ECONOMETRIC REVIEWS, 20(4), 485–505 (2001)

OPTIMAL IV ESTIMATION OF SYSTEMS WITH STOCHASTIC REGRESSORS AND VAR DISTURBANCES WITH APPLICATIONS TO DYNAMIC SYSTEMS

David M. Mandy[1] and Carlos Martins-Filho[2,*]

[1]Department of Economics, University of Missouri, 118 Professional Building, Columbia, MO 65211 USA
E-mail: mandy@econ.missouri.edu
[2]Department of Economics, Oregon State University, and EPGE-Graduate School of Economics, Fundação Getulio Vargas, Praia de Botafogo, 190-11o. Andar, Rio de Janeiro, RJ 22253-900 Brazil
E-mail: cmartins@fgv.br

ABSTRACT

This paper considers the general problem of Feasible Generalized Least Squares Instrumental Variables (FGLS IV) estimation using optimal instruments. First we summarize the sufficient conditions for the FGLS IV estimator to be asymptotically equivalent to an optimal GLS IV estimator. Then we specialize to stationary dynamic systems with stationary VAR errors, and use the sufficient conditions to derive new moment conditions for these models. These moment conditions produce useful IVs from the lagged endogenous variables, despite the correlation between errors and endogenous variables. This use of the information contained in the lagged endogenous variables expands the class of IV estimators under consideration and thereby potentially improves both asymptotic and small-sample efficiency of the optimal IV estimator in the class. Some Monte Carlo experiments compare the new methods with those of Hatanaka (1976). For the DGP used in the Monte Carlo experiments, asymptotic efficiency is strictly improved by the new IVs, and experimental small-sample efficiency is improved as well.

Key Words: Dynamic models; IV estimation; VAR errors

JEL Classification: C30.

1. INTRODUCTION

Mandy and Martins-Filho (1994 and 1997) provide sufficient conditions for asymptotic equivalence of feasible generalized least squares (FGLS) instrumental variables (IV) and generalized least squares (GLS) IV estimators. There the focus was on the structure of the error covariance matrix, and due to the generality of the model little attention was given to instrument design.

Here we note that a minor restatement of our earlier conditions is sufficient for asymptotic equivalence of FGLS IV and GLS IV when the instruments are optimal. This observation is new in that the conditions are general, applying to many familiar covariance structures, and the estimators are optimal in the class of IV estimators under consideration.

We then apply the conditions to stationary dynamic systems with stationary VAR errors. The sufficient conditions allow us to expand the class of IV estimators under consideration by identifying new moment conditions that enable use of (transformed) lagged endogenous variables as IVs, despite the presence of VAR errors in the dynamic system. This raises the prospect of both asymptotic and small sample efficiency gains relative to the IV estimators that have been considered to date. For a particular data generation process (DGP), we show that a strict improvement in asymptotic efficiency is obtained by using our new IVs, and experimental results for the DGP suggest that small-sample efficiency is improved as well. The results differ from earlier literature, notably Dhrymes and Taylor (1976) and Hatanaka (1976), in that the efficiency properties do not depend on specific distributional assumptions but instead are attained in a class of IV estimators.

Section 2 discusses sufficient conditions for FGLS IV estimators to be optimal IV estimators for some given set of IVs. Section 3 reviews estimation of error VAR(1) nuisance parameters that accommodates general stochastic regressors. Section 4 examines dynamic models and proposes IVs that satisfy the sufficient conditions from Section 2, thereby proposing new optimal IV estimators for dynamic models with VAR errors. Section 5 presents some Monte Carlo comparisons of the new efficient IV estimators with each other and with Hatanaka's (1976) methods.

2. SUFFICIENT CONDITIONS FOR ASYMPTOTIC EQUIVALENCE OF GLS IV AND FGLS IV USING OPTIMAL INSTRUMENTS

Consider the linear regression model

$$y = x\beta + u; \quad E(u) = 0; \quad E(uu') \equiv \Omega(\theta). \tag{1}$$

Here y is a $T \times 1$ stochastic observable vector, x is a $T \times K$ stochastic observable matrix, β is a $K \times 1$ nonstochastic unknown parameter vector to be estimated, u is a $T \times 1$ stochastic unobservable error vector, and θ is a $n \times 1$ nonstochastic vector of unknown nuisance parameters. When the conditional mean $E(u|x)$ is nonzero IV estimation is required. White (1984, Chapter VII) established the asymptotic equivalence of GLS IV and FGLS IV for some standard forms of $\Omega(\theta)$. Mandy and Martins-Filho (1994 and 1997) proposed general conditions on $\Omega(\theta)$ that are sufficient for this asymptotic equivalence, given by:

(A1.1) $\Omega(\theta)^{-1}$ has at most $W < \infty$ distinct nonzero elements for every T, denoted $g_{wT}(\theta)$ for $w = 1, \ldots, W$.
That is, there are $T^2 - W$ elements that are either zero or duplicates of other nonzero elements in $\Omega(\theta)^{-1}$. For each w, $g_{wT}(\theta)$ converges uniformly as $T \to \infty$ to a real-valued function $g_w(\theta)$ on an open set S containing the true value of θ, at which g_w is continuous.

(A1.2) The number of nonzero elements in each column (and row) of $\Omega(\theta)^{-1}$ is uniformly bounded by $N < \infty$ as $T \to \infty$.

Corresponding to a $T \times \bar{K} (\bar{K} \geq K)$ matrix D of IVs are the optimal instruments

$$z(\theta) \equiv \Omega(\theta)^{-1} D (D' \Omega(\theta)^{-1} D)^{-1} D' \Omega(\theta)^{-1} x,$$

which are used to form the GLS IV estimator $\hat{\beta}(\theta) \equiv (z(\theta)'x)^{-1} z(\theta)'y$. Standard asymptotic behavior of $\hat{\beta}(\theta)$ and asymptotic equivalence of the FGLS IV estimator $\hat{\beta}(\hat{\theta})$ (for some consistent estimator $\hat{\theta}$) with $\hat{\beta}(\theta)$ require that the IVs D possess some basic properties. From (A1), let I_{iwT} be the index set of elements in row i of $\Omega(\theta)^{-1}$ that are equal to g_w, for $w = 1, \ldots, W$. In this notation, the IVs D are assumed to satisfy:

(A2.1) $(1/\sqrt{T}) D' \Omega(\theta)^{-1} u \xrightarrow{d} N(0, Q_{Du})$ for some symmetric $(\bar{K} \times \bar{K})$ matrix Q_{Du}.

(A2.2) $\underset{T \to \infty}{\text{plim}} (1/T) D' \Omega(\theta)^{-1} D = Q_{DD}$, a finite nonsingular matrix.

(A2.3) $\underset{T \to \infty}{\text{plim}} (1/T) D' \Omega(\theta)^{-1} x = Q_{Dx}$, a finite matrix of full column rank.

(A2.4) Each IV D_{ih} can be expressed as $D_{ih} = \lambda_h' \eta_{ih}$, where λ_h is a vector of fixed finite dimension that is $O_p(1)$ and constant across i, and η_{ih} has uniformly bounded fourth moments and fourth cross moments with x_{jq} for $i, j = 1, 2, \ldots$; $h = 1, \ldots, \bar{K}$; and $q = 1, \ldots, K$.

(A2.5) $\sum_{i=1}^{T} \sum_{j \in I_{iwT}} D_{ih} u_j = O_p(T^{1/2})$ for $h = 1, \ldots, \bar{K}$ and $w = 1, \ldots, W$.

Under (A1) and (A2) there is a FGLS IV estimator that is optimal within the class of IV estimators that are based on D, provided we have available a consistent estimator of the nuisance parameters.

Lemma 1
Assume (A1) and (A2). If $\hat{\theta} \xrightarrow{p} \theta$ then $\sqrt{T}(\hat{\beta}(\hat{\theta}) - \hat{\beta}(\theta)) \xrightarrow{p} 0$.

The proof is an extension of Mandy and Martins-Filho (1994 and 1997) and is available on request from the authors. Condition (A1) applies to VAR errors and certain forms of heterocedastic (including random coefficients) and panel errors, but does not apply to moving average errors.

