinity (i.e. $c_{Tx}/c_T \to 1$ as $T \to \infty$ for $x > 0$). Thus $k/T^{2/3} + T^{1/4}/k \to 0$ and $k^4/T \to \infty$ as $T \to \infty$. Then

Theorem 3.1. *Under assumptions* A1–A4, *KL and BW*, *the modified* 2*SLS estimator* $\hat{\underset{\sim}{\delta}}{}^{*}_{g,\mathrm{m2SLS}} = D_g^{-1}\hat{\underset{\sim}{\delta}}{}^{**}_{g,\mathrm{m2SLS}}$ *is consistent. Furthermore*

$$\sqrt{T}\left(\hat{\underset{\sim}{\delta}}{}^{*}_{g1,\mathrm{m2SLS}} - \underset{\sim}{\delta}{}^{*}_{g1}\right) \Longrightarrow \mathrm{N}(\underset{\sim}{0}, \sigma_g^2(M^*_{z_{g1}x_1}M^{*-1}_{x_1x_1}M^*_{x_1z_{g1}})^{-1}) \tag{3.21}$$

and is independent of

$$T\left(\hat{\underset{\sim}{\delta}}{}^{*}_{g2,\mathrm{m2SLS}} - \underset{\sim}{\delta}{}^{*}_{g2}\right) \Longrightarrow (M^*_{z_{g_2}x_2}M^{*-1}_{x_2x_2}M^*_{x_2z_{g_2}})^{-1}M^*_{z_{g_2}x_2}M^{*-1}_{x_2x_2}\int B_{x_2^*}\,\mathrm{d}B_{\varepsilon_g\cdot x_2^*}, \tag{3.22}$$

which is a mixed normal of the form

$$\int_{M^*_{x_2x_2}>0}\mathrm{N}\left(\underset{\sim}{0}, \sigma^2_{g\cdot\triangledown x_2^*}(M^*_{z_{g2}x_2}M^{*-1}_{x_2x_2}M^*_{x_2z_{g2}})^{-1}\right)\mathrm{d}P(M^*_{x_2x_2}), \tag{3.23}$$

where $\sigma^2_{g\cdot\triangledown x_2^*} = \sigma_g^2 - \Omega_{\varepsilon_g\triangledown x_2^*}\Omega_{\triangledown x_2^*\triangledown x_2^*}\Omega_{\triangledown x_2^*\varepsilon_g}$.

Proof. See Appendix A. □

Corollary 3.1. *Under the assumptions of Theorem* 3.1, *when* $r = 0$, *we have*

$$T\left(\hat{\underset{\sim}{\delta}}{}^{*}_{g2,\mathrm{m2SLS}} - \underset{\sim}{\delta}{}^{*}_{g2}\right) \stackrel{\mathrm{p}}{\to} \underset{\sim}{0}, \tag{3.24}$$

i.e. $\hat{\underset{\sim}{\delta}}{}^{*}_{g2,\mathrm{m2SLS}}$ *is hyperconsistent in the sense that its rate of convergence is faster than* T. $M^*_{z_{g1}x_1} = \mathrm{plim}\ (1/T)Z^{*\prime}_{g1}X_1^*, M^*_{x_1x_1} = \mathrm{plim}\ (1/T)X_1^{*\prime}X_1^*, M^*_{z_{g2}x_2}$ *and* $M^*_{x_2x_2}$ *are* $b_g \times b$ *and* $b \times b$ *matrices of random variables that have the limiting distributions as that of* $(1/T^2)Z^{*\prime}_{g2}X_2^*$ *and* $(1/T^2)X_2^{*\prime}X_2^*$, *respectively.*

Proof. See Appendix A. $\square$

Remark 3.3. The modified 2SLS estimator of $\underset{\sim}{\delta}_g$ can be obtained as

$$\hat{\underset{\sim}{\delta}}_{g,\mathrm{m2SLS}} = \tilde{M}_g \hat{\underset{\sim}{\delta}}^{**}_{g,\mathrm{m2SLS}} = \tilde{M}_g D_g \hat{\underset{\sim}{\delta}}^{*}_{g,\mathrm{m2SLS}}, \tag{3.25}$$

where $\tilde{M}_g$ is a known matrix but in general, not D_g. However, although the modified 2SLS estimator of $\underset{\sim}{\delta}^*_g$ is either asymptotically normal or mixed normal, the Wald type test statistic

$$\frac{1}{\sigma_g^2}(P\hat{\underset{\sim}{\delta}}_{g,\mathrm{m2SLS}} - \underset{\sim}{c})'\{P[Z_g'X(X'X)^{-1}X'Z_g]P'\}^{-1}\left(P\hat{\underset{\sim}{\delta}}_{g,\mathrm{m2SLS}} - \underset{\sim}{c}\right) \tag{3.26}$$

does not always have the asymptotic chi-square distribution under the null hypothesis $P\underset{\sim}{\delta}_g = \underset{\sim}{c}$, where P is a known $k \times g_\Delta$ matrix of rank k. To see this, rewrite (3.26) in terms of $\hat{\underset{\sim}{\delta}}^*_{g,\mathrm{m2SLS}}$

$$\frac{1}{\sigma_g^2}\left(P^*H_g\hat{\underset{\sim}{\delta}}^*_{g,\mathrm{m2SLS}} - \underset{\sim}{c}\right)'\{P^*H_g[Z_g^{*\prime}X^*(X^{*\prime}X^*)^{-1}X^{*\prime}Z_g^*]H_g'P^{*\prime}\}^{-1} \times\left(P^*H_g\hat{\underset{\sim}{\delta}}^*_{g,\mathrm{m2SLS}} - \underset{\sim}{c}\right), \tag{3.27}$$

where

$$P^* = P\tilde{M}_g D_g H_g^{-1} \quad \text{and} \quad H_g = \begin{bmatrix} T^{-1/2}I_{lg} & 0 \\ 0 & T^{-1}I_{b_g} \end{bmatrix}.$$

The null hypothesis becomes $P^*H_g\hat{\underset{\sim}{\delta}}^*_{g,\mathrm{m2SLS}} = \underset{\sim}{c}$. Notice that the asymptotic covariance matrix of $H_g\hat{\underset{\sim}{\delta}}^*_{g,\mathrm{m2SLS}}$ converges to

$$\begin{pmatrix} \sigma_g^2(M^*_{z'_{g1}x_1}M^{*-1}_{x_1x_1}M^*_{x_1z_{g1}})^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \sigma^2_{g.\nabla x_2^*}(M^*_{z_{g2}x_2}M^{*-1}_{x_2x_2}M^*_{x_2z_{g2}})^{-1} \end{pmatrix},$$

while $H_g[Z_g^{*\prime}X^*(X^{*\prime}X^*)^{-1}X^{*\prime}Z_g^*]H_g'$ in (3.27) converges to

$$\sigma_g^2\begin{pmatrix} (M^*_{z_{g1}x_1}M^{*-1}_{x_1x_1}M^*_{x_1z_{g1}})^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & (M^*_{z_{g2}x_2}M^{*-1}_{x_2x_2}M^*_{x_2z_{g2}})^{-1} \end{pmatrix},$$

Wald statistic (3.26) (or equivalently (3.27)) is asymptotically chi-square distributed with k degrees of freedom if and only if $P\hat{\underset{\sim}{\delta}}_{g,\mathrm{m2SLS}}$ (or equivalently $P^*H_g\hat{\underset{\sim}{\delta}}^*_{g,\mathrm{m2SLS}}$) in the hypothesis does not involve the T-consistent component $\hat{\underset{\sim}{\delta}}^*_{g2,\mathrm{m2SLS}}$. Otherwise, $H_g[Z_g^{*\prime}X^*(X^{*\prime}X^*)^{-1}X^{*\prime}Z_g^*]H_g'$ would overestimate the asymptotic covariance matrix of $H_g \hat{\underset{\sim}{\delta}}^*_{g,\text{m2SLS}}$ because $\sigma^2_{g\cdot\nabla x_2^*} \leqslant \sigma^2_g$ for the submatrix corresponding to $\underset{\sim}{x}^*_2$ and $\underset{\sim}{z}^*_{g2}$. In general, the test statistic (3.26) is a conservative test, with its asymptotic distribution a weighted sum of k independent χ^2_1 variables with weights between 0 and 1.

4. An alternatively modified 2SLS estimator

Section 3 shows that without pretesting for or the prior knowledge of the cointegrating space, the modified 2SLS estimator is consistent and has the desired property that coefficient estimates of the transformed system are either $\sqrt{T}$-consistent and asymptotically normally distributed or T-consistent and mixed normally distributed in the limit. However, the construction of the modified 2SLS estimator requires nonparametric estimation of the long-run covariance matrix and the one-sided long-run covariance matrix. It is well known that kernel estimator and hence the finite sample performance of the modified 2SLS estimator could be affected substantially by the choice of the bandwidth parameter. In addition, since we cannot approximate the asymptotic covariance matrix of the modified 2SLS estimator properly, Wald test statistics based on the modified 2SLS estimator using the formula of (3.26) may not be chi-square distributed and critical values that are based on chi-square distributions can be used for conservative tests only. In this section, we propose an alternatively modified 2SLS estimator with the following properties: (1) it is fully parametric, (2) coefficient estimates of the transformed system are $\sqrt{T}$-convergence and asymptotically normally distributed in the stationary direction and T-convergence and asymptotically mixed normally distributed in the nonstationary direction, and (3) its asymptotic covariance matrix can be properly approximated so that Wald test statistics remain χ^2 distributed in the limit.

We note that (2.1) implies the existence and uniqueness of a vector autoregressive moving average process of order p and 1, respectively,

$$\nabla \underset{\sim}{w}_t = J_1 \nabla \underset{\sim}{w}_{t-1} + \cdots + J_p \nabla \underset{\sim}{w}_{t-p} + \underset{\sim}{\eta}_t, \tag{4.1}$$

where $\underset{\sim}{\eta}_t = (I - \Phi L)\underset{\sim}{v}_t$, and $\underset{\sim}{v}_t = A_0^{-1}\underset{\sim}{\varepsilon}_t$, subject to the constraint that the roots of $|I - J_1 z - \cdots - J_p z^p| = 0$ lie outside the unit circle and Φ is symmetric and idempotent. Let $w^+_{gt} = w_{gt} - \Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta}\underset{\sim}{\eta}_t$ and $\hat{w}^+_{gt} = w_{gt} - \hat{\Omega}_{\varepsilon_g\eta}\hat{\Omega}^{*-1}_{\eta\eta}\hat{\underset{\sim}{\eta}}_t$, where $\Omega_{\varepsilon_g\eta}$ and $\Omega_{\eta\eta}$ are the long-run covariance between ε_{gt} and $\underset{\sim}{\eta}_t$ and the long-run covariance matrix of $\underset{\sim}{\eta}_t$, respectively, $\hat{\Omega}_{\varepsilon_g\eta}, \hat{\Omega}_{\eta\eta}, \hat{\underset{\sim}{\eta}}$ denote their estimates, $\Omega^-_{\eta\eta}$ denotes the generalized inverse of $\Omega_{\eta\eta}$ and $\hat{\Omega}^*_{\eta\eta} = \hat{\Omega}_{\eta\eta} + T^{-d}I_m$, where $d \in (0, \frac{1}{2})$. The alternatively modified 2SLS estimator (A2SLS) is defined as

$$\hat{\underset{\sim}{\delta}}_{g,\text{a2SLS}} = \tilde{M}_g \hat{\underset{\sim}{\delta}}^{**}_{g,\text{a2SLS}} = \tilde{M}_g D_g \hat{\underset{\sim}{\delta}}^*_{g,\text{a2SLS}} \tag{4.2}$$

where

$$\hat{\underset{\sim}{\delta}}^{**}_{g,\text{a2SLS}} = [Z_g^{**\prime}X^{**}(X^{**\prime}X^{**})^{-1}X^{**\prime}Z_g^{**}]^{-1}\left[Z_g^{**\prime}X^{**}(X^{**\prime}X^{**})^{-1}\begin{pmatrix} X_1^{**\prime}\underset{\sim g}{w} \\ X_2^{**\prime}\hat{\underset{\sim g}{w}}^{+}\end{pmatrix}\right]. \tag{4.3}$$

The difference between the modified 2SLS and A2SLS is in the adjustment factor. The modified 2SLS uses $\hat{C}_g$ (3.16). The A2SLS adjusts $\underset{\sim gt}{w}$ by $-\hat{\Omega}_{\varepsilon_g\eta}\hat{\Omega}_{\eta\eta}^{*-1}\hat{\underset{\sim t}{\eta}}$. There is no serial correlation adjustment factor for A2SLS because $\underset{\sim}{\eta}$ is at most a moving average process of order 1. Furthermore, $\Omega_{\varepsilon_g\eta}$ and $\Omega_{\eta\eta}$ can be estimated parametrically. One such estimator is

$$\hat{\Omega}_{\varepsilon_g\eta} = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{gt}\hat{\underset{\sim t}{\eta}}' + T^{-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{gt}\hat{\underset{\sim t+1}{\eta}}' \tag{4.4}$$

and

$$\hat{\Omega}_{\eta\eta} = T^{-1}\sum_{t=1}^{T}\hat{\underset{\sim t}{\eta}}\hat{\underset{\sim t}{\eta}}' + T^{-1}\sum_{t=1}^{T-1}\hat{\underset{\sim t}{\eta}}\hat{\underset{\sim t+1}{\eta}}' + T^{-1}\sum_{t=2}^{T}\hat{\underset{\sim t}{\eta}}\hat{\underset{\sim t-1}{\eta}}', \tag{4.5}$$

where $\hat{\varepsilon}_{gt}$ and $\hat{\underset{\sim t}{\eta}}$ are the 2SLS residuals of (3.1) and the MLE residuals of (4.1), respectively. The estimators (4.4) and (4.5) converge to their true values, $\Omega_{\varepsilon_g\eta}$ and $\Omega_{\eta\eta}$ at the speed of $T^{1/2}$. However, since $\Omega_{\eta\eta}$ may be singular, $\hat{\Omega}_{\varepsilon_g\eta}\hat{\Omega}_{\eta\eta}^{-1}$ may not converge to $\Omega_{\varepsilon_g\eta}\Omega_{\eta\eta}^{-}$. Adding $T^{-d}I_m$ for $d \in (0, 1/2)$ to $\hat{\Omega}_{\eta\eta}$ does not affect the consistency property of $\hat{\Omega}_{\eta\eta}^{*}$, but ensures the convergence of $\hat{\Omega}_{\varepsilon_g\eta}\hat{\Omega}_{\eta\eta}^{*-1}$ to $\Omega_{\varepsilon_g\eta}\Omega_{\eta\eta}^{-}$. It is shown in Appendix B that the optimal value of $d = 1/4$.

