A Comparison of Some Recent Bayesian and Classical Procedures for Simultaneous Equation Models with Weak Instruments

Chuanming Gao, Kajal Lahiri*
Department of Economics, State University of New York at Albany, Albany, NY 12222

Abstract: We compare the finite sample performance of a number of Bayesian and classical procedures for limited information simultaneous equations models with weak instruments by a Monte Carlo study. We consider recent Bayesian approaches developed by Chao and Phillips (1998, CP), Geweke (1996), Kleibergen and van Dijk (1998, KVD), and Zellner (1998). Amongst the sampling theory methods, OLS, 2SLS, LIML, Fuller's modified LIML, and the jackknife instrumental variable estimator (JIVE) due to Angrist, Imbens and Krueger (1999) and Blomquist and Dahlberg (1999) are also considered. Since the posterior densities and their conditionals in CP and KVD are non-standard, we propose a "Gibbs within Metropolis-Hastings" algorithm, which only requires the availability of the conditional densities from the candidate generating density. Our results show that in cases with very weak instruments, there is no single estimator that is superior to others in all cases. When endogeneity is weak, Zellner's MELO does the best. When the endogeneity is not weak and $\rho w_{12} > 0$, where ρ is the correlation coefficient between the structural and reduced form errors, and w_{12} is the covariance between the unrestricted reduced form errors, BMOM outperforms all other estimators by a wide margin. When the endogeneity is not weak and $\beta\rho < 0$ (β being the structural parameter), KVD approach seems to work very well. Surprisingly, the performance of JIVE was disappointing in all our experiments.

JEL classification: C30, C11, C13, C15
Keywords: Limited information estimation, Metropolis-Hastings algorithm, Gibbs sampler, Monte Carlo method

* Corresponding author. E-mail: KL758@cnsvax.albany.edu. Tel: (518) 442 4758; fax: (518) 442 4736. An earlier version of this paper was presented at the 1999 Joint Statistical Meetings and the 2000 Winter Meetings of the Econometric Society. We are grateful to Ingolf Dittmann, Terrence Kinal, Yuichi Kitamura, Roberto Mariano, Arnold Zellner, and Eric Zivot for many helpful comments and suggestions. However, the responsibility for any remaining errors and shortcomings is solely ours.

1 Introduction

Recent research on Bayesian analysis of the simultaneous equations models addresses a problem, raised initially by Maddala (1976), and now recognized as related to the problem of local non-identification when diffuse/flat priors are used in traditional Bayesian analysis of such models, e.g., Drèze (1976), Drèze and Morales (1976), and Drèze and Richard (1983).[1] In this paper, we will examine the approaches developed by Chao and Phillips (1998, hereafter CP), Geweke (1996), Kleibergen and van Dijk (1998, hereafter KVD), and Zellner (1998). The idea in KVD is to treat an overidentified simultaneous equations model (SEM) as a linear model with nonlinear parameter restrictions. While KVD focuses mainly on resolving the problem of local nonidentification, CP explores further the consequences of using a Jeffreys prior. By deriving the exact and (asymptotically) approximate representations for the posterior density of the structural parameter, they show that the use of a Jeffreys prior brings Bayesian inference closer to classical inference in the sense that this prior choice leads to posterior distributions which exhibit Cauchy-like tail behavior like the LIML estimator. Geweke (1996), being aware of the potential problem of local nonidentification, suggests a shrinkage prior such that the posterior density is properly defined for each parameter. In another novel approach, Zellner (1998) has developed a finite sample Bayesian method of moments (BMOM) procedure based on given data without specifying a likelihood function or introducing any sampling

[1] Zellner (1998) provides the latest comprehensive review of the finite sample properties of SEM estimators, and emphasizes the need for finite sample optimal estimation procedure for such models.

assumptions.

For the Bayesian approaches considered, while Geweke (1996) proposes Gibbs sampling (GS) to evaluate the posterior density with a shrinkage prior, the posterior densities as well as their conditional densities resulting from CP and KVD are non-standard and cannot be readily simulated. In the category of "block-at-a-time" approach, we suggest a new MCMC procedure, which we call a "Gibbs within M-H" algorithm. The advantage of this algorithm is that it only requires the availability of the conditional densities from the candidate generating density. These conditional densities are used in a Gibbs sampler to simulate the candidate generating density, whose drawings, after convergence, are then weighted to generate drawings from the target density in a Metropolis-Hastings (M-H) algorithm. In this study, we will focus on weak instruments, where the classical approach has been particularly unsatisfactory.[2]

The main objective of the present paper is to compare the small sample performance of some Bayesian and classical approaches using Monte Carlo simulations. For the purpose of comparison, a number of classical methods including OLS, 2SLS, LIML, Fuller's modified LIML, and a recent jackknife instrumental variables estimator (JIVE) due to Angrist, Imbens and Krueger (1999) and Blomquist and Dahlberg (1999) are also computed from the generated data. Our simulation results from repeated sampling experiments provide some unambiguous guidelines for empirical practitioners.

The plan of the paper is as follows. In Section 2, we set up the model.

[2]There has been a growing interest in the estimation of LISEM with weak instruments. See Buse (1992), Bound, Jaeger and Baker (1995), Staiger and Stock (1997), Angrist, Imbens and Krueger (1999), Blomquist and Dahlberg (1999), among others.

Section 3 reviews in limited details the recent Bayesian approaches and JIVE. Section 4 suggests a new MCMC procedure for evaluating the posterior distributions for CP and KVD, and discusses the convergence diagnostics implemented. Section 5 presents simulation results and some discussions. Section 6 contains the main conclusions.

2 The Model

Consider the following limited information formulation of the m-equation simultaneous equations model (LISEM):

$$y_1 = Y_2\beta + Z_1\gamma + u, \tag{1}$$

$$Y_2 = Z_1\Pi_1 + Z_2\Pi_2 + V_2, \tag{2}$$

where $y_1 : (T \times 1)$ and $Y_2 : (T \times (m-1))$ are the m included endogenous variables; $Z_1 : (T \times k_1)$ is an observation matrix of exogenous variables included in the structural equation (1); $Z_2 : (T \times k_2)$ is an observation matrix of exogenous variables excluded from (1); and u and V_2 are, respectively, a $T \times 1$ vector and a $T \times (m-1)$ matrix of random disturbances to the system. We assume that $(u, V_2) \sim N(0, \Sigma \otimes I_T)$, where the $m \times m$ covariance matrix Σ is positive definite symmetric (pds) and is partitioned conformably with the rows of (u, V_2) as follows

$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{21}' \\ \sigma_{21} & \Sigma_{22} \end{pmatrix}$.

