Econometrics Journal (2007), volume **10**, pp. 49–81.
doi: 10.1111/j.1368-423X.2007.00199.x

Lag-augmented two- and three-stage least squares estimators for integrated structural dynamic models

CHENG HSIAO[*,†] AND SIYAN WANG[‡]

[*]*Department of Economics, University of Southern California, Los Angeles, CA 90089-0253, USA*
E-mail: chsiao@usc.edu

[†]*Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

[‡]*Department of Economics, University of Delaware, Newark, DE 19716*
E-mail: wangs@lerner.udel.edu

Received: February 2006

Summary We consider a lag-augmented two- or three-stage least-squares estimator for a structural dynamic model of non-stationary and possibly cointegrated variables without the prior knowledge of unit roots or rank of cointegration. We show that the conventional two- and three-stage least-squares estimators are consistent but contain non-standard distributions without the strict exogeneity assumption; hence the conventional Wald type test statistics may not be chi-square distributed. We propose a lag order augmented two- or three-stage least-squares estimator that is consistent and asymptotically normally distributed. Limited Monte Carlo studies are conducted to shed light on the finite sample properties of various estimators.

Keywords: *Structural vector autoregressions*, *Nonstationary time series*, *Cointegration*, *Hypothesis testing*, *Two- and three-stage least squares*.

1. INTRODUCTION

In this paper, we will consider the estimation of structural vector autoregressive models (SVAR) where there are possibly some unit roots and some cointegrating relations. The model is similar in spirit to the Cowles Commission structural equation specification in which each equation describes a behavioural or technological relation. However, unlike in Hsiao (1997a,b) in which strictly exogenous linearly independent $I(1)$ variables are assumed to be the driving force of the system, here, no exogeneity assumption has been imposed on any of the variables. In addition, our model does not assume prior knowledge of the direction of non-stationarity, the cointegrating relations, or rank of cointegration. The model is different from the reduced form VAR considered by Johansen (1988, 1991), Phillips (1995), Sims, Stock and Watson (1990), Toda and Yamamoto (1995) or Tsay and Tiao (1990) in that we allow more than one current variables to appear in each equation. It is also different from the currently popular approach of formulating a structural model in terms of the dynamic response to structural shocks in which the identification of a model is achieved through (i) the imposition of cointegration restrictions that impose constraints on the long-run multipliers to identify the long-run components of the shocks, (ii) assuming the innovations in the permanent components being uncorrelated with the innovations in the transitory components and (iii) the normalization conditions that decompose the components of permanent shocks (e.g. King *et al.* 1991). Since SVAR has been well established in the current literature for the approach that identifies the structural coefficients by placing restrictions on some of the contemporaneous coefficients and on the structure of the contemporaneous covariance matrix in a dynamic linear model, to avoid confusion we will follow a referee's suggestion to call our approach the structural dynamic model approach (SDM).

The SDM approach identifies the structural coefficients by placing restrictions on the coefficients on current and lagged variables in a dynamic model (e.g. see Lemma 1 in Section 2). The goal of interest is to identify behavioural relationships. It could also be to examine the contemporaneous and dynamic interactions among the joint dependent variables and to derive a dynamic-multiplier which could be the path of the effect of a change in a policy variable. On the other hand, the SVAR places restrictions with direct economic interpretation on the error terms (e.g. Blanchard and Quah 1989). The goal of interest is usually a structural impulse-response function which is the path of the effect of an unexpected change to an endogenous variables. However, there is a close relationship between the two. As pointed out by a referee, any identification condition of a SVAR can be obtained from some identification condition of a SDM. It is also possible to derive a structural impulse-response function from a SDM using the formula of Lütkepohl (2005). However, if the shocks are not identically distributed, it will be difficult to interpret the structural impulse-response function but the identification and estimation of a SDM would not be affected.

Phillips and Durlauf (1986) and Stock (1987) have shown that the least-squares estimator with integrated regressors remains consistent when the regressors and errors of the equations are correlated provided that the integrated regressors are not cointegrated. However, in a dynamic framework, even though regressors may be $I(1)$, the current and lagged variables are trivially cointegrated. Therefore, the least-squares estimator is not consistent (e.g. Hsiao 1997b). While the two- and three-stage least-squares estimators are consistent, their limiting distributions are functions of unit root distribution (Hsiao and Wang 2006). Hence, the Wald test statistics may not be asymptotically chi-square distributed.

Hsiao and Wang (2006), generalizing the approach of Kitamura and Phillips (1997) and Phillips (1995), propose two modified 2SLS estimators that are asymptotically normally or mixed normally distributed. Unfortunately, the limited Monte Carlo studies show that such estimators do not perform well in finite samples. In this paper, we propose a lag-augmented 2SLS or 3SLS estimator based on the lag augmentation idea of Toda and Yamamoto (1995) for estimating a reduced form VAR model (also see Dolado and Lütkepohl 1996). The limited Monte Carlos appear to indicate that the lag-augmented 2SLS and 3SLS estimators, besides being computationally simpler, outperform the two modified estimators proposed by Hsiao and Wang (2006).

In Section 2, we set up the basic model. Section 3 gives the condition that the test statistics constructed by 2SLS or 3SLS estimator remain asymptotically chi-square distributed despite the fact that some estimated coefficients have non-standard limiting distribution. However, since these conditions require very strong prior information, in Section 4 we suggest a lag order augmented 2SLS (LA2SLS) or 3SLS (LA3SLS) estimator that is asymptotically normally distributed. Section 5 provides some Monte Carlo simulation comparisons of OLS, 2SLS, 3SLS, LA2SLS and LA3SLS estimators as well as the modified estimators suggested by Hsiao and Wang (2006) that have the desirable asymptotic properties of being normally and mixed normally distributed. Conclusions are given in Section 6.

2. THE MODEL

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

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

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

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

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

$$\tilde{M} = \begin{bmatrix} I_m & I_m & \ldots & I_m \\ \underset{\sim}{0} & I_m & \ldots & I_m \\ \underset{\sim}{0} & \underset{\sim}{0} & \ldots & I_m \\ \ldots & \ldots & \ldots & \\ \underset{\sim}{0} & \ldots & \underset{\sim}{0} & I_m \end{bmatrix}. \tag{2}$$

Transforming (1) into an error-correction representation, we have

$$\sum_{j=0}^{p-1} A_j^* \nabla \underset{\sim}{w}_{t-j} + A_p^* \underset{\sim}{w}_{t-p} = \underset{\sim}{\epsilon}_t, \tag{3}$$

where $\nabla = (1 - L)$, $A_j^* = \sum_{\ell=0}^{j} A_\ell$, $j = 0, 1, \ldots, p$. Let $A^* = [A_1^*, \ldots, A_p^*] = [\tilde{A}_1^*, A_p^*]$, 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 (1) as defined in Hsiao (2001).

Model (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.[2] 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 any element of $\underset{\sim}{w}_t$ (e.g. Koopmans *et al.* 1950; Hsiao 1997a). Multiplying A_0^{-1} to (1) yields the conventional VAR which may be viewed as a reduced form representation of (1).

In this paper, we are concerned with statistical inference of (1). If the roots of $|A(L)| = 0$ are all outside the unit circle, $\underset{\sim}{w}_t$ is stationary. It is well known that the least-squares estimator

[1]One may incorporate time trends in (1) by treating w_t as deviations from the trends or add intercepts and time trends as additional terms to the left-hand side of (1). The limiting distributions of the intercepts and trends coefficients can be found in Hsiao and Wang (2006).

[2]Using the idea of indirect least squares, one can recover A from the reduced form estimates based on the Johansen approach by introducing prior restrictions of the form stated in Lemma 1 (see Lütkepohl (2005) and references cited therein).

(LS) is inconsistent. The 2SLS and 3SLS using lagged $\underset{\sim}{w}_t$ as instruments are consistent and asymptotically normally distributed (e.g. Malinvaud 1980). Therefore, we will 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, $0 \leq r \leq m-1$.
(b) $\underset{\sim}{\alpha}'_{\perp} J \underset{\sim}{\beta}_{\perp}$ is nonsingular, where $J = \sum_{j=0}^{p-1} A_j^*$, $\underset{\sim}{\alpha}_{\perp}$ and $\underset{\sim}{\beta}_{\perp}$ are $m \times (m-r)$ matrices of full column rank such that $\underset{\sim}{\alpha}'_{\perp}\underset{\sim}{\alpha} = \underset{\sim}{0} = \underset{\sim}{\beta}'_{\perp}\underset{\sim}{\beta}$. (If $r = 0$, then we take $\underset{\sim}{\alpha}_{\perp} = I_m = \underset{\sim}{\beta}_{\perp}$.)

Under A1–A4, $\underset{\sim}{w}_t$ has r cointegrating vectors (the columns of β) and $m - r$ unit roots. As shown by Johansen (1988, 1991) and Toda and Phillips (1993), Ã4 ensures that the Granger representation theorem (Engle and Granger 1987) applies, so that $\nabla \underset{\sim}{w}_t$ is stationary, $\beta' \underset{\sim}{w}_t$ is stationary and $\underset{\sim}{w}_t$ is a vector of $I(1)$ processes.

Suppose that the gth equation of (1) satisfies the prior restrictions $\underset{\sim}{a}'_g \Phi_g = \underset{\sim}{0}'$, where a'_g 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 $\underset{\sim}{a}'_g \Phi_g = \underset{\sim}{0}'$ is equivalent to the existence of prior restrictions $\underset{\sim}{a}_g^{*\prime}\Phi_g^* = \underset{\sim}{0}'$, where $\underset{\sim}{a}_g^{*\prime}$ is the gth row of A^*. Hsiao (2001) has shown the following.

Lemma 1 *Suppose that the gth equation of (1) is subject to the prior restrictions* $\underset{\sim}{a}'_g \Phi_g = \underset{\sim}{0}'$. *A necessary and sufficient condition for the identification of the gth equation of (1) or (2) is that*

$$\operatorname{rank}(A\Phi_g) = m - 1, \tag{4}$$

or

$$\operatorname{rank}\left(A^*\Phi_g^*\right) = m - 1. \tag{5}$$

Remark 1. While the direction of non-stationarity or rank of cointegration can provide information for identification, the identification condition (4) or (5) does not require such prior information. Many econometric models are identified using information (4) alone. For instance, the three behavioural equations of Klein (1950) interwar model are: consumption being a function of current private wage income and current and lagged profit; investment being a function of current and lagged profit and lagged capital stock; private wage being a function of current and lagged output and time trend. So are the medium-sized Klein and Goldberger (1955) model and the large-scale Wharton quarterly model (Klein and Evans 1969). The Klein–Goldberger model consists of 25 equations. Apart from accounting identities, the behavioural equations in each period are functions of current and lagged endogenous variables up to the fifth order, as well as some exogenous inputs. For instance, consumption is a function of current wage income, current profit, current farm income, current population, lagged consumption and lagged private liquid assets. The original Wharton model consists of 76 behavioural equations with lags up to nine quarters. While some of the behavioural equations use first differences for some variables and levels for others, none of these models are identified through prior information on the rank of cointegration or location of unit roots, but through condition (4).

3. 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 (1) be written as

$$\underset{\sim}{w}_g = Z_g \underset{\sim}{\delta}_g + \underset{\sim}{\epsilon}_g, \tag{6}$$

where $\underset{\sim}{w}_g$ and $\underset{\sim}{\epsilon}_g$ denote the $T \times 1$ vectors of $(w_{g1}, \ldots, w_{gT})'$ and $(\epsilon_{g1}, \ldots, \epsilon_{gT})'$, respectively, and Z_g denotes the $T \times [(p+1)g_\Delta - 1]$ matrix of g_Δ included current and lagged variables of $\underset{\sim}{w}_t$.

Phillips and Durlauf (1986) and Stock (1987) have shown that the least-squares estimator with integrated regressors is consistent even when the regressors and the errors are correlated. However, the basic assumption underlying their result is that the regressors are not cointegrated. In a dynamic framework even though $\underset{\sim}{w}_{t-j}$ are $I(1)$, the current and lagged variables are trivially cointegrated. It is shown in Hsiao (1997a) that when contemporaneous joint-dependent variables also appear as explanatory variables in (6), applying least-squares method to (6) does not always yield consistent estimator for $\underset{\sim}{\delta}_g$.

Using lagged variables as instruments, the 2SLS estimator of $\underset{\sim}{\delta}_g$ is given by

$$\hat{\underset{\sim}{\delta}}_{g,\text{2SLS}} = \left[Z_g' X(X'X)^{-1}X'Z_g\right]^{-1}\left[Z_g' X(X'X)^{-1}X'\underset{\sim}{w}_g\right], \tag{7}$$

where $X = (W_{-1}, W_{-2}, \ldots, W_{-p})$ and W_{-j} denotes a $T \times m$ matrix representation of $\underset{\sim}{w}_{t-j}'$. To derive the limiting properties of 2SLS, we note that we can transform (6) into a regression model that involves $I(0)$ and $I(1)$ regressors. Let M_g be the non-singular 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 b_g-dimensional full rank $I(1)$ variables,[3] then

$$\begin{aligned} \underset{\sim}{w}_g &= Z_g M_g M_g^{-1} \underset{\sim}{\delta}_g + \underset{\sim}{\epsilon}_g \\ &= Z_g^* \underset{\sim}{\delta}_g^* + \underset{\sim}{\epsilon}_g \end{aligned}, \tag{8}$$

where $\underset{\sim}{\delta}_g^* = M_g^{-1}\underset{\sim}{\delta}_g = (\underset{\sim}{\delta}_{g1}^{*\prime}, \underset{\sim}{\delta}_{g2}^{*\prime})'$ with $\underset{\sim}{\delta}_{g1}^*$ and $\underset{\sim}{\delta}_{g2}^*$ denoting the $\ell_g \times 1$ and $b_g \times 1$ vector, respectively. Such transformation always exists. For instance, if no cointegrating relation exists among the g_Δ included variables, say $\tilde{\underset{\sim}{w}}_{gt}$, then $b_g = g_\Delta$, Z_{g1}^* consists of the current and $p-1$ lagged $\nabla\tilde{\underset{\sim}{w}}_{gt}$, and Z_{g2}^* is simply the $T \times b_g$ (or $T \times g_\Delta$) matrix of the included $\tilde{\underset{\sim}{w}}_{gt}$ lagged by p periods, $\tilde{\underset{\sim}{w}}_{g,t-p}$. On the other hand, if there exist $g_\Delta - b_g$ linearly independent cointegrating relations among the g_Δ included variables, $\tilde{\underset{\sim}{w}}_{gt}$, then Z_{g1}^* consists of the current and $p-1$ lagged $\nabla\tilde{\underset{\sim}{w}}_{gt}$ and cointegrating relations $\tilde{\underset{\sim}{w}}_{g,t-p}'\underset{\sim}{\pi}_g$, where $\underset{\sim}{\pi}_g$ is a $g_\Delta \times (g_\Delta - b_g)$ matrix of constants, and Z_{g2}^* is the $T \times b_g$ matrix of the full rank $I(1)$ variables $\tilde{\underset{\sim}{w}}_{g2,t-p}$. Similarly, there exists a non-singular transformation matrix M_x such that $XM_x = X^* = (X_1^*, X_2^*)$ where X_1^* consists of the linearly independent $I(0)$ variables and X_2^* consists of the full rank $I(1)$ variables.

Remark 2. The precise form of M_g and M_x need not be known a priori. They are introduced to facilitate derivation of the limiting properties of the 2SLS or augmented 2SLS or 3SLS. They are not needed for the implementation of the actual estimator.

[3]By full rank $I(1)$ variables, we mean that there is no cointegrating relation among Z_{g2}^*.