The reason for adjusting $\underset{\sim gt}{w}$ by $-\Omega_{\varepsilon_g\eta}\Omega_{\eta\eta}^{-}\underset{\sim t}{\eta}$ is that the elements of the long-run covariance matrix between $\underset{\sim g}{\varepsilon}$ and $\underset{\sim}{\eta}$ that correspond to the stationary directions are zero because the corresponding elements of $\underset{\sim}{\eta}$ are in the form of $\underset{\sim}{\alpha}'(\underset{\sim t}{v} - \underset{\sim t-1}{v})$ with zero long-run covariance. Only the elements of $\underset{\sim}{\eta}$ that drive the nonstationary direction ($\underset{\sim\perp}{\alpha}'\underset{\sim t}{v}$) will have nonzero long-run covariance. They are the only elements that enter into the adjustment, hence establishes the orthogonality between the conditional error $\underset{\sim gt}{\varepsilon}^{+} = \underset{\sim gt}{\varepsilon} - \hat{\Omega}_{\varepsilon_g\eta}\hat{\Omega}_{\eta\eta}^{*-1}\underset{\sim t}{\eta}$ of the gth equation and the innovations driving the common trends.

Let

$$\hat{\underset{\sim}{\delta}}^*_{g,\mathrm{a2SLS}} = [Z_g^{*\prime} X^* (X^{*\prime} X^*)^{-1} X^{*\prime} Z_g^*]^{-1} \left[Z_g^{*\prime} X^* (X^{*\prime} X^*)^{-1} D_x' \begin{pmatrix} X_1^{**\prime} \underset{\sim}{w}_g \\ X_2^{**\prime} \hat{\underset{\sim}{w}}_g^+ \end{pmatrix} \right], \quad (4.6)$$

where $X^* = X^{**} D_x$. It follows that

Theorem 4.1. *When $p \geqslant 2$, the alternatively modified 2SLS estimator $\hat{\underset{\sim}{\delta}}^*_{g,\mathrm{a2SLS}}$ is consistent. Furthermore,*

$$\begin{bmatrix} \sqrt{T}(\hat{\underset{\sim}{\delta}}^*_{g1,\mathrm{a2SLS}} - \underset{\sim}{\delta}^*_{g1}) \\ T(\hat{\underset{\sim}{\delta}}^*_{g2,\mathrm{a2SLS}} - \underset{\sim}{\delta}^*_{g2}) \end{bmatrix} \Longrightarrow \begin{pmatrix} \underset{\sim}{\phi}_{g1} \\ \underset{\sim}{\phi}_{g2} \end{pmatrix} \sim \begin{pmatrix} \mathrm{N}(0, \Sigma_{g1}^*) \\ \int_{M^*_{x_2x_2} > 0} \mathrm{N}(0, \Sigma_{g2}^*)\, \mathrm{d}P(M^*_{x_2x_2}) \end{pmatrix}, \quad (4.7)$$

where $\underset{\sim}{\phi}_{g1}$ and $\underset{\sim}{\phi}_{g2}$ are independent, and

$$\Sigma_{g1}^* = (M^*_{z_{g1}x_1} M^{*-1}_{x_1x_1} M^*_{x_1z_{g1}})^{-1} M^*_{z_{g1}x_1} M^{*-1}_{x_1x_1} \tilde{\Sigma}_{g1} M^{*-1}_{x_1x_1} M^*_{x_1z_{g1}} (M^*_{z_{g1}x_1} M^{*-1}_{x_1x_1} M^*_{x_1z_{g1}})^{-1},$$

$$\Sigma_{g2}^* = \sigma_{g+}^2 (M^*_{z_{g2}x_2} M^{*-1}_{x_2x_2} M^*_{x_2z_{g2}})^{-1},$$

$$\sigma_{g+}^2 = \sigma_g^2 - \Omega_{\varepsilon_g\eta} \Omega_{\eta\eta}^- \Omega_{\eta\varepsilon_g},$$

$$\tilde{\Sigma}_{g1} = \begin{bmatrix} \sigma_g^2 M^{**}_{x_1x_1} & \sigma_{g+}^2 M^{**}_{x_1\tilde{w}_{g1}} + \Theta_2' \\ \sigma_{g+}^2 M^{**}_{\tilde{w}_{g1}x_1} + \Theta_2 & \Sigma_{g1} \end{bmatrix},$$

where

$$M^{**}_{x_1x_1} = \mathrm{plim}\ \frac{1}{T} X_1^{**\prime} X_1^{**},$$

$$M^{**}_{x_1\tilde{w}_{g1}} = \mathrm{plim}\ \frac{1}{T} X_1^{**\prime} \tilde{W}^*_{g1,-p},$$

$$\Sigma_{g1} = \sigma_{g+}^2 M^*_{\tilde{w}_{g1}\tilde{w}_{g1}} + (\Omega_{\varepsilon_g\eta}\Omega_{\eta\eta}^- \otimes M^{**}_{\tilde{w}_{g1}x_1}) \mathrm{Cov}(\hat{\underset{\sim}{\theta}})(\Omega_{\eta\eta}^-\Omega_{\eta\varepsilon_g} \otimes M^{**}_{x_1\tilde{w}_{g1}}) + \Theta_1 + \Theta_1',$$

$$M^*_{\tilde{w}_{g1}\tilde{w}_{g1}} = \mathrm{plim} \frac{1}{T} \tilde{W}^{*\prime}_{g1,-p} \tilde{W}^*_{g1,-p},$$

$$\Theta_1 = \mathrm{E}[T^{-1/2} \tilde{W}^{*\prime}_{g1,-p} (I_T \otimes \Omega_{\varepsilon_g\eta}\Omega_{\eta\eta}^-) \tilde{X} (\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}) \cdot T^{-1/2} \underset{\sim}{\varepsilon}_g' \tilde{W}^*_{g1,-p}],$$

$$\Theta_2 = \mathrm{E}[T^{-1/2} \tilde{W}^{*\prime}_{g1,-p} (I_T \otimes \Omega_{\varepsilon_g\eta}\Omega_{\eta\eta}^-) \tilde{X} (\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}) \cdot T^{-1/2} \underset{\sim}{\varepsilon}_g' X_1^{**}]$$

with

$$\underset{\sim}{\theta} = \text{vech}\,(J^*), \quad J^* = (J_1, \ldots, J_p), \quad \tilde{X} = \begin{pmatrix} I_m \otimes \nabla X_1' \\ \vdots \\ I_m \otimes \nabla X_T' \end{pmatrix},$$

$$\nabla X_t' = \left(\nabla \underset{\sim t-1}{w'}, \ldots, \nabla \underset{\sim t-p}{w'} \right)$$

so that (4.1) *is rewritten as* $\nabla \underset{\sim}{w} = \tilde{X}\,\underset{\sim}{\theta} + \underset{\sim}{\eta}$, *where* $\nabla \underset{\sim}{w'} = (\nabla \underset{\sim 1}{w'}, \ldots, \nabla \underset{\sim T}{w'})$.

Proof. See Appendix B. □

The alternative 2SLS estimator (4.2) is related to $\hat{\underset{\sim g,\text{a2SLS}}{\delta}}^*$ by $\hat{\underset{\sim g,\text{a2SLS}}{\delta}} = M_g \hat{\underset{\sim g,\text{a2SLS}}{\delta}}^*$. The limiting distribution of $\hat{\underset{\sim g,\text{a2SLS}}{\delta}}$ is determined by the component that has the slower rate of convergence. Therefore, if none of the rows of M_g are identically zero in its first ℓ_g columns, $\hat{\underset{\sim g,\text{a2SLS}}{\delta}}$ converges to $\underset{\sim g}{\delta}$ at the speed of $T^{1/2}$ and its limiting distribution is singular normal. On the other hand, if for some rows of M_g, the first ℓ_g columns are identically zero, then the corresponding components of $\hat{\underset{\sim g,\text{a2SLS}}{\delta}}$ converges to their true values at the speed of T. Let M_{g+} and M_{g++} denote the submatrix of M_g that the first ℓ_g columns of each row are not and are identically zero, respectively, and $\underset{\sim g+}{\delta}$ and $\underset{\sim g++}{\delta}$ denote the subvectors of $\underset{\sim g}{\delta}$ that correspond to M_{g+} and M_{g++}, respectively. Then

Theorem 4.2. *When* $p \geqslant 2$, *the alternatively modified 2SLS estimator* (4.2) *is consistent. Furthermore*

$$\sqrt{T}\left(\hat{\underset{\sim g+,\text{a2SLS}}{\delta}} - \underset{\sim g+}{\delta} \right) \Longrightarrow \text{N}\left(\underset{\sim}{0}, M_{g+} \begin{pmatrix} \Sigma_{g1}^* & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix} M_{g+}' \right), \tag{4.8}$$

and is independent of

$$T\left(\hat{\underset{\sim g++,\text{a2SLS}}{\delta}} - \underset{\sim g++}{\delta} \right) \Longrightarrow \int_{M_{x_2x_2}^* > 0} \text{N}\left(\underset{\sim}{0}, M_{g++} \begin{pmatrix} \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \Sigma_{g2}^* \end{pmatrix} M_{g++}' \right) \text{d}P(M_{x_2x_2}^*), \tag{4.9}$$

which is mixed normal with mean $\underset{\sim}{0}$ *and conditional covariance matrix*

$$M_{g++} \begin{pmatrix} \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \Sigma_{g2}^* \end{pmatrix} M_{g++}'.$$

Given that the limiting distribution of $\hat{\underset{\sim}{\delta}}_{g,\text{a2SLS}}$ is either asymptotic normal or mixed normal, the conventional Wald-style test statistic can be approximated by the chi-square distribution with appropriate degree of freedom. For instance, suppose that the null hypothesis is

$$\mathrm{H}_0 : P\underset{\sim}{\delta}_g = \underset{\sim}{c}, \tag{4.10}$$

where P is a known $k \times (\ell_g + b_g)$ matrix with rank k and $\underset{\sim}{c}$ is a known $k \times 1$ vector. Under the null,

$$\begin{aligned}
&\left(\hat{\underset{\sim}{\delta}}_{g,\text{a2SLS}} - \underset{\sim}{\delta}_g\right)' P' \,\mathrm{Cov}\left(P\hat{\underset{\sim}{\delta}}_{g,\text{a2SLS}}\right)^{-1} P\left(\hat{\underset{\sim}{\delta}}_{g,\text{a2SLS}} - \underset{\sim}{\delta}_g\right) \\
&= \left\{\begin{bmatrix} \tilde{P}_{11} & \tilde{P}_{12} \\ \underset{\sim}{0} & \tilde{P}_{22} \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{\delta}}^*_{g1,\text{a2SLS}} \\ \hat{\underset{\sim}{\delta}}^*_{g2,\text{a2SLS}} \end{bmatrix} - L\underset{\sim}{c} \right\}' \mathrm{Cov}\,(LP\hat{\underset{\sim}{\delta}}_{g,\text{a2SLS}})^{-1} \\
&\quad \times \left\{\begin{bmatrix} \tilde{P}_{11} & \tilde{P}_{12} \\ \underset{\sim}{0} & \tilde{P}_{22} \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{\delta}}^*_{g1,\text{a2SLS}} \\ \hat{\underset{\sim}{\delta}}^*_{g2,\text{a2SLS}} \end{bmatrix} - L\underset{\sim}{c} \right\}
\end{aligned} \tag{4.11}$$

$$\begin{aligned}
&\Longrightarrow T\left(\tilde{P}_{11}\hat{\underset{\sim}{\delta}}^*_{g1,\text{a2SLS}} + \tilde{P}_{12}\hat{\underset{\sim}{\delta}}^*_{g2,\text{a2SLS}} - \tilde{\underset{\sim}{c}}_1\right)' \mathrm{Cov}\left(\sqrt{T}\tilde{P}_{11}\hat{\underset{\sim}{\delta}}^*_{g1,\text{a2SLS}}\right)^{-1} \\
&\quad \left(\tilde{P}_{11}\hat{\underset{\sim}{\delta}}^*_{g1,\text{a2SLS}} + \tilde{P}_{12}\hat{\underset{\sim}{\delta}}^*_{g2,\text{a2SLS}} - \tilde{\underset{\sim}{c}}_1\right) \\
&\quad + T^2\left(\tilde{P}_{22}\hat{\underset{\sim}{\delta}}^*_{g2,\text{a2SLS}} - \tilde{\underset{\sim}{c}}_2\right)' \mathrm{Cov}\left(T\tilde{P}_{22}\hat{\underset{\sim}{\delta}}^*_{g2,\text{a2SLS}}\right)^{-1} \left(\tilde{P}_{22}\hat{\underset{\sim}{\delta}}^*_{g2,\text{a2SLS}} - \tilde{\underset{\sim}{c}}_2\right),
\end{aligned} \tag{4.12}$$

where L is a nonsingular matrix that transforms $LP\underset{\sim}{\delta}_g$ into the form (3.7) and $L\underset{\sim}{c} = (\tilde{\underset{\sim}{c}}'_1, \tilde{\underset{\sim}{c}}'_2)'$. Since $\sqrt{T}\hat{\underset{\sim}{\delta}}^*_{g1,\text{a2SLS}}$ is asymptotically normal, $T\hat{\underset{\sim}{\delta}}^*_{g2,\text{a2SLS}}$ is asymptotically mixed normal, and the two limiting distributions are independent, (4.12) converges to a χ^2 distribution with k degrees of freedom.