The likelihood function for the model described by (1) and (2) can be

written as

$$L(\beta, \gamma, \Pi_1, \Pi_2, \Sigma | Y, Z) = (2\pi)^{-Tm/2} |\Sigma|^{-T/2} \exp\{-\frac{1}{2} tr[\Sigma^{-1}(u, V_2)'(u, V_2)]\}, \tag{3}$$

where $Y = (y_1, Y_2)$ and $Z = (Z_1, Z_2)$.

The structural model described by (1) and (2) can alternatively be written in its reduced form

$$\begin{pmatrix} y_1 & Y_2 \end{pmatrix} = \begin{pmatrix} Z_1 & Z_2 \end{pmatrix} \begin{pmatrix} \pi_1 & \Pi_1 \\ \Pi_2\beta & \Pi_2 \end{pmatrix} + \begin{pmatrix} \xi_1 & V_2 \end{pmatrix}, \tag{4}$$

where $\pi_1 = \gamma + \Pi_1\beta$, $\xi_1 = u + V_2\beta$, $(\xi_1, V_2) \sim N(0, \Omega \otimes I_T)$, $\Sigma = C'\Omega C$, $C = \begin{pmatrix} 1 & 0 \\ -\beta & I_{m-1} \end{pmatrix}$. The likelihood function corresponding to this alternative representation is:

$$L(\beta, \gamma, \Pi_1, \Pi_2, \Omega | Y, Z) = (2\pi)^{-Tm/2} |\Omega|^{-T/2} \exp\{-\frac{1}{2} tr[\Omega^{-1}(\xi_1, V_2)'(\xi_1, V_2)]\}. \tag{5}$$

The likelihood functions (3) and (5) are equivalent since the Jacobian between Ω and Σ is unity.

Geweke (1996) considers the following reduced rank regression specification[3]

$$Y = Z_1 A + Z_2 \Theta + E, \tag{6}$$

where $A = (\Pi_1, \pi_1)$, $\Theta = \Pi_2\underline{\Phi}$, and $\underline{\Phi} = (I_{m-1}, \beta)$, $E = (V_2, \xi_1) \sim N(0, \underline{\Sigma} \otimes I_T)$ with $\underline{\Sigma}^{-1} = \begin{pmatrix} \underline{\Sigma}^{11} & \underline{\Sigma}^{12} \\ \underline{\Sigma}^{21} & \underline{\Sigma}^{22} \end{pmatrix}$ partitioned conformably with the rows of (V_2, ξ_1). Obviously, (6) is equivalent to (4) and the corresponding likelihood function is similar to (5).

Note that in the absence of restrictions on the covariance structure, (1) is fully identified if and only if $\text{rank}(\Pi_2) = (m-1) \leq k_2$.

[3]Geweke (1996) considered a more general specification. To facilitate comparison, for Geweke approach only, we have denoted $Y = (Y_2, y_1)$.

3 Review of some recent formulations

Among the most recent Bayesian approaches, Geweke (1996) used a shrinkage prior such that all parameters are identified (in the sense that a proper posterior distribution exists) even when Π_2 has reduced rank. KVD treated overidentified SEMs as linear models with nonlinear parameter restrictions using the singular value decomposition. A diffuse or natural conjugate prior for the parameters of the embedding linear model results in the posterior for the parameters of the SEM having zero weight in the region of parameter space where Π_2 has reduced rank. This is a feature of the Jacobian of transformation from the multivariate linear model to the SEM. CP used a prior by applying Jeffreys principle on the model described by (1) and (2) and the assumptions regarding the disturbances. An important quality of the Jeffreys prior in the present context is that it places no weight in the region of the parameter space where $\text{rank}(\Pi_2) < (m-1)$ and relatively low weight in close neighborhoods of this region where the model is nearly unidentified.

3.1 Zellner's Bayesian method of moments approach (BMOM)

Among the various Bayesian treatments of SEM proposed by Zellner (1971, 1978, 1986, 1994, 1998), the recent Bayesian method of moments approach applies the principle of maximum entropy and generates optimal estimates which can be evaluated by double K-class estimators. Given the unrestricted reduced form equation $y_1 = Z\overline{\pi}_1 + \xi_1$, Zellner (1998) considered a balanced

loss function,

$$\begin{aligned} L_b &= \omega L_g + (1-\omega)L_p \\ &= \omega(y_1 - X\widehat{\delta})'(y_1 - X\widehat{\delta}) + (1-\omega)(Z\overline{\pi}_1 - X\widehat{\delta})'(Z\overline{\pi}_1 - X\widehat{\delta}), \text{ for } 0 \leq \omega \leq 1 \end{aligned}$$

where $X = (Y_2, Z_1), \delta = (\beta', \gamma')'$, and $\widehat{\delta}$ is an estimate of δ. The BMOM estimate that minimizes EL_b, where the expectation is taken with respect to a probability density function of the π matrices appeared in unrestricted reduced form equations, is given by

$$\begin{pmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} = \begin{bmatrix} Y_2'Y_2 - K_1\widehat{V}_2'\widehat{V}_2 & Y_2'Z_1 \\ Z_1'Y_2 & Z_1'Z_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_2 - K_2\widehat{V}_2)'y_1 \\ Z_1'y_1 \end{bmatrix}, \qquad (7)$$

where

$$K_1 = 1 - k/(T-k), K_2 = 1 - (1-\omega)k/(T-k) \text{ with } 0 \leq \omega \leq 1,$$

and $\widehat{V}_2 = (I - Z(Z'Z)^{-1}Z')Y_2$.