3. VAR ERRORS

Now we specialize the model to a system of T observations on G equations with VAR(1) errors,[1] $Y = XB + U$, where Y is a $T \times G$ matrix of endogenous variables, X is a $T \times K$ matrix of (possibly stochastic) regressors, B is a $K \times G$ matrix of unknown parameters that incorporates any exclusion restrictions, and U is a $T \times G$ matrix of VAR(1) errors. Letting the ith row and column of an arbitrary matrix M be $M_{i.}$ and $M_{.i}$, respectively, we assume that:

(A3) $U_{t.} = U_{t-1.}R + V_{t.}$ for $t = 0, \pm 1, \pm 2, \ldots,$ where

(A3.1) R is a $G \times G$ matrix of nuisance parameters with absolute eigenvalues less than one,

(A3.2) $V'_{t.} \sim IID(0, \Sigma)$ for some symmetric finite positive definite $G \times G$ nuisance matrix Σ,

(A3.3) $V_{t.}$ has finite absolute fourth moments, and

(A3.4) there exists a bound $\bar{B}$ such that $E(X_{t-h,i}X_{tj}X_{\tau-h,i}X_{\tau j}) \leq \bar{B}$ and $E(X_{t-h,i}X_{\tau-h,i}U_{t\ell}U_{\tau\ell}) \leq \bar{B}$; for every $t, \tau = 1, 2, \ldots;$ $h = 0, 1, 2, \ldots;$ $i, j = 1, \ldots, K;$ and $\ell = 1, \ldots, G$.

The autocovariance function of $U_{t.}$ is denoted $\Gamma(h) = E(U'_{t-h.}U_{t.})$ for $h = 0, \pm 1, \pm 2, \ldots$. Assumption (A3) implies $U_{t.}$ is both strictly stationary and covariance stationary (Anderson (1971), pp. 372–378). Henceforth we express the error VAR as $U = \Delta UR + \Delta\Delta' V + (I_T - \Delta\Delta')U$ for

$$\Delta = \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ & 1 & \ddots & & \\ & & \ddots & \ddots & \\ & & & 1 & 0 \end{bmatrix}.$$

In this model, the nuisance vector θ of the previous section consists of vec R and the unique elements of Σ. Given that $E(U|X) \neq 0$, consistent estimation of θ requires a $T \times K$ instrument matrix C (possibly different from D) which we assume satisfies

(A4.1) the eigenvalues of $(1/T)C'X$ possess a uniform lower bound in probability, and

(A4.2) $C'U = O_p(T^{1/2})$.

[1] It is unnecessary to consider higher order error VARs since a VAR(p) process can always be written as a VAR(1) process Anderson (1971), p. 177.

Estimation of the nuisance parameters is straightforward from (A3) and (A4).

Lemma 2

Assume (A3) and (A4), and let $\tilde{\Gamma}(h) = (1/T)\sum_{t=h+1}^{T} \tilde{U}'_{t-h\cdot}\tilde{U}_{t\cdot}$, where $\tilde{U} = (I_T - X(C'X)^{-1}C')Y$. Then $\tilde{\Gamma}(h) \xrightarrow{p} \Gamma(h)$. In particular, $\tilde{R} - R = O_p(T^{-1/2})$ and $\tilde{\Sigma} \xrightarrow{p} \Sigma$, where $\tilde{R} = \tilde{\Gamma}(0)^{-1}\tilde{\Gamma}(1)$ and $\tilde{\Sigma} = \tilde{\Gamma}(0) - \tilde{R}'\tilde{\Gamma}(1)$.

The standard proof from, for example, Fuller (1976) must be modified only slightly to accommodate the nonzero conditional mean. Details are available on request from the authors. Note that $\tilde{R}$ converges faster than $o_p(1)$. This fact is used in the design of new IVs in Section 4.

For checking assumptions (A1) and (A2) it is convenient to stack the system by observation, thereby expressing the model in the notation of Section 2 with dependent vector $y = \mathrm{vec}Y'$, regressor matrix $x = X \otimes I_G$, parameter vector $\beta = \mathrm{vec}B'$, and error vector $u = \mathrm{vec}U'$. In this notation the VAR(1) inverse error covariance is $\Omega(\theta)^{-1} = P'(I_T \otimes \Sigma^{-1})P$, where P is the VAR(1) transformation matrix given by

$$P = \begin{bmatrix} A & 0 & \dots & \dots & 0 \\ -R' & I_G & 0 & \dots & \vdots \\ 0 & -R' & I_G & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -R' & I_G \end{bmatrix} = ((I_T - \Delta\Delta') \otimes A) - (\Delta \otimes R') + (\Delta\Delta' \otimes I_G).$$

Here A is a lower triangular matrix such that $AE(U'_{1\cdot}U_{1\cdot})A' = \Sigma$. From P and the definition of Δ we can derive an explicit expression for $\Omega(\theta)^{-1}$ that is useful for verifying (A1) and (A2):

$$\begin{aligned} \Omega(\theta)^{-1} &= P'(I_T \otimes \Sigma^{-1})P \\ &= [(I_T - \Delta\Delta') \otimes A'\Sigma^{-1}A] + [\Delta\Delta' \otimes \Sigma^{-1}] - [\Delta \otimes \Sigma^{-1}R'] - [\Delta' \otimes R\Sigma^{-1}] + [\Delta'\Delta \otimes R\Sigma^{-1}R'] \\ &= \begin{bmatrix} A'\Sigma^{-1}A + R\Sigma^{-1}R' & -R\Sigma^{-1} & 0 & \dots & 0 \\ -\Sigma^{-1}R' & \Sigma^{-1} + R\Sigma^{-1}R' & -R\Sigma^{-1} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -\Sigma^{-1}R' & \Sigma^{-1} + R\Sigma^{-1}R' & -R\Sigma^{-1} \\ 0 & \dots & 0 & -\Sigma^{-1}R' & \Sigma^{-1} \end{bmatrix}. \end{aligned} \tag{2}$$

4. IVS FOR ESTIMATION OF DYNAMIC SIMULTANEOUS EQUATION MODELS WITH VAR ERRORS

The availability of IVs is context-dependent and can only be investigated under some assumption on the source of the nonzero conditional mean. Perhaps

the most obvious context is a simultaneous equation model. The usual 2SLS and 3SLS IVs can be expected to satisfy (A2) under traditional assumptions when there is only contemporaneous error correlation, an observation that only adds to traditional results for simultaneous equation models by including them as special cases of well-behaved stochastic regressors. When errors satisfy (A3) three problems emerge. First, (A1) must be checked. As noted in Mandy and Martins-Filho (1994), inspection of (2) verifies that (A1) is indeed satisfied by VAR errors. Second, we must derive IVs that satisfy (A2). The usual 2SLS and 3SLS IVs suffice if there are no lagged endogenous variables,[2] but if $Y = XB + U$ is a dynamic simultaneous equation model then, as is well-known from Wallis (1967 and 1972), the lagged endogenous variables as IVs interact with the VAR errors to make traditional 2SLS and 3SLS inconsistent. We demonstrate here the utility of stating the sufficient conditions generally, in the form of (A1) and (A2), by using these conditions to develop new IVs that overcome this problem. The new IVs are derived from lagged endogenous variables and thereby provide a heretofore unnoticed way of using some information from the lagged endogenous variables for optimal IV estimation of B. Third, we must derive IVs that satisfy (A4). Following Wallis (1967 and 1972), this is usually straightforward provided the structural model contains exogenous variables.

Letting $X = [S\ Y_0\ Y_{-1}\ \cdots\ Y_{-q}]$ and $B' = [\Theta'\ B_0'\ B_1'\ \cdots B_q']$, we have a dynamic simultaneous equation model written in normalized structural form $Y = S\Theta + \sum_{L=0}^{q} Y_{-L}B_L + U$, where subscripts on the endogenous matrix Y indicate lags, S is a $T \times K'$ matrix of variables not determined endogenously by the system under consideration (but perhaps stochastic), and Θ and B_L are $K' \times G$ and $G \times G$ unknown parameter matrices, respectively, that incorporate normalization and exclusion restrictions (so $K = K' + (q+1)G$). We assume throughout that all G equations are identified by these exclusion restrictions, the stochastic process $S_{t\cdot}$ is covariance stationary with $S_{t\cdot}$ independent of $V_{\tau\cdot}$ for $t \leq \tau$, the autoregressive process in $Y_{t\cdot}$ is covariance stationary, and that (A3) holds. Since (A1) is satisfied by VAR errors and S (and lags thereof) can be used for (A4), the central issue is deriving IVs that satisfy (A2).