Corollary 4.1. *When prior restrictions are in the form of exclusion restrictions and the structural VAR model has order $p>1$, then $M_{g+} \equiv M_g$, $\hat{\underset{\sim}{\delta}}_{g,\text{a2SLS}} \equiv \hat{\underset{\sim}{\delta}}_{g+,\text{a2SLS}}$, i.e., each element of the alternatively modified 2SLS estimator $\hat{\underset{\sim}{\delta}}_{g,\text{a2SLS}}$ converges to $\underset{\sim}{\delta}_g$ at the rate of $\sqrt{T}$.*

Corollary 4.2. *When rank of cointegration* $r = 0$,

$$T\left(\hat{\underset{\sim}{\delta}}^*_{g2,\text{a2SLS}} - \underset{\sim}{\delta}^*_{g2}\right) \xrightarrow{\text{p}} \underset{\sim}{0}.$$

Remark 4.1. The asymptotic efficiency of $\hat{\underset{\sim}{\delta}}^*_{g1,\text{a2SLS}}$ is given by the asymptotic efficiency of the first stage estimator, $\hat{\underset{\sim}{\theta}}$. Since the reduced-form specification (4.1) ignores overidentification restrictions of (2.1), the MLE of $\underset{\sim}{\theta}$ is not as efficient as the MLE of $\underset{\sim}{\theta}$ that incorporates the overidentification restrictions. Therefore, unless the system is exactly identified, the estimator of $\hat{\underset{\sim}{\delta}}^*_{g1,\text{a2SLS}}$ is in general less efficient than the 2SLS of $\underset{\sim}{\delta}^*_{g1}$. What it implies is that although alternatively modified 2SLS estimator allows one to get rid of the nonstandard distribution of the part of the level coefficients associated with estimating unit roots either explicitly or implicitly, it pays a cost of efficiency loss.

Remark 4.2. Both estimators (3.20) and (4.3) have the desirable property of being consistent and asymptotically normally or mixed normally distributed. However, estimator (3.20) requires the nonparametric estimation of the long-run covariance matrix ((3.17) and (3.18)), but estimator (4.3) does not because it is known that the error of (4.1) is at most a first-order moving average process. This difference can have implication on the finite sample performance of the two estimators. Moreover, the asymptotic conditional covariance matrix of (4.2) can be properly approximated so that the Wald-type test statistic can be approximated by a chi-square distribution. But the chi-square approximation of the test statistic (3.26) may only give a conservative bound if the null hypothesis $P\underset{\sim}{\delta} = \underset{\sim}{c}$ isolates the coefficients that are T convergent.

5. Structural VAR containing intercepts

For ease of exposition, we have formulated the data generating process (2.1) as having no intercept term. In this section, we briefly illustrate that the basic messages of previous sections remain unchanged when we add an intercept term. Let

$$A(L)\underset{\sim}{w}_t = \underset{\sim}{\gamma} + \underset{\sim}{\varepsilon}_t, \tag{5.1}$$

where γ denotes the $G \times 1$ intercept term, which may or may not be equal to zero. Writing the gth equation of (5.1) in the form of (3.1) yields

$$\underset{\sim}{w}_g = Z_g\underset{\sim}{\delta}_g + \underset{\sim}{e}\gamma_g + g, \tag{5.2}$$

where $\underset{\sim}{e}$ is a $T \times 1$ vector with all elements equal to one. The 2SLS of (5.2) then takes the form

$$\begin{pmatrix} \hat{\underset{\sim}{\delta}}_{g,2SLS} \\ \hat{\gamma}_{g,2SLS} \end{pmatrix} = \left\{ \begin{pmatrix} Z_g' \\ \underset{\sim}{e}' \end{pmatrix} (X, \underset{\sim}{e}) \left[(X, \underset{\sim}{e})'(X, \underset{\sim}{e}) \right]^{-1} \begin{pmatrix} X' \\ \underset{\sim}{e}' \end{pmatrix} (Z_g, \underset{\sim}{e}) \right\}^{-1} \times \left\{ \begin{pmatrix} Z_g' \\ \underset{\sim}{e}' \end{pmatrix} (X, \underset{\sim}{e}) \left[\begin{pmatrix} X' \\ \underset{\sim}{e}' \end{pmatrix} (X, \underset{\sim}{e}) \right]^{-1} \begin{pmatrix} X' \\ \underset{\sim}{e}' \end{pmatrix} \underset{\sim}{w}_g \right\}. \tag{5.3}$$

The limiting distribution of the 2SLS estimator (and the modified 2SLS estimators) depends on whether the $I(1)$ process $\underset{\sim}{w}$ is with or without drift. We shall first consider the case that there is no drift ($\underset{\sim}{\gamma} = \underset{\sim}{0}$). Then we can transform (5.2) in the form of (3.3),

$$\underset{\sim}{w}_g = Z_g^* \underset{\sim}{\delta}_g^* + \underset{\sim}{e}\,\gamma_g^* + \underset{\sim}{\varepsilon}_g, \tag{5.4}$$

where $Z_g^* = Z_g M_g = (Z_{g1}^*, Z_{g2}^*)$, $\underset{\sim}{\delta}_g^* = (\underset{\sim}{\delta}_{g1}^{*\prime}, \underset{\sim}{\delta}_{g2}^{*\prime})' = M_g^{-1} \underset{\sim}{\delta}_g$ and $\gamma_g^* = \gamma_g$. Similarly transform $X = XM_x = (X_1^*, X_2^*)$ as those defined after (3.3), then the 2SLS of (5.2) is equal to

$$\begin{pmatrix} \hat{\underset{\sim}{\delta}}_{g,2SLS} \\ \hat{\gamma}_{g,2SLS} \end{pmatrix} = \begin{pmatrix} M_g^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{\underset{\sim}{\delta}}^*_{g,2SLS} \\ \hat{\underset{\sim}{\gamma}}^*_{g,2SLS} \end{pmatrix}, \tag{5.5}$$

where

$$\begin{pmatrix} \hat{\underset{\sim}{\delta}}^*_{g1,2SLS} \\ \hat{\underset{\sim}{\delta}}^*_{g2,2SLS} \\ \hat{\gamma}^*_{g,2SLS} \end{pmatrix} = \left\{ \begin{pmatrix} Z_{g1}^{*\prime} \\ Z_{g2}^{*\prime} \\ \underset{\sim}{e}' \end{pmatrix} (X_1^*, X_2^*, \underset{\sim}{e}) \begin{pmatrix} X^{*\prime}X^* & X^{*\prime}\underset{\sim}{e} \\ \underset{\sim}{e}'X^* & T \end{pmatrix}^{-1} \begin{pmatrix} X^{*\prime} \\ \underset{\sim}{e}' \end{pmatrix} (Z_g^*, \underset{\sim}{e}) \right\}^{-1} \times \left\{ \begin{pmatrix} Z_g^{*\prime} \\ \underset{\sim}{e}' \end{pmatrix} (X^*, \underset{\sim}{e}) \begin{pmatrix} X^{*\prime}X^* & X^{*\prime}\underset{\sim}{e} \\ \underset{\sim}{e}'X^* & T \end{pmatrix}^{-1} \begin{pmatrix} X_1^{*\prime}\underset{\sim}{w}_g \\ X_2^{*\prime}\underset{\sim}{w}_g \\ \underset{\sim}{e}'\underset{\sim}{w}_g \end{pmatrix} \right\}. \tag{5.6}$$

It follows that

Lemma 5.1.

$$\sqrt{T}\left(\hat{\underset{\sim}{\delta}}^*_{g1,2SLS} - \underset{\sim}{\delta}^*_{g1} \right) \Longrightarrow \mathrm{N}(\underset{\sim}{0}, \sigma_g^2 (M^*_{z_{g1}x_1} M^{*-1}_{x_1x_1} M^*_{x_1z_{g1}})), \tag{5.7}$$

and are asymptotically independent of

$$\begin{bmatrix} T(\hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}} - \underset{\sim}{\delta}^*_{g2}) \\ \sqrt{T}\hat{\gamma}^*_{g,2\text{SLS}} \end{bmatrix} \Longrightarrow (RS^{-1}R')^{-1}RS^{-1} \begin{bmatrix} \int B_{x_2^*}\,\mathrm{d}B_{\varepsilon_g} \\ \mathrm{N}(0,\sigma_g^2) \end{bmatrix}, \tag{5.8}$$

where

$$R = \begin{pmatrix} \int B_{z^*_{g2}} B'_{x^*_2}\,\mathrm{d}r & \int B_{z^*_{g2}}\,\mathrm{d}r \\ \int B'_{x^*_2}\,\mathrm{d}r & 1 \end{pmatrix},$$

$$S = \begin{pmatrix} \int B_{x^*_2} B'_{x^*_2}\,\mathrm{d}r & \int B_{x^*_2}\,\mathrm{d}r \\ \int B'_{x^*_2}\,\mathrm{d}r & 1 \end{pmatrix}.$$

Since B_{ε_g} is not asymptotically independent of $B_{x_2^*}$, the 2SLS estimator of (5.2) has the same problem as the 2SLS estimator (3.1), namely, the limiting distribution of $\hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}}$ is nonstandard because of the long-run endogeneities between X_2^* and $\underset{\sim}{\varepsilon}_g$. Therefore, the Wald test statistic of the form (3.8) may not be asymptotically χ^2 distributed.

Transform (5.2) in the form of (3.11),

$$\underset{\sim}{w}_g = Z_g^{**}\underset{\sim}{\delta}^{**}_g + \underset{\sim}{e}\,\gamma_g^{**} + \underset{\sim}{\varepsilon}_g, \tag{5.9}$$

where Z_g^{**} and $\underset{\sim}{\delta}^{**}_g$ are defined after (3.11) and $\gamma_g^{**} = \gamma_g$. The modified 2SLS for (5.2) takes the form

$$\begin{pmatrix} \hat{\underset{\sim}{\delta}}_{g,\text{m2SLS}} \\ \hat{\gamma}_{g,\text{m2SLS}} \end{pmatrix} = \begin{pmatrix} \tilde{M}_g & \underset{\sim}{0} \\ \underset{\sim}{0}' & 1 \end{pmatrix} \begin{pmatrix} \hat{\underset{\sim}{\delta}}^{**}_{g,\text{m2SLS}} \\ \hat{\gamma}^{**}_{g,\text{m2SLS}} \end{pmatrix}, \tag{5.10}$$

where

$$\begin{pmatrix} \hat{\underset{\sim}{\delta}}^{**}_{g1,\text{m2SLS}} \\ \hat{\underset{\sim}{\delta}}^{**}_{g2,\text{m2SLS}} \\ \hat{\gamma}^{**}_{g,\text{m2SLS}} \end{pmatrix} = \left\{ \begin{pmatrix} Z_{g1}^{**\prime} \\ Z_{g2}^{**\prime} \\ \underset{\sim}{e} \end{pmatrix} (X^{**},\underset{\sim}{e}) \begin{pmatrix} X^{**\prime}X^{**} & X^{**\prime}\underset{\sim}{e} \\ \underset{\sim}{e}'X^{**} & T \end{pmatrix}^{-1} \begin{pmatrix} X^{**\prime} \\ \underset{\sim}{e}' \end{pmatrix} (Z^{**},\underset{\sim}{e}) \right\}^{-1}$$
$$\times \left\{ \begin{pmatrix} Z_g^{**\prime} \\ \underset{\sim}{e}' \end{pmatrix} (X^{**},\underset{\sim}{e}) \begin{pmatrix} X^{**\prime}X^{**} & X^{**\prime}\underset{\sim}{e} \\ \underset{\sim}{e}'X^{**\prime} & T \end{pmatrix}^{-1} \times \begin{pmatrix} X_1^{**\prime}\underset{\sim}{w}_g \\ X_2^{**\prime}\underset{\sim}{w}_g - \hat{C}_g \\ \underset{\sim}{e}'\underset{\sim}{w}_g \end{pmatrix} \right\}. \tag{5.11}$$

The limiting distribution of (5.11) can be derived from

$$\begin{pmatrix} \hat{\underset{\sim}{\delta}}^*_{g,\text{m2SLS}} \\ \hat{\underset{\sim}{\gamma}}^*_{g,\text{m2SLS}} \end{pmatrix} = \begin{pmatrix} D_g^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0}' & 1 \end{pmatrix} \begin{pmatrix} \hat{\underset{\sim}{\delta}}^{**}_{g,\text{m2SLS}} \\ \hat{\gamma}^{**}_{g,\text{m2SLS}} \end{pmatrix}. \tag{5.12}$$

Using similar manipulations as Section 3, it can be shown that

Lemma 5.2. *The limiting distribution of* $\sqrt{T}\left(\hat{\underset{\sim}{\delta}}^*_{g1,\text{m2SLS}} - \underset{\sim}{\delta}^*_{g1}\right)$ *is of the form* (3.21) *and is asymptotically independent of*

$$\begin{bmatrix} T\left(\hat{\underset{\sim}{\delta}}^*_{g2,\text{m2SLS}} - \underset{\sim}{\delta}^*_{g2}\right) \\ \sqrt{T}\hat{\gamma}^*_{g,\text{m2SLS}} \end{bmatrix} \Longrightarrow (RS^{-1}R')^{-1}RS^{-1}\begin{bmatrix} \int B_{x_2^*}\,\mathrm{d}B_{\varepsilon_g.x_2^*} \\ \mathrm{N}(0,\sigma_g^2) \end{bmatrix}. \tag{5.13}$$

Since $B_{\varepsilon_g.x_2^*}$ is asymptotically independent of $B_{x_2^*}$, the modified 2SLS is either normally distributed or mixed normally distributed.