BMOM estimate will vary depending on the value of ω. When $\omega = 1$, it is the optimal estimate resulting from a "goodness of fit" loss function L_g. When $\omega = 0$, it is the optimal estimate given by a precision of estimation loss function L_p. Meanwhile, the well-known minimum expected loss (MELO) estimator is derived using a precision of estimation loss function and may be evaluated as a K-class estimator with

$$K_1 = K_2 = 1 - k/(T-k-m-1).$$

3.2 The Geweke (1996) approach

Geweke (1996) assumes the following reference prior

$$|\underline{\Sigma}|^{-(m+\nu+1)/2} \exp[-\frac{1}{2} tr S\underline{\Sigma}^{-1}] \exp[-\frac{\tau^2}{2}(\beta'\beta + tr\Pi_2'\Pi_2 + tr A'A)], \qquad (8)$$

which is the product of an independent inverted Wishart distribution for Σ with ν degrees of freedom and scale matrix S, and an independent $N(0, \tau^2)$ shrinkage priors for each element of β and Π_2. Geweke derived the respective conditional posterior distributions, which may be used to generate drawings through Gibbs sampling from the joint posterior distribution. Regarding the vector of parameters $(\underline{\Sigma}^{-1}, A, \Pi_2, \beta)$, we obtain the full conditional densities as follows:

(1) Conditional density of $\underline{\Sigma}^{-1}$

$$\underline{\Sigma}^{-1}|(\Pi_2, \beta, A, Z, Y) \sim \text{Wishart}(T + \nu, G^{-1}), \tag{9}$$

where $G = S + (Y - Z_1 A - Z_2\Theta)'(Y - Z_1 A - Z_2\Theta)$.

(2) Conditional density of A

$$\begin{aligned} & \text{vec}(A)|(\Pi_2, \beta, \underline{\Sigma}^{-1}, Z, Y) \\ \sim \; & N([\underline{\Sigma}^{-1} \otimes Z_1'Z_1 + \tau^2 I_{mk_1}]^{-1}[\underline{\Sigma}^{-1} \otimes Z_1'Z_1]\text{vec}(\widehat{A}), \\ & [\underline{\Sigma}^{-1} \otimes Z_1'Z_1 + \tau^2 I_{mk_1}]^{-1}), \end{aligned} \tag{10}$$

where $\widehat{A} = (Z_1'Z_1)^{-1}Z_1'(Y - Z_2\Theta)$.

(3) Conditional density of Π_2[4]

$$\begin{aligned} & \text{vec}(\Pi_2)|(\beta, \underline{\Sigma}^{-1}, A, Z, Y) \\ \sim \; & N([\widetilde{\Sigma}^{11} \otimes Z_2'Z_2 + \tau^2 I_{k_2(m-1)}]^{-1}[\widetilde{\Sigma}^{11} \otimes Z_2'Z_2]\text{vec}(\widehat{\Pi}_2), \\ & [\widetilde{\Sigma}^{11} \otimes Z_2'Z_2 + \tau^2 I_{k_2(m-1)}]^{-1}), \end{aligned} \tag{11}$$

[4]The expressions for the conditional densities of Π_2 and β given in Geweke (1996, expressions (11) and (13)) contain some typographical errors and are corrected here in (11) and (12).

where $\widehat{\Pi}_2 = \widehat{\Theta}[\underline{\Phi}^+ + \underline{\Phi}^0\widetilde{\Sigma}^{21}(\widetilde{\Sigma}^{11})^{-1}]$, $\widehat{\Theta} = (Z_2'Z_2)^{-1}Z_2'(Y - Z_1A)$. $\underline{\Phi}^+$ is the Moore-Penrose generalized inverse of $\underline{\Phi}$ and the columns of $\underline{\Phi}^+$ and $\underline{\Phi}^0$ are orthogonal, and $C \equiv (\underline{\Phi}^+, \underline{\Phi}^0)$ is an $m \times m$ nonsingular matrix. Finally, $\widetilde{\Sigma}^{ij}$ denotes the partitioning of $\widetilde{\Sigma}^{-1} = (C'\underline{\Sigma}C)^{-1}$conformably with $Y = (Y_2, y_1)$.

(4) Conditional density of β

$$\begin{aligned} & \beta|(\Pi_2, \underline{\Sigma}^{-1}, A, Z, Y) \\ \sim\ & N([\underline{\Sigma}^{22} \otimes \Pi_2'Z_2'Z_2\Pi_2 + \tau^2 I_{m-1}]^{-1}[\underline{\Sigma}^{22} \otimes \Pi_2'Z_2'Z_2\Pi_2]\widehat{\beta}, \\ & [\underline{\Sigma}^{22} \otimes \Pi_2'Z_2'Z_2\Pi_2 + \tau^2 I_{m-1}]^{-1}), \end{aligned} \tag{12}$$

where

$$\begin{aligned} \widehat{\beta} \ =\ & (\Pi_2'Z_2'Z_2\Pi_2)^{-1}\Pi_2'Z_2'(Y_2 - Z_1\Pi_1)\underline{\Sigma}^{12}(\underline{\Sigma}^{22})^{-1} \\ & -\underline{\Sigma}^{12}(\underline{\Sigma}^{22})^{-1} + (\Pi_2'Z_2'Z_2\Pi_2)^{-1}\Pi_2'Z_2'(y_1 - Z_1\pi_1). \end{aligned}$$

3.3 The Chao and Phillips (1998) approach

Using Jeffreys prior, CP obtains exact and approximate analytic expressions for the posterior density of the structural coefficient β in the LISEM (1) and (2). Their formulas are found to exhibit Cauchy-like tails analogous to comparable results in the classical literature on LIML estimation. For the model (1) and (2) under normality assumption for the disturbances, a Jeffreys prior on the parameters, $\theta = (\beta, \gamma, \Pi_1, \Pi_2, \Sigma)$, is of the form

$$\begin{aligned} p(\beta, \gamma, \Pi_1, \Pi_2, \Sigma) \ &\propto\ \left| -E\left\{ \frac{\partial^2}{\partial\theta\partial\theta'} \ln L(\theta|Y, Z) \right\} \right|^{1/2} \\ &\propto\ |\sigma_{11}|^{(k_2-m+1)/2}|\Sigma|^{-(k+m+1)/2}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2}, \end{aligned} \tag{13}$$

where $\ln L(\theta|Y,Z)$ is the log-likelihood function as specified in (3), and $Q_X = I_T - P_X$, $P_X = X(X'X)^{-1}X'$. As first noted by Poirier (1996), the prior in (13) places no weight where $\text{rank}(\Pi_2) < (m-1)$ through the factor $|\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{1/2}$.