4.1. Assumption (A2)

Usually the IVs in a simultaneous equation model take the form $H \otimes I_G$ for some $(T \times \hat{K})$ matrix H, where $\bar{K} \equiv \hat{K}G$. For example, $H = [S\ Y_{-1}\ \cdots\ Y_{-q}]$ yields the usual 3SLS IVs. This along with (2) allows us to state (A2) more simply for the

[2]This statement may appear at odds with the finding by Turkington (1998) that when R is unknown a different IV set must be used to perform 3SLS than when R is known. Turkington encounters this difficulty because his IV set is derived from the structural form *after* it has been transformed to possess white noise errors. What we are calling the 3SLS IVs herein are derived from the original structural form.

present model in terms of H rather than D. To obtain the simplification note first from (2) that there are five distinct nonzero $G \times G$ blocks in $\Omega(\theta)^{-1}$, and that the upper left block appears only once and therefore is of no consequence in (A2). Ignoring this block, we can substitute $D = H \otimes I_G$, $x = X \otimes I_G$, (2) for $\Omega(\theta)^{-1}$, and $Pu = P\text{vec}U' = \text{vec}V'$ in (A2) to rewrite the assumption as

(A2.1′) A Central Limit Theorem (CLT) applies to $(1/\sqrt{T})(H' \otimes \Sigma^{-1} - H'_{-1} \otimes R\Sigma^{-1})\text{vec}V'$.

(A2.2′) $\underset{T\to\infty}{\text{plim}}(1/T)H'H_L$ is a finite nonsingular matrix for lags $L = -1, 0, 1$.

(A2.3′) $\underset{T\to\infty}{\text{plim}}(1/T)H'X_L$ is a finite matrix of full column rank[3] for lags $L = -1, 0, 1$.

(A2.4′) Each element of H can be expressed in the specified form.

(A2.5′) $H'U_L = O_p(T^{1/2})$ for lags $L = -1, 0, 1$.

Assumptions (A2.1′)–(A2.4′) hold for the usual 3SLS IVs in dynamic models satisfying the assumptions of the present section. The problem with including lags of Y in the IV set is that they violate (A2.5′). So the central problem in designing new IVs based on lags of Y is finding a transformation for the lags that destroys the asymptotic correlation between them and U_L, thereby introducing additional orthogonality conditions that can be exploited in IV estimation.

4.2. Transforming Lagged Endogenous Variables

Consider the set Φ of all $T \times T$ matrices that can be written as a finite sum $\sum_i r_i \Delta^{t_i}$, where r_i are real numbers, t_i are nonnegative integers, and $\Delta^0 \equiv I_T$. This set is a commutative ring whose addition and multiplication operations are standard matrix addition and multiplication and whose multiplicative identity is I_T. Stacking the error VAR (by equation) yields

$$[(I_G \otimes I_T) - (R' \otimes \Delta)]\text{vec}U = (I_G \otimes \Delta\Delta')\text{vec}V + (I_G \otimes (I_T - \Delta\Delta'))\text{vec}U, \tag{3}$$

and $A \equiv (I_G \otimes I_T) - (R' \otimes \Delta)$ is a $G \times G$ matrix over Φ. Although A converts vec$U$ into white noise, (A2.5′) requires that *each column* of U be asymptotically uncorrelated with the IVs. So the problem of constructing IVs from lags of Y can be stated as a problem of finding a transformation Π over Φ such that ΠU is asymptotically uncorrelated with Y_{-L}. This is accomplished by solving (3) over Φ for $U_{\cdot i}$, which involves the determinant of A over Φ and therefore includes Δ^{ℓ}

[3]When the model is a simultaneous equation model (A2.3′) must be restated to reflect the identifying exclusion restrictions. Using a regressor matrix obtained from the generic form $X \otimes I_G$ by dropping the known zeros in β and the corresponding columns of $X \otimes I_G$. Then the product in (A2.3′) is of full column rank since we have assumed the model is identified by exclusion restrictions.

for $\ell = 0, \ldots, G$. To obtain zero correlation, we must ensure that the time index on the lags of Y are prior to the time index on V_t. Thus for IVs $\Pi' Y_{-L}$ we require $L > G$.

To see this formally let $\mathcal{I}_G = I_G \otimes I_T$ denote the $G \times G$ identity matrix over Φ and, in order to avoid confusing matrix multiplication between matrices over Φ with ordinary matrix multiplication, use $\odot$ to denote matrix multiplication between matrices over Φ and $\cdot$ to denote multiplication between a scalar in Φ (i.e., a $T \times T$ matrix) and a matrix over Φ. We continue to denote ordinary matrix multiplication by the absence of an operator. Since Φ is commutative, the ordinary rules for matrix inversion apply to matrices over Φ (see Bourbaki (1989) II §8.7). In particular, $(\det A) \cdot \mathcal{I}_G = A^* \odot A$, where $(\det A)$ and A^* are the determinant and adjoint of A over Φ, respectively. It is straightforward to verify that $B \odot C = BC$ for any conformable matrices B and C over Φ, and also that $(\det A) \cdot \mathcal{I}_G = I_G \otimes (\det A)$, so $I_G \otimes (\det A) = A^* A$. Ordinary matrix premultiplication of (3) by A^* and then substitution of the last equality yields $(I_G \otimes (\det A))\mathrm{vec}U = A^*[(I_G \otimes \Delta\Delta')\mathrm{vec}V + (I_G \otimes (I_T - \Delta\Delta'))\mathrm{vec}U]$. In terms of the individual columns of U that appear in (A2.5′), this is

$$\begin{aligned}(\det A)U_{\cdot i} &= [K_{1i} \ \cdots K_{Gi}][(I_G \otimes \Delta\Delta')\mathrm{vec}V + (I_G \otimes (I_T - \Delta\Delta'))\mathrm{vec}U] \\ &= \sum_{j=1}^{G} K_{ji}[\Delta\Delta' V_{\cdot j} + (I_T - \Delta\Delta')U_{\cdot j}] \text{ for } i = 1, \ldots, G, \end{aligned} \tag{4}$$

where K_{ji} is the (j, i) cofactor of A over Φ. Equation (4) expresses $U_{\cdot i}$ in terms of only the white noise V (except for the presence of $(I_T - \Delta\Delta')U_{\cdot j}$, which only involves the first observation $U_{1\cdot}$). Since the coefficient $(\det A)$ transforms $U_{\cdot i}$ into white noise, an estimate of this matrix using $\tilde{R}$ provides a transformation to be used on lags of Y that results in no asymptotic correlation between the IVs and $U_{\cdot i}$. This is exactly what is required by (A2.5′).

Thus the proposal is to include $\tilde{\Pi}' Y_{-L}$ in H for lags $L > G$, where $\tilde{\Pi}$ is $(\det A)$ with $\tilde{R}$ replacing R. These IVs can be calculated by noting that each $T \times T$ element of $(I_G \otimes I_T) - (\tilde{R}' \otimes \Delta)$ in Φ is either of the form $I_T - \tilde{R}_{\ell\ell}\Delta$ or of the form $-\tilde{R}_{\ell k}\Delta$, so from the definition of the determinant we may express the transformation matrix as $\tilde{\Pi} = \det((I_G \otimes I_T) - (\tilde{R}' \otimes \Delta)) = \sum_{\ell=0}^{G} (-1)^\ell \tilde{r}_\ell \Delta^\ell$, where $\tilde{r}_\ell$ is the sum of all ℓth order principal minors of $\tilde{R}'$ (and $\tilde{r}_0 \equiv 1$ for notational convenience). Hence the new IVs are

$$\tilde{\Pi}' Y_{-L} = \sum_{\ell=0}^{G} (-1)^\ell \tilde{r}_\ell (\Delta')^\ell Y_{-L} = \sum_{\ell=0}^{G} (-1)^\ell \tilde{r}_\ell Y_{\ell - L}. \tag{5}$$

These IVs satisfy (A2′) for $L > G$.

Theorem

Assume (A3) and (A4). Then $H = \tilde{\Pi}' Y_{-L}$ satisfies (A2′) for L > G.

Proof

In the Appendix.

The theorem shows that it is possible to incur no asymptotic cost to not knowing the VAR error parameters while still using some information from the lagged endogenous variables, thereby expanding the class of IVs used in estimation and potentially improving both asymptotic and small-sample efficiency. Note that the FGLS IV estimators considered in this section are possibly asymptotically efficient relative to the estimators in Dhrymes and Taylor (1976) or Hatanaka (1976), since the efficiency properties of their estimators rely on more stringent distributional assumptions than those made herein.

Note finally that, while $P'y_{-L}$ or $P'(Y_{-L} \otimes I_G)$ have intuitive appeal as IVs since Pu_h is white noise, these IVs do not in fact satisfy (A2.5′). For (A2.5′), D must embody a transformation of lagged endogenous variables that results in zero asymptotic correlation with *each individual column* of U_h. The P transformation does not meet this requirement since P only transforms the *entire vector* $u_h = \text{vec}U_h'$ into white noise.