Similarly, one can derive the alternatively modified 2SLS in the form similar to that of (4.3) and its limiting distribution is either normal or mixed normal.

When $\underset{\sim}{\gamma} \neq 0$, then some or all elements of $\underset{\sim}{w}_t$ are $I(1)$ with drift. As $T \to \infty$, those $I(1)$ elements of $\underset{\sim}{w}_t$ with nonzero drift will be dominated by the trend term $\underset{\sim}{h}\,t$, where $\underset{\sim}{h} = A_0^{-1}\underset{\sim}{\gamma}$. However, as noted by Sims et al. (1990) those elements of $\underset{\sim}{w}_t$ with nonzero drifts will be perfectly collinear. To derive the limiting distribution of 2SLS or modified 2SLS or alternatively modified 2SLS, we can follow the Sims et al. (1990) to transform $\underset{\sim}{w}_t$ into $\underset{\sim}{w}^*_t = H\underset{\sim}{w}_t$, where H is an $m \times m$ nonsingular matrix of the form

$$H = \begin{bmatrix} 1 & \cdot & \ldots & 0 & -(h_1/h_m) \\ 0 & 1 & \ldots & \cdot & -(h_2/h_m) \\ \ldots & \cdot & \ldots & \cdot & \ldots \\ 0 & \cdot & \ldots & 1 & -(h_{m-1}/h_m) \\ 0 & \cdot & \ldots & 0 & 1 \end{bmatrix}, \tag{5.14}$$

and there is no loss of generality in assuming $h_m \neq 0$. The resulting $w^*_{gt} = w_{gt} - (h_g/h_m)w_{mt}, g = 1, \ldots, m-1$, becomes $I(1)$ without drift and $w^*_{mt} = w_{mt}$ remains $I(1)$ with drift. Similarly, (5.1) can be expressed in terms of $\underset{\sim}{w}^*_t$

$$A(L)H^{-1}\underset{\sim}{w}^*_t = \underset{\sim}{\gamma} + \underset{\sim}{\varepsilon}_t, \tag{5.15}$$

and the gth equation of (5.15) can be expressed in the form

$$\underset{\sim}{w}^*_g = \tilde{Z}_g\tilde{\underset{\sim}{\delta}}_g + \underset{\sim}{e}\gamma_g + \underset{\sim}{\varepsilon}_g, \tag{5.16}$$

where $\tilde{Z}_g$ denotes the matrix of T observed current and lagged $\underset{\sim t}{w}^*$ that appear in the gth equation. We can transform (5.16) into the form in terms of $I(0)$, $I(1)$ without drift and $I(1)$ with drift variables:

$$\underset{\sim g}{w}^* = \tilde{Z}^*_g \underset{\sim g}{\tilde{\delta}}^* + \underset{\sim}{e}\gamma^*_g + \underset{\sim g}{\varepsilon}, \tag{5.17}$$

where $\gamma^*_g = \gamma_g$, $\tilde{Z}^*_g = \tilde{Z}_g \tilde{M}^*_g = (Z^*_{g1}, Z^*_{g2}, Z^*_{g3})$, with Z^*_{g1} denoting the ℓ_g-dimensional linearly independent zero mean $I(0)$ variables, Z^*_{g2} denoting the b_g linearly independent $I(1)$ variables without drift, and Z^*_{g3} denoting the $I(1)$ variable with drift, $\underset{\sim m,-p}{w}$, and $\left(\underset{\sim g1}{\tilde{\delta}}^{*\prime}, \underset{\sim g2}{\tilde{\delta}}^{*\prime}, \tilde{\delta}^*_{g3}\right)$ the corresponding partition of the transformed parameter vector $\underset{\sim g}{\tilde{\delta}}^* = \tilde{M}^{*-1}_g \underset{\sim g}{\tilde{\delta}}$.

Similarly, we can transform X into $X^* = X\tilde{M}^*_x = (X^*_1, X^*_2, X^*_3)$, where X^*_1, X^*_2 and X^*_3 consist of linearly independent $I(0)$, $I(1)$ without drift, and $I(1)$ with drift $\underset{\sim m,-p}{w}$, variables, respectively. Then the 2SLS of (5.16) can be written as the transformation of the 2SLS of $\underset{\sim g,\text{2SLS}}{\hat{\tilde{\delta}}^*}$,

$$\begin{pmatrix} \underset{\sim g,\text{2SLS}}{\hat{\tilde{\delta}}} \\ \underset{\sim g,\text{2SLS}}{\hat{\gamma}} \end{pmatrix} = \begin{pmatrix} \tilde{M}^*_g & \underset{\sim}{0} \\ \underset{\sim}{0} & 1 \end{pmatrix} \begin{pmatrix} \underset{\sim g,\text{2SLS}}{\hat{\tilde{\delta}}^*} \\ \underset{\sim g,\text{2SLS}}{\hat{\gamma}^*} \end{pmatrix}. \tag{5.18}$$

Lemma 5.3. *The limiting distribution of* $\sqrt{T}\left(\underset{\sim g1,\text{2SLS}}{\hat{\tilde{\delta}}^*} - \underset{\sim g1}{\tilde{\delta}^*}\right)$ *is asymptotically normally distributed with mean zero and variance covariance matrix of the form similar to* (3.21), *and is asymptotically independent of*

$$\begin{bmatrix} T(\underset{\sim g2,\text{2SLS}}{\hat{\tilde{\delta}}^*} - \underset{\sim g2}{\tilde{\delta}^*}) \\ T^{3/2}(\hat{\tilde{\delta}}^*_{g3,\text{2SLS}} - \tilde{\delta}^*_{g3}) \\ T^{1/2}(\hat{\gamma}^*_{g,\text{2SLS}} - \gamma^*_g) \end{bmatrix} \Longrightarrow (R^* S^{*-1} R^*)^{-1} R^* S^{*-1} \begin{bmatrix} \underset{\sim 1}{q} \\ q_2 \\ q_3 \end{bmatrix}, \tag{5.19}$$

where

$$R^* = \begin{bmatrix} \int B_{z^*_{g2}} B'_{x^*_2}\,\mathrm{d}r & h_m \int r B_{z^*_{g2}}\,\mathrm{d}r & \int B_{z^*_{g2}}\,\mathrm{d}r \\ h_m \int r B'_{x^*_2}\,\mathrm{d}r & h_m^2/3 & h_m/2 \\ \int B'_{x^*_2}\,\mathrm{d}r & h_m/2 & 1 \end{bmatrix},$$

$$S^* = \begin{bmatrix} \int B_{x_2^*} B'_{x_2^*}\, \mathrm{d}r & h_m \int r B_{x_2^*}\, \mathrm{d}r & \int B_{x_2^*}\, \mathrm{d}r \\ h_m \int r B'_{x_2^*}\, \mathrm{d}r & h_m^2/3 & h_m/2 \\ \int B'_{x_2^*}\, \mathrm{d}r & h_m/2 & 1 \end{bmatrix},$$

$\underset{\sim 1}{q} = \int B_{x_2^*}\, \mathrm{d}B_{\varepsilon_g}, q_2 \sim \mathrm{N}(0, \frac{1}{3}\sigma_g^2 h_m^2)$, *and* $q_3 \sim \mathrm{N}(0, \sigma_g^2)$.

Although $\underset{\sim g1,2\mathrm{SLS}}{\hat{\tilde{\delta}}^*}$ *and* $\underset{\sim g2,2\mathrm{SLS}}{\hat{\tilde{\delta}}^*}$ are asymptotically normal, $\underset{\sim g3,2\mathrm{SLS}}{\hat{\tilde{\delta}}^*}$ is not asymptotically mixed normal. Since the 2SLS of (5.1) (or (5.15)) is a linear combination of $\underset{\sim g1,2\mathrm{SLS}}{\hat{\tilde{\delta}}^*}$, $\underset{\sim g2,2\mathrm{SLS}}{\hat{\tilde{\delta}}^*}$ and $\underset{\sim g3,2\mathrm{SLS}}{\hat{\tilde{\delta}}^*}$, the Wald test statistic (3.8) again may not be asymptotically chi-squaredistributed. To ensure that the Wald test statistic be asymptotically chi-square distributed, the modified 2SLS or the alternatively modified 2SLS can be applied to ensure the asymptotic mixed normality of the estimated $\underset{\sim g2}{\tilde{\delta}^*}$.

6. Monte Carlo comparisons

In this section, a small simulation study is conducted to compare the finite sample performance of the 2SLS, M2SLS and A2SLS estimators. For each estimator, we compute its bias, root mean square estimation error, the size of the Wald test where critical values are derived from the conventional chi-square distributions. All computations are performed in MATLAB. It is hoped that this simulation study will shed some light on the choice of the estimators in finite sample.

We consider a three variable vector time series $\{\underset{\sim t}{w}\}_{t=-1}^{T}$ generated by a second-order structural VAR model of the form

$$A_0 \underset{\sim t}{w} = A_1 \underset{\sim t-1}{w} + A_2 \underset{\sim t-2}{w} + \underset{\sim t}{\varepsilon}, \tag{6.1}$$

where $\underset{\sim t}{\varepsilon} \sim \mathrm{N}(0, \Sigma_{\varepsilon\varepsilon})$. We let (6.1) be identified by the exclusion restrictions of the form

$$A_0 = \begin{pmatrix} 1 & a_{0,12} & 0 \\ 0 & 1 & a_{0,23} \\ a_{0,31} & 0 & 1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} a_{1,11} & a_{1,12} & 0 \\ 0 & a_{1,22} & a_{1,23} \\ a_{1,31} & 0 & a_{1,33} \end{pmatrix} \quad \text{and}$$

$$A_2 = \begin{pmatrix} a_{2,11} & a_{2,12} & 0 \\ 0 & a_{2,22} & a_{2,23} \\ a_{2,31} & 0 & a_{2,33} \end{pmatrix}.$$

To generate the time series $\{\underset{\sim t}{w}\}_{t=-1}^{T}$, we initialize the system at $t = -51$ with $(\underset{\sim -50}{w}, \underset{\sim -51}{w}) = (\underset{\sim}{0}, \underset{\sim}{0})$. A sequence of independent trivariate standard normal random variables $\{\underset{\sim}{e}_t\}_{t=-49}^{T}$ is generated by the RANDN function of MATLAB. Let

$$\Gamma = \begin{pmatrix} 1 & -0.5 & 0.3 \\ -0.5 & 0.9 & 0.4 \\ 0.3 & 0.4 & 2.5 \end{pmatrix}^{1/2} \quad \text{and} \quad \underset{\sim}{\varepsilon}_t = \Gamma \underset{\sim}{e}_t,$$

so that $\{\underset{\sim}{\varepsilon}_t\}_{t=-49}^{T}$ is a sequence of independent normal random variables with mean $\underset{\sim}{0}$ and covariance matrix Γ. To generate $\{\underset{\sim}{w}_t\}_{t=-49}^{T}$, we use the following parameter values of (A_0, A_1, A_2):

$$A_0 = \begin{pmatrix} 1 & -0.4 & 0 \\ 0 & 1 & 0.8 \\ 0.6 & 0 & 1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0.2 & -0.1 & 0 \\ 0 & 0.7 & 0.6 \\ 0.2 & 0 & 0.4 \end{pmatrix} \quad \text{and}$$

$$A_2 = A_0 - A_1 + \underset{\sim}{\alpha}' \underset{\sim}{\beta},$$

$$DGP1 : \underset{\sim}{\alpha} = \underset{\sim}{\beta} = (0 \quad 0 \quad 0),$$

$$DGP2 : \underset{\sim}{\alpha} = (0 \quad -0.4 \quad 0), \underset{\sim}{\beta} = (0 \quad 1 \quad 2),$$

$$DGP3 : \underset{\sim}{\alpha} = \begin{pmatrix} -0.5 & 0 & -0.3 \\ 0.25 & -0.4 & 0 \end{pmatrix} \quad \text{and} \quad \underset{\sim}{\beta} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix},$$

It is easy to check that $|A_0| \neq 0$ and that DGP1–DGP3 satisfy the rank condition for identification. In addition, DGP1 represents a system of full-rank $I(1)$ variables, DGP2 represents a system of $I(1)$ variables that has one linearly independent cointegrating relation, and DGP3 represents a system of $I(1)$ variables that has two linearly independent cointegrating relations.