The joint posterior of the parameters of LISEM (1) and (2) is constructed as proportional to the product of the prior (13) and the likelihood function (3),

$$\begin{aligned} p(\beta,\gamma,\Pi_1,\Pi_2,\Sigma|Y,Z) &\propto p(\beta,\gamma,\Pi_1,\Pi_2,\Sigma)L(\beta,\gamma,\Pi_1,\Pi_2,\Sigma|Y,Z) \\ &\propto |\sigma_{11}|^{(k_2-m+1)/2}|\Sigma|^{-(T+k+m+1)/2}|\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{1/2} \\ &\quad \times \exp\{-\frac{1}{2} tr[\Sigma^{-1}(u,V_2)'(u,V_2)]\}, \end{aligned} \tag{14}$$

where (u, V_2) is defined in (1) and (2). Note that (14) or its conditionals do not belong to any standard class of probability density functions.

3.4 The Kleibergen and van Dijk (1998) approach

To solve the problem of local nonidentification and also to avoid the so called Borel-Kolmogorov paradox [see Billingsley (1986) and Poirier (1995)], KVD considered (4) as a multivariate linear model with nonlinear parameter restrictions:

$$\begin{pmatrix} y_1 & Y_2 \end{pmatrix} = \begin{pmatrix} Z_1 & Z_2 \end{pmatrix} \begin{pmatrix} \pi_1 & \Pi_1 \\ \phi_1 & \Phi_2 \end{pmatrix} + \begin{pmatrix} \xi_1 & V_2 \end{pmatrix}, \tag{15}$$

where ϕ_1 is a $k_2 \times 1$ vector, Φ_2 is a $k_2 \times (m-1)$ matrix. Denote $\Phi = (\phi_1, \Phi_2)$. The reduced form model (4) is obtained if a reduced rank restriction is imposed on the linear model (15) such that $\text{rank}(\Phi) = (m-1)$ instead of

m. Using a singular value decomposition (SVD) of Φ, they show that (15) is identical to the so-called unrestricted reduced form (URF) model,[5]

$$\begin{pmatrix} y_1 & Y_2 \end{pmatrix} = Z_1 \begin{pmatrix} \pi_1 & \Pi_1 \end{pmatrix} + Z_2\Pi_2 B + Z_2\Pi_{2\perp}\lambda B_\perp + \begin{pmatrix} \xi_1 & V_2 \end{pmatrix}, \qquad (16)$$

where $B = \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}$, λ is a $(k_2-m+1)\times 1$ vector. $\Pi_{2\perp}$and $B_\perp$ are the orthogonal complements of $\Pi_2$and B respectively, such that $\Pi_2'\Pi_{2\perp} \equiv 0$, $BB_\perp' \equiv 0$, and $\Pi_{2\perp}'\Pi_{2\perp} \equiv I_{k_2-m+1}$, $B_\perp B_\perp' \equiv 1$ (i.e. $\Pi_{2\perp} = \begin{pmatrix} -\Pi_{22}\Pi_{21}^{-1} & I_{k_2-m+1} \end{pmatrix}'(I_{k_2-m+1}+ \Pi_{22}\Pi_{21}^{-1}\Pi_{21}^{-1\prime}\Pi_{22})^{-1/2}$, where $\Pi_2 = \begin{pmatrix} \Pi_{21}' & \Pi_{22}' \end{pmatrix}'$, $\Pi_{21} : (m-1)\times(m-1)$, $\Pi_{22} : (k_2-m+1)\times(m-1)$, and $B_\perp = (1+\beta'\beta)^{-1/2}\begin{pmatrix} 1 & -\beta' \end{pmatrix}$).

There is one-to-one correspondence between the parameters in (15) and (16). The SVD of Φ is,

$$\Phi = USV', \qquad (17)$$

where $U : k_2 \times k_2$, $U'U = I_{k_2}$; $V : m \times m$, $V'V = I_m$; and $S : k_2 \times m$ is a rectangular matrix containing the (nonnegative) singular values (in decreasing order) on its main diagonal ($= (s_{11}, s_{22}, ..., s_{mm})$). Rewrite

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}, S = \begin{pmatrix} S_1 & 0 \\ 0 & s_2 \end{pmatrix} \text{ and } V = \begin{pmatrix} v_{11} & v_{12} \\ V_{21} & v_{22} \end{pmatrix}, \qquad (18)$$

where U_{11}, S_1, $V_{21} : (m-1)\times(m-1)$; $v_{12} : 1\times 1$; v_{11}', $v_{22} : (m-1)\times 1$; $U_{12} : (m-1)\times(k_2-m+1)$; $U_{21} : (k_2-m+1)\times(m-1)$; $U_{22} : (k_2-m+1)\times(k_2-m+1)$; $s_2 : (k_2-m+1)\times 1$, then the following relationship between (Π_2, β, λ) and (U, S, V) results,

$$\Pi_2 = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 V_{21}', \ \beta = V_{21}'^{-1} v_{11}', \text{ and}$$

[5]Note that this formulation or the singular value decomposition does not change the identification status of the LISEM specified by (1) and (2). If rank$(\Pi_2) < (m-1)$, β is locally nonidentified.

$$\lambda = (U_{22}U_{22}')^{-1/2}U_{22}s_2v_{12}'(v_{12}v_{12}')^{-1/2}. \tag{19}$$

Note that λ is obtained through pre- and postmultiplication of s_2 by orthogonal matrices while s_2 contains the smallest singular values of Φ and is invariant with respect to the ordering of variables contained in Y and Z_2.

According to KVD, the above shows that the model described by (1) and (2) can be considered as equivalent to the linear model (16) with a nonlinear (reduced rank) restriction $\lambda = 0$ on the parameters. Therefore the priors and posteriors of the parameters of the LISEM (1) and (2) may be constructed as proportional to the priors and posteriors of the parameters of the linear model (16) evaluated at $\lambda = 0$.