The 2SLS estimator $\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}}$ can be written as a linear transformation of the 2SLS estimator of (8), $\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}} = M_g \hat{\underset{\sim}{\delta}}^*_{g,2\text{SLS}}$, where

$$\hat{\underset{\sim}{\delta}}^*_{g,2\text{SLS}} = \left[Z_g^{*\prime} X^* (X^{*\prime} X^*)^{-1} X^{*\prime} Z_g^*\right]^{-1} \left[Z_g^{*\prime} X^* (X^{*\prime} X^*)^{-1} X^{*\prime} \underset{\sim}{w}_g\right] \tag{9}$$

Since under fairly general conditions (e.g. Phillips and Durlauf 1986; Chan and Wei 1988) $\frac{1}{T} Z_{g1}^{*\prime} X_1^* \longrightarrow M^*_{z_{g1}x_1}$, $\frac{1}{T^2} Z_{g1}^{*\prime} X_2^* \longrightarrow 0$, $\frac{1}{T} Z_{g2}^{*\prime} X_1^* \Longrightarrow M^*_{z_{g2}x_1}$ a.s., $\frac{1}{T^2} Z_{g2}^{*\prime} X_1^* \longrightarrow 0$, $\frac{1}{T} X_1^{*\prime} X_1^* \longrightarrow M^*_{x_1x_1}$, $\frac{1}{T} X_1^{*\prime} X_2^* \Longrightarrow M^*_{x_1x_2}$ a.s., $\frac{1}{T^2} X_1^{*\prime} X_2^* \longrightarrow 0$, $\frac{1}{T^2} X_2^{*\prime} X_2^* \Longrightarrow M^*_{x_2x_2}$ a.s., $\frac{1}{T} X_1^{*\prime} \underset{\sim}{\epsilon}_g \longrightarrow 0$ and $\frac{1}{T^2} X_2^{*\prime} \underset{\sim}{\epsilon}_g \longrightarrow 0$, where $\longrightarrow$ and $\Longrightarrow$ denote convergence in probability and in distribution of the associated probability measures, respectively, and $M^*_{x_1x_1}$ and $M^*_{x_2x_2}$ are non-singular, it follows that $\hat{\underset{\sim}{\delta}}^*_{g,2\text{SLS}}$ converges to $\underset{\sim}{\delta}^*_{g,2\text{SLS}}$. Hence the 2SLS estimator of $\underset{\sim}{\delta}_g$ is consistent.

To derive the limiting distribution of the 2SLS estimator, we let $H_g = \left[\begin{smallmatrix} T^{-\frac{1}{2}} I_{\ell_g} & \underset{\sim}{0} \\ \underset{\sim}{0} & T^{-1} I_{b_g} \end{smallmatrix}\right]$ and $H_x = \left[\begin{smallmatrix} T^{-\frac{1}{2}} I_{\ell^*} & \underset{\sim}{0} \\ \underset{\sim}{0} & T^{-1} I_{b^*} \end{smallmatrix}\right]$, where ℓ^* and b^* are the column dimensions of X_1^* and X_2^*, respectively. Since $T^{-3/2} X_1^{*\prime} Z_{g2}^* \longrightarrow 0$, $T^{-3/2} X_2^{*\prime} Z_{g1}^* \longrightarrow 0$ and $T^{-3/2} X_1^{*\prime} X_2^* \longrightarrow 0$,

$$\begin{aligned}
H_g^{-1}\left(\hat{\underset{\sim}{\delta}}^*_{g,2\text{SLS}} - \underset{\sim}{\delta}^*_g\right) &= \begin{bmatrix} \sqrt{T}\left(\hat{\underset{\sim}{\delta}}^*_{g1,2\text{SLS}} - \underset{\sim}{\delta}^*_{g1}\right) \\ T\left(\hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}} - \underset{\sim}{\delta}^*_{g2}\right) \end{bmatrix} \\
&= \left\{ \begin{pmatrix} \frac{1}{T} Z_{g1}^{*\prime} X_1^* & \frac{1}{T^{3/2}} Z_{g1}^{*\prime} X_2^* \\ \frac{1}{T^{3/2}} Z_{g2}^{*\prime} X_1^* & \frac{1}{T^2} Z_{g2}^{*\prime} X_2^* \end{pmatrix} \begin{pmatrix} \frac{1}{T} X_1^{*\prime} X_1^* & \frac{1}{T^{3/2}} X_1^{*\prime} X_2^* \\ \frac{1}{T^{3/2}} X_2^{*\prime} X_1^* & \frac{1}{T^2} X_2^{*\prime} X_2^* \end{pmatrix}^{-1} \right. \\
&\quad \left. \begin{pmatrix} \frac{1}{T} X_1^{*\prime} Z_{g1}^* & \frac{1}{T^{3/2}} X_1^{*\prime} Z_{g2}^* \\ \frac{1}{T^{3/2}} X_2^{*\prime} Z_{g1}^* & \frac{1}{T^2} X_2^{*\prime} Z_{g2}^* \end{pmatrix} \right\}^{-1} \\
&\quad \left\{ \begin{pmatrix} \frac{1}{T} Z_{g1}^{*\prime} X_1^* & \frac{1}{T^{3/2}} Z_{g1}^{*\prime} X_2^* \\ \frac{1}{T^{3/2}} Z_{g2}^{*\prime} X_1^* & \frac{1}{T^2} Z_{g2}^{*\prime} X_2^* \end{pmatrix} \begin{pmatrix} \frac{1}{T} X_1^{*\prime} X_1^* & \frac{1}{T^{3/2}} X_1^{*\prime} X_2^* \\ \frac{1}{T^{3/2}} X_2^{*\prime} X_1^* & \frac{1}{T^2} X_2^{*\prime} X_2^* \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{T^{1/2}} X_1^{*\prime} \underset{\sim}{\epsilon}_g \\ \frac{1}{T} X_2^{*\prime} \underset{\sim}{\epsilon}_g \end{pmatrix} \right\}
\end{aligned}$$

$$\Longrightarrow \begin{bmatrix} \left(M^*_{z_{g1}x_1} M^{*-1}_{x_1x_1} M^*_{x_1z_{g1}}\right)^{-1} \left(M^*_{z_{g1}x_1} M^{*-1}_{x_1x_1} \cdot T^{-1/2} X_1^{*\prime} \underset{\sim}{\epsilon}_g\right) \\ \left(M^*_{z_{g2}x_2} M^{*-1}_{x_2x_2} M^*_{x_2z_{g2}}\right)^{-1} \left(M^*_{z_{g2}x_2} M^{*-1}_{x_2x_2} \cdot T^{-1} X_2^{*\prime} \underset{\sim}{\epsilon}_g\right) \end{bmatrix}. \tag{10}$$

Under assumption (A3), we have

$$\frac{1}{\sqrt{T}} X_1^{*\prime} \underset{\sim}{\epsilon}_g \Longrightarrow N\left(\underset{\sim}{0}, \sigma_g^2 M^*_{x_1x_1}\right). \tag{11}$$

However, as shown by Phillips and Durlauf (1986),

$$\frac{1}{T} X_2^{*\prime} \underset{\sim}{\epsilon}_g \Longrightarrow \int B_{x_2^*} dB_{\epsilon_g}, \tag{12}$$

where B_{ε_g} denotes the Brownian motion of ϵ_{gt} with variance σ_g^2, $B_{x_2^*}$ denotes a $b^* \times 1$ vector Brownian motion of $\triangledown \underset{\sim}{x}^*_{2t}$ with covariance matrix $\Omega_{\triangledown x_2^* \triangledown x_2^*}$ with $\Omega_{\triangledown x_2^* \triangledown x_2^*}$ being the long-run covariance matrix of $\triangledown \underset{\sim}{x}^*_{2t}$. Following Phillips (1991), we can decompose the right-hand side of

(12) into two terms as

$$\int B_{x_2^*} dB_{\epsilon_g \cdot x_2^*} + \int B_{x_2^*} \Omega_{\epsilon_g \nabla x_2^*} \Omega^{-1}_{\nabla x_2^* \nabla x_2^*} dB_{x_2^*}, \tag{13}$$

where $B_{\epsilon_g \cdot x_2^*} = B_{\epsilon_g} - \Omega_{\epsilon_g \nabla x_2^*} \Omega^{-1}_{\nabla x_2^* \nabla x_2^*} B_{x_2^*} \equiv BM(\sigma^2_{g \cdot \nabla x_2^*})$ with $\sigma^2_{g.\nabla x_2^*} = \sigma^2_g - \Omega_{\epsilon_g \nabla x_2^*} \Omega^{-1}_{\nabla x_2^* \nabla x_2^*} \Omega_{\nabla x_2^* \epsilon_g}$, and $\Omega_{\epsilon_g \nabla x_2^*}$ denotes the long-run covariance between $\underset{\sim}{\epsilon}_g$ and $\nabla \underset{\sim}{x}_2^*$. The first term of (13) is a mixed normal. The second term involves a matrix unit root distribution that arises from using lagged $\underset{\sim}{w}$ as instruments when $\underset{\sim}{w}$ is $I(1)$ and the contemporaneous correlation between $\underset{\sim}{\epsilon}_{gt}$ and $\underset{\sim}{w}_t$ is non-zero. The 'long-run endogeneity' of the non-stationary instruments X_2^* results in a skewness of the limiting distribution of $\hat{\underset{\sim}{\delta}}^*_{g,2\text{SLS}}$ and its dependence on nuisance parameters that are impossible to eliminate by the 2SLS. Therefore,

Lemma 2. *The 2SLS estimator of $\underset{\sim}{\delta}_g^*$ is consistent and*

$$\sqrt{T}\big(\hat{\underset{\sim}{\delta}}^*_{g1,2\text{SLS}} - \underset{\sim}{\delta}^*_{g1}\big) \Longrightarrow N\big(\underset{\sim}{0}, \sigma_g^2 \big(M^*_{z_{g1} x_1} M^{*-1}_{x_1 x_1} M^*_{x_1 z_{g1}}\big)^{-1}\big), \tag{14}$$

$$\begin{aligned} T\big(\hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}} - \underset{\sim}{\delta}^*_{g2}\big) \Longrightarrow & \\ & \left\{ \int B_{z^*_{g2}} B'_{x_2^*} dr \left(\int B_{x_2^*} B'_{x_2^*} dr \right)^{-1} \int B_{x_2^*} B'_{z^*_{g2}} dr \right\}^{-1} \\ & \left\{ \int B_{z^*_{g2}} B'_{x_2^*} dr \left(\int B_{x_2^*} B'_{x_2^*} dr \right)^{-1} \left[\int B_{x_2^*} dB_{\epsilon_g \cdot x_2^*} + \int B_{x_2^*} \Omega_{\epsilon_g \nabla x_2^*} \Omega^{-1}_{\nabla x_2^* \nabla x_2^*} dB_{x_2^*} \right] \right\}, \end{aligned} \tag{15}$$

where $B_{z^*_{g2}}$ denotes a $b_g \times 1$ vector Brownian motion of $\nabla \underset{\sim}{z}^*_{g2,t}$ which appears in the gth equation.

The conventional formula $\sigma_g^2 [Z_g^{*\prime} X^* (X^{*\prime} X^*)^{-1} X^{*\prime} Z_g^*]^{-1}$ can no longer be used to approximate the asymptotic covariance matrix of $\hat{\underset{\sim}{\delta}}^*_{g,2\text{SLS}}$ as in the stationary case. It can be shown that $\sigma_g^2 H_g [Z_g^{*\prime} X^* \, (X^{*\prime} X^*)^{-1} X^{*\prime} Z_g^*]^{-1} H_g'$ converges to

$$\sigma_g^2 \begin{bmatrix} \big(M^*_{z_{g1} x_1} M^{*-1}_{x_1 x_1} M^*_{x_1 z_{g1}}\big)^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \big(M^*_{z_{g2} x_2} M^{*-1}_{x_2 x_2} M^*_{x_2 z_{g2}}\big)^{-1} \end{bmatrix}. \tag{16}$$

While the top left block of (16) is the asymptotic covariance matrix of $\sqrt{T}(\hat{\underset{\sim}{\delta}}^*_{g1,2\text{SLS}} - \underset{\sim}{\delta}^*_{g1})$, the lower right block of (16) is not the asymptotic covariance matrix of $T(\hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}} - \underset{\sim}{\delta}^*_{g2})$ as indicated by (15).

Since $\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}} = M_g \hat{\underset{\sim}{\delta}}^*_{g,2\text{SLS}}$, the limiting distribution of $\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}}$ is given by the components that have slower rate of convergence. Therefore, if none of the rows of M_g are identically zero in the first ℓ_g columns, $\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}}$ converges to $\underset{\sim}{\delta}_g$ 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 elements of $\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}}$ converge to their true values at the speed of T. However, their limiting distribution is non-standard. Let M_{g+} and M_{g++} denote the submatrices of M_g that the first ℓ_g columns of each row are not and are identically zero, respectively, and $\underset{\sim}{\delta}_{g+}$ and $\underset{\sim}{\delta}_{g++}$ denote the subvectors of $\underset{\sim}{\delta}_g$ that correspond to M_{g+} and M_{g++}, respectively. Then

Lemma 3. *The 2SLS estimator of (1) is consistent. The limiting distribution of $\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}}$ has the form*

$$\sqrt{T}(\hat{\underset{\sim}{\delta}}_{g+,2\text{SLS}} - \underset{\sim}{\delta}_{g+}) \Longrightarrow N\left(\underset{\sim}{0}, \sigma_g^2 M_{g+}\begin{pmatrix} \left(M^*_{z_{g1}x_1}M^{*-1}_{x_1x_1}M^*_{x_1z_{g1}}\right)^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0}\end{pmatrix} M'_{g+}\right) \tag{17}$$

$$T\left(\hat{\underset{\sim}{\delta}}_{g++,2\text{SLS}} - \underset{\sim}{\delta}_{g++}\right) = T\tilde{M}_{g++}\left(\hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}} - \underset{\sim}{\delta}^*_{g2}\right), \tag{18}$$

where $\tilde{M}_{g++}$ denotes the last b_g columns of M_{g++}.

When $p > 1$, under the exclusion restrictions, each element of $\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}}$ is either a linear combination of $\hat{\underset{\sim}{\delta}}^*_{g1,2\text{SLS}}$ alone or a linear combination of both $\hat{\underset{\sim}{\delta}}^*_{g1,2\text{SLS}}$ and $\hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}}$. Therefore, all the elements of $\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}}$ converge to $\underset{\sim}{\delta}_g$ at the speed of $T^{1/2}$.