5. MONTE CARLO COMPARISONS

This section presents a set of Monte Carlo experiments that compare optimal FGLS IV estimators based on various sets of IVs with each other and with Hatanaka's (1976) estimators in a dynamic simultaneous equation model with VAR(1) errors. The DGP is Ericsson's (1991) two-equation model with the error structure modified to be VAR(1), as in Hendry and Harrison (1974). Hence

$$
\begin{aligned}
Y_{t1} &= \Theta_{11}S_{t1} + Y_{t2}B_{0_{21}} + Y_{t-1,1}B_{1_{11}} + U_{t1} \\
Y_{t2} &= \Theta_{22}S_{t2} + \Theta_{32}S_{t3} + \Theta_{42}S_{t4} + Y_{t1}B_{0_{12}} + U_{t2} \\
U_{t\cdot} &= U_{t-1\cdot}R + V_{t\cdot},
\end{aligned}
$$

where $V_{t\cdot}$ is generated as the bivariate mixed normal $V_{t\cdot}' = \lambda_t a_t + (1 - \lambda_t)b_t$.[4] Here, λ_t is binomial with $E(\lambda_t) = \frac{1}{2}$; $b_t \sim N(0, \Xi)$; $a_t \sim N(0, 31\Xi)$; λ_t, a_t, and b_t independent; and $\Xi = \begin{bmatrix} \frac{\Sigma_{11}}{16} & \Xi_{12} \\ \Xi_{12} & \frac{1}{16} \end{bmatrix}$.

Following Ericsson and Hendry and Harrison $S_{t\cdot}$ is generated by $s_{t\cdot} = s_{t-1\cdot}\Lambda' + E_{t\cdot}$, where $E_{t\cdot}' \sim NID(0, \Psi)$, and the following parameter values are fixed for all experiments: $\Lambda = \text{diag}\{0.8,\ 0.7,\ 0.4,\ 0.2\}$, $\Theta_{22} = \Theta_{32} = \Theta_{42} = \Theta_{11} = 1$, $B_{0_{12}} = 0.3$, and

$$
\Psi = \begin{bmatrix} 0.25 & 0.237 & 0 & 0 \\ 0.237 & 0.25 & 0 & 0 \\ 0 & 0 & 0.49 & 0 \\ 0 & 0 & 0 & 0.49 \end{bmatrix}.
$$

[4]Normality of $V_{t\cdot}$ is avoided so that the data generating process is not favorable by design to Hatanaka's MLE estimators.

We also use Ericsson's values for $B_{0_{21}} \in \{-0.5, 0.3\}$, $B_{1_{11}} \in \{-0.4, 0.2, 0.7\}$, and $T \in \{20, 40, 80\}$. The values chosen for the error VAR parameters are $R \in \left\{ \begin{bmatrix} 0.6 & 0.6 \\ 0.2 & 0.2 \end{bmatrix}, \begin{bmatrix} 0.6 & 0.2 \\ 0.2 & 0.6 \end{bmatrix}, \begin{bmatrix} 0.2 & 0.6 \\ 0.6 & 0.2 \end{bmatrix} \right\}$, taken from Guilkey and Schmidt (1973), and $\Sigma_{11} \in \{0.25, 1, 4\}$, taken from Hendry and Harrison (1974). For each value of Σ_{11}, Ξ_{12} is set so that the correlation between V_{t1} and V_{t2} is 0.5, as in Ericsson, and $V_{t\cdot}$ is designed so that $\Sigma_{22} = 1$ for all experiments, also as in Ericsson.[5]

Thus, there are 54 experiments for each sample size T, and for each experiment we use Ericsson's specification of $N = 80{,}000/T$ replications. Note that $E_{t\cdot}$ is independent of λ_t, a_t, and b_t; all of the "primitives" of the model have moments of all orders; and $S_{t\cdot}$, $Y_{t\cdot}$, and $U_{t\cdot}$ are stationary VAR processes for all parameter sets; so (A3) is clearly satisfied.

5.1. Pseudorandom Number Generation

Generating multivariate normal pseudorandom vectors with a general covariance matrix involves several algorithms. All univariate normal random number generators utilize a uniform generator and a transformation of the simulated uniform random numbers to produce simulated standard normal random numbers. Multivariate normal generators add an additional step by transforming the simulated univariate normal random numbers into simulated multivariate normal random vectors that possess the specified covariance matrix. Hence, three algorithms are involved, and problems can arise at any step in the process.

Noticeable improvements in methodology for all three of these steps have appeared in the statistical computing literature. Unfortunately, these advances have not all made their way into the econometric literature. We incorporated these advances by using Fushimi's (1990) generalized feedback shift register recurrence formula to produce uniform pseudorandom numbers that perform well in a series of tests for randomness, including a test suggested by Marsaglia (1985). Fushimi's algorithm was implemented with a seed value of 1589 by the IMSL V2.0 routine DRNUN. This seed value was used successfully by Fushimi in testing the algorithm.

Univariate standard normals were obtained using the algorithm proposed by Kinderman and Ramage (1976) as implemented by the IMSL V2.0 routine DRNNOA. This algorithm has been shown to perform better than the traditional one based on Box and Muller (1958), at least when used with traditional mixed congruential uniform generators.

The method we used to transform the univariate standard normal pseudorandom numbers into multivariate normal pseudorandom vectors with a particular covariance structure was the triangular factorization method (Cholesky's) as implemented by the IMSL V2.0 routine DRNMVN. To obtain the Cholesky

[5]Interpretations of these parameter values are given by Ericsson and Guilkey and Schmidt.

factorization of the covariances matrices, the IMSL V2.0 routine DCHFAC is used. We suggest that future Monte Carlo experimenters consider using some of these improvements in methodology where appropriate.

Following Ericsson, separate $E_{t\cdot}$ vectors were generated for each observation of each replication in each experiment (i.e., the $S_{t\cdot}$ vectors are truly exogenous). However, to reduce interexperiment variability (Hendry, 1984; Davidson and Mackinnon, 1993, pp. 738–743) only enough a_t, b_t, and λ_t were generated for one experiment and then these were reused across experiments (for variations in Σ_{11}, the same underlying uniform deviates were reused to produce a_t and b_t with the specified covariances). The λ_t were produced from pseudo-random uniform numbers by setting $\lambda_t = 1$ if the uniform variate is less than $\frac{1}{2}$ and $\lambda_t = 0$ otherwise.

5.2. Data Generation and Estimators

With the simulated $V_{t\cdot}$ and $E_{t\cdot}$ vectors in place, GAUSS v3.2.13 was used to generate observations and calculate estimators. Following Ericsson, the above VAR's in $U_{t\cdot}$ and $S_{t\cdot}$, and the structural model, were used to generate $T + 30$ observations from initial values $U_{0\cdot}$, $S_{0\cdot}$, and $Y_{0\cdot}$ set at the unconditional means of zero. Then the first 30 observations were discarded in an effort to obtain stationarity. This process uses four times as many initial values of $V_{t\cdot}$ when $T = 20$ than when $T = 80$, since N varies from 4,000 to 1,000. For the $T = 40$ and $T = 80$ experiments we skipped through the databank to maintain alignment of the $V_{t\cdot}$ values, again reducing interexperiment variability.

For estimation purposes it is more convenient to stack the system by equation rather than by observation, and we do so henceforth. The first step in estimation is to specify instruments C for generating $\tilde{R}$ and $\tilde{\Sigma}$. We follow Wallis (1967 and 1972) in using S and lags thereof as IVs in the first step. These IVs satisfy (A4) as long as S is truly exogenous because (A2.3′) implies (A4.1), (A2.5′) implies (A4.2), and (A2′) is standard for exogenous S and lags thereof.[6] This leads to optimal first-step instruments $C = M(M'M)^{-1}M'(I_G \otimes X)$ that ignore the nonspherical errors, from the IVs $M = I_G \otimes [S\ S_{-1}\ \ldots\ S_{-L}]$, where L is large enough to identify the model. Since S alone identifies our Monte Carlo DGP, the first nine estimators we consider use $M = I_2 \otimes S$ as first-step IVs to obtain instruments C for the structural matrix (including exclusion restrictions) $\begin{bmatrix} Y_{\cdot 2} & Y_{\cdot -1,1} \\ & Y_{\cdot 1} \end{bmatrix}$. These IVs satisfy (A4) because $S_{t\cdot}$ is independent of $U_{t\cdot}$ in the DGP described above. The residuals from this IV estimation were used to

[6] As with (A2.3′), (A4.1) must be restated to reflect the exclusion restrictions in a simultaneous equation model.

calculate $\tilde{R}, \tilde{P}, \tilde{\Sigma}$, and $\tilde{\Pi}$. Since adding more IVs can only improve asymptotic efficiency,[7] the second nine estimators we consider are the same as the first nine estimators in the second step but use $M = I_2 \otimes [S\ S_{-1}]$ in the first step. This permits us to study the small-sample effects of estimating the error VAR parameters from an asymptotically more precise estimate of B.