To see if there are distortions of using normal approximation in hypothesis testing, we consider the following hypotheses: (A) (Test for the value of $a_{0,12}$ alone), $H_{\mathrm{A}} : a_{0,12} = c_0$; (B) (A joint test) H_{B}: $a_{0,12} = c_0, a_{1,12} = c_1, a_{2,12} = c_2$, where c_0, c_1 and c_2 denote the true values of $a_{0,12}, a_{1,12}$ and $a_{2,12}$, respectively.

Our analysis shows that the standard normal distribution provides a good approximation for the conventional t-statistic for H_{A}, be the estimator as 2SLS, M2SLS, A2SLS. On the other hand, chi-square distribution may or may not be a good approximation for the Wald-type statistic for H_{B}. For instance, Wald test of H_{B} for DGP3 involves standard limiting distribution, but not for DGP1 or DGP2. For DGP1, DGP2 and DGP3, we can transform $\mathrm{H_B}$ into the form of (3.7), then test B becomes a joint test of $a_{0,12} = c_0, a_{1,12} - a_{0,12} = c_1 - c_0$ and $a_{2,12} + a_{1,12} - a_{0,12} = c_2 + c_1 - c_0$. For DGP1 and DGP2, test B isolates the coefficient of the $I(1)$ regressor, $w_{2,t-2}, a_{2,12} + a_{1,12} - a_{0,12}$. For DGP3, it only involves the coefficients of $I(0)$ regressors, $\nabla w_{2,t}, \nabla w_{2,t-1}$ and $w_{2,t-2} - 2w_{1,t-2}$, hence the Wald statistic is asymptotically chi-square distributed. In other words, chi-square approximation is

Table 1
Average percentage estimation bias (Bias)

	2SLS	A2SLS	M2SLS (Parzen)			M2SLS (Tukey–Hanning)			M2SLS (quadratic)		
			$k=0.3$	$k=0.5$	$k=0.66$	$k=0.3$	$k=0.5$	$k=0.66$	$k=0.3$	$k=0.5$	$k=0.66$
DGP1											
$T=50$	0.3424	0.7615	0.2500	0.4133	1.0228	0.1916	0.6275	23.0699	0.4333	1.0467	16.1004
100	0.1854	0.0865	0.1363	0.0983	0.0899	0.1202	0.0893	0.1601	0.0984	0.0994	0.5187
200	0.0878	0.0327	0.0853	0.0654	0.0437	0.0802	0.0576	0.0406	0.0705	0.0461	0.5743
400	0.0405	0.0164	0.0388	0.0372	0.0352	0.0384	0.0368	0.0335	0.0374	0.0347	0.0282
DGP2											
$T=50$	0.2950	0.3477	0.3060	0.2637	0.1042	0.3054	0.6037	0.8564	0.2627	0.4577	0.8069
100	0.1372	0.1099	0.1463	0.1361	0.1079	0.1463	0.1229	0.6249	0.1417	0.1452	0.6046
200	0.0696	0.0696	0.0703	0.0637	0.0538	0.0685	0.0606	0.0452	0.0655	0.0539	0.6793
400	0.0399	0.0309	0.0401	0.0370	0.0335	0.0389	0.0361	0.0305	0.0375	0.0340	0.0189
DGP3											
$T=50$	0.3728	0.1120	0.3290	0.2352	0.2804	0.2817	0.2275	0.9336	0.2321	7.1467	14.3012
100	0.1821	0.1139	0.1370	0.1001	0.1298	0.1117	0.1153	0.1996	0.0956	0.1304	0.1188
200	0.0897	0.1420	0.0614	0.0520	0.0622	0.0541	0.0556	0.0530	0.0506	0.0630	0.3021
400	0.0470	0.0730	0.0199	0.0232	0.0353	0.0148	0.0283	0.0405	0.0172	0.0340	0.2077

not appropriate for DGP1 or DGP2, but is appropriate for DGP3 if the sample is of reasonable size.

Although the true DGP (6.1) has no constant term, in practice one usually estimates a VAR with an intercept. It therefore seems more appropriate in this study to include an intercept in the estimated structural VAR model. Sample sizes are fixed at $T=50$, 100, 200 and 400. The number of repetition is 1000.

Tables 1 and 2 present the average percentage estimation bias (Bias) and the average percentage root mean square estimation error (RMSE), respectively.[2] In terms of Bias, the 2SLS, A2SLS and M2SLS are of similar magnitude. In terms of RMSE, 2SLS seems to be the best for $T\leqslant 200$. However, RMSE of A2SLS and M2SLS decrease rapidly with sample size and are comparable to the RMSE of 2SLS at $T=400$.

Table 3 presents the actual sizes of tests A and B where the critical values are derived from the chi-square distribution with appropriate degrees of freedom. For 2SLS, actual sizes of test A are close to nominal sizes for all three data generating processes, which is consistent with the asymptotic results. Size distortions of test B are severe if the limiting distribution of Wald statistics involves the unit root distribution (DGP1 and DGP2); otherwise, chi-square distribution approximates

[2]The average percentage estimation bias (BIAS) is the absolute value of the percentage estimation bias averaged over the five coefficients in the first equation. The average percentage root mean square estimation error (RMSE) is the absolute value of the percentage root mean square estimation error averaged over the five coefficients in the first equation.

Table 2
Average percentage root mean square estimation error (RMSE)

	2SLS	A2SLS	M2SLS (Parzen)			M2SLS (Tukey–Hanning)			M2SLS (quadratic)		
			$k=0.3$	$k=0.5$	$k=0.66$	$k=0.3$	$k=0.5$	$k=0.66$	$k=0.3$	$k=0.5$	$k=0.66$
DGP1											
$T=50$	1.7706	9.6121	2.7496	11.3405	62.6826	4.4299	16.6485	764.1035	12.8586	50.9528	480.6787
100	5.8644	2.7339	4.3117	3.1076	2.8417	3.8020	2.8231	5.0624	3.1111	3.1419	16.4015
200	0.4835	0.7795	0.5044	0.5863	0.8735	0.5225	0.6564	1.2814	0.5622	0.8773	19.9671
400	0.3330	0.3796	0.3370	0.3436	0.3677	0.3389	0.3482	0.3913	0.3415	0.3611	1.2031
DGP2											
$T=50$	0.5835	3.3762	0.7141	1.0627	8.1326	2.5182	12.1853	34.4814	1.1544	15.181	61.472
100	4.3390	3.4753	4.6274	4.3047	3.4108	4.6254	3.8877	19.7622	4.4809	4.5921	19.1205
200	0.2041	0.7074	0.2173	0.2130	0.2262	0.2177	0.2156	0.3074	0.2149	0.2297	23.4914
400	0.1361	0.1470	0.1471	0.1399	0.1397	0.1460	0.1385	0.1502	0.1427	0.1389	0.2692
DGP3											
$T=50$	1.0346	1.1079	1.0849	1.1022	1.5102	1.1076	1.3548	33.7875	1.1217	229.4205	451.0891
100	0.6646	1.6121	0.7057	0.7058	0.7193	0.7107	0.7220	2.3839	0.7147	0.7975	2.3263
200	0.4499	1.2146	0.4800	0.4691	0.4637	0.4815	0.4665	0.6184	0.4772	0.4720	11.9911
400	0.3109	0.8831	0.3370	0.3212	0.3159	0.3354	0.3178	0.3552	0.3292	0.3157	6.5968

Table 3
Finite-sample size

	2SLS	A2SLS	M2SLS (Parzen)			M2SLS (Tukey–Hanning)			M2SLS (quadratic)		
			$k=0.3$	k 0.5	k 0.66	$k=0.3$	k 0.5	k 0.66	$k=0.3$	k 0.5	k 0.66
Finite-sample size: DGP1											
Test A: *test a single coefficient parameter*											
$\alpha=0.01$											
$T=50$	0.003	0.153	0.014	0.025	0.036	0.020	0.030	0.049	0.022	0.047	0.056
100	0.005	0.075	0.011	0.020	0.034	0.015	0.028	0.040	0.022	0.033	0.070
200	0.001	0.051	0.005	0.011	0.019	0.007	0.014	0.038	0.009	0.023	0.058
400	0.010	0.014	0.012	0.014	0.023	0.012	0.014	0.031	0.013	0.022	0.045
$\alpha=0.05$											
$T=50$	0.033	0.220	0.053	0.087	0.119	0.067	0.101	0.130	0.089	0.129	0.152
100	0.044	0.139	0.057	0.085	0.114	0.066	0.092	0.136	0.084	0.123	0.157
200	0.043	0.103	0.047	0.069	0.091	0.051	0.076	0.104	0.060	0.094	0.155
400	0.055	0.078	0.057	0.057	0.074	0.056	0.061	0.086	0.057	0.074	0.109
$\alpha=0.1$											
$T=50$	0.075	0.292	0.099	0.155	0.192	0.134	0.184	0.211	0.153	0.195	0.251
100	0.094	0.195	0.112	0.137	0.175	0.125	0.148	0.197	0.141	0.182	0.231
200	0.082	0.157	0.100	0.122	0.148	0.109	0.130	0.155	0.119	0.143	0.201
400	0.104	0.137	0.109	0.113	0.138	0.114	0.116	0.148	0.113	0.135	0.172
Test B: *joint test of several coefficient parameters*											
$\alpha=0.01$											
$T=50$	0.043	0.301	0.083	0.166	0.224	0.130	0.196	0.267	0.160	0.258	0.300
100	0.060	0.206	0.098	0.160	0.217	0.112	0.187	0.277	0.149	0.229	0.371

Table 3 (*continued*)

	2SLS	A2SLS	M2SLS (Parzen)			M2SLS (Tukey–Hanning)			M2SLS (quadratic)		
			$k=0.3$	k 0.5	k 0.66	$k=0.3$	k 0.5	k 0.66	$k=0.3$	k 0.5	k 0.66
200	0.053	0.143	0.074	0.113	0.173	0.083	0.130	0.215	0.095	0.181	0.299
400	0.058	0.074	0.069	0.089	0.152	0.075	0.107	0.189	0.083	0.133	0.265
$\alpha=0.05$											
$T=50$	0.159	0.402	0.196	0.265	0.320	0.228	0.304	0.378	0.269	0.363	0.424
100	0.144	0.311	0.201	0.267	0.337	0.233	0.293	0.394	0.260	0.357	0.489
200	0.159	0.240	0.172	0.215	0.277	0.184	0.235	0.313	0.204	0.277	0.404
400	0.176	0.160	0.174	0.214	0.267	0.190	0.225	0.307	0.208	0.225	0.363
$\alpha=0.1$											
$T=50$	0.238	0.469	0.277	0.337	0.401	0.304	0.371	0.458	0.341	0.441	0.502
100	0.243	0.390	0.291	0.350	0.426	0.315	0.376	0.478	0.351	0.446	0.564
200	0.262	0.316	0.268	0.299	0.359	0.282	0.324	0.409	0.292	0.365	0.499
400	0.279	0.219	0.292	0.314	0.353	0.292	0.319	0.385	0.300	0.344	0.436
Finite-sample size: DGP2											
Test A: test a single coefficient parameter											
$\alpha=0.01$											
$T=50$	0.032	0.069	0.055	0.060	0.069	0.056	0.054	0.082	0.061	0.077	0.091
100	0.013	0.041	0.035	0.034	0.031	0.042	0.031	0.042	0.042	0.033	0.057
200	0.015	0.046	0.029	0.029	0.020	0.030	0.027	0.023	0.030	0.024	0.047
400	0.012	0.026	0.055	0.021	0.012	0.043	0.017	0.014	0.026	0.014	0.018
$\alpha=0.05$											
$T=50$	0.097	0.138	0.126	0.141	0.157	0.141	0.149	0.179	0.139	0.170	0.209
100	0.066	0.100	0.113	0.106	0.092	0.114	0.097	0.108	0.107	0.101	0.149
200	0.052	0.095	0.107	0.087	0.073	0.103	0.081	0.077	0.097	0.078	0.107
400	0.046	0.073	0.138	0.075	0.049	0.118	0.061	0.063	0.106	0.049	0.075
$\alpha=0.1$											
$T=50$	0.141	0.208	0.189	0.212	0.228	0.222	0.219	0.252	0.216	0.250	0.275
100	0.123	0.169	0.182	0.166	0.149	0.193	0.154	0.159	0.173	0.156	0.225
200	0.103	0.158	0.180	0.154	0.130	0.185	0.137	0.129	0.166	0.131	0.107
400	0.095	0.121	0.180	0.132	0.101	0.179	0.119	0.098	0.156	0.109	0.122
Test B: joint test of several coefficient parameters											
$\alpha=0.01$											
$T=50$	0.175	0.275	0.264	0.329	0.393	0.317	0.353	0.450	0.330	0.427	0.527
100	0.123	0.206	0.203	0.224	0.254	0.206	0.241	0.300	0.220	0.262	0.387
200	0.094	0.152	0.137	0.151	0.193	0.144	0.165	0.209	0.146	0.190	0.280
400	0.119	0.158	0.138	0.122	0.130	0.132	0.116	0.122	0.122	0.118	0.187
$\alpha=0.05$											
$T=50$	0.366	0.424	0.408	0.461	0.515	0.446	0.495	0.566	0.465	0.540	0.642
100	0.282	0.331	0.348	0.359	0.381	0.346	0.365	0.411	0.365	0.387	0.496
200	0.248	0.295	0.305	0.291	0.302	0.298	0.293	0.325	0.296	0.315	0.382
400	0.272	0.288	0.312	0.240	0.237	0.300	0.228	0.254	0.272	0.213	0.274
$\alpha=0.1$											
$T=50$	0.479	0.512	0.514	0.541	0.594	0.528	0.560	0.629	0.546	0.611	0.696
100	0.418	0.446	0.461	0.456	0.463	0.462	0.462	0.473	0.455	0.471	0.560
200	0.380	0.390	0.422	0.389	0.388	0.435	0.389	0.410	0.412	0.401	0.444
400	0.402	0.362	0.399	0.321	0.326	0.399	0.315	0.315	0.370	0.303	0.375