Corollary 1. *When all the elements of $\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}}$ converge to $\underset{\sim}{\delta}_g$ at the speed of $T^{1/2}$, then $\sqrt{T}(\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}} - \underset{\sim}{\delta}_g)$ converges to a singular normal distribution with mean $\underset{\sim}{0}$ and covariance matrix*

$$\Sigma_{g,2\text{SLS}} = \sigma_g^2 M_g \begin{bmatrix} \left(M^*_{z_{g1}x_1}M^{*-1}_{x_1x_1}M^*_{x_1z_{g1}}\right)^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0}\end{bmatrix} M'_g, \tag{19}$$

where $\Sigma_{g,2\text{SLS}}$ may be approximated by the usual formula of the covariance matrix of the 2SLS estimator,

$$T\sigma_g^2[Z'_g X(X'X)^{-1}X'Z_g]^{-1}. \tag{20}$$

Proof. Equation (20) is identical to

$$T\sigma_g^2 M_g\left[Z^{*\prime}_g X^*(X^{*\prime}X^*)^{-1}X^{*\prime}Z^*_g\right]^{-1}M'_g \Longrightarrow$$

$$\begin{aligned} &\sigma_g^2 M_g \left\{\begin{pmatrix} M^*_{z_{g1}x_1} & T^{-1/2}M^*_{z_{g1}x_2} \\ M^*_{z_{g2}x_1} & T^{1/2}M^*_{z_{g2}x_2}\end{pmatrix}\begin{pmatrix} M^{*-1}_{x_1x_1} & \underset{\sim}{0} \\ \underset{\sim}{0} & M^{*-1}_{x_2x_2}\end{pmatrix}\begin{pmatrix} M^*_{x_1z_{g1}} & M^*_{x_1z_{g2}} \\ T^{-1/2}M^*_{x_2z_{g1}} & T^{1/2}M^*_{x_2z_{g2}}\end{pmatrix}\right\}^{-1} M'_g \\ &= \sigma_g^2 M_g \begin{bmatrix} \left(M^*_{z_{g1}x_1}M^{*-1}_{x_1x_1}M^*_{x_1z_{g1}}\right)^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0}\end{bmatrix} M'_g, \end{aligned} \tag{21}$$

where the last equality follows from the use of partitioned inverse formula and $T^{-1/2}M^*_{z_{g1}x_2} \to 0$, $T^{-1}[M^*_{z_{g2}x_2}M^{*-1}_{x_2x_2}M^*_{x_2z_{g2}}]^{-1} \to 0$.

Therefore, when all the coefficients converge at the speed of $\sqrt{T}$ and if the null hypothesis is in the form of testing a single coefficient, say δ_{gk} equals a specific value c_k,

$$H_0 : \delta_{gk} = c_k,$$

then there is no problem of using the 2SLS estimator to estimate $\underset{\sim}{\delta}_g$ and construct a t-statistic using the conventional formula, $\frac{\sqrt{T}(\hat{\delta}_{gk}-c_k)}{sd(\sqrt{T}\hat{\delta}_{gk})}$, which will converge to a standard normal, where $sd(\sqrt{T}\hat{\delta}_{gk})$ denotes the standard error of $\sqrt{T}\hat{\delta}_{gk}$, the square root of the kth diagonal element of (19). [4]

However, Lemma 3 suggests that inference about the joint hypothesis $P\underset{\sim}{\delta}_g = \underset{\sim}{c}$ can be tricky, where P and $\underset{\sim}{c}$ are a known matrix and a known vector of the proper dimensions. If $\sqrt{T}P(\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}} - \underset{\sim}{\delta}_g)$ has a non-singular covariance matrix, it means that the Wald test statistic

$$\left(\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}} - \underset{\sim}{\delta}_g\right)' P' \operatorname{Cov}\left(P\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}}\right)^{-1} P\left(\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}} - \underset{\sim}{\delta}_g\right) \tag{22}$$

will be asymptotically chi-square distributed. On the other hand, if $\sqrt{T}P(\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}} - \underset{\sim}{\delta}_g)$ has a singular covariance matrix, it means that there exists a non-singular matrix L such that

$$LP\underset{\sim}{\delta}_g = LP^*\underset{\sim}{\delta}^*_g = \begin{bmatrix} \tilde{P}_{11} & \tilde{P}_{12} \\ \underset{\sim}{0} & \tilde{P}_{22} \end{bmatrix} \begin{bmatrix} \underset{\sim}{\delta}^*_{g1} \\ \underset{\sim}{\delta}^*_{g2} \end{bmatrix} \tag{23}$$

with non-zero $\tilde{P}_{22}$. Then

$$\begin{aligned}
&(P\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}} - \underset{\sim}{c})' \operatorname{Cov}(P\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}})^{-1}(P\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}} - \underset{\sim}{c}) \\
= \;& \left\{\begin{bmatrix} \tilde{P}_{11} & \tilde{P}_{12} \\ \underset{\sim}{0} & \tilde{P}_{22} \end{bmatrix}\begin{bmatrix} \hat{\underset{\sim}{\delta}}^*_{g1,2\text{SLS}} \\ \hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}} \end{bmatrix} - L\underset{\sim}{c}\right\}' \operatorname{Cov}(LP\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}})^{-1} \left\{\begin{bmatrix} \tilde{P}_{11} & \tilde{P}_{12} \\ \underset{\sim}{0} & \tilde{P}_{22} \end{bmatrix}\begin{bmatrix} \hat{\underset{\sim}{\delta}}^*_{g1,2\text{SLS}} \\ \hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}} \end{bmatrix} - L\underset{\sim}{c}\right\} \\
\Longrightarrow\;& T\left(\tilde{P}_{11}\hat{\underset{\sim}{\delta}}^*_{g1,2\text{SLS}} + \tilde{P}_{12}\hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}} - \tilde{\underset{\sim}{c}}_1\right)' \operatorname{Cov}\left(\sqrt{T}\tilde{P}_{11}\hat{\underset{\sim}{\delta}}^*_{g1,2\text{SLS}}\right)^{-1}\left(\tilde{P}_{11}\hat{\underset{\sim}{\delta}}^*_{g1,2\text{SLS}} + \tilde{P}_{12}\hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}} - \underset{\sim}{c}_1\right) \\
&+T^2\left(\tilde{P}_{22}\hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}} - \tilde{\underset{\sim}{c}}_2\right)' \operatorname{Cov}\left(T\tilde{P}_{22}\hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}}\right)^{-1}\left(\tilde{P}_{22}\hat{\underset{\sim}{\delta}}^*_{g2,2\text{SLS}} - \tilde{\underset{\sim}{c}}_2\right),
\end{aligned} \tag{24}$$

where $L\underset{\sim}{c} = (\tilde{\underset{\sim}{c}}'_1, \tilde{\underset{\sim}{c}}'_2)'$. The first term on the right-hand side of (24) is asymptotically chi-square distributed. The second term, according to Lemma 2, has a non-standard distribution. Hence (24) is not asymptotically chi-square distributed.

Remark 3. Equations (23) and (24) demonstrate that in testing the null hypothesis of $P\underset{\sim}{\delta}_g = \underset{\sim}{c}$ whether the limiting distribution of $\sqrt{T}P(\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}} - \underset{\sim}{\delta}_g)$ has a non-singular covariance matrix is critical in determining if the chi-square distribution provides a good approximation of the Wald test statistic (22). Unfortunately, without the prior knowledge of the location of unit roots, we are not able to know a priori if $\sqrt{T}P(\hat{\underset{\sim}{\delta}}_{g,2\text{SLS}} - \underset{\sim}{\delta}_g)$ has a non-singular covariance matrix in the limit.

Remark 4. The critical difference between the model formulated here and those of Hsiao (1997a,b) is that in Hsiao (1997a,b), there exists a set of strictly exogenous variables that drive the system. In other words, there exists prior information that allows one to decompose $\underset{\sim}{w}$ into $(\underset{\sim}{w}'_1, \underset{\sim}{w}'_2)'$ such that (1) takes the form

$$A_{11}(L)\underset{\sim}{w}_{1t} + A_{12}(L)\underset{\sim}{w}_{2t} = \underset{\sim}{\epsilon}_{1t}, \tag{25}$$

$$A_{22}(L)\underset{\sim}{w}_{2t} = \underset{\sim}{\epsilon}_{2t}, \tag{26}$$

[4] $\sqrt{T}(\hat{\underset{\sim}{\delta}}_g - \underset{\sim}{\delta}_g)$ being singular normal only implies that some linear combination of $\sqrt{T}\hat{\underset{\sim}{\delta}}_g$ will have zero variance. On the other hand, $\sqrt{T}(\hat{\delta}_{gk} - \delta_{gk})$ remains to follow a normal limiting distribution.

and

$$E\underset{\sim}{\epsilon}_{1t}\underset{\sim}{\epsilon}'_{2t} = \underset{\sim}{0}. \tag{27}$$

That is, $A_{21}(L) \equiv \underset{\sim}{0}$ and (27) imply that $\underset{\sim}{w}_{2t}$ can be treated as strictly exogenous, applying 3SLS to (25) yields asymptotically normal or mixed normal distribution.

Without the strict exogeneity assumption on some of the variables as commonly formulated in the Cowles Commission structural equation modelling (e.g. Koopmans *et al.* 1950; Hsiao 1997a) or the prior knowledge of the direction of non-stationarity, the system estimation method such as the three-stage least-squares estimator (3SLS), although consistent, will have the same kind of inference problems as those associated with the presence of a matrix unit root distribution. However, if strict exogeneity of $\underset{\sim}{w}_2$ does not hold, as long as the system satisfies the Wold (1954) block causal ordering or Granger (1969) non-causality so that the corresponding partition of $A(L)$ has the form that $A_{21}(L) \equiv \underset{\sim}{0}$, but not $E\underset{\sim}{\epsilon}_{1t}\underset{\sim}{\epsilon}'_{2t} = 0$, and $\underset{\sim}{w}_2$ contains the common trends driving the system in the sense that $\underset{\sim}{x}_2^*$ is a subset of $\underset{\sim}{w}_2$, although the limiting distribution of the 2SLS estimator remains skewed, the 3SLS estimator applied to (25) and (26) together will yield asymptotically normal or mixed normal distribution for $A_{11}(L)$ and $A_{12}(L)$. In other words, hypothesis testing involving coefficients of (25) alone can be conducted using Wald type test statistics which are asymptotically chi-square distributed provided their estimates are obtained by applying 3SLS to the complete system of (25) and (26). But if 3SLS is applied to the system (25) alone, then the limiting distributions of the resulting estimates remain to be functions of unit root distribution.

4. THE LAG ORDER AUGMENTED 2SLS AND 3SLS ESTIMATORS

The problem of applying the 2SLS or 3SLS estimator is not that of the consistency of the estimator, but that the limiting distribution of the estimator of the coefficients on the full rank $I(1)$ variables, Z_{g2}^*, $\underset{\sim}{\delta}_{g2}^*$, involves a matrix unit root distribution that arises from using lagged dependent variables as instruments when dependent variables are integrated of order 1. In a system of reduced form VAR, i.e., $A_0 = I_m$, Toda and Yamamoto (1995) note that if we augment the order of a pth order autoregressive process by the maximum order of integration, then the skewness of the limiting distribution of the least-squares estimator will be concentrated on the coefficient matrices associated with the augmented lags which are known a priori to be zero. Therefore, standard inference procedure can still be applied to the coefficients in the first p coefficient matrices. In this section, we consider a lag-augmented 2SLS or 3SLS (LA2SLS or LA3SLS) estimator following the suggestion of Toda and Yamamoto (1995) and Yamada and Toda (1998).

Under the assumption that $\underset{\sim}{w}_t$ is $I(1)$, we augment the pth order SDM (1) to a $(p+1)$th order SDM

$$A_0\underset{\sim}{w}_t + A_1\underset{\sim}{w}_{t-1} + \cdots + A_p\underset{\sim}{w}_{t-p} + A_{p+1}\underset{\sim}{w}_{t-p-1} = \underset{\sim}{\epsilon}_t, \tag{28}$$

where $A_{p+1} \equiv \underset{\sim}{0}$. Transforming (28) into an error-correction form, we have

$$\sum_{j=0}^{p} A_j^*\nabla\underset{\sim}{w}_{t-j} + A_{p+1}^*\underset{\sim}{w}_{t-p-1} = \underset{\sim}{\epsilon}_t, \tag{29}$$

where $A_j^* = \sum_{\ell=0}^{j} A_\ell$, $j = 0, 1, \ldots, p$ and $A_{p+1}^* = A_p^*$. It follows that $A = [A_0, \ldots, A_p] = [A_0^*, \ldots, A_p^*]\tilde{M}^{-1}$.

Let the gth equation of (28) be written as

$$\underset{\sim}{w}_g = Z_g^A \underset{\sim}{\delta}_g^A + \underset{\sim}{\epsilon}_g, \tag{30}$$

where $Z_g^A = (Z_g, \tilde{W}_{g,-(p+1)})$, $\underset{\sim}{\delta}_g^A = (\underset{\sim}{\delta}_g', -\underset{\sim}{a}_{g,p+1}')'$ with $\tilde{W}_{g,-(p+1)}$ denoting the $T \times g_\Delta$ matrix of included $\underset{\sim}{\tilde{w}}_{gt}$ lagged by $(p+1)$ periods and $\underset{\sim}{a}_{g,p+1}'$ being the gth row of A_{p+1} excluding those elements that are subject to exclusion restrictions. Just like (6), there exists a non-singular transformation matrix M_g^A that transforms Z_g^A into $Z_g^{*A} = Z_g^A M_g^A = (Z_{g1}^{*A}, Z_{g2}^{*A})$, and $\underset{\sim}{\delta}_g^A$ into $\underset{\sim}{\delta}_g^{*A} = (M_g^A)^{-1}\underset{\sim}{\delta}_g^A = (\underset{\sim}{\delta}_{g1}^{*A\prime}, \underset{\sim}{\delta}_{g2}^{*A\prime})'$ where $Z_{g1}^{*A} = (\nabla Z_g, \tilde{W}_{g,-(p+1)}\underset{\sim}{\pi}_g)$ is stationary and Z_{g2}^{*A} is the $T \times \mathrm{b}_g$ matrix of the full rank $I(1)$ variables, $\underset{\sim}{\tilde{w}}_{g2,t-(p+1)}$. We rewrite (30) in terms of the transformed variables

$$\underset{\sim}{w}_g = Z_g^A M_g^A \left(M_g^A\right)^{-1} \underset{\sim}{\delta}_g^A + \underset{\sim}{\epsilon}_g = \begin{pmatrix} Z_{g1}^{*A} & Z_{g2}^{*A} \end{pmatrix} \begin{pmatrix} \underset{\sim}{\delta}_{g1}^{*A} \\ \underset{\sim}{\delta}_{g2}^{*A} \end{pmatrix} + \underset{\sim}{\epsilon}_g. \tag{31}$$

Let $X^A = (X, W_{-(p+1)})$. The 2SLS estimator of (30) is defined as

$$\underset{\sim}{\hat{\delta}}_{g,2\mathrm{SLS}}^A = \left[Z_g^{A\prime} X^A (X^{A\prime} X^A)^{-1} X^{A\prime} Z_g^A\right]^{-1} \left[Z_g^{A\prime} X^A (X^{A\prime} X^A)^{-1} X^{A\prime} \underset{\sim}{w}_g\right]. \tag{32}$$

The LA2SLS of (6) is defined as

$$\underset{\sim}{\hat{\delta}}_{g,LA2\mathrm{SLS}} = Q_g^A \underset{\sim}{\hat{\delta}}_{g,2\mathrm{SLS}}^A, \tag{33}$$

where $Q_g^A = (I_{(p+1)g_\Delta - 1}, \underset{\sim}{0}_{g_\Delta})$ with $\underset{\sim}{0}_{g_\Delta}$ denoting a $[(p+1)g_\Delta - 1] \times g_\Delta$ matrix of zeros. Since $\underset{\sim}{\hat{\delta}}_{g,2\mathrm{SLS}}^A = M_g^A \underset{\sim}{\hat{\delta}}_{g,2\mathrm{SLS}}^{*A}$, we have

$$\begin{aligned} \underset{\sim}{\hat{\delta}}_{g,LA2\mathrm{SLS}} &= Q_g^A M_g^A \underset{\sim}{\hat{\delta}}_{g,2\mathrm{SLS}}^{*A} \\ &= (\tilde{M}_g, \underset{\sim}{0}_{g_\Delta}) \underset{\sim}{\hat{\delta}}_{g,2\mathrm{SLS}}^{*A} \\ &= (\tilde{M}_g, \underset{\sim}{0}_g) \underset{\sim}{\hat{\delta}}_{g1,2\mathrm{SLS}}^{*A}, \end{aligned} \tag{34}$$

where $\tilde{M}_g$ is a $[(p+1)g_\Delta - 1]$-dimensional square matrix of the form,[5]

$$\tilde{M}_g = \begin{pmatrix} I_{g_\Delta - 1} & \underset{\sim}{0} & \ldots & \ldots & \ldots & \underset{\sim}{0} \\ \begin{pmatrix} -I_{g_\Delta - 1} \\ \underset{\sim}{0}' \end{pmatrix} & I_{g_\Delta} & \ldots & \ldots & \ldots & \ldots \\ \ldots & -I_{g_\Delta} & I_{g_\Delta} & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & I_{g_\Delta} & \underset{\sim}{0} \\ \ldots & \ldots & \ldots & \ldots & -I_{g_\Delta} & I_{g_\Delta} \end{pmatrix}, \tag{35}$$

[5]For ease of notation, we assume all the included variables appear with the same lag order.

with w_{gt} being ordered as the last element of $\underset{\sim}{\tilde{w}}_{gt}$, and $\underset{\sim}{0}_g$ is a $[(p+1)g_\Delta - 1] \times r_g$ matrix of zeros with r_g denoting the rank of cointegration of $\underset{\sim}{\tilde{w}}_{gt}$ such that $\underset{\sim}{\tilde{w}}'_{gt}\underset{\sim}{\pi}_g$ is $I(0)$. Then $\underset{\sim}{\delta}_g = (\tilde{M}_g, \underset{\sim}{0}_g)\underset{\sim}{\delta}^{*A}_{g1}$.