The nineteen estimators calculated for each replication are summarized in the table below. In the second step, estimators 1–6 and 10–15 are FGLS IV estimators with all exclusion restrictions imposed. Their IVs are listed in the table and all satisfy (A2′) by the Theorem because, as discussed above, the underlying $S_{t\cdot}$ and $V_{t\cdot}$ processes satisfy (A3) and the first-step IVs satisfy (A4). Estimators 7–9 and 16–18 are the three estimators proposed by Hatanaka. Finally, to get a sense of the small-sample loss associated with estimating R and Σ we calculated the GLS IV estimator that uses the true values of R and Σ, and the same IVs as estimators 6 and 15 (except that the true value of Π is used) since this is asymptotically the most efficient set of IVs we consider.

Estimator	Step 1 IVs (M)	Step 2 IVs (H)
1	$I_2 \otimes S$	S
2	$I_2 \otimes S$	$\left[S\ \tilde{\Pi}' Y_{\cdot,-3,1}\right]$
3	$I_2 \otimes S$	$\left[S\ S_{\cdot,-1,1}\right]$
4	$I_2 \otimes S$	$\left[S\ S_{\cdot,-1,1}\ \tilde{\Pi}' Y_{\cdot,-3,1}\right]$
5	$I_2 \otimes S$	$\left[S\ S_{-1}\right]$
6	$I_2 \otimes S$	$\left[S\ S_{-1}\ \tilde{\Pi}' Y_{\cdot,-3,1}\right]$
7	$I_2 \otimes S$	Hatanaka method 1
8	$I_2 \otimes S$	Hatanaka method 2
9	$I_2 \otimes S$	Hatanaka method 3
10	$I_2 \otimes \left[S\ S_{-1}\right]$	S
11	$I_2 \otimes \left[S\ S_{-1}\right]$	$\left[S\ \tilde{\Pi}' Y_{\cdot,-3,1}\right]$
12	$I_2 \otimes \left[S\ S_{-1}\right]$	$\left[S\ S_{\cdot,-1,1}\right]$
13	$I_2 \otimes \left[S\ S_{-1}\right]$	$\left[S\ S_{\cdot,-1,1}\ \tilde{\Pi}' Y_{\cdot,-3,1}\right]$
14	$I_2 \otimes \left[S\ S_{-1}\right]$	$\left[S\ S_{-1}\right]$
15	$I_2 \otimes \left[S\ S_{-1}\right]$	$\left[S\ S_{-1}\ \tilde{\Pi}' Y_{\cdot,-3.1}\right]$
16	$I_2 \otimes \left[S\ S_{-1}\right]$	Hatanaka method 1
17	$I_2 \otimes \left[S\ S_{-1}\right]$	Hatanaka method 2
18	$I_2 \otimes \left[S\ S_{-1}\right]$	Hatanaka method 3
19	True R and Σ	$\left[S\ S_{-1}\ \Pi' Y_{\cdot,-3,1}\right]$

[7]Asymptotic efficiency may be enhanced by including more IVs than are needed for identification, although the small-sample consequences of this practice are not well-understood (Bowden and Turkington, 1984, p. 38; Davidson and MacKinnon, 1993, p. 222; West and Wilcox, 1996).

5.3. Results

Since these are IV estimators with non-normal errors, small sample moments may not exist. Hence we compare them only by examining their asymptotic covariances and their empirical distribution functions.

Turning first to the asymptotic covariances, larger sets of IVs can be expected to yield a true variance reduction in our DGP since all lags are theoretically relevant. To date, no optimal set of IVs has been derived for a model with both lagged endogenous variables and VAR errors, and as noted in Section 4 under these conditions the 3SLS IVs do not even yield a feasible estimator that is asymptotically equivalent to its infeasible counterpart. Hence for the model considered here we must directly calculate the asymptotic covariances and then compare them.

There are 54 different parameter sets considered in our experiments and six different sets of second-step IVs. The asymptotic covariance for IVs D is

$$\left[E\left[\frac{1}{T}F'\Omega^{-1}D\right]\left[E\left[\frac{1}{T}D'\Omega^{-1}D\right]\right]^{-1}E\left[\frac{1}{T}D'\Omega^{-1}F\right]\right]^{-1}$$

for any T, where

$$F=\begin{bmatrix} S_{.1} & Y_{.2} & Y_{.-1,1} & & \\ & & & S_{.2} & S_{.3} & S_{.4} & Y_{.1} \end{bmatrix}$$

is the regressor matrix $I_G \otimes X$ with columns dropped to reflect exclusion restrictions. The asymptotic covariance is calculated from the autocovariance function of the $[Y\ S]$ process viewed as one 6-dimensional VAR(1), and differs for each parameter and IV set, so there are far too many comparisons to present in detail. We calculated and studied all 54×6 versions, but indicate here only some broad trends.

There are large gains in asymptotic efficiency when the IV set is extended beyond S, irrespective of whether the additional IVs are $S_{.-1,1}, S_{-1}$, or $\tilde{\Pi}'Y_{.-3,1}$. These gains can reduce the variances by as much as several thousand percent for some elements of B and some parameter sets. For example, when $R=\begin{bmatrix} 0.6 & 0.2 \\ 0.2 & 0.6 \end{bmatrix}$, $\Sigma_{11}=4, B_{0_{21}}=0.3$, and $B_{1_{11}}=-0.4$ the asymptotic variance of $\hat{B}_{1_{11}}$ improves from 165.4 to 7.5 when the IV set expands from S to $\left[S\ S_{-1}\right]$. The gains are generally largest on the coefficients of the endogenous and lagged endogenous variables, and when the true value of $B_{1_{11}}$ is -0.4, but they vary widely across parameters. Neither $S_{.-1,1}$ nor S_{-1} dominates $\tilde{\Pi}'Y_{.-3,1}$ for all parameter sets and all elements of B, nor vice-versa. Adding lagged values of S, even just $S_{.-1,1}$, is better than adding $\tilde{\Pi}'Y_{.-3,1}$ alone for many parameter sets but there are parameter sets for which adding $\tilde{\Pi}'Y_{.-3,1}$ yields smaller variances for all elements of B than adding S_{-1}. Moreover, when $\left[S\ S_{-1}\right]$ is extended further to include $\tilde{\Pi}'Y_{.-3,1}$ there sometimes remain large additional efficiency gains, with variances decreasing by several hundred percent in a few cases but usually

improving by less than 5 percent. For the experiment mentioned above, adding $\tilde{\Pi}' Y_{\cdot -3,1}$ to the IV set further reduces the asymptotic variance of $\hat{B}_{1_{11}}$ from 7.5 to 1.8. This verifies that the optimality of the $[S\ S_{-1}]$ IV set shown by Turkington (1998) does not carry-over to dynamic structural models, and demonstrates that our new moment conditions indeed contribute to asymptotic efficiency. Again, the gains are usually largest on the coefficients of the endogenous and lagged endogenous variables and when $B_{1_{11}} = -0.4$. We also studied adding further lags of S, up to S_{-3}, and found further but less dramatic efficiency gains. Finally, we verified that the usual 3SLS IVs are definitely not asymptotically optimal for this model (our expression for the asymptotic covariance in this case is only valid when R and Σ are known), but do yield smaller variances than $[S\ S_{-1}\ \tilde{\Pi}' Y_{\cdot -3,1}]$ for some parameter sets.

Turning now to the Monte Carlo results, due to the size of the study we focus attention on the performance of the estimators for the endogenous and lagged endogenous coefficients $B_{0_{21}}$ and $B_{1_{11}}$. We begin by identifying some general tendencies. First, confirming the asymptotic properties of the estimators under study, as T increases from 20 to 40 and to 80 we observe less disperse and better centered estimates for all parametric specifications. The effects of increased sample size are strong enough to make the estimators increasingly similar in performance and thus the distinctions that can be made are much more pronounced when $T = 20$. Second, the larger first-step IV set, $I_2 \otimes [S\ S_{-1}]$, improves both dispersion and centrality of all estimators across all parametric specifications, but the improvement is modest and is most noticeable for small sample sizes. This improved performance is less acute and at times barely perceptible for the estimation methods proposed by Hatanaka. Third, increases in Σ_{11} produce much more disperse and worse centered estimates for all parametric specifications and all estimators, and the noncentrality becomes severe for $\Sigma_{11} = 4$. Fourth, the specification of R does not materially affect estimator performance.