Table 3 (*continued*)

	2SLS	A2SLS	M2SLS (Parzen)			M2SLS (Tukey–Hanning)			M2SLS (quadratic)		
			$k=0.3$	k 0.5	k 0.66	$k=0.3$	k 0.5	k 0.66	$k=0.3$	k 0.5	k 0.66
Finite-sample size: DGP3											
Test A: test a single coefficient parameter											
$\alpha=0.01$											
$T=50$	0.019	0.030	0.027	0.024	0.020	0.025	0.023	0.023	0.025	0.020	0.029
100	0.018	0.028	0.027	0.025	0.021	0.030	0.022	0.027	0.029	0.021	0.041
200	0.011	0.024	0.018	0.015	0.013	0.017	0.014	0.022	0.016	0.012	0.040
400	0.013	0.035	0.036	0.014	0.013	0.029	0.016	0.019	0.019	0.011	0.034
$\alpha=0.05$											
$T=50$	0.070	0.088	0.081	0.069	0.080	0.085	0.071	0.079	0.073	0.081	0.105
100	0.062	0.076	0.097	0.076	0.068	0.091	0.072	0.084	0.079	0.078	0.117
200	0.059	0.080	0.087	0.065	0.055	0.080	0.058	0.070	0.072	0.057	0.110
400	0.056	0.093	0.107	0.062	0.057	0.100	0.054	0.073	0.081	0.059	0.091
$\alpha=0.1$											
$T=50$	0.125	0.157	0.140	0.131	0.137	0.137	0.132	0.142	0.144	0.146	0.166
100	0.120	0.136	0.152	0.138	0.125	0.151	0.134	0.142	0.144	0.137	0.184
200	0.106	0.138	0.157	0.125	0.107	0.158	0.115	0.124	0.138	0.113	0.161
400	0.116	0.152	0.170	0.120	0.118	0.167	0.122	0.125	0.139	0.116	0.153
Test B: joint test of several coefficient parameters											
$\alpha=0.01$											
$T=50$	0.035	0.059	0.050	0.050	0.053	0.058	0.055	0.084	0.054	0.082	0.138
100	0.018	0.069	0.037	0.038	0.035	0.043	0.041	0.058	0.042	0.051	0.095
200	0.011	0.152	0.032	0.025	0.018	0.037	0.021	0.042	0.031	0.019	0.101
400	0.011	0.093	0.035	0.019	0.016	0.037	0.017	0.030	0.027	0.018	0.077
$\alpha=0.05$											
$T=50$	0.116	0.141	0.136	0.127	0.152	0.138	0.146	0.171	0.132	0.181	0.251
100	0.076	0.175	0.110	0.113	0.111	0.122	0.116	0.153	0.122	0.126	0.187
200	0.065	0.276	0.106	0.092	0.073	0.104	0.085	0.103	0.103	0.076	0.155
400	0.061	0.147	0.130	0.082	0.067	0.119	0.072	0.090	0.107	0.068	0.135
$\alpha=0.1$											
$T=50$	0.186	0.206	0.214	0.208	0.244	0.223	0.222	0.258	0.225	0.263	0.333
100	0.133	0.283	0.185	0.179	0.187	0.195	0.179	0.210	0.195	0.202	0.257
200	0.111	0.350	0.187	0.154	0.126	0.183	0.144	0.165	0.181	0.137	0.229
400	0.113	0.198	0.198	0.147	0.118	0.192	0.128	0.134	0.171	0.123	0.186

well as sample size increases. For test B, the 2SLS seems to have smaller size distortions than A2SLS and M2SLS for $T \leqslant 200$. However, for DGP1 and DGP2 the size distortion for 2SLS remains largely unchanged as T increases. On the other hand the performance of A2SLS and M2SLS appear to rapidly improve with T.

It is worth noticing that the results of M2SLS are sensitive to the choice of the bandwith parameter and the kernel function. Our results does not corroborate the findings in Yamada and Toda (1998), in which Monte Carlo experiments was conducted to examine the size distortions of Granger causality test in the standard VAR framework. Yamada and Toda studied the fully modified VAR estimator (FM-VAR) with various kernel functions and bandwidth parameters and found that

Parzen kernel with bandwidth parameter being the closest integer to $T^{0.66}$ gives the least size distortions for most combinations of parameter values and sample sizes ranging from 50 to 200. Our simulation results of test B (which is a Granger causality test in the structural VAR model) indicate that setting bandwidth parameter to the closest integer to $T^{0.66}$ produces larger size distortion whether we use Parzen or Tukey–Hanning or quadratic kernel. In addition, setting bandwidth parameter to the integer closest to $T^{0.66}$ seems to produce substantially large Bias and RMSE for small samples ($T = 50$). Our results appear to indicate that Parzen kernel with $k = 0.3$ or 0.5 does better than $k = 0.66$ on Tukey–Hanning or quadratic kernel.

7. Conclusions

In this paper, we consider the single equation estimation of a structural VAR model of nonstationary and possibly cointegrated variables without the prior knowledge of unit roots or rank of cointegration. When all variables are integrated of order 1, the conventional 2SLS and 3SLS estimators are consistent. However, some coefficient estimates of the transformed system are $\sqrt{T}$-convergent and asymptotically normally distributed while others are T-convergent and involve unit root distribution in the limit. Thus, Wald-type test statistics for the joint hypotheses may not be chi-square distributed. We propose a modified 2SLS estimator and an alternatively modified 2SLS estimator that have the desirable large sample property that coefficient estimates of the transformed system are either $\sqrt{T}$-consistent and asymptotically normally distributed or T-consistent and mixed normally distributed in the limit. The modified estimators also have the nice property that both $I(0)$ and $I(1)$ variables are allowed in the model and we can therefore avoid the error in testing the stationarity of the variables. Between the two, the modified 2SLS estimator requires nonparametric estimation of the long-run covariance matrix and the one-sided long-run covariance matrix, so its finite sample performance could be affected by the choice of the kernel function and the bandwidth parameter. In addition, since we can not approximate the asymptotic covariance matrix of the modified 2SLS estimator properly, the resulting Wald type test statistics may not be chi-square distributed and critical values that are based on chi-square distributions can be used to construct conservative tests only. In comparison, the alternatively modified 2SLS estimator does not require nonparametric estimation of the long-run covariance matrix or the one-sided long-run covariance matrix and its asymptotic covariance matrix can be properly approximated so that Wald test statistics remain chi-square distributed. On the other hand, the constrained maximum likelihood estimation in the first stage may be computationally more demanding.

Monte Carlo studies are also conducted to evaluate the finite sample performance of various estimators. Unfortunately, the desirable properties of A2SLS and M2SLS in large sample do not appear to carry over in finite sample. In general, we find that 2SLS, M2SLS and A2SLS have similar order of bias and RMSE. On the other hand, if the null hypothesis involves transformations of unit root components, the actual size of the Wald type test statistic based on the 2SLS estimates is severely distorted, so are

M2SLS or A2SLS in finite sample despite that their limiting distributions no longer involve the unit root distribution. However, the size distortion of the Wald test statistic based on M2SLS or A2SLS appears to diminish as sample size increases, while the conventional 2SLS remains the same as T increases. Therefore, if T is less than 200, it is probably more desirable to just use 2SLS, in particular, if the hypothesis an investigator is concerned with only involves a single parameter. One may attempt to use the M2SLS or A2SLS only when T is large and one's primary focus is not just in estimating unknown parameters, but also in testing joint hypotheses.

Acknowledgements

We would like to thank the editors and five referees for helpful comments and suggestions. We would also like to thank Echu Liu for computational assistance.

Appendix A. Proof of Theorem 3.1

Let D_{g2} be the submatrix of D_g that transforms Z_{g2}^{**} into $\tilde{W}_{g,-p}D_{g2} = \tilde{W}^*_{g,-p} = [\tilde{W}^*_{g1,-p}, \tilde{W}^*_{g2,-p}] = [\tilde{W}^*_{g1,-p}, Z^*_{g2}]$, where Z^*_{g2} consists of linearly independent $I(1)$ variables of $\underset{\sim gt}{\tilde{w}}$, and $\tilde{W}^*_{g1}$ consists of the remaining $I(1)$ variables that has been transformed into cointegrating relations. Let D_{w2} be the transformation matrix that transform W_{-p} into $W^*_{-p} = W_{-p}D_{w_2} = [W^*_{1,-p}, W^*_{2,-p}] = [W^*_{1,-p}, X^*_2]$, where X^*_2 denotes the $(m-r)$ linearly independent $I(1)$ variables of $\underset{\sim t-p}{w}$ and $W^*_{1,-p}$ denotes the $T \times r$ cointegrating relations of $\underset{\sim t-p}{w}$. Let $C^*_g = (W^{*\prime}_{-p} \bigtriangledown W^*_{-p} - T\Delta_{\bigtriangledown w^* \bigtriangledown w^*})$ $\Omega^{-1}_{\bigtriangledown w^* \bigtriangledown w^*}\Omega_{\bigtriangledown w^* \varepsilon_g}$, then $C^*_g = D'_{w2}C_g$ and $\hat{C}^*_g = D'_{w2}\hat{C}_g$. Partition

$$C^*_g = \begin{bmatrix} C^*_{g1} \\ C^*_{g2} \end{bmatrix},$$

where $C^*_{g_i} = (W^{*\prime}_{i,-p} \bigtriangledown W^*_{-p} - T\Delta_{\bigtriangledown w^*_i \bigtriangledown w^*})\Omega^{-1}_{\bigtriangledown w^* \bigtriangledown w^*}\Omega_{\bigtriangledown w^* \varepsilon_g}, i = 1, 2,$ and similarly for $\hat{C}^*_g$. Then $\underset{\sim g,\text{m2SLS}}{\hat{\delta}^*} = D_g^{-1}\underset{\sim g,\text{m2SLS}}{\hat{\delta}^{**}}$ can be written as

$$\underset{\sim g,\text{m2SLS}}{\hat{\delta}^*} = \{Z^{*\prime}_g X^*(X^{*\prime}X^*)^{-1}X^{*\prime}Z^*_g\}^{-1}\left\{Z^{*\prime}_g X^*(X^{*\prime}X^*)^{-1}\begin{pmatrix} X_1^{**\prime}\underset{\sim g}{w} \\ W^{*\prime}_{1,-p}\underset{\sim g}{w} - \hat{C}^*_{g1} \\ W^{*\prime}_{2,-p}\underset{\sim g}{w} - \hat{C}^*_{g2} \end{pmatrix}\right\}. \tag{A.1}$$

Under A1–A4, KL and BW, following the arguments of Phillips (1995), one can show that

$$H_g \begin{bmatrix} X_1^{**\prime} \underset{\sim g}{\varepsilon} \\ W_{1,-p}^{*\prime} \underset{\sim g}{\varepsilon} - \hat{C}_{g1}^* \\ W_{2,-p}^{*\prime} \underset{\sim g}{\varepsilon} - \hat{C}_{g2}^* \end{bmatrix} = \begin{bmatrix} T^{-1/2} \begin{pmatrix} X_1^{**\prime} \underset{\sim g}{\varepsilon} \\ W_{1,-p}^{*\prime} \underset{\sim g}{\varepsilon} - \hat{C}_{g1}^* \end{pmatrix} \\ T^{-1}(X_2^{*\prime} \underset{\sim g}{\varepsilon} - \hat{C}_{g_2}^*) \end{bmatrix}$$

$$\Longrightarrow \begin{pmatrix} \underset{\sim g1}{\xi} \\ \underset{\sim g2}{\xi} \end{pmatrix} \sim \begin{pmatrix} \mathrm{N}(\underset{\sim}{0}, \sigma_g^2 M_{x_1x_1}^*) \\ \int_0^1 B_{x_2^*}(r)\,\mathrm{d}B_{\varepsilon_g \cdot x_2^*}(r) \end{pmatrix}, \tag{A.2}$$

with $\underset{\sim g1}{\xi}$ independent of $\underset{\sim g2}{\xi}$, where $B_{\varepsilon_g \cdot x_2^*}(r) = B_{\varepsilon_g}(r) - \Omega_{\varepsilon_g \, \nabla x_2^*} \Omega^{-1}_{\nabla x_2^* \, \nabla x_2^*} B_{x_2^*}(r)$, which is independent of $B_{x_2^*}(r)$. The convergence is due to the fact that under assumptions *KL* and *BW*,

$$\hat{C}_{g1}^* = \mathrm{O_p}(k^{-2}) + \mathrm{O_p}((kT)^{-1/2})$$

and

$$\hat{C}_{g2}^* = T \int_0^1 B_{x_2^*}(r)\,\mathrm{d}B_{x_2^*}(r) \Omega^{-1}_{\nabla x_2^* \nabla x_2^*} \Omega_{\nabla x_2^* \varepsilon_g} + \mathrm{O_p}(T^{-1/2}) + \mathrm{O_p}(k^{3/2} T^{-1}) + \mathrm{o_p}(1).$$

Therefore $T^{-1/2} \hat{C}_{g1}^* = \mathrm{o_p}(1)$ and $T^{-1} \hat{C}_{g2}^* = \int_0^1 B_{x_2^*}(r)\,\mathrm{d}B_{x_2^*}(r) \Omega^{-1}_{\nabla x_2^* \nabla x_2^*} \Omega_{\nabla x_2^* \varepsilon_g} + \mathrm{o_p}(1)$. Theorem 4.1 follows from (A.2).