Since

$$\sqrt{T}\big(\underset{\sim}{\hat{\delta}}^{*A}_{g1,2\text{SLS}} - \underset{\sim}{\delta}^{*A}_{g1}\big) \Longrightarrow N\big[\underset{\sim}{0}, \sigma_g^2\big(M^{A*}_{z_{g1}x_1} M^{A*-1}_{x_1x_1} M^{A*}_{x_1z_{g1}}\big)^{-1}\big], \tag{36}$$

where $M^{A*}_{z_{g1}x_1} = \text{plim}\ \frac{1}{T} Z^{*A\prime}_{g1} X^{*A}_1$, $M^{A*}_{x_1x_1} = \text{plim}\ \frac{1}{T} X^{*A\prime}_1 X^{*A}_1$, with $X^{*A}_1 = (\nabla X, W_{-(p+1)}\underset{\sim}{\pi})$ being the $T \times (mp + r)$ matrix of the linearly independent $I(0)$ variables. It follows that

Theorem 1. *The LA2SLS of $\underset{\sim}{\delta}_g$ is consistent and*

$$\sqrt{T}(\underset{\sim}{\hat{\delta}}_{g,LA2\text{SLS}} - \underset{\sim}{\delta}_g) \Longrightarrow N\left\{\underset{\sim}{0}, \sigma_g^2(\tilde{M}_g \quad \underset{\sim}{0}_g)\big[M^{A*}_{z_{g1}x_1} M^{A*-1}_{x_1x_1} M^{A*}_{x_1z_{g1}}\big]^{-1}\begin{pmatrix}\tilde{M}'_g \\ \underset{\sim}{0}'_g\end{pmatrix}\right\}. \tag{37}$$

Remark 5. The LA2SLS estimator of the coefficients of the original structural dynamic model (1) converges to the true value at the speed of $T^{1/2}$ and is asymptotically normally distributed with non-singular covariance matrix. Therefore, Wald type test statistics based on the LA2SLS estimates are asymptotically chi-square distributed. Compared to the conventional 2SLS, the LA2SLS estimator loses the T-convergent component of 2SLS and ignores the prior restrictions that the coefficients on $\underset{\sim}{\tilde{w}}_{g,t-(p+1)}$ are zero, hence may lose some efficiency.[6]

Remark 6. The LA2SLS estimator remains consistent even when the errors are not identically distributed. When the errors are heteroscedastic, we may replace the conventional formula for the asymptotic covariance matrix by the robust estimator

$$\hat{\sum}^{\text{R}}_{\text{g,LA2SLS}} = Q^A_g\big(Z^{A\prime}_g P_X Z^A_g\big)^{-1} Z^{A\prime}_g P_X \hat{\Omega}_g P_X Z^A_g \big(Z^{A\prime}_g P_X Z^A_g\big)^{-1} Q^{A\prime}_g$$

where $P_X = X^A(X^{A\prime}X^A)^{-1}\ X^{A\prime}$ and $\hat{\Omega}_g$ is a $T \times T$ diagonal matrix with the tth element given by any of the following:

$$C_1 : \hat{\epsilon}^2_{g,t}$$

$$C_2 : \frac{T}{T - [(p+2)g_\Delta - 1]}\hat{\epsilon}^2_{g,t}$$

$$C_3 : \frac{1}{1 - \tilde{h}_t}\hat{\epsilon}^2_{g,t}$$

$$C_4 : \frac{1}{(1 - \tilde{h}_t)^2}\hat{\epsilon}^2_{g,t}$$

where $\tilde{h}_t$ is the tth diagonal element of P_X. Based on the simulation results in MacKinnon and White (1985), we speculate that C_3 and C_4 may have the best finite sample performance while C_1 is probably the worst.

[6] Although the 2SLS estimator has a T-convergent component, when $p > 1$, under the exclusion restrictions, each element of the 2SLS estimator is $\sqrt{T}$-convengent. Therefore, the loss of efficiency in estimating $\underset{\sim}{\delta}_g$ by LA2SLS estimator may not be significant (see Tables 1 and 2).

Table 1. Absolute percentage estimation bias (Bias).

								M2SLS (Parzen)			M2SLS (Tukey–Hanning)			M2SLS (quadratic)		
		OLS	2SLS	3SLS	LA2SLS	LA3SLS	A2SLS	$k=0.3$	$k=0.5$	$k=0.66$	$k=0.3$	$k=0.5$	$k=0.66$	$k=0.3$	$k=0.5$	$k=0.66$
DGP1	$T=50$	0.5005	0.3805	0.7507	0.3967	0.3907	1.8459	1.9701	6.6137	7.5595	4.6819	10.3942	24.1918	6.4504	3.6284	6.6729
	100	0.3991	0.1950	0.2411	0.2366	0.2274	0.1286	0.1282	0.1008	0.0982	0.1124	0.0865	0.1373	0.1028	0.1068	5.0929
	200	0.3381	0.0794	0.1056	0.1300	0.1401	0.0397	0.0894	0.0929	0.1032	0.0901	0.0959	0.0959	0.0909	0.1056	0.0524
	400	0.3054	0.0390	0.0564	0.0524	0.0687	0.0130	0.0378	0.0365	0.0328	0.0371	0.0362	0.0298	0.0363	0.0336	0.0313
DGP2	$T=50$	0.4025	0.2955	0.2691	0.3305	0.3034	0.2306	0.3183	0.2835	0.2621	0.3196	0.2477	0.3778	0.2872	0.3015	2.8317
	100	0.2669	0.1524	0.1412	0.1803	0.1507	0.1702	0.1502	0.1388	0.1197	0.1476	0.1281	0.0890	0.1409	0.0913	0.0876
	200	0.1924	0.0756	0.0712	0.0893	0.0824	0.0569	0.0734	0.0669	0.0639	0.0724	0.0679	0.0581	0.0706	0.0648	0.1305
	400	0.1577	0.0376	0.0378	0.0555	0.0491	0.0281	0.0381	0.0376	0.0354	0.0378	0.0370	0.0345	0.0376	0.0359	0.0274
DGP3	$T=50$	0.6037	0.3454	0.3290	0.3378	0.3106	0.1615	0.2920	0.2186	0.2336	0.2601	0.2142	0.5555	0.2179	0.2048	1.0129
	100	0.4437	0.1872	0.1668	0.1707	0.1494	0.1792	0.1329	0.1017	0.1246	0.1137	0.1096	0.3737	0.0979	0.1010	0.6111
	200	0.3608	0.0905	0.0767	0.0742	0.0599	0.2822	0.0460	0.0420	0.0642	0.0371	0.0501	0.0908	0.0371	0.0647	0.0462
	400	0.3269	0.0414	0.0293	0.0338	0.0213	0.2823	0.0163	0.0185	0.0291	0.0115	0.0224	0.0130	0.0127	0.0276	0.0322

Table 2. Root mean squared percentage estimation error (RMSPE).

								M2SLS (Parzen)			M2SLS (Tukey-Hanning)			M2SLS (quadratic)		
		OLS	2SLS	3SLS	LA2SLS	LA3SLS	A2SLS	$k = 0.3$	$k = 0.5$	$k = 0.66$	$k = 0.3$	$k = 0.5$	$k = 0.66$	$k = 0.3$	$k = 0.5$	$k = 0.66$
DGP1	$T = 50$	0.7069	2.0761	11.6241	1.1603	1.7093	60.0008	55.4400	184.7781	212.4558	117.9583	300.9338	503.4322	178.1836	79.1171	161.3943
	100	0.5183	1.0702	1.2221	0.8306	0.9171	1.4459	2.3323	3.4137	3.8105	2.9653	3.5651	6.6248	3.5396	4.5912	142.6248
	200	0.4104	0.5332	0.5205	0.5462	0.5450	0.6760	0.5357	0.5887	0.7327	0.5461	0.6345	0.9251	0.5690	0.7425	4.1131
	400	0.3472	0.3180	0.3093	0.3599	0.3507	0.3409	0.3204	0.3302	0.3589	0.3222	0.3369	0.3856	0.3249	0.3522	0.5468
DGP2	$T = 50$	0.6061	0.5848	0.5485	0.6608	0.6215	1.2000	0.6946	0.8793	3.2018	0.9254	1.5528	15.1530	0.9270	4.9084	76.3155
	100	0.3915	0.3469	0.3286	0.4078	0.3880	2.1915	0.3801	0.3938	0.5031	0.3865	0.4234	1.1395	0.3959	1.2287	3.1080
	200	0.2690	0.2160	0.2076	0.2655	0.2561	0.2603	0.2268	0.2239	0.2298	0.2265	0.2245	0.2462	0.2253	0.2308	3.1986
	400	0.2081	0.1409	0.1384	0.1798	0.1728	0.1579	0.1495	0.1443	0.1426	0.1489	0.1431	0.1436	0.1464	0.1430	0.1700
DGP3	$T = 50$	1.0574	1.0370	1.0092	1.1126	1.0774	5.3364	1.0934	1.1791	1.7692	1.1491	1.3017	19.3915	1.2090	3.3080	23.8907
	100	0.7508	0.6697	0.6559	0.7106	0.6932	0.7358	0.7064	0.6980	0.7440	0.7127	0.7010	6.0492	0.7093	1.1691	31.6062
	200	0.5614	0.4538	0.4409	0.4855	0.4724	0.6578	0.4893	0.4755	0.4725	0.4892	0.4723	0.8772	0.4836	0.4753	1.0173
	400	0.4499	0.3152	0.3045	0.3396	0.3304	0.7576	0.3361	0.3235	0.3206	0.3363	0.3224	0.8706	0.3294	0.3227	0.6869

Similarly, we can stack all m equations of (28) in the form of (30)

$$\underset{\sim}{w} = Z^A \underset{\sim}{\delta}^A + \underset{\sim}{\epsilon}, \tag{38}$$

where $Z^A = \text{diag}\ (Z_1^A, \ldots, Z_m^A)$, $\underset{\sim}{\delta}^A = (\underset{\sim}{\delta}_1^{A\prime}, \ldots, \underset{\sim}{\delta}_m^{A\prime})'$. The lag-augmented 3SLS (LA3SLS) estimator is defined as

$$\hat{\underset{\sim}{\delta}}_{LA3SLS} = Q^A \hat{\underset{\sim}{\delta}}_{3SLS}^A, \tag{39}$$

where $\hat{\underset{\sim}{\delta}}_{3SLS}^A$ denotes the 3SLS estimator of (38), $Q^A = \text{diag}(Q_1^A, \ldots, Q_m^A)$. Then

Theorem 2. *The LA3SLS of $\underset{\sim}{\delta}$ is consistent and*

$$\sqrt{T}(\hat{\underset{\sim}{\delta}}_{LA3SLS} - \underset{\sim}{\delta}) \Longrightarrow N(\underset{\sim}{0},\ \Sigma_{\text{LA3SLS}}), \tag{40}$$

where Σ_{LA3SLS} is a full rank matrix of the form

$$\Sigma_{LA3SLS} = M^* \tilde{\Sigma}_{11}^A M^{*\prime}, \tag{41}$$

where

$$M^* = \text{diag}\big(M_1^*, \ldots, M_m^*\big),\ M_g^* = (\tilde{M}_g, \underset{\sim}{0}_g),$$
$$\tilde{\Sigma}_{11}^A = \big(\sigma^{gh} R_{gh.1}^A\big),\ R_{gh.1}^A = M_{z_{g1}x_1}^{A*} M_{x_1x_1}^{A*-1} M_{x_1z_{h1}}^{A*},\ \Sigma_{\epsilon\epsilon}^{-1} = (\sigma^{gh}).$$

5. MONTE CARLO COMPARISONS

In this section, a small simulation study is conducted to compare the finite sample performance of the OLS, 2SLS, 3SLS, LA2SLS and LA3SLS estimators as well as the two modified estimators suggested by Hsiao and Wang (2006) that have asymptotic normal and mixed normal distributions.[7] To see how sensitive the M2SLS estimator is to the choice of the kernel function and the bandwidth parameter, we consider three kernel functions (Parzen, Turkey–Hanning and quadratic spectral) and three bandwidth expansion rates ($k = 0.3, 0.5, 0.66$).

For each estimator, we compute its bias, root mean square estimation error, the size and size-adjusted power of the t- and Wald tests where critical values are derived from the assumption of asymptotic normality. In addition, we compare the theoretical and actual histograms of the test statistics. All computation is performed by MATLAB. It is hoped that this simulation study will shed some light on the choice of the estimators in finite sample.

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

$$A_0 \underset{\sim}{w}_t = A_1 \underset{\sim}{w}_{t-1} + A_2 \underset{\sim}{w}_{t-2} + \underset{\sim}{\epsilon}_t, \tag{42}$$

where $\underset{\sim}{\epsilon}_t \sim N(0, \sum_{\epsilon\epsilon})$.

[7]We did not include the Johansen estimator in the comparison because it requires the prior knowledge of the rank of cointegration.

We let (42) be identified by the exclusion restrictions of the form that

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

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 $\{\underset{\sim}{w}_t\}_{t=-1}^{T}$, we initialize the system at $t = -51$ with $(\underset{\sim}{w}_{-50}, \underset{\sim}{w}_{-51}) = (\underset{\sim}{0}, \underset{\sim}{0})$. 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

$$\Gamma = \begin{pmatrix} 1 & -0.5 & 0.3 \\ -0.5 & 0.9 & 0.4 \\ 0.3 & 0.4 & 2.5 \end{pmatrix}^{1/2}$$

and $\underset{\sim}{\epsilon}_t = \Gamma \underset{\sim}{e}_t$, so that $\{\underset{\sim}{\epsilon}_t\}_{t=-49}^{T}$ is a sequence of independent trivariate normal random variables with zero mean and covariance matrix $\Sigma_{\epsilon\epsilon} = \Gamma\Gamma'$. To generate $\{\underset{\sim}{w}_t\}_{t=-49}^{T}$, we use the following parameter values of (A_0, A_1, A_2):

$$A_0 = \begin{pmatrix} 1 & -0.4 & 0 \\ 0 & 1 & 0.8 \\ 0.6 & 0 & 1 \end{pmatrix}, \qquad A_1 = \begin{pmatrix} 0.2 & -0.1 & 0 \\ 0 & 0.7 & 0.6 \\ 0.2 & 0 & 0.4 \end{pmatrix} \text{ and } A_2 = A_0 - A_1 + \underset{\sim}{\alpha}\underset{\sim}{\beta}';$$

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

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

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

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 in using normal approximation in hypothesis testing, we consider the following hypotheses: (A) test the value of $a_{0,12}$ alone, i.e. $H_A : a_{0,12} = c_0$; (B) a joint test of $H_B : a_{0,12} = c_0$, $a_{1,12} = c_1$ and $a_{2,12} = c_2$, where c_0, c_1 and c_2 are known constants.