Among the estimators proposed by Hatanaka, methods 1 and 3 are superior to method 2 overall in these experiments. Method 2 shows a marked tendency to produce very disperse estimates. In estimating $B_{0_{21}}$, method 1 overestimates slightly more often than method 3 when the true value is $B_{0_{21}} = -0.5$. When $B_{0_{21}} = 0.3$ method 3 is much better centered than method 1 while their dispersions are comparable. Hence, these experiments suggest that Hatanaka's method 3 is preferable to his other two methods.

Among the FGLS IV estimators 1–6, estimator 6 displays the tightest dispersion in estimating both $B_{0_{21}}$ and $B_{1_{11}}$, suggesting that increasing the number of IVs, including the new IVs derived here, brings benefits even in small samples. Interestingly, estimators 2 and 3, which use the same number of IVs and differ only in the use of a transformed lagged endogenous IV rather than a lagged exogenous IV, have almost identical dispersions across experiments. This is mildly different from the asymptotics, which showed modestly smaller variances for estimator 2 for most, but not all, parameter sets. The marginal gains from adding more IVs decrease with the number of IVs used. Thus, estimators 2 and 3

are significantly less disperse than estimator 1 but estimator 6 is only slightly less disperse than estimator 5.

The effect on central tendency of using more IVs is much less definitive. Since all the FGLS IV estimators are asymptotically $\sqrt{T}$-normal centered at the true value, this is consistent with theory. All feasible estimators tend to overestimate both $B_{0_{21}}$ and $B_{1_{11}}$, and there is neither clear improvement nor deterioration within any parameter set as the number of IVs increases.

For a fixed value of $B_{0_{21}}$, dispersion and central tendency of the FGLS IV estimators of $B_{0_{21}}$ are not sensitive to the value of $B_{1_{11}}$, but dispersion is usually slightly smaller when $B_{0_{21}} = 0.3$. In contrast, recall that the asymptotic variances are improved more substantially by adding IVs when $B_{1_{11}} = -0.4$, so asymptotically the dispersion of at least some of the estimators is indeed sensitive to the value of $B_{1_{11}}$.

Interestingly, there is considerable improvement in both the dispersion and central tendency of all FGLS IV estimators of $B_{1_{11}}$ when $B_{0_{21}}$ increases to 0.3. However, their relative performance is unaffected.

Given these general observations, it seems most useful to focus on Hatanaka's method 3 (estimator 18), the FGLS IV estimator using a large IV set (estimator 15), and the infeasible estimator 19 for a closer comparison. Figures 1 and 2 show cumulative frequencies for these three estimators of $B_{0_{21}}$ and $B_{1_{11}}$, respectively, for the following six representative experiments:

$$\text{All Experiments: } T = 20,\ R = \begin{bmatrix} 0.6 & 0.6 \\ 0.2 & 0.2 \end{bmatrix},\ \Sigma_{11} = 0.25,$$

Experiment	$B_{0_{21}}$	$B_{1_{11}}$
1	−0.5	−0.4
2	−0.5	0.2
3	−0.5	0.7
4	0.3	−0.4
5	0.3	0.2
6	0.3	0.7.

Figure 1 shows that all three estimators usually overestimate $B_{0_{21}}$, but this tendency is mild for the infeasible estimator 19, followed by Hatanaka's estimator 18 and then the FGLS IV estimator 15. Likewise, the infeasible estimator is most tightly distributed, followed by Hatanaka's estimator and then the FGLS IV estimator. Hence ignorance of R and Σ is indeed costly is small samples, and Hatanaka's method 3 performed best among the feasible estimators of $B_{0_{21}}$ in these experiments.

Figure 2 shows that these conclusions do not apply to estimation of $B_{1_{11}}$. The infeasible estimator tends to underestimate when $B_{1_{11}} = -0.5$, and this tendency becomes more pronounced as $B_{0_{21}}$ increases, so that the feasible estimators are actually better centered than the infeasible estimator in experiments 2 and 3. The infeasible estimator continues to display smaller dispersion in most experiments, but surprisingly does not dominate as it does in estimating $B_{0_{21}}$ because in

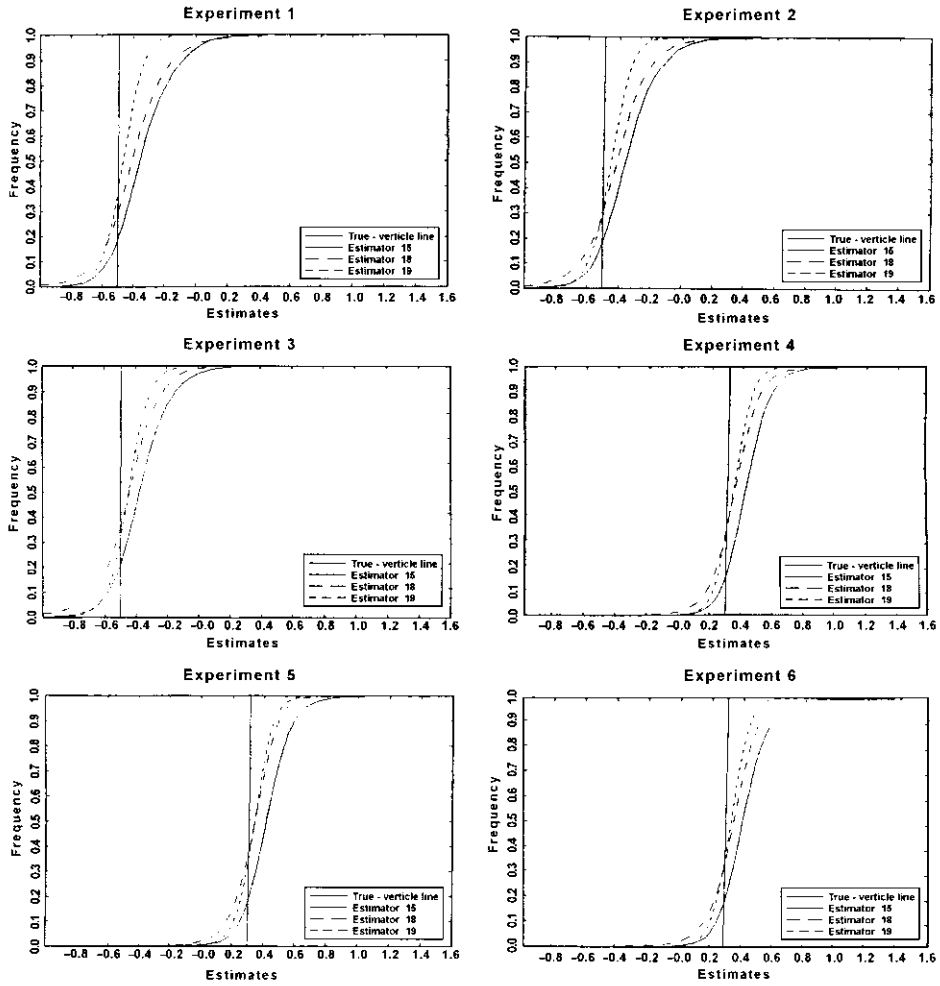

Figure 1. Cumulative Frequency for $B_{0_{21}}$.

experiment 6 of Figure 2 both feasible estimators are more tightly distributed than the infeasible estimator. Also unlike Figure 1, Hatanaka's estimator does not emerge as the clear choice between the feasible estimators, as the FGLS IV estimator has less dispersion and comparable or better centering in experiments 2, 3, 5, and 6.