When the rank of cointegration, $r = 0$, the structural VAR model (2.1) implies that $\nabla \underset{\sim t}{w}$ follows a stationary VAR$(p-1)$ process of the form (B.3) with $\Pi^* \equiv 0$. When $r = 0, X_2^* = W_{-p}$, then $\Omega_{\nabla w^* \nabla w^*} = (I_m - \Sigma_{j=1}^{p-1} \Pi_j^*)^{-1} A_0^{-1} \Sigma_{\varepsilon\varepsilon} A_0^{\prime -1} (I_m - \Sigma_{j=1}^{p-1} \Pi_j^*)^{\prime -1}, \Omega_{\varepsilon_g \nabla w^*} = \Sigma_{\varepsilon\varepsilon, g} A_0^{\prime -1} (I_m - \Sigma_{j=1}^{p-1} \Pi_j^*)^{\prime -1}$, where $\Sigma_{\varepsilon\varepsilon,g}$ denotes the gth row of $\Sigma_{\varepsilon\varepsilon}$. Therefore

$$\sigma^2_{g.\nabla x_2^*} = \sigma_g^2 - \Omega_{\varepsilon_g \, \nabla x_2^*} \Omega^{-1}_{\nabla x_2^* \, \nabla x_2^*} \Omega_{\nabla x_2^* \varepsilon_g} = 0.$$

Corollary 4.2 follows from $\sigma^2_{g.\nabla x_2^*} = 0$.

Appendix B. Proof of Theorem 4.1

We first show that there exists a unique VARMA$(p, 1)$ representation (4.1) given (2.1) under A.1–A.4. We then show that the errors of the conditional equation

$$\underset{\sim g}{w^+} = Z_g \underset{\sim g}{\delta} + \underset{\sim g}{\varepsilon^+} \tag{B.1}$$

is independent of the innovations driving the common trends.

Multiplying A_0^{-1} to (2.1) yields the reduced form

$$\underset{\sim}{w}_t = \sum_{j=1}^{p} \Pi_j \underset{\sim}{w}_{t-j} + \underset{\sim}{v}_t, \tag{B.2}$$

where $\Pi_j = -A_0^{-1}A_j$ and $\underset{\sim}{v}_t = A_0^{-1}\underset{\sim}{\varepsilon}_t$. Expressing (B.1) in the error correction form, we have

$$\nabla \underset{\sim}{w}_t = \sum_{j=1}^{p-1} \Pi_j^* \nabla \underset{\sim}{w}_{t-j} + \Pi^* \underset{\sim}{w}_{t-p} + \underset{\sim}{v}_t, \tag{B.3}$$

where $\Pi_j^* = \sum_{\ell=1}^{j} \Pi_\ell - I$ and $\Pi^* = \sum_{\ell=1}^{p} \Pi_\ell - I$, Suppose that rank $(\Pi^*) = r$, i.e. there are r linearly independent cointegrating relations among $\underset{\sim}{w}_t$, we can write $\Pi^* = \underset{\sim}{\alpha}\,\underset{\sim}{\beta}'$, where $\underset{\sim}{\alpha}, \underset{\sim}{\beta}$ are $m \times r$ matrices of rank r. Let $\underset{\sim}{\alpha}_\perp$ be an $m \times (m-r)$ full column rank matrix such that $\underset{\sim}{\alpha}'_\perp \underset{\sim}{\alpha} = \underset{\sim}{0}$. We normalize $\underset{\sim}{\alpha}$ and $\underset{\sim}{\alpha}_\perp$ so that they are orthonormal matrices.

Let $R = [\underset{\sim}{\alpha}, \underset{\sim}{\alpha}_\perp]$. Then R is an $m \times m$ orthogonal matrix, i.e., $RR' = R'R = I_m$. Premultiplying (B.3) by R', we have

$$\begin{pmatrix} \underset{\sim}{\alpha}' \nabla \underset{\sim}{w}_t \\ \underset{\sim}{\alpha}'_\perp \nabla \underset{\sim}{w}_t \end{pmatrix} = \sum_{j=1}^{p-1} \begin{bmatrix} \underset{\sim}{\alpha}' \Pi_j^* \\ \underset{\sim}{\alpha}'_\perp \Pi_j^* \end{bmatrix} \nabla \underset{\sim}{w}_{t-j} + \begin{pmatrix} \underset{\sim}{\beta}' \\ \underset{\sim}{0} \end{pmatrix} \underset{\sim}{w}_{t-p} + \begin{pmatrix} \underset{\sim}{\alpha}' \underset{\sim}{v}_t \\ \underset{\sim}{\alpha}'_\perp \underset{\sim}{v}_t \end{pmatrix}. \tag{B.4}$$

Note that (B.4) is identical to

$$\begin{pmatrix} \underset{\sim}{\alpha}' \underset{\sim}{w}_t \\ \underset{\sim}{\alpha}'_\perp \nabla \underset{\sim}{w}_t \end{pmatrix} = \sum_{j=1}^{p} \begin{bmatrix} \underset{\sim}{\alpha}' \Pi_j \underset{\sim}{w}_{t-j} \\ \underset{\sim}{\alpha}'_\perp \Pi_j^* \nabla \underset{\sim}{w}_{t-j} \end{bmatrix} + \begin{bmatrix} \underset{\sim}{\alpha}' \underset{\sim}{v}_t \\ \underset{\sim}{\alpha}'_\perp \underset{\sim}{v}_t \end{bmatrix}, \tag{B.5}$$

where $\Pi_p^* \equiv 0$, which implies that

$$\begin{pmatrix} \underset{\sim}{\alpha}' \nabla \underset{\sim}{w}_t \\ \underset{\sim}{\alpha}'_\perp \nabla \underset{\sim}{w}_t \end{pmatrix} = \sum_{j=1}^{p} \begin{pmatrix} \underset{\sim}{\alpha}' \Pi_j \\ \underset{\sim}{\alpha}'_\perp \Pi_j^* \end{pmatrix} \nabla \underset{\sim}{w}_{t-j} + \begin{pmatrix} \underset{\sim}{\alpha}'(\underset{\sim}{v}_t - \underset{\sim}{v}_{t-1}) \\ \underset{\sim}{\alpha}'_\perp \underset{\sim}{v}_t \end{pmatrix}, \tag{B.6}$$

Multiplying R to (B.6) yields

$$\nabla \underset{\sim}{w}_t = \sum_{j=1}^{p} R \begin{pmatrix} \underset{\sim}{\alpha}' \Pi_j \\ \underset{\sim}{\alpha}'_\perp \Pi_j^* \end{pmatrix} \nabla \underset{\sim}{w}_{t-j} + \left[I_m - R \begin{pmatrix} I_r & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix} R'L \right] \underset{\sim}{v}_t. \tag{B.7}$$

Let $J(L) = I - J_1 L - \cdots - J_p L^p$, and $\Phi(L) = I - \Phi L$, where

$$J_j = R \begin{bmatrix} \underset{\sim}{\alpha}' \Pi_j \\ \underset{\sim}{\alpha}'_\perp \Pi_j^* \end{bmatrix}, \Phi = R \begin{bmatrix} I_r & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix} R'.$$

Then (B.7) can be rewritten as

$$J(L) \nabla \underset{\sim}{w}_t = \Phi(L) \underset{\sim}{v}_t, \tag{B.8}$$

with the properties that (i) the roots of $|J(L)| = 0$ lie outside the unit circle, and (ii) Φ is symmetric and idempotent. Property (i) follows from

$$RM(L)R'J(L) = \left[I_m - R \begin{pmatrix} \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & I_{m-r} \end{pmatrix} R'L \right] J(L) = \Pi(L), \tag{B.9}$$

where

$$M(L) = \begin{bmatrix} I_r & \underset{\sim}{0} \\ \underset{\sim}{0} & (1-L)I_{m-r} \end{bmatrix}.$$

Since $|\Pi(L)| = |I - \Pi_1 L - \cdots - \Pi_p L^p| = 0$ has $m - r$ unit roots and $m(p-1) + r$ roots outside the unit circle and $|M(L)| = 0$ has $m - r$ unit roots, clearly, all the roots of $|J(L)| = 0$ lie outside the unit circle. Therefore (B.8) is a stationary VARMA$(p, 1)$ model. However (B.8) is not invertible because $|\Phi(L)| = 0$ contains r unit roots, unless $r = 0$. However, the restriction that Φ is symmetric idempotent is sufficient for (B.8) to be the unique stationary VARMA$(p, 1)$ representation of $\nabla \underset{\sim}{w}_t$. To see this, we make the following observations.

First, since (B.2) is the true data generating process of $\nabla \underset{\sim}{w}_t$, for any stationary VARMA$(p, 1)$ representation of $\nabla \underset{\sim}{w}_t$, $C(L) \nabla \underset{\sim}{w}_t = \underset{\sim}{\eta}_t$, where $C(L) = I_m - \sum_{i=1}^{p} C_i L^i$ and $\underset{\sim}{\eta}_t$ is a MA(1) process, there exists a lag polynomial $\phi(L) = I_m - \phi L$ such that

$$(1 - L)C(L) = \phi(L)\Pi(L) \tag{B.10}$$

and $\underset{\sim}{\eta}_t = \phi(L)\underset{\sim}{v}_t$. Then (B.9) and (B.10) imply that $(1 - L)C(L) = \phi(L)RM(L) R'J(L)$, or equivalently

$$R'\phi(L)RM(L) = (1 - L)R'C(L)J(L)^{-1}R = (1 - L)D(L), \tag{B.11}$$

where $D(L) \equiv R'C(L)J(L)^{-1}R$. Since the left-hand side of (B.11) is a lag polynomial of maximum order 2, $D(L)$ must be a lag polynomial of maximum order 1. Let $D(L) = I_m - DL$ and $\tilde{\phi}(L) \equiv R'\phi(L)R = I_m - \tilde{\phi}L$. Some simple calculation indicates that (B.11) holds if and only if

$$\tilde{\phi} = \begin{pmatrix} I_r & \tilde{\phi}_{12} \\ \underset{\sim}{0} & \tilde{\phi}_{22} \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} \underset{\sim}{0} & \tilde{\phi}_{12} \\ \underset{\sim}{0} & \tilde{\phi}_{22} \end{pmatrix}.$$

So we have

$$D(L) = \begin{pmatrix} I_r & \tilde{\phi}_{12}L \\ \underset{\sim}{0} & I_{m-r} - \tilde{\phi}_{22}L \end{pmatrix} \quad \text{and}$$

$$\phi(L) = R\tilde{\phi}(L)R' = I_m - R\begin{pmatrix} I_r & \tilde{\phi}_{12} \\ \underset{\sim}{0} & \tilde{\phi}_{22} \end{pmatrix} R'L.$$

Second, the VARMA$(p, 1)$ representation $C(L)\nabla \underset{\sim t}{w} = \underset{\sim t}{\eta}$ is stationary if and only if roots of $|C(L)| = 0$ are outside unit circle. Since $C(L) = RD(L)R'J(L)$, this condition is equivalent to that all roots of $|D(L)| = |I_{m-r} - \tilde{\phi}_{22}L| = 0$ are outside the unit circle. In particular, it requires $|I_{m-r} - \tilde{\phi}_{22}| \neq 0$. Third, for

$$\phi = R\begin{pmatrix} I_r & \tilde{\phi}_{12} \\ \underset{\sim}{0} & \tilde{\phi}_{22} \end{pmatrix} R',$$

the restriction that ϕ is symmetric leads to

$$\phi = R\begin{pmatrix} I_r & \underset{\sim}{0} \\ \underset{\sim}{0} & \tilde{\phi}_{22} \end{pmatrix} R'$$

and $\tilde{\phi}_{22}$ being symmetric. When ϕ is further restricted to be idempotent, i.e. $\phi^2 = \phi$, we must have $\tilde{\phi}_{22} = \tilde{\phi}_{22}^2$, i.e., $\tilde{\phi}_{22}$ is idempotent. Then we can decompose $\tilde{\phi}_{22}$ as $\tilde{\phi}_{22} = EFE'$, where E is a $(m-r) \times (m-r)$ orthogonal matrix,

$$F = \begin{pmatrix} I_{R_\phi} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix}$$

and R_ϕ is the rank of $\tilde{\phi}_{22}$ (Judge et al., 1985, A.2.11, p. 942). Therefore, we have

$$I_{m-r} - \tilde{\phi}_{22}L = E\begin{pmatrix} (1-L)I_{R_\phi} & \underset{\sim}{0} \\ \underset{\sim}{0} & I_{m-r-R_\phi} \end{pmatrix} E',$$

and $|I_{m-r} - \tilde{\phi}_{22}L| = (1-L)^{R_\phi}$. Since the stationarity of $C(L)\nabla \underset{\sim t}{w} = \underset{\sim t}{\eta}$ requires that $|I_{m-r} - \tilde{\phi}_{22}| \neq 0$, we must have $R_\phi = 0$, and hence

$$\tilde{\phi}_{22} = 0 \quad \text{and} \quad \phi = R\begin{pmatrix} \underset{\sim r}{I} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix} R' = \Phi.$$

We have therefore proved the following lemma.