A diffuse (Jeffreys) prior for the parameters $(\pi_1, \Pi_1, \Phi, \Omega)$ of the linear model[6]

$$\begin{aligned} p(\pi_1, \Pi_1, \Phi, \Omega) &\propto |\Omega|^{-(k+m+1)/2} \\ &\propto |\Omega|^{-(m+1)/2}|\Omega^{-1} \otimes Z'Z|^{1/2}, \end{aligned} \tag{20}$$

where $k = k_1 + k_2$, implies the prior for the parameters $(\beta, \pi_1, \Pi_1, \Pi_2, \Omega)$ of the LISEM (4)

$$\begin{aligned} & p(\beta, \pi_1, \Pi_1, \Pi_2, \Omega) \\ \propto\ & p(\pi_1, \Pi_1, \Phi(\Pi_2, \beta, \lambda), \Omega)|_{\lambda=0}|J(\Phi, (\Pi_2, \beta, \lambda))|_{\lambda=0}| \\ \propto\ & |\Omega|^{-(m+1)/2}|\Omega^{-1} \otimes Z'Z|^{1/2}|J(\Phi, (\Pi_2, \beta, \lambda))|_{\lambda=0}| \\ \propto\ & |\Omega|^{-(m+1)/2}|\Omega^{-1} \otimes Z'Z|^{1/2} \\ & \times \left|\begin{pmatrix} B' \otimes I_{k_2} & e_1 \otimes \Pi_2 & B_\perp' \otimes \Pi_{2\perp} \end{pmatrix}\right|, \end{aligned} \tag{21}$$

[6]This is the prior suggested in Drèze (1976). Zellner (1971) and Zellner, Bauwens and van Dijk (1988) used a similar prior with $-(m+1)/2$ in the exponent.

where $e_1 = (1, 0, 0, ..., 0)'$. Note that the prior (21) is the Jeffreys prior of the unrestricted reduced form (16) evaluated at $\lambda = 0$. Most importantly, $|J(\Phi, (\Pi_2, \beta, \lambda))|_{\lambda=0}| = 0$ when Π_2 has reduced rank. This feature in KVD approach eliminates the potential impact of local nonidentification.

The joint posterior of the parameters of the LISEM (4) is readily constructed as proportional to the product of the prior (21) and the likelihood function (5),

$$\begin{aligned}
& p(\beta, \pi_1, \Pi_1, \Pi_2, \Omega | Y, Z) \\
\propto\ & p(\beta, \pi_1, \Pi_1, \Pi_2, \Omega) L^*(\beta, \gamma, \Pi_1, \Pi_2, \Omega | Y, Z) \\
\propto\ & |\Omega|^{-(T+m+1)/2} |\Omega^{-1} \otimes Z'Z|^{1/2} \\
& \times \left| \begin{pmatrix} B' \otimes I_{k_2} & e_1 \otimes \Pi_2 & B'_\perp \otimes \Pi_{2\perp} \end{pmatrix} \right| \\
& \times \exp\{-\frac{1}{2} tr[\Omega^{-1}(\begin{pmatrix} y_1 & Y_2 \end{pmatrix} - \begin{pmatrix} Z_1 & Z_2 \end{pmatrix} \begin{pmatrix} \pi_1 & \Pi_1 \\ \Pi_2\beta & \Pi_2 \end{pmatrix})' \\
& (\begin{pmatrix} y_1 & Y_2 \end{pmatrix} - \begin{pmatrix} Z_1 & Z_2 \end{pmatrix} \begin{pmatrix} \pi_1 & \Pi_1 \\ \Pi_2\beta & \Pi_2 \end{pmatrix})]\}, \qquad (22)
\end{aligned}$$

Unfortunately, the above posterior or its conditional densities do not belong to a known class of probability density functions.

3.5 The Jackknife Instrumental Variable Estimator (JIVE)

Motivated by split sample instrumental variables estimators, Angrist, Imbens and Krueger (1999), and Blomquist and Dahlberg (1999) independently suggested a jackknife instrumental variables estimator (JIVE). For model (1) and (2), JIVE is given by

$$\widehat{\delta}_{jive} = (\widehat{X}'_{jive} X)^{-1} (\widehat{X}'_{jive} y_1), \qquad (23)$$

where $\widehat{X}_{jive}$ is the $T \times (m-1+k_1)$ matrix with t-th row defined by

$$Z_t\widehat{\Pi}_{-t} = Z_t(Z'_{-t}Z_{-t})^{-1}(Z'_{-t}X_{-t}) = \frac{Z_t\widehat{\Pi} - h_tX_t}{1-h_t},$$

Z_{-t} and X_{-t} are $(T-1)\times k$ and $(T-1)\times(m-1+k_1)$ matrices obtained after eliminating the t-th rows of Z and X matrics respectively, $\widehat{\Pi} = (Z'Z)^{-1}Z'X$, and $h_t = Z_t(Z'Z)^{-1}Z'_t$. In JIVE, the instrument is independent of the disturbances even in finite samples, which is achieved by using a 'leave-one-out' jackknife-type fitted value in place of the usual unrestricted reduced form predictions.

Angrist, Imbens and Krueger (1999) also proposed a second jackknife estimator that is a slight modification of (23). Similar to their study, we found that its performance is very similar to JIVE, and is not reported here.

4 Posterior simulator: "Gibbs within M-H" algorithm

Given the full conditional densities in (9) through (12) for the four blocks of parameters, evaluating the joint posterior densities in Geweke (1996) by Gibbs sampling is straightforward, see Geweke (1996) for a detailed description. Although Geweke's (1996) shrinkage prior does not meet the argument in KVD that the implied prior/posterior on the parameters of an embedding linear model should be well-behaved, we found that the use of Geweke's shrinkage prior does not lead to a reducible Markov Chain. With the specification of a shrinkage prior, when Π_2 has reduced rank, the joint posterior density still depends on β and will not exhibit any asymptotic cusp. In the following we only discuss the posterior simulation for CP and KVD.

KVD suggested two simulation algorithms for the posterior (22): an Importance sampler and an Metropolis-Hastings algorithm. We found that their M-H algorithm performs unsatisfactorily with low acceptance rate even for reasonable parameter specifications. As mentioned earlier, since the posteriors (14) and (22) as well as their conditional posteriors do not belong to any standard class of probability density functions, Gibbs sampling can not be used. In this section, we suggest an alternative simulation algorithm which combines Gibbs sampling [see Casella and George (1992), and Chib and Greenberg (1996)] and Metropolis-Hastings algorithm [see Metropolis *et al.* (1953), Hastings (1970), Smith and Roberts (1993), Tierney (1994), Chib and Greenberg (1995)]. Our algorithm is different from the "M-H within Gibbs" algorithm, and can find its usefulness in other applications as well.