The qualitative and comparison conclusions derived from the analysis of Figures 1 and 2 are essentially unaltered for $T = 40$ and $T = 80$. However, two observations are in order. First, as T grows the estimators become better centered and less disperse to the point that drawing distinctions at $T = 80$ becomes rather difficult for some parameters. Second, the positive effects of increased sample size are much more pronounced on the feasible estimators than on the infeasible

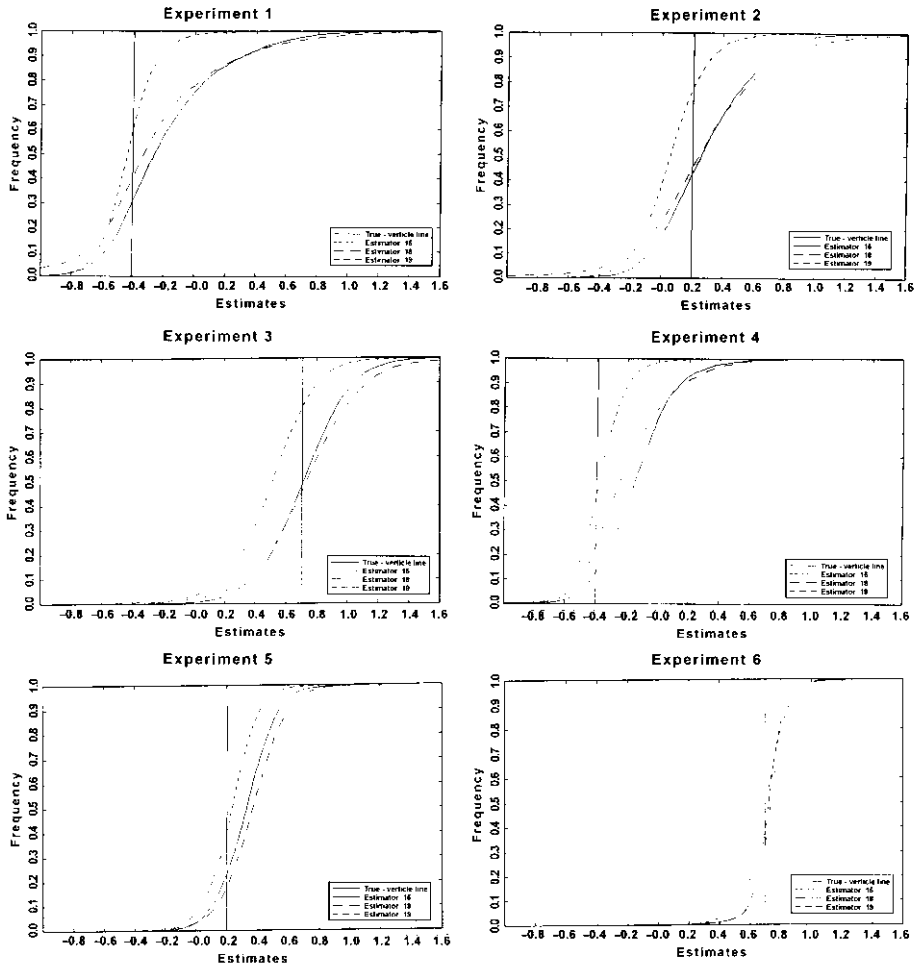

Figure 2. Cumulative Frequency for $B_{1_{11}}$.

estimator 19. A result to be expected, as the increased sample size benefits the estimation of the nuisance, as well as the parameters of interest.[8]

In summary, for the sample sizes and parameter configurations considered in our Monte Carlo study, neither the best FGLS IV nor the best Hatanaka method emerges as a clear best choice. In fact, these estimators perform rather similarly and become more similar as the sample size increases. The experiments show that introducing additional IVs when constructing FGLS IV estimators, such as those proposed in Section 4 above, is beneficial both asymptotically and in small samples. More IVs in the first step is beneficial as well. The benefits of additional

[8]Figures comparable to 1 and 2 for $T = 40$ and $T = 80$ are available from the authors on request.

IVs diminish, however, as the number of IVs expands. Finally, it is somewhat disturbing that there is a marked tendency of all estimators to overestimate $B_{0_{21}}$. Centering is better in the estimation of $B_{1_{11}}$, but there is still a tendency for the feasible estimators to overestimate.

APPENDIX

Proof of Theorem

(A2.1′) Let $\Pi = (\det A)$ (i.e., with the true R rather than $\tilde{R}$). The counterpart to (5) based on Π is

$$\Pi' Y_{-L} = \sum_{\ell=0}^{G} (-1)^{\ell} r_{\ell} Y_{\ell-L}. \tag{P1}$$

Suppose momentarily that H is given by (P1) rather than the feasible version of (P1) based on $\tilde{\Pi}$. Note that

$$(H' \otimes \Sigma^{-1} - H'_{-1} \otimes R\Sigma^{-1})\mathrm{vec}V' = \begin{bmatrix} \sum_{t=1}^{T} (H_{t1}\Sigma^{-1} - H_{t-1,1}R\Sigma^{-1})(V_{t.})' \\ \vdots \\ \sum_{t=1}^{T} (H_{t\hat{K}}\Sigma^{-1} - H_{t-1,\hat{K}}R\Sigma^{-1})(V_{t.})' \end{bmatrix}. \tag{P2}$$

Hence, using the Cramér-Wold Device, we need only apply a CLT to

$$\frac{1}{\sqrt{T}}\lambda'(H' \otimes \Sigma^{-1} - H'_{-1} \otimes R\Sigma^{-1})\mathrm{vec}V' = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \times \left[\sum_{i=1}^{\hat{K}} \lambda_i' (H_{ti}\Sigma^{-1} - H_{t-1,i}R\Sigma^{-1})(V_{t.})' \right], \tag{P3}$$

where $\lambda' = (\lambda_1' \ \ldots \ \lambda_{\hat{K}}')$ is an arbitrary nonstochastic vector composed of $(G \times 1)$ subvectors λ_i. Sufficient conditions for the m-dependent CLT to apply to (P3) for the special case of $m = 0$ are (Schmidt, 1976; p. 258):

(CLT1) $E\left[\sum_{t=1}^{T} \lambda_i'(H_{ti}\Sigma^{-1} - H_{t-1,i}R\Sigma^{-1})(V_{t.})\right] = 0.$

(CLT2) $E\left[\left|\sum_{t=1}^{T} \lambda_i'(H_{ti}\Sigma^{-1} - H_{t-1,i}R\Sigma^{-1})(V_{t.})\right|^3\right]$ is bounded across t.

(CLT3) $\sigma^2 \equiv \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} A_{t+n}$ exists and is independent of t, where

$$A_t = E\left[\left(\sum_{i=1}^{\hat{K}} \lambda_i'(H_{ti}\Sigma^{-1} - H_{t-1,i}R\Sigma^{-1})(V_{t.})'\right)^2\right].$$

(CLT1) is immediate since the time index on row t of each lag of Y in (P1) is less than t (since $L > G$), making the lag of Y independent of $V_{t.}$. (CLT2) follows from (A3) and (A2.4′). (CLT3) holds because

$$A_t = \sum_{i=1}^{\hat{K}} \sum_{j=1}^{\hat{K}} \lambda_i' \left[\Upsilon_{ij}(0)(\Sigma^{-1} + R\Sigma^{-1}R') - \Upsilon_{ij}(1)(R\Sigma^{-1} + \Sigma^{-1}R') \right] \lambda_j$$

for every t, where $\Upsilon_{ij}(h) = E[H_{ti}H_{t-h,j}]$ is independent of t by covariance stationarity of $Y_{t\cdot}$, and we have again made use of independence between $H_{t\cdot}$ or $H_{t-1,\cdot}$ and $V_{t\cdot}$ for the lags that are present in A_t. Thus (A2.1′) holds when the IVs are based on Π rather than $\tilde{\Pi}$.

Now we show that (P3) converges in probability to the analogous expression based on $\tilde{\Pi}$, so that using $\tilde{\Pi}$ does not affect the asymptotic distribution. From (P3), the difference between using $\tilde{\Pi}$ and Π in (A2.1′) is

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{G} \lambda_i' [(\tilde{\Pi}_{\cdot,t} - \Pi_{\cdot,t})' Y_{\cdot-L,i} \Sigma^{-1} - (\tilde{\Pi}_{\cdot,t-1} - \Pi_{\cdot,t-1})' Y_{\cdot-L,i} R\Sigma^{-1}](V_{t\cdot})'. \tag{P4}$$

From (5),

$$\begin{aligned} [\tilde{\Pi}_{\cdot,t-h} - \Pi_{\cdot,t-h}]' Y_{\cdot-L,i} &= \sum_{\ell=1}^{G} (-1)^{\ell} (\tilde{r}_{\ell} - r_{\ell}) ((\Delta^{\ell})_{\cdot,t-h})' Y_{\cdot-L,i} \\ &= \sum_{\ell=1}^{G} (-1)^{\ell} (\tilde{r}_{\ell} - r_{\ell}) Y_{t-L-h+\ell,i}. \end{aligned} \tag{P5}$$

Since $t - L - h + \ell < t$ when $L > G$, $\ell \leq G$, and $h \geq 0$, by (A3) we can apply Chebyshev's Inequality to show

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Y_{t-L+\ell,i} (\lambda_i' \Sigma^{-1} (V_{t\cdot})') = O_p(1),$$

$$\text{and } \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Y_{t-L-1+\ell,i} (\lambda_i' R\Sigma^{-1} (V_{t\cdot})') = O_p(1). \tag{P6}$$

Substituting (P5) and (P6) into (P4) yields

$$\sum_{i=1}^{G} \sum_{\ell=1}^{G} (-1)^{\ell} (\tilde{r}_{\ell} - r_{\ell}) [O_p(1) - O_p(1)]. \tag{P4′}$$

Since $\tilde{r}_{\ell} \xrightarrow{p} r_{\ell}$ by Lemma 2, we have (P4) $\xrightarrow{p} 0$.