Lemma. *Suppose* (B.2) *is the true data generating process of* $\nabla \underset{\sim t}{w}$. *Consider a* VARMA$(p,1)$ *specification of* $\nabla \underset{\sim t}{w}$,

$$C(L)\nabla \underset{\sim t}{w} = \phi(L)\underset{\sim t}{v}, \tag{B.12}$$

where $\phi(L) = I_m - \phi L$. *The constraint that* ϕ *is symmetric idempotent is sufficient and necessary for* (B.7)/(B.8) *to be the unique stationary representation of* $\nabla \underset{\sim t}{w}$.

Let

$$\underset{\sim t}{\xi^*} = \begin{pmatrix} \underset{\sim 1t}{\xi^*} \\ \underset{\sim 2t}{\xi^*} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{\alpha'}(\underset{\sim t}{v} - \underset{\sim t-1}{v}) \\ \underset{\sim\perp}{\alpha'}\underset{\sim t}{v} \end{pmatrix}$$

then $\underset{\sim t}{\eta} = R\underset{\sim t}{\xi^*}$ and $\Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta}\underset{\sim t}{\eta} = \Sigma_{\varepsilon_g\xi_2^*}\Sigma^{-1}_{\xi_2^*\xi_2^*}\underset{\sim 2t}{\xi^*}$. Hence $\varepsilon^+_{gt} = \varepsilon_{gt} - \Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta}\underset{\sim t}{\eta} = \varepsilon_{gt} - \Sigma_{\varepsilon_g\xi_2^*}\Sigma^{-1}_{\xi_2^*\xi_2^*}\underset{\sim 2t}{\xi^*}$ is i.i.d. and uncorrelated with $\underset{\sim 2t}{\xi^*}$.

Furthermore, since (B.8) is stationary, we can rewrite it as

$$\nabla \underset{\sim t}{w} = J(L)^{-1}R\begin{pmatrix} \underset{\sim 1t}{\xi^*} \\ \underset{\sim 2t}{\xi^*} \end{pmatrix}.$$

It follows that $\nabla \underset{\sim t}{w}$ and ε^+_{gt} has zero long-run covariance, so is $\nabla \underset{\sim 2t}{x^*}$ and ε^+_{gt}. Therefore,

The process $(\nabla \underset{\sim 2t}{x^*}, \varepsilon^+_{gt})$ satisfies the multivariate invariance principle, i.e.

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{[Tr]} \nabla \underset{\sim 2t}{x^*} \\ T^{-1/2}\sum_{t=1}^{[Tr]} \varepsilon^+_{gt} \end{bmatrix} \Longrightarrow \begin{pmatrix} B_{x_2^*}(r) \\ B_{\varepsilon_g^+}(r) \end{pmatrix},$$

where $B_{x_2^*}(r)$ and $B_{\varepsilon_g^+}(r)$ are independent vectors of Brownian motion.

The maximum likelihood estimator of (B.8) is consistent and asymptotically normally distributed (for detail, see Wang, 2001). Therefore, we can use the estimated residuals, $\underset{\sim t}{\hat{\eta}} = \hat{\Phi}(L)\underset{\sim t}{\hat{v}}$ to construct $\underset{\sim gt}{\hat{w}^+}$.

Decompose $\underset{\sim g}{\hat{\varepsilon}^+}$ as

$$\underset{\sim g}{\hat{\varepsilon}^+} = \underset{\sim g}{\varepsilon^+} + [I_T \otimes (\Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta} - \hat{\Omega}_{\varepsilon_g\eta}\hat{\Omega}^{*-1}_{\eta\eta})]\underset{\sim}{\hat{\eta}} + [I_T \otimes \Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta}](\underset{\sim}{\eta} - \underset{\sim}{\hat{\eta}}). \tag{B.13}$$

Then,

$$T^{-1/2}W^{*\prime}_{1,-p}\underset{\sim g}{\varepsilon^+} \Longrightarrow \mathrm{N}(\underset{\sim}{0}, \sigma^2_{g+}M^*_{w_1w_1}), \tag{B.14}$$

and

$$T^{-1}X_2^{*\prime}\underset{\sim g}{\varepsilon^+} \Longrightarrow \int_0^1 B_{x_2^*}(r)\,\mathrm{d}B_{\varepsilon_g^+}(r). \tag{B.15}$$

The former (B.14) is asymptotically normal. The latter (B.15) is a mixed normal of the form $\int_{M^*_{x_2x_2}>0} \mathrm{N}(\underset{\sim}{0}, \sigma^2_{g+} M^*_{x_2x_2})\,\mathrm{d}P(M^*_{x_2x_2})$, because $B_{x_2^*}(r)$ and $B_{\varepsilon_g^+}(r)$ are independent Brownian motions.

Because $\underset{\sim}{\eta} - \underset{\sim}{\hat{\eta}} = \tilde{X}(\underset{\sim}{\hat{\theta}} - \underset{\sim}{\theta})$, as $T \to \infty$,

$$
\begin{aligned}
&T^{-1/2} W^{*\prime}_{1,-p}(I_T \otimes \Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta})(\underset{\sim}{\eta} - \underset{\sim}{\hat{\eta}}) \\
&\quad = (\Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta} \otimes T^{-1} W^{*\prime}_{1,-p}\tilde{X}) \cdot \sqrt{T}(\underset{\sim}{\hat{\theta}} - \underset{\sim}{\theta}) \\
&\quad \Longrightarrow (\Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta} \otimes M_{w_1^*\tilde{x}}) \cdot \mathrm{N}(\underset{\sim}{0},\ \mathrm{cov}\ (\underset{\sim}{\hat{\theta}}))
\end{aligned} \tag{B.16}
$$

which is a normal with mean 0 and covariance $(\Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta} \otimes M_{w_1^*\tilde{x}})$ Cov $(\underset{\sim}{\hat{\theta}})(\Omega^-_{\eta\eta}\Omega_{\eta\varepsilon_g} \otimes M'_{w_1^*\tilde{x}})$ with $M_{w_1^*\tilde{x}} = \operatorname{plim}\,(1/T)W^{*\prime}_{1,-p}\tilde{X}$.

$$
T^{-1}X_2^{*\prime}(I_T \otimes \Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta})(\underset{\sim}{\eta} - \underset{\sim}{\hat{\eta}}) = (\Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta} \otimes T^{-3/2}X_2^{*\prime}\tilde{X}) \cdot \sqrt{T}(\underset{\sim}{\hat{\theta}} - \underset{\sim}{\theta}) \xrightarrow{\mathrm{p}} \underset{\sim}{0}\,. \tag{B.17}
$$

Since $\hat{\Omega}_{\varepsilon_g\eta} \xrightarrow{\mathrm{p}} \Omega_{\varepsilon_g\eta}$ and $\hat{\Omega}^{*-1}_{\eta\eta} \xrightarrow{\mathrm{p}} \Omega^-_{\eta\eta}$ at rate $T^{1/2}$ and T^d, respectively, it follows that $\hat{\Omega}_{\varepsilon_g\eta}\hat{\Omega}^{*-1}_{\eta\eta} - \Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta} = (\mathrm{O}(T^{-d}), \mathrm{O}(T^{-1/2+d}))R$ (for detail, see Wang, 2001). To ensure the maximum rate of convergence, we let $d = \frac{1}{4}$. Then

$$
T^{-1/2}W^{*\prime}_{1,-p}[I_T \otimes (\Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta} - \hat{\Omega}_{\varepsilon_g\eta}\hat{\Omega}^{*-1}_{\eta\eta})]\underset{\sim}{\hat{\eta}} \xrightarrow{\mathrm{p}} \underset{\sim}{0},\ \text{for } p \geqslant 2, \tag{B.18}
$$

and

$$
T^{-1}X_2^{*\prime}[I_T \otimes (\Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta} - \hat{\Omega}_{\varepsilon_g\eta}\hat{\Omega}^{*-1}_{\eta\eta})]\underset{\sim}{\hat{\eta}} \xrightarrow{\mathrm{p}} \underset{\sim}{0}\,. \tag{B.19}
$$

at the rate $T^{1/4}$. Substituting (B.13)–(B.19) into (4.6) yields Theorem 4.1. Corollary 7.1 follows from the argument that the limiting distribution of $\underset{\sim g,\mathrm{a2SLS}}{\hat{\delta}}$ is given by the component that has a slower rate of convergence.

When rank of cointegration $r = 0, \Phi = \underset{\sim}{0}$ and $\underset{\sim t}{\eta} = \underset{\sim t}{v} = A_0^{-1}\underset{\sim t}{\varepsilon}$. It follows that $\Omega_{\varepsilon_g\eta}\Omega^-_{\eta\eta} = \Sigma_{\varepsilon_g\varepsilon}\Sigma^{-1}_{\varepsilon\varepsilon}A_0 = \underset{\sim og}{a'}$, where $\underset{\sim og}{a'}$ is the gth row of A_0. Then $\underset{\sim g,t}{\varepsilon^+} = \underset{\sim}{0}, \sigma^2_{g_+} = 0$ for $g = 1\ldots, m$. Corollary 4.2 follows. Theorem 4.1 and Corollary 4.2 imply that $\sqrt{T}(\underset{\sim g1,\mathrm{a2SLS}}{\hat{\delta}^*} - \underset{\sim g1}{\delta^*}) \Longrightarrow \mathrm{N}(\underset{\sim}{0}, \Sigma^*_{g1})$and $T(\underset{\sim g2,\mathrm{a2SLS}}{\hat{\delta}^*} - \underset{\sim g2}{\delta^*}) \xrightarrow{\mathrm{p}} \underset{\sim}{0}$, where Σ^*_{g1} is defined in Theorem 4.1 except that now Σ^*_{g1} becomes $\sigma^2_g(M^*_{z_{g1}x_1}M^{*-1}_{x_1x_1}M^*_{x_1z_{g1}})^{-1}$.

When $r > 0$, it is also possible for $\underset{\sim g2,\mathrm{a2SLS}}{\hat{\delta}^*}$ to be hyperconsistent for some g if $\underset{\sim og}{a'} = \underset{\sim}{d'}\underset{\sim\perp}{\alpha'}$. This follows from $\underset{\sim gt}{\varepsilon^+} = \varepsilon_{gt} - \mathrm{E}(\varepsilon_{gt}|\underset{\sim 2t}{\xi^*}) = \underset{\sim og}{a'}\underset{\sim t}{v} - \mathrm{E}(\underset{\sim og}{a'}\underset{\sim t}{v}|\underset{\sim\perp}{\alpha'}\underset{\sim t}{v})$ equaling zero if and only if $\underset{\sim og}{a}$ is a linear combination of $\underset{\sim\perp}{\alpha}$. From $\underset{\sim}{\alpha}\underset{\sim}{\beta'} = \Pi^* = -A_0^{-1}A_p^*$ where $A_p^* = \Sigma_{j=0}^p A_j, \underset{\sim og}{a'} = \underset{\sim}{d'}\underset{\sim\perp}{\alpha'}$ holds if and only if $\underset{\sim og}{a'}A_0^{-1}A_p^* = \underset{\sim}{0}$, i.e., the gth row of

$A_p^*, \underset{\sim p,g}{a^{*\prime}} = \underset{\sim}{0}'$. Therefore, $\underset{\sim g2,\mathrm{a2SLS}}{\hat{\delta}^*}$ is hyperconsistent if the gth equation is lying on the nonstationary direction with $\underset{\sim p,g}{a^{*\prime}} = \underset{\sim}{0}'$.

References

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

Hsiao, C., 1997a. Cointegration and dynamic simultaneous equations models. Econometrica 65, 647–670.

Hsiao, C., 1997b. Statistical properties of the two stage least squares estimator under cointegration. Review of Economic Studies 64, 385–398.

Hsiao, C., 2001. Identification and dichotomization of long- and short-run relations of cointegrated vector autoregressive models. Econometric Theory 17, 889–912.

Hsiao, C., Wang, S., 2004. Statistical inference of structural vector autoregressive process. Mimeo.

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

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

Judge, G.G., Griffiths, W.E., Carter Hill, R., Lûtkepohl, H., Lee, T.C., 1985. The Theory and Practice of Econometrics, second ed. Wiley, New York.

King, R.G., Plosser, C.I., Stock, J.H., Watson, M.W., 1991. Stochastic trends and economic fluctuations. American Economic Review 81, 819–840.

Klein, L.R., Evans, M.K., 1969. Econometric Gaming. MacMillan, New York.

Klein, L.R., Evans, M.K., Goldberger, A.S., 1955. An Econometric Model of the United States. North-Holland, Amsterdam.

Koopmans, T.C., Rubin, H., Leipnik, R.B., 1950. Measuring the equation systems of dynamic economics. In: Koopman, T.C. (Ed.), Statistical Inference in Dynamic Economic Models. Wiley, New York, pp. 53–237.

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

Sims, C.A., 1980. Macroeconomics and reality. Econometrica 48, 1–48.

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

Toda, H.Y., Phillips, P.C.B., 1993. Vector autoregression and causality. Econometrica 61, 1367–1393.

Tsay, R.S., Tiao, G.C., 1990. Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. Annals of Statistics 18, 220–250.

Wang, S., 2001. Two essays on time series econometrics—multiple time series analysis of the Asian currency crisis and estimation and inference of the nonstationary structural vector autoregressive model. Ph.D. dissertation submitted to the University of Southern California.

Yamada, H., Toda, H.Y., 1998. Inferences in possibly integrated vector autoregressive models: some finite sample evidence. Journal of Econometrics 86, 55–96.