To generate drawings from the target density $p(x)$, we use a candidate-generating density $r(x)$. An Independence sampler, which is a special case of the M-H sampler, in algorithmic form is as follows:

0. Choose starting values x^0

1. Draw x^i from $r(x)$

2. Accept x^i with probability

$$\alpha(x^{i-1}, x^i) = \begin{cases} \min\left(\frac{p(x^i)r(x^{i-1})}{p(x^{i-1})r(x^i)},\ 1\right), & \text{if } p(x^{i-1})r(x^i) > 0 \\ 1, & \text{if } p(x^{i-1})r(x^i) = 0 \end{cases} \tag{24}$$

otherwise $x^i = x^{i-1}$.

3. $i = i + 1$. Go to 1.

It is generally not feasible to draw all elements of the vector x simultaneously. A block-at-a-time possibility was first discussed in Hastings (1970, sec. 2.4) and then in Chib and Greenberg (1995) along with an example. Chib

and Greenberg (1995) considered applying the M-H algorithm in turn to subblocks of the vector x, which presumes that the target density $p(x)$ may be manipulated to generate full conditional densities for each of the subblocks of x, conditioning on other elements of x. However the full conditionals are sometimes not readily available from the target density for empirical investigators. The posteriors (14) and (22) happen to fall in this category. In this latter case, problems come up at step 1 while trying to generate drawings from the joint marginal density $r(x)$. Note that these drawings, whether accepted or rejected at step 2, satisfy the necessary reversibility condition if step 1 is performed successfully.

To simplify the notation, we consider a vector x which contains two blocks, $x = (x_1, x_2)$. KVD used the fact that

$$r(x_1, x_2) = r(x_1)r(x_2|x_1) \tag{25}$$

and suggested to draw x_1^i from $r(x_1)$ and then draw x_2^i from $r(x_2|x_1^i)$. The pair (x_1^i, x_2^i) is then taken as a drawing from $r(x)$. It turns out that this strategy gives very low acceptance rate at step 2 in simulation studies for various reasonable parameter values. Sometimes the move never take place and the posterior has all its mass at the parameter values of the first drawing. The reason for the failure is that information is not updated at subsequent drawings and the transition kernel of (25) is static.

If the full conditionals $r(x_1|x_2)$ and $r(x_2|x_1)$ are available, which is usually true for many standard densities, we propose to use them in a Gibbs sampler to make independent drawings from the invariant density $r(x)$ after the Markov chain has converged.

The combined algorithm is thus as follows, which we call "Gibbs within M-H":

0. Choose starting values $x^0 = (x_1^0, x_2^0)$.

1. Draw x_1^i from $r(x_1|x_2^{i-1})$, draw x_2^i from $r(x_2|x_1^i)$.

2. Accept $x^i = (x_1^i, x_2^i)$ with probability $\alpha(x^{i-1}, x^i)$ as defined in (24), otherwise $x^i = x^{i-1}$.

3. $i = i + 1$. Go to 1.

As explained, step 2 is the Gibbs step and step 3 is the M-H step in our combined algorithm. In the following subsections, we describe the steps for implementing the above procedure to generate drawings from the posteriors (14) and (22).[7]

4.1 Implementing the CP approach

Note that the posterior in the CP approach is proportional to the product of the prior, which is uniformly bounded, and the likelihood function, which can be sampled by a Gibbs sampler. Therefore we choose the candidate-generating density the way suggested by Chib and Greenberg (1995): we use the likelihood function, $L(\beta, \gamma, \Pi_1, \Pi_2, \Sigma|Y, Z)$, as the candidate generating density for the posterior (14). Using precision matrix Σ^{-1}, the simulation steps are as follows,

0. Choose starting values $(\beta^0, \gamma^0, \Pi_1^0, \Pi_2^0, \Sigma^{-1,0})$

1. Draw $\Sigma^{-1,i}$ from $p(\Sigma^{-1}|\beta^{i-1}, \gamma^{i-1}, \Pi_1^{i-1}, \Pi_2^{i-1}, Y, Z)$

   Draw $(\beta^i, \gamma^i, \Pi_1^i, \Pi_2^i)$ from $p(\beta, \gamma, \Pi_1, \Pi_2|\Sigma^{-1,i}, Y, Z)$

[7] The algorithm has been illustrated with a simple labor supply model in Gao and Lahiri (2000).

2. Accept $(\beta^i, \gamma^i, \Pi_1^i, \Pi_2^i, \Sigma^{-1,i})$ as a drawing from the posterior (14) with probability,

$$\min\left(\frac{|\sigma_{11}^i|^{(k_2-m+1)/2}|\Sigma^{-1,i}|^{(k+m+1)/2}|\Pi_2^{i\prime}Z_2'Q_{Z_1}Z_2\Pi_2^i|^{1/2}}{|\sigma_{11}^{i-1}|^{(k_2-m+1)/2}|\Sigma^{-1,(i-1)}|^{(k+m+1)/2}|\Pi_2^{i-1\prime}Z_2'Q_{Z_1}Z_2\Pi_2^{i-1}|^{1/2}}, 1\right),$$

otherwise, $(\beta^i, \gamma^i, \Pi_1^i, \Pi_2^i, \Sigma^{-1,i}) = (\beta^{i-1}, \gamma^{i-1}, \Pi_1^{i-1}, \Pi_2^{i-1}, \Sigma^{-1,(i-1)})$.

3. $i = i + 1$. Go to 1.