(A2.2′) From (5), $H_h = \sum_{\ell=0}^{G} (-1)^{\ell} \tilde{r}_{\ell} Y_{\ell-L+h}$. So

$$\frac{1}{T} H'H_h = \frac{1}{T} \sum_{\ell=0}^{G} \sum_{i=0}^{G} (-1)^{\ell+1} \tilde{r}_{\ell} \tilde{r}_i Y_{\ell-L}' Y_{i-L+h}. \tag{P7}$$

Since $\tilde{r}_{\ell} \xrightarrow{p} r_{\ell}$ by Lemma 2 and the fourth moments of the $Y_{t\cdot}$ process exist by (A3.4), (P7) converges to

$$\sum_{\ell=0}^{G} \sum_{i=0}^{G} (-1)^{\ell+1} r_{\ell} r_i E\left(Y_{\ell-L}' Y_{i-L+h}\right) \text{ for } h = -1, 0, 1.$$

(A2.3′) The convergence argument is identical to (A2.2′).

(A2.4′) From (5),

$$H_{th} = \sum_{\ell=0}^{G} (-1)^{\ell} \tilde{\boldsymbol{r}}_{\ell} Y_{t+\ell-L,h}.$$

Since $(-1)^{\ell}\tilde{\boldsymbol{r}}_{\ell}$ is independent of t and is $O_p(1)$ by Lemma 2, and the fourth moments of $Y_{t+\ell-L,h}$ have the specified property by (A3.4), H_{th} takes the form required by (A2.4′).

(A2.5′) From (5),

$$H'U_h = \sum_{\ell=1}^{G} (-1)^{\ell} (\tilde{\boldsymbol{r}}_{\ell} - r_{\ell}) Y'_{\ell-L} U_h + Y'_{-L} \Pi U_h. \tag{P8}$$

By Lemma 2, the sum is $O_p(1)$ if $Y'_{\ell-L}U_h = O_p(T)$ (note the importance of $\tilde{R}$ converging faster than $o_p(1)$). A routine application of Chebyshev's Inequality establishes this because the moments involved are bounded by (A3.4). From (4),

$$K_{ji} = (-1)^{j+i} \sum_{\ell=0}^{G-1} (-1)^{\ell} r_{\ell ji} \Delta^{\ell}, \tag{P9}$$

where $r_{\ell ji}$ is the sum of all ℓth order principal minors of the submatrix of R' obtained by deleting row j and column i (and $r_{0ji} \equiv 1$ for notational convenience). Substituting (P9) into (4) yields

$$Y'_{-L}\Pi U_{\cdot+h,i} = Y'_{-L}(\det A) U_{\cdot+h,i} = \sum_{j=1}^{G} (-1)^{j+1} \sum_{\ell=0}^{G-1} (-1)^{\ell} r_{\ell ji} Y'_{-L} \Delta^{\ell} [\Delta\Delta' V_{\cdot+h,j} + (I_T - \Delta\Delta') U_{\cdot+h,j}]. \tag{P10}$$

Since the only nonzero element of $(I_T - \Delta\Delta')U_{\cdot+h,j}$ is $U_{1+h,j}$, this term is clearly $O_p(T^{1/2})$. For the remaining term, note that

$$\frac{1}{\sqrt{T}} Y'_{-L} \Delta^{\ell} \Delta\Delta' V_{\cdot+h,j} = \frac{1}{\sqrt{T}} Y'_{\ell-L} V_{\cdot+h,j}$$

except for the first row of V. As shown above, this satisfies the conditions for the m-dependent CLT (and thus is $O_p(1)$) since the index on row t of each lag of Y is less than $t + h$ when $\ell \leq G - 1,\ L > G$, and $h \leq 1$. □

ACKNOWLEDGMENT

We thank Aman Ullah, an anonymous referee and Lee Adkins for helpful comments. The authors retain responsibility for any remaining errors. Research support for C. Martins-Filho from CNPq, PRONEX and the Federal University of Ceará, Brazil is greatfully acknowledged.

REFERENCES

Anderson, T.W. *The Statistical Analysis of Time Series*; Wiley: New York, 1971.

Bourbaki, N. *Elements of Mathematics, Algebra I, Chapters 1–3*; Vol. 1, Springer-Verlag: Berlin, 1989.

Bowden, R.J.; Turkington, D.A. *Instrumental Variables*; Cambridge University Press: Cambridge, 1984.

Box, G.E.P.; Muller, M.E. A Note on Generation of Normal Deviates. *Annals of Mathematical Statistics* **1958**, *28*, 610–611.

Davidson, R.; Mackinnon, J.G. *Estimation and Inference in Econometrics*; Oxford University Press: Oxford, 1993.

Dhrymes, P.J.; Taylor, J.B. On an Efficient Two-Step Estimator for Dynamic Simultaneous Equations Models with Autoregressive Errors. *International Economic Review* **1976**, *17*, 362–376.

Ericsson, N.R. Monte Carlo Methodology and the Finite Sample Properties of Instrumental Variables Statistics for Testing Nested and Non-Nested Hypotheses. *Econometrica* **1991**, *59*, 1249–1277.

Fuller, W.A. *Introduction to Statistical Time Series*; Wiley: New York, 1976.

Fushimi, M. Random Number Generation with the Recursion $X_t = X_{t-3p} \otimes X_{t-3q}$. *Journal of Computational and Applied Mathematics* **1990**, *31*, 105–118.

Guilkey, D.K.; Schmidt, P. Estimation of Seemingly Unrelated Regressions with Vector Autoregressive Errors. *Journal of the American Statistical Association* **1973**, *68*, 642–647.

Hatanaka, M. Several Efficient Two-Step Estimators for the Dynamic Simultaneous Equations Model with Autoregressive Disturbances. *Journal of Econometrics* **1976**, *4*, 189–204.

Hendry, D.F. Monte Carlo Experimentation in Econometrics. In *Handbook of Econometrics*, *Chapter 16*; Grilliches, Z., Intrilligator, M.D., Ed.; Vol. 2, North-Holland, Amsterdam, 1984.

Hendry, D.F. and R. W. Harrison, *Monte Carlo Methodology and the Small Sample Behavior of Ordinary and Two-Stage Least Squares*, Journal of Econometrics **2** (1974), 151–174.

Kinderman, A.; Ramage, J. Computer Generation of Normal Random Variables. *Journal of the American Statistical Association* **1976**, *71*, 893–896.

Mandy, D.M.; Martins-Filho, C. A Unified Approach to Asymptotic Equivalence of Aitken and Feasible Aitken Instrumental Variables Estimators. *International Economic Review* **1994**, *35*, 957–979.

Mandy, D.M., Martins-Filho, C. A Note on A Unified Approach to Asymptotic Equivalence of Aitken and Feasible Aitken Instrumental Variables Estimators. *International Economic Review* **1997**, *38*, 479.

Marsaglia, G. *A Current View of Random Number Generators*. In *Computer Science and Statistics*, Proceedings of the 16th Symposium on the Interface; Billard, L. Ed.; North-Holland, Amsterdam, 1985.

Schmidt, P. *Econometrics*; Marcel Dekker: New York, 1976.

Turkington, D.A. Efficient Estimation in the Linear Simultaneous Equations Model with Vector Autoregressive Disturbances. *Journal of Econometrics* **1998**, *85*, 51–74.

Wallis, K.F. The Efficiency of the Two-Step Estimator. *Econometrica* **1972**, *40*, 769–770.

Wallis, K.F. Lagged Dependent Variables and Serially Correlated Errors: A Reappraisal of Three-Pass Least Squares. *Review of Economics and Statistics* **1967**, *49*, 555–567.

West, K.; Wilcox, D.W. A Comparison of Alternative Instrumental Variables Estimators of a Dynamic Linear Model. *Journal of the American Statistical Association* **1996**, *14*, 281–293.

White, H. *Asymptotic Theory for Econometrician*; Academic Press: Orlando, Fl, 1984.