Our analysis shows that the standard normal distribution provides a good approximation for the conventional t-statistics for H_A, be the estimator as 2SLS, 3SLS, LA2SLS, LA3SLS, or the modified estimators. On the other hand, the chi-square distribution may or may not be a good approximation for testing H_B even asymptotically if the Wald statistics are constructed from the 2SLS or 3SLS estimates. For instance, Wald statistics for H_B have a limiting chi-square distribution for DGP3, but not for DGP1 or DGP2. For all three data generating processes, we can transform H_B into the form of (23), then test B becomes a joint test of $a_{0,12} = c_0$, $a_{1,12} - a_{0,12} = c_1 - c_0$ and $a_{2,12} + a_{1,12} - a_{0,12} = c_2 + c_1 - c_0$. While for DGP1 and DGP2, test B isolates the coefficient on the $I(1)$ regressor, $w_{2,t-2}$, $(a_{2,12} + a_{1,12} - a_{0,12})$, for DGP3, it only involves the coefficients on the $I(0)$ regressors, $\nabla w_{2,t}$, $\nabla w_{2,t-1}$ and $w_{2,t-2} - 2w_{1,t-2}$. Hence, for

Wald test based on the 2SLS or 3SLS estimator, chi-square approximation is not appropriate for DGP1 or DGP2, but is appropriate for DGP3 if the sample is of reasonable size.

Although the true DGP (42) has no constant term, in practice one usually estimates a VAR with an intercept. It therefore seems more appropriate to include an intercept in the estimated structural dynamic model. Sample sizes are fixed at T=50, 100, 200 and 400. The number of repetitions is 1,000.

The consistency property of the alternative estimators can be examined using Table 1, which reports the absolute value of the percentage estimation bias (Bias), averaged over the five coefficients in the first equation of model (42). To compare the asymptotic efficiency of the estimators, we report in Table 2 the root mean squared percentage estimation error (RMSPE), averaged over the five coefficients in the first equation of model (42).

As expected, the OLS is severely biased, but not for 2SLS, 3SLS, LA2SLS, LA3SLS, or the modified estimators for moderate T. The Bias and RMSPE of the LA2SLS and LA3SLS estimators are about the same as the conventional 2SLS and 3SLS.[8] Hence, the loss of efficiency of the lag-augmented estimators does not appear to be significant even for small T.

It is interesting to note that although the two modified estimators proposed by Hsiao and Wang (2006) have desirable large sample properties, they tend to have larger Bias and RMSPE than conventional 2SLS and 3SLS for $T < 200$. In addition, the M2SLS estimator appears to be very sensitive to the choice of the kernel function and bandwidth parameter for small T. However, the modified estimators improve significantly with the sample size. When $T = 200$, their Bias and RMSPE are of similar order as the 2SLS.

Table 3 presents the actual sizes of the tests where the critical values are derived from standard normal for test A and chi-square distribution with three degrees of freedom for test B. For test A, normal approximation appears to be reasonable for all the consistent estimators. For test B, if the limiting distribution of Wald statistics does not involve the unit root distribution (DGP3), chi-square distribution approximates well as T increases for conventional and lag-augmented 2SLS and 3SLS. But if the limiting distribution of Wald statistics involves the unit root distribution (DGP1 and DGP2), then size distortions for 2SLS and 3SLS are severe. The actual size of a 1% test, 5% test and 10% test can be more than 5%, 15% and 30%, respectively. On the other hand, the actual sizes of LA2SLS and LA3SLS are very close to the nominal sizes for even moderate T.

As shown in Table 3, although the modified estimators have desirable large sample properties, they seem to have larger size distortions than the LA2SLS for $T \leq 400$. The results on the M2SLS estimator are also sensitive to the bandwidth choice. However, the performance of the modified estimators appears to improve as T increases.

Table 4 compares the size-adjusted power of different estimators. Although we do not know how to adjust for size distortion in practice, it is possible to find the critical values of different test statistics in a simulation framework. These results suggest that for test B Wald statistics

[8] Readers may note that the RMSPE of the OLS estimator is smaller than that of the 2SLS for DGP1 but not for DGP2 or DGP3. This observation corroborates the findings in Hurd (1972), in which Monte Carlo studies were conducted to compare the efficiency of the OLS and 2SLS estimators of a structural equation. Hurd found that for many parameter values, the OLS estimator has smaller mean squared error than the 2SLS if sample size is small ($T \leq 24$). However, the relative efficiency of the 2SLS estimator increases rapidly with sample size, the correlation between the endogenous regressor and the error term, and the variation (second moment) of the endogenous regressor. In the first equation of model (42), the correlation between $\nabla\omega_{2,t}$ and $\epsilon_{1,t}$ is 0.53, same for all three data generating processes. But the second moment of $\nabla\omega_{2,t}$ is about 3.5 for DGP1, 5.1 for DGP2 and 6.6 for DGP3. However, as noted by a referee, the RMSPE results for DGP1/3SLS and DGP1/A2SLS indicate severe instability for small sample ($T = 50$).

Table 3. Finite-sample size: DGP1.

								M2SLS (Parzen)			M2SLS (Tukey–Hanning)			M2SLS (quadratic)		
		OLS	2SLS	3SLS	LA2SLS	LA3SLS	A2SLS	$k = 0.3$	$k = 0.5$	$k = 0.66$	$k = 0.3$	$k = 0.5$	$k = 0.66$	$k = 0.3$	$k = 0.5$	$k = 0.66$
Test A: test a single coefficient parameter																
Alpha = 0.01	$T = 50$	0.190	0.002	0.026	0.012	0.043	0.147	0.010	0.018	0.032	0.014	0.027	0.038	0.017	0.037	0.038
	100	0.367	0.015	0.029	0.010	0.039	0.075	0.016	0.016	0.036	0.016	0.026	0.042	0.017	0.036	0.072
	200	0.706	0.007	0.019	0.009	0.021	0.039	0.010	0.014	0.027	0.011	0.015	0.040	0.013	0.026	0.052
	400	0.952	0.002	0.008	0.007	0.012	0.012	0.004	0.006	0.010	0.005	0.008	0.018	0.005	0.008	0.038
Alpha = 0.05	$T = 50$	0.365	0.045	0.081	0.063	0.109	0.238	0.062	0.090	0.119	0.086	0.099	0.132	0.099	0.127	0.143
	100	0.602	0.046	0.078	0.059	0.090	0.137	0.055	0.075	0.105	0.062	0.088	0.124	0.070	0.106	0.171
	200	0.863	0.045	0.059	0.046	0.071	0.091	0.049	0.061	0.094	0.052	0.071	0.098	0.059	0.092	0.131
	400	0.989	0.043	0.047	0.050	0.063	0.057	0.042	0.051	0.063	0.041	0.054	0.075	0.047	0.061	0.098
Alpha = 0.1	$T = 50$	0.481	0.094	0.137	0.113	0.160	0.295	0.107	0.166	0.180	0.145	0.171	0.202	0.167	0.211	0.229
	100	0.719	0.080	0.120	0.100	0.134	0.189	0.097	0.137	0.170	0.110	0.149	0.203	0.126	0.183	0.255
	200	0.912	0.094	0.112	0.104	0.123	0.142	0.105	0.115	0.147	0.114	0.123	0.158	0.116	0.149	0.200
	400	0.995	0.087	0.092	0.097	0.107	0.112	0.091	0.100	0.127	0.088	0.108	0.138	0.092	0.123	0.153
Test B: joint test of several coefficient parameters																
Alpha = 0.01	$T = 50$	0.232	0.053	0.145	0.032	0.090	0.298	0.093	0.185	0.245	0.154	0.213	0.270	0.185	0.265	0.307
	100	0.366	0.060	0.138	0.031	0.073	0.172	0.087	0.166	0.235	0.104	0.205	0.243	0.147	0.230	0.345
	200	0.629	0.058	0.099	0.017	0.032	0.134	0.089	0.133	0.202	0.102	0.147	0.228	0.120	0.206	0.292
	400	0.922	0.048	0.073	0.013	0.017	0.069	0.047	0.080	0.131	0.050	0.088	0.166	0.064	0.113	0.242
Alpha = 0.05	$T = 50$	0.445	0.149	0.266	0.089	0.180	0.404	0.191	0.292	0.348	0.257	0.359	0.373	0.289	0.368	0.410
	100	0.592	0.169	0.278	0.084	0.153	0.280	0.208	0.278	0.349	0.236	0.319	0.370	0.268	0.363	0.472
	200	0.822	0.168	0.223	0.063	0.092	0.208	0.185	0.236	0.300	0.198	0.256	0.327	0.219	0.301	0.407
	400	0.970	0.144	0.180	0.060	0.068	0.136	0.152	0.188	0.233	0.157	0.192	0.270	0.047	0.291	0.355
Alpha = 0.1	$T = 50$	0.561	0.236	0.348	0.151	0.257	0.472	0.284	0.364	0.428	0.354	0.400	0.442	0.368	0.440	0.481
	100	0.699	0.255	0.373	0.153	0.226	0.341	0.294	0.363	0.424	0.318	0.392	0.442	0.362	0.442	0.547
	200	0.886	0.276	0.320	0.122	0.145	0.263	0.279	0.327	0.360	0.297	0.346	0.428	0.318	0.392	0.484
	400	0.986	0.246	0.263	0.113	0.118	0.218	0.230	0.266	0.325	0.242	0.282	0.356	0.258	0.308	0.427

Table 3. Finite-sample size: DGP2.

		OLS	2SLS	3SLS	LA2SLS	LA3SLS	A2SLS	M2SLS (Parzen) $k=0.3$	M2SLS (Parzen) $k=0.5$	M2SLS (Parzen) $k=0.66$	M2SLS (Tukey–Hanning) $k=0.3$	M2SLS (Tukey–Hanning) $k=0.5$	M2SLS (Tukey–Hanning) $k=0.66$	M2SLS (quadratic) $k=0.3$	M2SLS (quadratic) $k=0.5$	M2SLS (quadratic) $k=0.66$
Test A: test a single coefficient parameter																
Alpha = 0.01	$T=50$	0.069	0.019	0.025	0.020	0.022	0.067	0.036	0.041	0.050	0.039	0.050	0.058	0.044	0.060	0.081
	100	0.149	0.016	0.018	0.016	0.016	0.045	0.045	0.040	0.033	0.044	0.032	0.049	0.041	0.035	0.076
	200	0.238	0.017	0.016	0.014	0.013	0.043	0.036	0.035	0.024	0.038	0.027	0.026	0.034	0.024	0.047
	400	0.547	0.009	0.011	0.004	0.005	0.036	0.039	0.026	0.011	0.039	0.020	0.013	0.032	0.012	0.023
Alpha = 0.05	$T=50$	0.208	0.085	0.079	0.072	0.075	0.152	0.108	0.117	0.125	0.122	0.123	0.149	0.125	0.145	0.185
	100	0.311	0.071	0.068	0.066	0.069	0.114	0.125	0.117	0.124	0.130	0.116	0.129	0.125	0.130	0.171
	200	0.460	0.047	0.070	0.052	0.056	0.107	0.121	0.089	0.083	0.113	0.085	0.085	0.101	0.085	0.119
	400	0.750	0.065	0.063	0.052	0.052	0.086	0.120	0.086	0.069	0.110	0.072	0.073	0.090	0.067	0.091
Alpha = 0.1	$T=50$	0.294	0.145	0.142	0.129	0.131	0.223	0.172	0.186	0.189	0.199	0.184	0.210	0.199	0.209	0.264
	100	0.409	0.137	0.140	0.124	0.128	0.179	0.200	0.203	0.197	0.210	0.202	0.209	0.213	0.207	0.242
	200	0.570	0.132	0.132	0.100	0.104	0.174	0.192	0.158	0.133	0.184	0.141	0.137	0.171	0.143	0.177
	400	0.829	0.110	0.113	0.112	0.113	0.149	0.191	0.134	0.121	0.187	0.125	0.119	0.146	0.120	0.140
Test B: joint test of several coefficient parameters																
Alpha = 0.01	$T=50$	0.230	0.149	0.169	0.086	0.094	0.278	0.243	0.283	0.335	0.282	0.311	0.393	0.290	0.381	0.503
	100	0.265	0.143	0.316	0.048	0.043	0.203	0.214	0.247	0.274	0.228	0.265	0.325	0.242	0.284	0.412
	200	0.331	0.123	0.108	0.028	0.031	0.154	0.182	0.170	0.206	0.182	0.172	0.219	0.180	0.208	0.257
	400	0.554	0.104	0.077	0.023	0.021	0.126	0.135	0.137	0.134	0.130	0.138	0.147	0.128	0.137	0.183
Alpha = 0.05	$T=50$	0.445	0.338	0.341	0.208	0.200	0.426	0.400	0.417	0.455	0.428	0.441	0.501	0.424	0.495	0.596
	100	0.494	0.331	0.296	0.114	0.134	0.332	0.390	0.370	0.427	0.392	0.386	0.445	0.379	0.427	0.523
	200	0.568	0.309	0.269	0.106	0.102	0.311	0.339	0.332	0.331	0.341	0.327	0.346	0.337	0.331	0.395
	400	0.751	0.256	0.228	0.087	0.088	0.252	0.309	0.273	0.268	0.300	0.273	0.276	0.288	0.269	0.311
Alpha = 0.1	$T=50$	0.568	0.472	0.446	0.296	0.289	0.516	0.493	0.512	0.530	0.507	0.528	0.575	0.507	0.568	0.664
	100	0.622	0.449	0.409	0.223	0.224	0.423	0.488	0.478	0.500	0.488	0.481	0.526	0.483	0.500	0.592
	200	0.676	0.431	0.386	0.174	0.174	0.471	0.464	0.436	0.429	0.458	0.420	0.429	0.452	0.427	0.477
	400	0.841	0.368	0.338	0.156	0.154	0.351	0.423	0.376	0.361	0.414	0.370	0.358	0.387	0.362	0.405

Table 3. Finite-sample size: DGP3.