The conditional densities used in the first step are constructed as follows (see Percy (1992) and Chib and Greenberg (1996)): Rewrite the model (1) and (2) as a SUR model,

$$y_t = W_t\delta + \begin{pmatrix} u_t \\ V_{2,t} \end{pmatrix}, \tag{26}$$

where

$y_t = (\ y_{1,t} \quad Y_{2,t}'\ )'$,

$W_t = \begin{bmatrix} (\ Y_{2,t}' \quad Z_{1,t}'\ ) & \\ & (I_{m-1} \otimes Z_t') \end{bmatrix}$,

$\delta = (\beta', \gamma', (\mathrm{vec}\begin{pmatrix} \Pi_1 \\ \Pi_2 \end{pmatrix})')'$. Then

$$p(\Sigma^{-1}|\delta, Y, Z) \propto |\Sigma^{-1}|^{(T-2(m+1))/2}\exp[-\frac{1}{2}tr(\Sigma^{-1}H)], \tag{27}$$

which follows a Wishart distribution with $(T - m - 1)$ degrees of freedom, where $H = \sum_{t=1}^T (y_t - W_t\delta)(y_t - W_t\delta)'$, and

$$p(\delta|\Sigma^{-1}, Y, Z) = N\left(\left(\sum_{t=1}^T W_t'\Sigma^{-1}W_t\right)^{-1}\left(\sum_{t=1}^T W_t'\Sigma^{-1}y_t\right), \left(\sum_{t=1}^T W_t'\Sigma^{-1}W_t\right)^{-1}\right). \tag{28}$$

4.2 Implementing the KVD approach

KVD proposed to use the posterior of the unrestricted linear model (16), $p(\beta, \lambda, \Pi_2, \Omega|Y, Z)$, as the candidate generating density of the posterior (22), $p(\beta, \Pi_2, \Omega|Y, Z)$, where the parameters (π_1, Π_1) have been concentrated out. First (Φ, Ω) is generated from $p(\Phi, \Omega|Y, Z)$, and then (β, λ, Π_2) is obtained from Φ using (19). However, λ is also sampled which is not present in the posterior $p(\beta, \Pi_2, \Omega|Y, Z)$. Therefore KVD assume that λ is generated by a conditional density of the form,

$$\begin{aligned} & g(\lambda|\beta, \Pi_2, \Omega) \\ = \; & (2\pi)^{-(k_2-m+1)/2}|B_\perp \Omega^{-1} B_\perp'|^{(k_2-m+1)/2}|\Pi_{2\perp}' Z_2' M_{z_1} Z_2 \Pi_{2\perp}|^{1/2} \\ & \times \exp[-\frac{1}{2} tr(B_\perp \Omega^{-1} B_\perp' (\lambda - \widehat{\lambda})' \Pi_{2\perp}' Z_2' M_{z_1} Z_2 \Pi_{2\perp} (\lambda - \widehat{\lambda}))], \end{aligned} \quad (29)$$

where $\widehat{\lambda} = (\Pi_{2\perp}' Z_2' M_{z_1} Z_2 \Pi_{2\perp})^{-1} \Pi_{2\perp}' Z_2' M_{z_1} (Y - Z_2 \Pi_2 B) \Omega^{-1} B_\perp' (B_\perp \Omega^{-1} B_\perp')^{-1}$. Therefore the density $p(\beta, \lambda, \Pi_2, \Omega|Y, Z)$ is used to approximate the posterior $g(\lambda|\beta, \Pi_2, \Omega) p(\beta, \Pi_2, \Omega|Y, Z)$. The weight function, defined as the ratio of the posterior and the candidate generating density, becomes

$$\begin{aligned} w(\beta, \lambda, \Pi_2, \Omega) &= \frac{g(\lambda|\beta, \Pi_2, \Omega) p(\beta, \Pi_2, \Omega|Y, Z)}{p(\beta, \lambda, \Pi_2, \Omega|Y, Z)} \\ &= \frac{|J(\Phi, (\Pi_2, \beta, \lambda))|_{\lambda=0}|}{|J(\Phi, (\Pi_2, \beta, \lambda))|} g(\lambda|\beta, \Pi_2, \Omega)|_{\lambda=0}, \end{aligned} \quad (30)$$

where the Jacobian matrix $J(\Phi, (\Pi_2, \beta, \lambda))$ as well as $J(\Phi, (\Pi_2, \beta, \lambda))|_{\lambda=0}$ have been carefully derived in KVD.[8] Note that $w(\cdot) = p(\cdot)/r(\cdot)$, so (30) may be used in the "GS within M-H" algorithm to simplify (24).

[8] See also Kleibergen (1997, 1998). Note that their claimed relationship that $|J(\Phi, (\Pi_2, \beta, \lambda))| \geq |J(\Phi, (\Pi_2, \beta, \lambda))|_{\lambda=0}|$ is analytically incorrect; see the Appendix for proof.

Similar to the way we implemented the CP approach, it is more convenient to work with the precision matrix Ω^{-1}in the conditional densities. Applying the procedure outlined above, the steps involved in constructing the Markov chain for the posterior (22) are summarized as follows,

0. Choose starting values $(\Phi^0, \Omega^{-1,0})$
1. Draw $\Omega^{-1,i}$ from $p(\Omega^{-1}|\Phi^{i-1}, Y, Z)$

   Draw Φ^i from $p(\Phi|\Omega^{-1,i}, Y, Z)$
2. Perform a singular value decomposition of $\Phi^i = U^i S^i V^{i\prime}$
3. Compute $\beta^i, \lambda^i, \Pi_2^i$ according to (18)-(19)
4. Compute $w(\beta^i, \lambda^i, \pi_1^i, \Pi_1^i, \Pi_2^i, \Omega^{-1,i})$ according to (29)-(30)
5. Draw (π_1^i, Π_1^i) from $p(\pi_1, \Pi_1|\Omega^{-1,i}, \Phi^i(\Pi_2^i, \beta^i, \lambda), Y, Z)|_{\lambda=0}$
6. Accept $(\beta^i, \pi_1^i, \Pi_1^i, \Pi_2^i, \Omega^{-1,i})$ as a drawing from the posterior with probability,

$$\min\left(\frac{w(\beta^i, \lambda^i, \Pi_2^i, \Omega, ^{-1,i})}{w(\beta^{i-1}, \lambda^{i-1}, \Pi_2^{i-1}, \Omega^{-1,(i-1)})},\ 1\right),$$

   otherwise, $(\beta^i, \pi_1^i, \Pi_1^i, \Pi_2^i, \Omega^{-1,i}) = (\beta^{i-1}, \pi_1^{i-1}, \Pi_1^{i-1}, \Pi_2^{i-1}, \Omega^{-1,(i-1)})$.
7. $i = i + 1$. Go to 1.