								M2SLS (Parzen)			M2SLS (Tukey–Hanning)			M2SLS (quadratic)		
		OLS	2SLS	3SLS	LA2SLS	LA3SLS	A2SLS	$k = 0.3$	$k = 0.5$	$k = 0.66$	$k = 0.3$	$k = 0.5$	$k = 0.66$	$k = 0.3$	$k = 0.5$	$k = 0.66$
Test A: test a single coefficient parameter																
Alpha = 0.01	$T = 50$	0.116	0.022	0.031	0.023	0.030	0.033	0.026	0.021	0.024	0.030	0.024	0.029	0.020	0.025	0.044
	100	0.266	0.012	0.011	0.009	0.009	0.017	0.021	0.010	0.010	0.023	0.010	0.021	0.015	0.014	0.036
	200	0.416	0.018	0.010	0.006	0.009	0.023	0.029	0.011	0.014	0.026	0.011	0.027	0.015	0.013	0.037
	400	0.758	0.007	0.006	0.006	0.006	0.022	0.021	0.015	0.009	0.023	0.015	0.021	0.018	0.011	0.032
Alpha = 0.05	$T = 50$	0.260	0.070	0.073	0.076	0.088	0.085	0.086	0.091	0.080	0.091	0.087	0.085	0.095	0.087	0.099
	100	0.414	0.052	0.072	0.061	0.072	0.083	0.098	0.094	0.075	0.104	0.082	0.078	0.098	0.073	0.104
	200	0.636	0.043	0.049	0.040	0.048	0.076	0.093	0.061	0.049	0.090	0.054	0.066	0.078	0.051	0.088
	400	0.893	0.053	0.059	0.054	0.055	0.094	0.092	0.061	0.055	0.094	0.057	0.074	0.072	0.056	0.099
Alpha = 0.1	$T = 50$	0.379	0.132	0.125	0.135	0.140	0.153	0.164	0.160	0.131	0.174	0.142	0.155	0.176	0.139	0.171
	100	0.539	0.122	0.131	0.121	0.134	0.144	0.173	0.160	0.134	0.178	0.149	0.157	0.170	0.142	0.182
	200	0.750	0.093	0.093	0.084	0.090	0.132	0.171	0.126	0.109	0.156	0.106	0.121	0.136	0.109	0.148
	400	0.940	0.106	0.112	0.100	0.108	0.163	0.164	0.109	0.108	0.158	0.110	0.115	0.117	0.112	0.154
Test B: joint test of several coefficient parameters																
Alpha = 0.01	$T = 50$	0.104	0.042	0.065	0.030	0.038	0.059	0.061	0.061	0.072	0.070	0.071	0.098	0.064	0.096	0.156
	100	0.156	0.020	0.032	0.016	0.020	0.056	0.045	0.033	0.033	0.048	0.030	0.064	0.043	0.049	0.103
	200	0.280	0.014	0.022	0.011	0.012	0.216	0.035	0.032	0.032	0.041	0.033	0.050	0.035	0.034	0.101
	400	0.616	0.013	0.016	0.004	0.009	0.480	0.038	0.016	0.016	0.035	0.018	0.035	0.029	0.019	0.062
Alpha = 0.05	$T = 50$	0.233	0.109	0.146	0.102	0.120	0.127	0.141	0.166	0.161	0.166	0.166	0.185	0.173	0.180	0.240
	100	0.329	0.080	0.105	0.079	0.081	0.169	0.123	0.106	0.097	0.123	0.105	0.146	0.115	0.123	0.184
	200	0.479	0.063	0.071	0.054	0.057	0.395	0.110	0.093	0.083	0.114	0.081	0.118	0.112	0.083	0.171
	400	0.794	0.056	0.056	0.059	0.063	0.549	0.100	0.076	0.069	0.105	0.072	0.089	0.097	0.075	0.135
Alpha = 0.1	$T = 50$	0.345	0.174	0.212	0.184	0.199	0.213	0.226	0.238	0.230	0.242	0.230	0.256	0.241	0.252	0.320
	100	0.438	0.143	0.163	0.126	0.145	0.259	0.196	0.180	0.175	0.201	0.177	0.222	0.196	0.212	0.261
	200	0.608	0.122	0.117	0.099	0.112	0.502	0.187	0.152	0.143	0.190	0.144	0.175	0.177	0.154	0.232
	400	0.867	0.111	0.108	0.107	0.116	0.591	0.177	0.137	0.122	0.186	0.121	0.146	0.169	0.132	0.195

Table 4 Finite-sample size-adjusted power: DGP1.

								M2SLS (Parzen)			M2SLS (Tukey–Hanning)			M2SLS (quadratic)		
		OLS	2SLS	3SLS	LA2SLS	LA3SLS	A2SLS	$k = 0.3$	$k = 0.5$	$k = 0.66$	$k = 0.3$	$k = 0.5$	$k = 0.66$	$k = 0.3$	$k = 0.5$	$k = 0.66$
Test A: test a single coefficient parameter																
Alpha = 0.01	$T = 50$	0.076	0.086	0.023	0.057	0.027	0.010	0.067	0.042	0.024	0.032	0.028	0.026	0.038	0.023	0.021
	100	0.223	0.102	0.032	0.101	0.029	0.013	0.091	0.086	0.070	0.095	0.069	0.062	0.087	0.071	0.027
	200	0.477	0.352	0.194	0.279	0.190	0.114	0.302	0.270	0.219	0.284	0.270	0.158	0.282	0.195	0.108
	400	0.844	0.724	0.676	0.649	0.605	0.644	0.710	0.686	0.659	0.707	0.679	0.544	0.700	0.683	0.375
Alpha = 0.05	$T = 50$	0.292	0.203	0.166	0.167	0.144	0.047	0.185	0.143	0.112	0.143	0.120	0.190	0.146	0.088	0.082
	100	0.463	0.307	0.206	0.253	0.226	0.130	0.274	0.232	0.183	0.267	0.211	0.171	0.245	0.185	0.101
	200	0.770	0.580	0.509	0.506	0.466	0.384	0.575	0.527	0.445	0.554	0.499	0.375	0.536	0.447	0.303
	400	0.957	0.862	0.849	0.810	0.783	0.815	0.853	0.839	0.809	0.853	0.826	0.780	0.849	0.813	0.702
Alpha = 0.1	$T = 50$	0.416	0.299	0.269	0.271	0.280	0.140	0.300	0.227	0.193	0.238	0.226	0.164	0.225	0.186	0.160
	100	0.611	0.450	0.371	0.411	0.354	0.297	0.242	0.350	0.297	0.385	0.325	0.257	0.366	0.291	0.186
	200	0.837	0.689	0.633	0.622	0.596	0.603	0.660	0.637	0.567	0.658	0.630	0.550	0.651	0561	0.467
	400	0.980	0.918	0.914	0.888	0.877	0.897	0.916	0.902	0.870	0.914	0.890	0.859	0.909	0.873	0.812
Test B: joint test of several coefficient parameters																
Alpha = 0.01	$T = 50$	0.102	0.054	0.000	0.052	0.000	0.008	0.018	0.010	0.011	0.021	0.009	0.007	0.008	0.000	0.006
	100	0.258	0.067	0.001	0.097	0.005	0.013	0.029	0.017	0.010	0.036	0.009	0.007	0.017	0.012	0.014
	200	0.528	0.239	0.052	0.343	0.258	0.025	0.090	0.031	0.015	0.062	0.020	0.012	0.043	0.013	0.009
	400	0.923	0.751	0.734	0.844	0.827	0.435	0.709	0.591	0.091	0.686	0.400	0.028	0.639	0.089	0.009
Alpha = 0.05	$T = 50$	0.235	0.128	0.043	0.145	0.077	0.048	0.101	0.052	0.045	0.004	0.050	0.050	0.063	0.045	0.047
	100	0.434	0.201	0.105	0.252	0.132	0.070	0.158	0.101	0.053	0.134	0.079	0.051	0.111	0.054	0.052
	200	0.789	0.490	0.324	0.635	0.562	0.268	0.411	0.223	0.103	0.332	0.108	0.056	0.259	0.127	0.049
	400	0.983	0.895	0.874	0.929	0.917	0.824	0.894	0.835	0.696	0.893	0.797	0.448	0.873	0.707	0.141
Alpha = 0.1	$T = 50$	0.373	0.218	0.115	0.228	0.172	0.098	0.093	0.138	0.129	0.153	0.123	0.114	0.123	0.114	0.105
	100	0.583	0.326	0.203	0.416	0.324	0.162	0.282	0.202	0.146	0.266	0.188	0.117	0.208	0.130	0.107
	200	0.892	0.655	0.559	0.735	0.717	0.482	0.596	0.489	0.308	0.555	0.398	0.247	0.499	0.315	0.116
	400	0.992	0.944	0.922	0.962	0.953	0.925	0.943	0.917	0.869	0.937	0.904	0.784	0.924	0.884	0.467

Table 4. Finite-sample size-adjusted power: DGP2

		OLS	2SLS	3SLS	LA2SLS	LA3SLS	A2SLS	M2SLS (Parzen) $k = 0.3$	M2SLS (Parzen) $k = 0.5$	M2SLS (Parzen) $k = 0.66$	M2SLS (Tukey–Hanning) $k = 0.3$	M2SLS (Tukey–Hanning) $k = 0.5$	M2SLS (Tukey–Hanning) $k = 0.66$	M2SLS (quadratic) $k = 0.3$	M2SLS (quadratic) $k = 0.5$	M2SLS (quadratic) $k = 0.66$
Test A: test a single coefficient parameter																
Alpha = 0.01	$T = 50$	0.895	0.941	0.939	0.874	0.894	0.489	0.872	0.904	0.714	0.835	0.762	0.610	0.791	0.657	0.458
	100	1.000	1.000	1.000	1.000	0.999	0.983	0.997	0.994	0.978	0.996	0.987	0.943	0.993	0.970	0.843
	200	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	0.974
	400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
Alpha = 0.05	$T = 50$	0.956	0.979	0.983	0.964	0.966	0.916	0.951	0.886	0.814	0.912	0.839	0.714	0.887	0.736	0.598
	100	1.000	1.000	1.000	1.000	1.000	0.997	0.998	0.996	0.986	0.998	0.993	0.962	0.996	0.973	0.887
	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.984
	400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Alpha = 0.1	$T = 50$	0.982	0.989	0.990	0.979	0.985	0.962	0.972	0.924	0.862	0.943	0.893	0.779	0.913	0.802	0.651
	100	1.000	1.000	1.000	1.000	1.000	0.997	0.998	0.996	0.987	0.998	0.994	0.966	0.996	0.977	0.902
	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.988
	400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Test B: joint test of several coefficient parameters																
Alpha = 0.01	$T = 50$	0.824	0.842	0.767	0.824	0.785	0.009	0.469	0.337	0.256	0.397	0.303	0.198	0.343	0.212	0.182
	100	0.999	1.000	1.000	0.998	0.998	0.020	0.977	0.938	0.522	0.961	0.799	0.389	0.935	0.461	0.233
	200	1.000	1.000	1.000	1.000	1.000	0.938	1.000	1.000	0.998	1.000	1.000	0.988	1.000	0.996	0.524
	400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Alpha = 0.05	$T = 50$	0.949	0.924	0.909	0.907	0.912	0.303	0.817	0.607	0.408	0.658	0.474	0.317	0.588	0.337	0.256
	100	1.000	1.000	1.000	0.999	1.000	0.973	0.996	0.990	0.942	0.994	0.982	0.801	0.988	0.910	0.497
	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	1.000	0.999	0.985
	400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Alpha = 0.1	$T = 50$	0.963	0.962	0.961	0.945	0.945	0.729	0.897	0.787	0.637	0.827	0.679	0.468	0.775	0.528	0.367
	100	1.000	1.000	1.000	1.000	1.000	0.999	1.000	0.997	0.995	0.999	0.995	0.966	0.996	0.985	0.784
	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.994
	400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4 Finite-sample size-adjusted power: DGP3.

		OLS	2SLS	3SLS	LA2SLS	LA3SLS	A2SLS	M2SLS (Parzen) $k = 0.3$	M2SLS (Parzen) $k = 0.5$	M2SLS (Parzen) $k = 0.66$	M2SLS (Tukey–Hanning) $k = 0.3$	M2SLS (Tukey–Hanning) $k = 0.5$	M2SLS (Tukey–Hanning) $k = 0.66$	M2SLS (quadratic) $k = 0.3$	M2SLS (quadratic) $k = 0.5$	M2SLS (quadratic) $k = 0.66$
Test A: test a single coefficient parameter																
Alpha = 0.01	$T = 50$	0.578	0.659	0.621	0.640	0.618	0.452	0.662	0.651	0.635	0.664	0.626	0.586	0.664	0.603	0.382
	100	0.917	0.989	0.988	0.985	0.985	0.977	0.965	0.989	0.984	0.967	0.986	0.927	0.985	0.972	0.815
	200	1.000	1.000	1.000	1.000	1.000	0.999	0.999	1.000	1.000	0.999	1.000	0.992	1.000	1.000	0.970
	400	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.994
Alpha = 0.05	$T = 50$	0.808	0.883	0.864	0.834	0.816	0.844	0.846	0.844	0.838	0.823	0.842	0.758	0.827	0.782	0.628
	100	0.987	0.997	0.996	0.994	0.997	0.995	0.991	0.994	0.991	0.992	0.994	0.968	0.994	0.983	0.930
	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	1.000	1.000	0.989
	400	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.995
Alpha = 0.1	$T = 50$	0.892	0.930	0.926	0.904	0.898	0.919	0.891	0.890	0.894	0.873	0.889	0.842	0.880	0.853	0.755
	100	0.994	0.999	0.998	0.998	0.998	0.995	0.995	0.996	0.995	0.995	0.996	0.975	0.996	0.988	0.947
	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.993
	400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.995
Test B: joint test of several coefficient parameters																
Alpha = 0.01	$T = 50$	0.800	0.812	0.720	0.826	0.773	0.338	0.748	0.644	0.571	0.667	0.554	0.397	0.615	0.443	0.226
	100	0.998	0.998	0.996	0.998	0.998	0.976	0.997	0.994	0.990	0.994	0.993	0.775	0.994	0.979	0.420
	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.993	1.000	0.999	0.972
	400	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	0.989
Alpha = 0.05	$T = 50$	0.926	0.942	0.928	0.903	0.891	0.837	0.891	0.870	0.825	0.858	0.844	0.709	0.865	0.739	0.526
	100	1.000	1.000	1.000	1.000	1.000	0.994	1.000	1.000	0.999	1.000	1.000	0.972	0.999	0.992	0.922
	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.990
	400	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999
Alpha = 0.1	$T = 50$	0.969	0.977	0.964	0.951	0.945	0.925	0.948	0.918	0.888	0.925	0.894	0.837	0.909	0.834	0.692
	100	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	0.980	1.000	0.993	0.958
	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.993
	400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999

constructed from LA2SLS or LA3SLS perform favourably compared to other estimators that make use of T-convergence properties in the non-stationary direction. In other words, they appear to further enhance the implication of Tables 1 and 2 that although LA2SLS and LA3SLS are only $\sqrt{T}$-convergent, in finite sample the loss of efficiency, if any, is not of significant concern for a small-scale structural dynamic model.

To see the extent to which the finite sample distributions of the test statistics deviate from the asymptotic ones, for each estimator, we compare the actual versus theoretical histogram of its t-statistics for test A and Wald statistics for test B. The theoretical histograms are derived from the standard normal distribution for test A and chi-square distribution with three degrees of freedom for test B. To conserve space, we report the actual histograms only for $T = 200$. The actual histograms for $T = 50$, 100 and 400 are available upon request.

Figures 1–3 plot the actual and theoretical histograms of the t-statistics for test A for DGP1, DGP2 and DPG3, respectively. The dotted line is used for the theoretical histogram and the solid line for the actual histogram. The OLS is severely biased downward. The actual histograms of the 2SLS, 3SLS, LA2SLS, LA3SLS and the modified estimators are similar and close to the theoretical one. Thus, as we inferred from Table 3, for test A, normal approximation works well for all the consistent estimators when $T = 200$.

Figures 4–6 plot the actual and theoretical histograms of the Wald statistics for test B for DGP1, DGP2 and DPG3, respectively. For DGP1 and DGP2, the actual histograms of the lag-augmented estimators are very close to the theoretical ones while the actual histograms of the conventional 2SLS and 3SLS estimators have fatter tails. For DGP3, the actual histograms of both the conventional and lag-augmented estimators are very close to the theoretical ones. This is expected since, as we discussed earlier, the Wald statistics based on 2SLS and 3SLS have a chi-square limiting distribution for DGP3, but not for DGP1 or DGP2. We also noted large spikes at the right end of some of the actual histograms of the modified estimators, indicating that Wald statistics based on the modified estimators, especially the M2SLS with bandwidth expansion rate k=0.66, tend to have a mass at large values.

Since, a priori, one does not know if the null hypothesis involves estimates in the non-stationary direction,[9] the limited Monte Carlo results appear to suggest that if the null hypothesis involves a single coefficient parameter and efficiency of the point estimates is a concern, 2SLS or 3SLS should be used. On the other hand, if the null hypothesis involves a number of coefficient parameters, then one probably should use LA2SLS or LA3SLS if T is moderate. For T smaller than 200, both 2SLS and LA2SLS are subject to size distortions for test B. However, size distortion for 2SLS is about twice as large as for LA2SLS. For $T = 400$, the actual sizes of the LA2SLS and LA3SLS estimators are almost identical to the nominal sizes.

6. CONCLUSIONS

In this paper, we consider inference of a structural dynamic model of non-stationary and possibly cointegrated variables without the prior knowledge of unit roots or rank of cointegration. Despite

[9] Sims *et al.* (1990) show that in a trivariate reduced form VAR model, the F-statistics of Granger causality test have a chi-square limiting distribution if the system has a single cointegrating vector. However, this result may not hold in our structural dynamic model. Test B in this paper can be viewed as a causality test from y_2 to y_1 if we set $c_0 = c_1 = c_2 = 0$. In DGP2, although the system has a single cointegrating vector, it involves y_2 and y_3, not y_1 and y_2. Wald statistics for test B in this case do not have a chi-square limiting distribution.

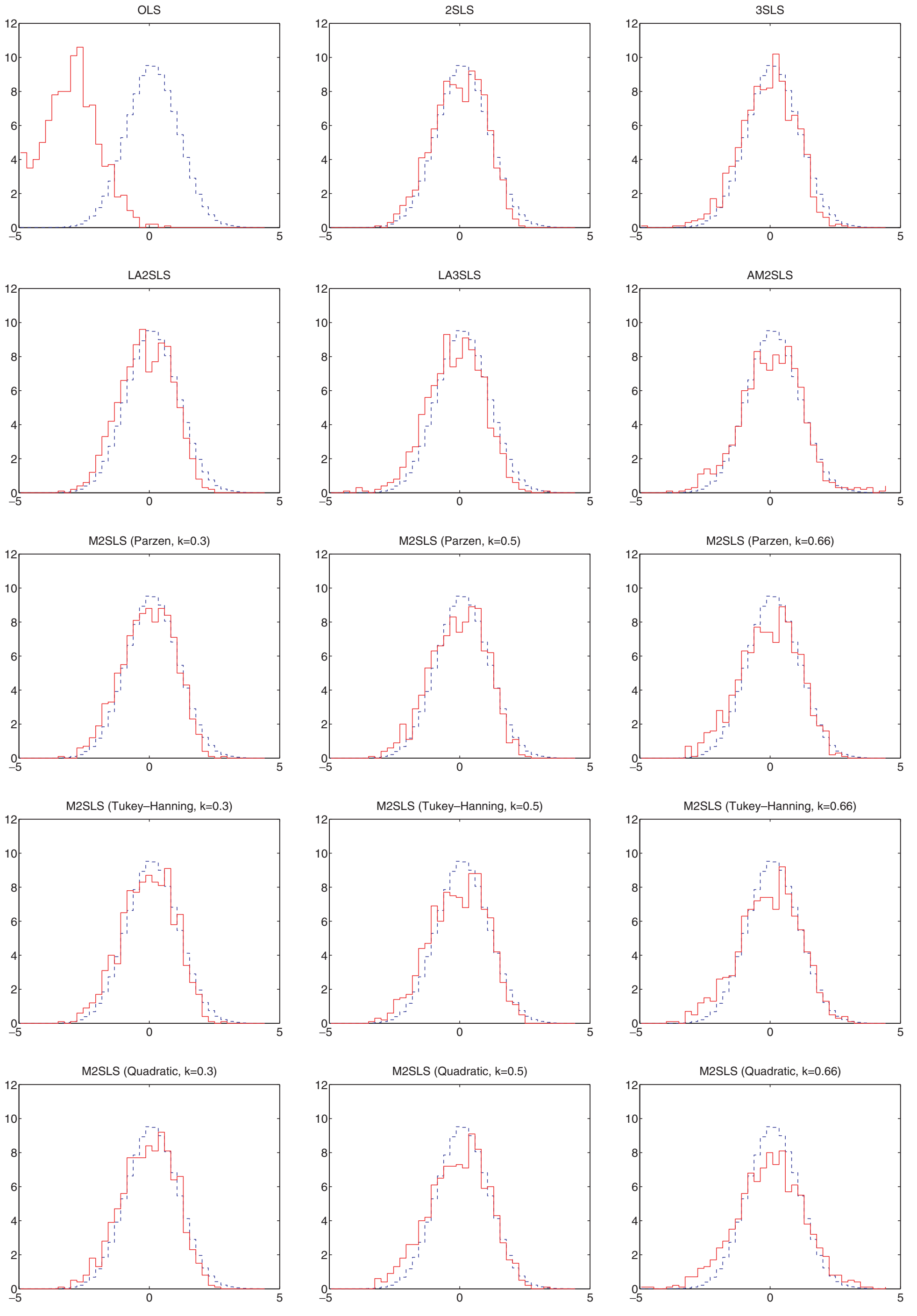

Figure 1. Histogram of t-statistics of test A. DGP1, $T = 200$. - - - theoretical, — acutal.

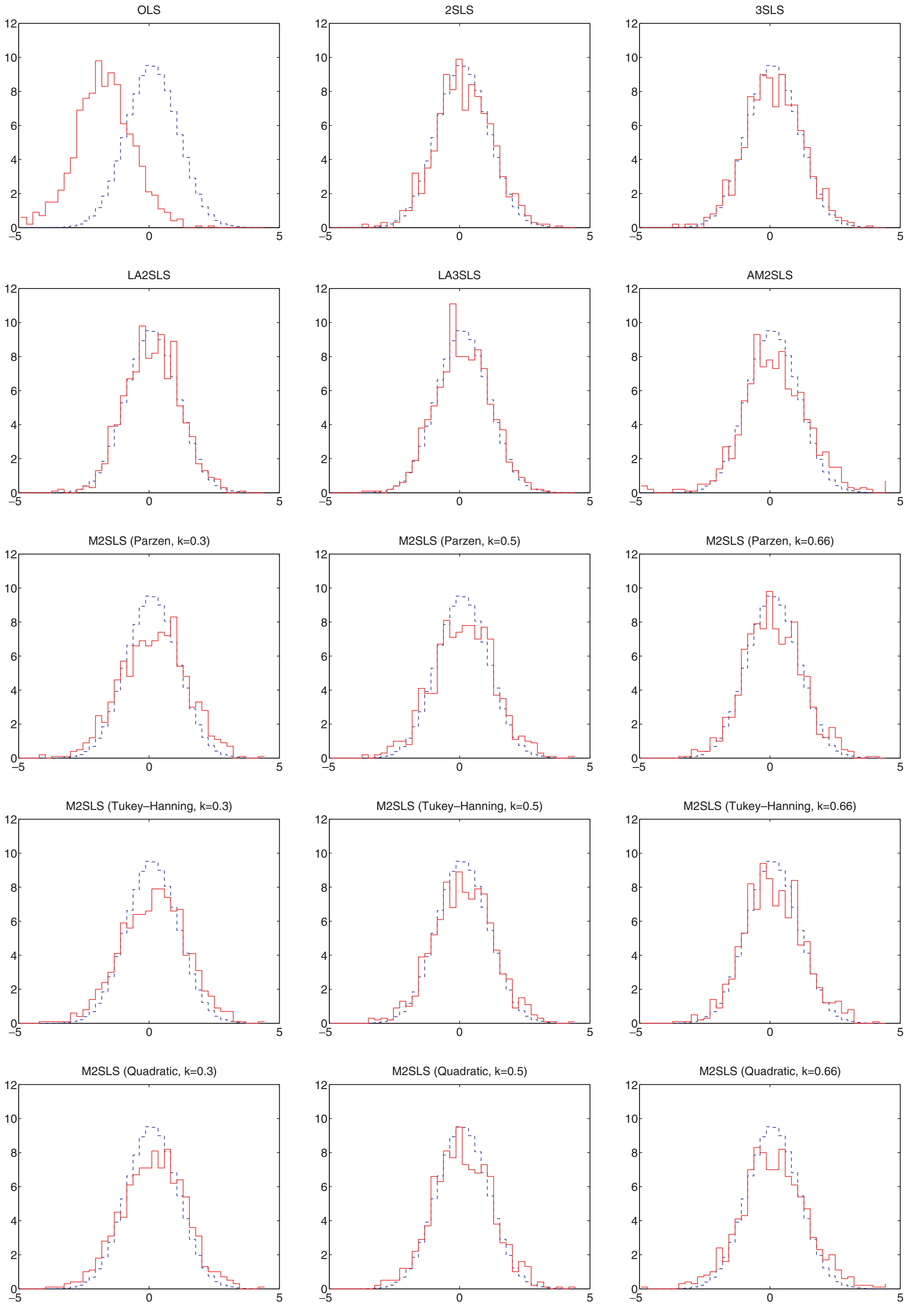

Figure 2. Histogram of t-statistics of test A. DGP2, $T = 200$. - - - theoretical, — acutal.

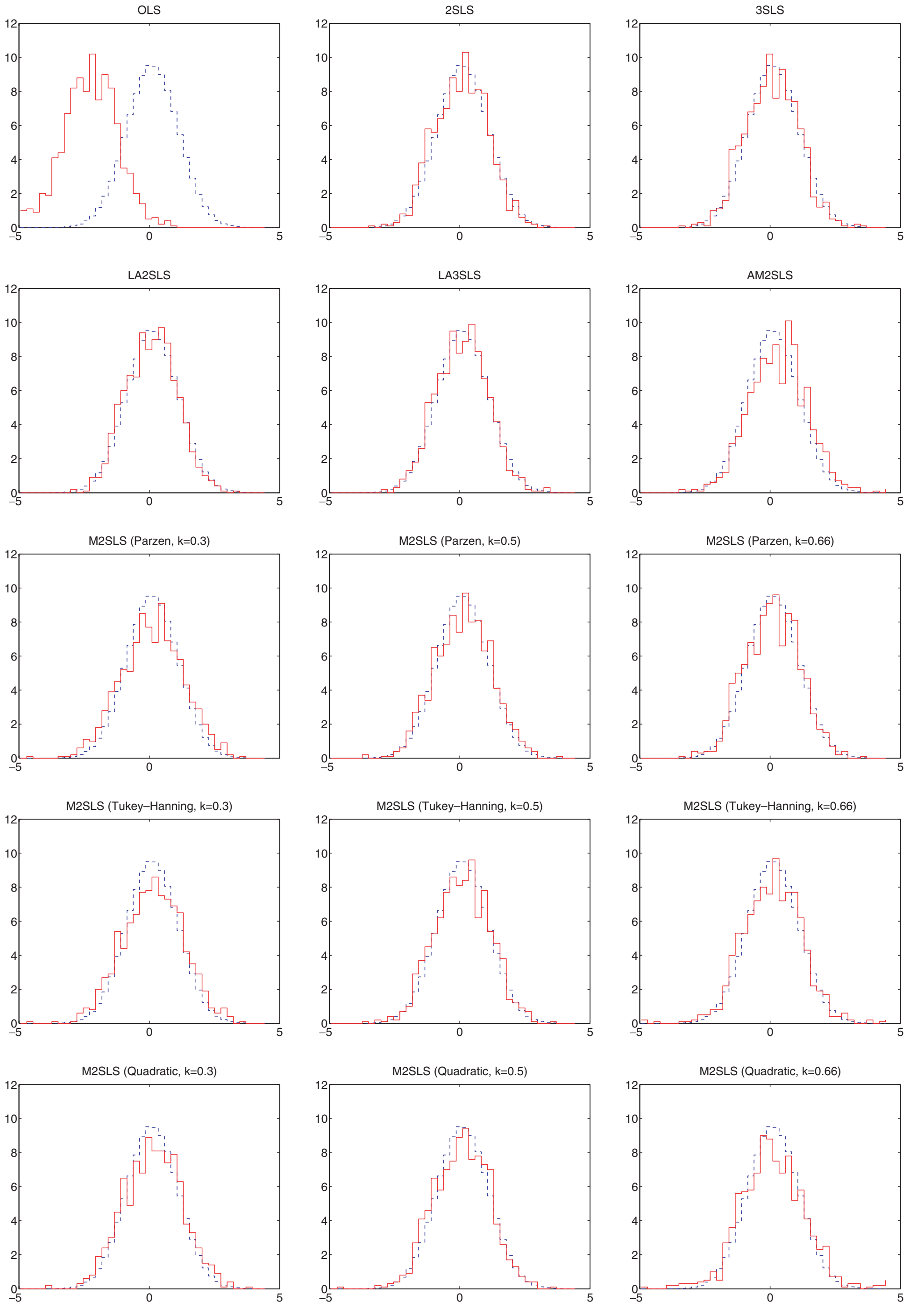

Figure 3. Histogram of t-statistics of test A. DGP3, $T = 200$. - - - theoretical, — acutal.

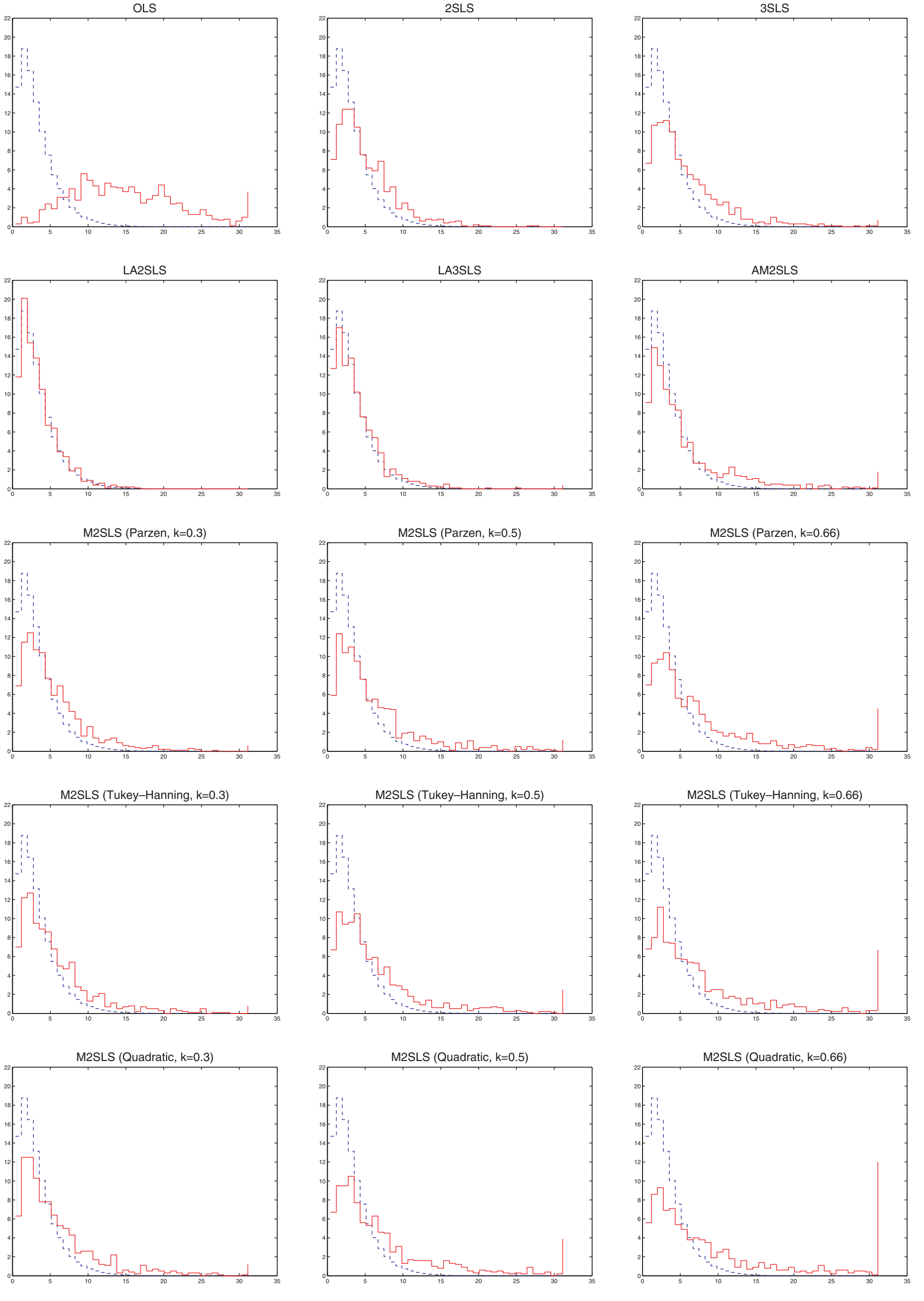

Figure 4. Histogram of Wald statistics of test B. DGP1, $T = 200$. - - - theoretical, — acutal.

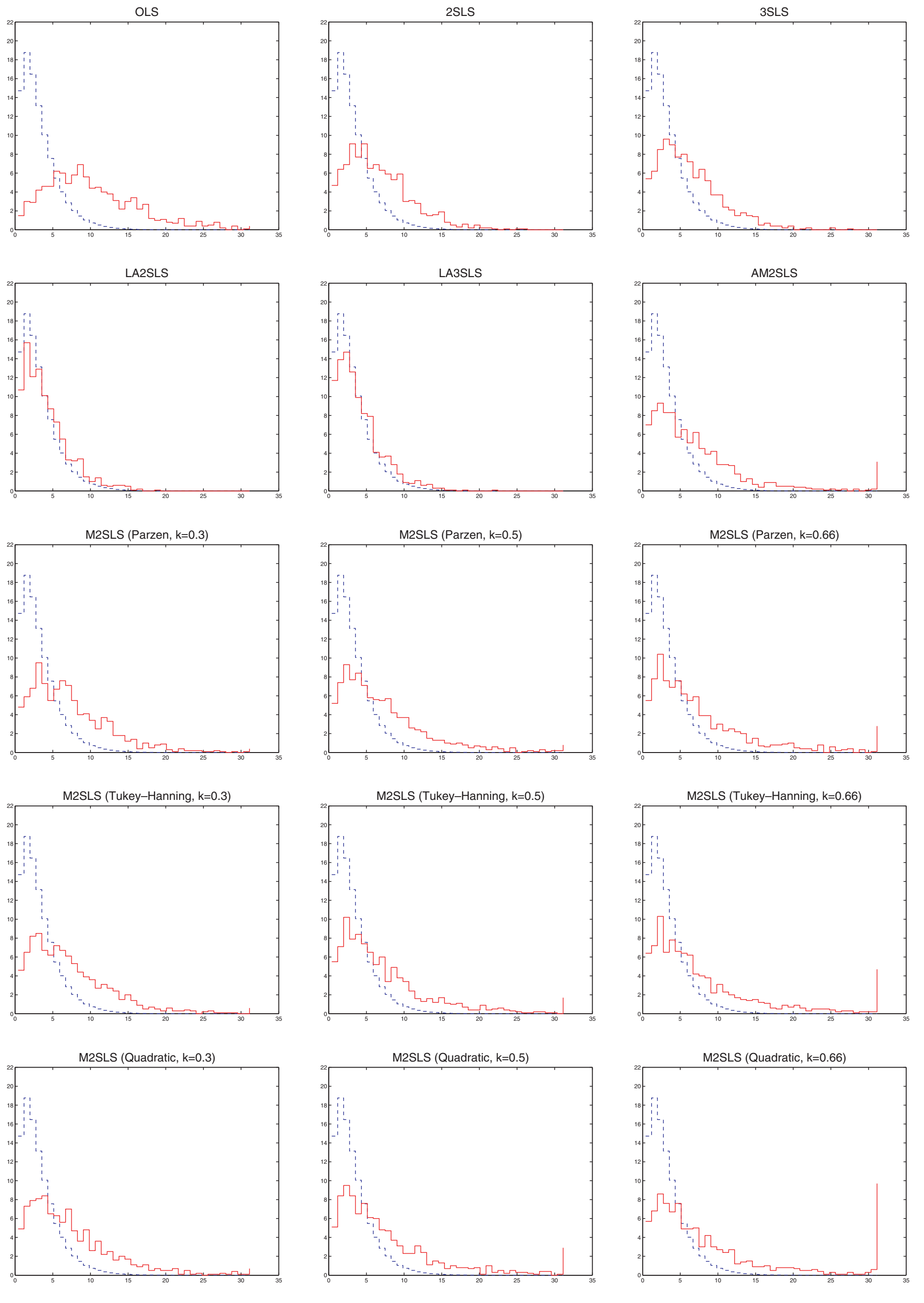

Figure 5. Histogram of Wald statistics of test B. DGP2, $T = 200$. - - - theoretical, — acutal.

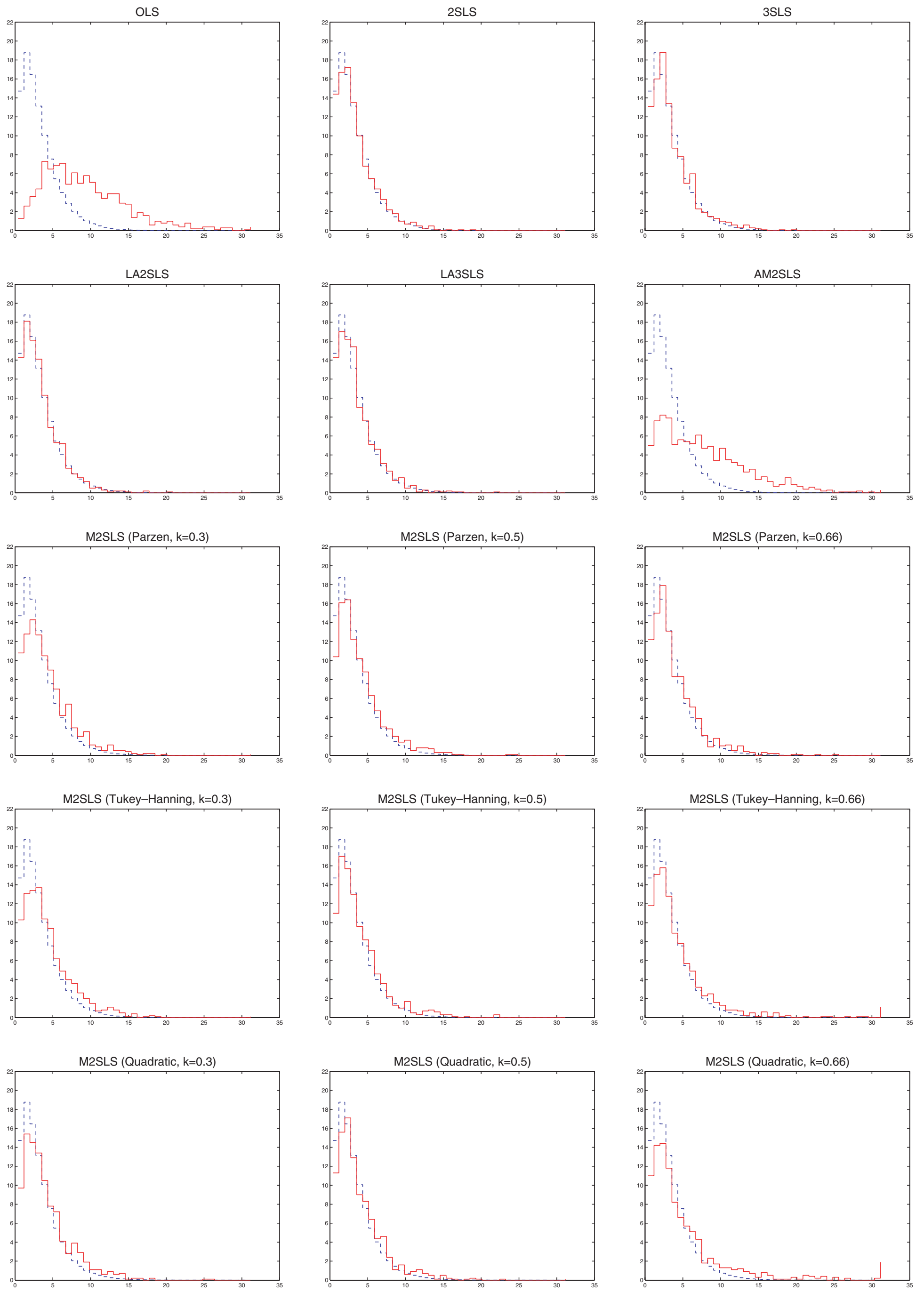

Figure 6. Histogram of Wald statistics of test B. DGP3, $T = 200$. - - - theoretical, — acutal.

all variables being integrated of order 1, the OLS estimator of the structural parameters is not always consistent. The conventional 2SLS and 3SLS estimators are consistent. 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 may not be chi-square distributed. However, when prior restrictions are in the form of exclusion restrictions and the structural dynamic model has lag order greater than 1, the t-statistics of individual coefficient parameters remain asymptotically $N(0, 1)$ distributed.

We have also suggested a lag-augmented 2SLS or 3SLS estimator. The resulting estimators are $\sqrt{T}$-consistent and asymptotically normally distributed. Compared to the 2SLS estimator, the LA2SLS estimator is likely to be less efficient asymptotically because it ignores some of the prior restrictions on the model and loses the T-convergent components that appear in 2SLS. However, in addition to the desirable large sample properties, the LA2SLS and LA3SLS are simple to implement and have closed form asymptotic covariance matrices. On the other hand, 2SLS and Phillips–Hansen type modified estimators, although may be more efficient asymptotically, do not have closed form formulae for their asymptotic covariance matrices.

Monte Carlo studies are conducted to shed light on the finite sample performance of various estimators. In general, we find that in a small system LA2SLS and LA3SLS have comparable bias and root mean squared error with the 2SLS and 3SLS. Their actual size is also close to the nominal size even for moderate T whether the null hypothesis involves single or multiple restrictions while the two modified estimators proposed by Hsiao and Wang (2006), despite their desirable large sample properties, do not work very well in small sample.

An alternative approach would be first to test for the rank of cointegration, then impose the reduced rank condition to derive the limited and full information maximum likelihood estimator. Unfortunately, even ignoring the issues of unreliability of cointegration rank test in finite sample (e.g. Ho and Sorensen 1996; Gonzalo and Pitarakis 1999), unlike the case of reduced form reduced rank models, computationally feasible one stage limited and full information maximum likelihood estimators that impose the prior cointegrating rank restriction are yet to be derived for the structural models.

ACKNOWLEDGEMENTS

We wish to thank the two referees for helpful comments on an early version entitled ‘A Lag-Augmented Two- and Three-Stage Least Squares Estimator for Structural Vector Autoregressive Integrated Processes’. As a matter of fact, the current title follows a suggestion of a referee to avoid the confusion of the term associated with Structural Vector Autoregressive Process. We also wish to thank H. Lawrence Hotchkiss for technical assistance in Matlab.

REFERENCES

Blanchard, O. J. and D. Quah (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review 79*, 655–73.

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

Dolado, J. J. and H. Lütkepohl (1996). Making Wald tests work for cointegrated VAR systems. *Econometric Reviews 15*, 369–86.

Engle, R. F. and C. W. J. Granger (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica 55*, 251–76.

Gonzalo, J. and J. Y. Pitarakis (1999). Dimensionality effect in cointegrated systems, in R. Engle and H. White (Eds.), *Granger Festschrift*, Oxford: Oxford University Press.

Granger, C. W. J. (1969). Investigating causal relations by econometric eodels and cross-spectral methods. *Econometrica 37*, 424–38.

Ho, M. and B. Sorensen (1996). Finding cointegration rank in high dimensional systems using the johansen test: An illustration using data based on Monte Carlo simulations. *Review of Economics and Statistics 78*, 726–32.

Hsiao, C. (1997a). Cointegration and dynamic dimultaneous equations models. *Econometrica 65*, 647–70.

Hsiao, C. (1997b). Statistical properties of the two stage least-squares estimator under cointegration. *Review of Economic Studies 64*, 385–98.

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

Hsiao, C. and S. Wang (2006). Modified two stage least squares estimator for the estimation of a structural vector autoregressive integrated process. *Journal of Econometrics 135*, 427–63.

Hurd, M. (1972). Small-sample estimation of a structural equation with autocorrelated errors. *Journal of the American Statistical Association 67*, 567–73.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control 12*, 231–54.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica 59*, 1551–80.

King, R. G., C. I. Plosser, J. H. Stock and M. W. Watson (1991). Stochastic trends and economic flucations. *American Economic Review 81*, 819–40.

Kitamura, Y. and P. C. B. Phillips (1997). Fully modified IV, GIVE and GMM estimation with possibly non-stationary regressors and instruments. *Journal of Econometrics 80*, 85–124.

Klein, L. R. (1950). *Economic Fluctuations in the United States, 1921–1941*, Cowles Commission Monograph 11, New York: Wiley and Sons, Inc.

Klein, L. R. and M. K. Evans (1969). *Econometric Gaming*,. New York: MacMillan.

Klein, L. R. and A. S. Goldberger (1955). *An Econometric Model of the United States*,. Amsterdam: North Holland.

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

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. New York: Springer-Velag.

MacKinnon, J. G. and H. White (1985). Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics 29*, 53–70.

Malinvaud, E. (1980). *Statistical Methods of Econometrics*, 3rd ed., Amsterdam: North-Holland.

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

Phillips, P. C. B. (1995). Fully modified least-squares and vector autoregression. *Econometrica 63*, 1023–78.

Phillips, P. C. B. and S. N. Durlauf (1986). Multiple time series regression with integrated processes. *Review of Economic Studies 53*, 473–95.

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

Sims, C. A., Stock, J. H. and M. W. Watson (1990). Inference in linear time series models with some unit roots. *Econometrica 58*, 113–44.

Stock, J. H. (1987). Asymptotic properties of least-squares estimators of cointegrating vectors. *Econometrica 55*, 1035–56.

Toda, H. Y. and P. C. B. Phillips (1993). Vector autoregression and causality. *Econometrica 61*, 1367–93.

Toda, H. Y. and T. Yamamoto (1995). Statistical Inference in Vector Autoregressions with Possibly Integrated Processes. *Journal of Econometrics*, *66*, 225–50.

Tsay, R. S., P. C. B. Phillips and G. C. Tiao (1990). Asymptotic properties of multivariate non-stationary processes with applications to autoregressions. *Annals of Statistics 18*, 220–50.

Wold, H. (1954). *A Study in the Analysis of Stationary Time Series*, 2nd ed., Uppsala: Almquist and Wicksell.

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