Note that the conditional densities used in the first step are as follows:

$$p(\Omega^{-1}|\Phi, Y, Z) \propto |\Omega^{-1}|^{(T+k_2-m-1)/2} \exp[-\frac{1}{2} tr(\Omega^{-1} G)], \tag{31}$$

which follows a Wishart distribution $W_m(T + k_2, G^{-1})$ with $(T + k_2)$ degrees of freedom, where $G = Y'Q_z Y + (\Phi - \widehat{\Phi})' Z_2' M_{Z_1} Z_2 (\Phi - \widehat{\Phi})$, and $\widehat{\Phi} = (Z_2' M_{Z_1} Z_2)^{-1} Z_2' M_{Z_1} Y$. In addition,

$$p(\Phi|\Omega^{-1}, Y, Z) \propto |\Omega^{-1}|^{k_2/2} \exp\{-\frac{1}{2} tr[\Omega^{-1}(\Phi - \widehat{\Phi})' Z_2' M_{Z_1} Z_2 (\Phi - \widehat{\Phi})]\}, \tag{32}$$

which is a matric-variate normal density.

The conditional density used in step 5 is

$$p(\pi_1, \Pi_1|\Omega^{-1}, \Phi(\Pi_2, \beta, \lambda), Y, Z) \propto |\Omega^{-1}|^{k_1/2} \exp\{-\frac{1}{2} tr[\Omega^{-1}(\Lambda - \widehat{\Lambda})' Z_1' Z_1 (\Lambda - \widehat{\Lambda})]\}, \tag{33}$$

evaluated at $\lambda = 0$, where $\Lambda = (\ \pi_1 \quad \Pi_1\ )$, $\widehat{\Lambda} = (Z_1' Z_1')^{-1} Z_1' (Y - Z_2 \Phi)$.

4.3 Convergence Diagnosis

One important implementation issue associated with MCMC methods is that of determining the number of iterations required. There are various informal or formal methods for the diagnosis of convergence, see Cowles and Carlin (1996) and Brooks and Roberts (1999) for recent comprehensive reviews and recommendations. Since the posterior densities in (14) and (22) resulting from CP and KVD do not have moments of any positive integer order, most of the methods proposed in the MCMC literature which require the existence of at least the first moment (posterior mean) are ruled out. We are left with a very few alternatives that can be used in our context.

First, the popular Raftery and Lewis (1992) method has been recognized as the best for estimating the convergence rate of the Markov chain if quantiles of the posterior density are of major interest, although the method does not provide any information as to the convergence rate of the chain as a whole. Because we are interested in the posterior modes and medians for β associated with the Bayesian approaches, we may largely rely on Raftery and Lewis' method to determine the number of burn-ins, and the subsequent number of iterations required to attain specified accuracy (e.g., estimating the 0.50 quantile in any posterior within ± 0.05 with probability 0.95). But

we do not adopt their suggested skip-interval. MacEachern and Berliner (1994) showed that estimation quality is always degraded by discarding samples. We once experimented with using the skip-intervals and found that the results are basically the same if a sufficient number of iterations are run. This seems to be inefficient and sometimes infeasible in terms of computation time.

For each specification in our Monte Carlo study with repeated experiments, we determined the number of burn-ins and subsequent number of iterations by running the publicly available Fortran code gibbsit on MCMC output of 10,000 iterations from three or more testing replications. For KVD and CP approaches, the numbers of burn-ins for both the GS step and the M-H algorithm were estimated. It was found that the number of burn-ins in the GS step is negligible for most cases. However, we discarded more iterations as the transient phase than the estimated number of burn-ins.[9] The estimated number of subsequent iterations across testing replications was stable for the Gibbs sampler (in both Geweke approach and the GS step for KVD and CP approaches), but it varied a lot for the M-H procedures, which is also demonstrated by the variation in acceptance rates over repeated experiments. We used a generous value for the number of subsequent iterations when feasible.

Second, for MCMC output from each testing replication, we also applied other convergence diagnostic methods, including percentiles derived from every quarter of the long chain, Yu and Mykland (1994)'s CUSUM plot, and

[9]In practice, there is often a concern about possible underestimation of true length of the burn-in period using the Raftery and Lewis method if the quantile of interest is not properly pre-prescribed, see Brooks and Roberts (1999).

Brooks' (1996) D sequence statistic. While the CUSUM partial sums actually involve averaging over sampling drawings, the computation of Brooks' statistic is justified on the basis that it is designed to measure the frequency of back-forth movement in the MCMC algorithm. However, these diagnostics may sometimes provide contradictory outcomes so that one has to be extra careful in interpreting them before making a judgment on convergence.

5 Simulation results and discussions

In this section, we present results of Monte Carlo experiments and discuss some of the findings. As mentioned before, for the purpose of comparison, we also computed a number of single K-class estimators including OLS, 2SLS, LIML and Fuller's modified LIML. In summary, the set of K-class estimator for the structural coefficients in model (1) and (2) is given by:

$$\begin{pmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} = \begin{bmatrix} Y_2'Y_2 - K_1\widehat{V}_2'\widehat{V}_2 & Y_2'Z_1 \\ Z_1'Y_2 & Z_1'Z_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_2 - K_2\widehat{V}_2)'y_1 \\ Z_1'y_1 \end{bmatrix}, \qquad (7)$$

where $\widehat{V}_2 = Q_Z Y_2$.

The following LISEM estimators have been considered:

(1) Ordinary least squares (OLS)

$$K_1 = K_2 = 0.$$

(2) Two stage least squares (2SLS)

$$K_1 = K_2 = 1.$$

(3) Zellner's (1978) Bayesian minimum expected loss estimator (MELO)

$$K_1 = K_2 = 1 - k/(T - k - m - 1).$$

(4) Zellner's Bayesian method of moments relative to balanced loss function (BMOM)[10]

$$K_1 = 1 - k/(T - k), K_2 = 1 - (1 - \omega)k/(T - k) \text{ with } \omega = 0.75$$

(5) Classical LIML. We compute classical LIML as an iterated Aitken estimator [see Pagan (1979), Gao and Lahiri (1999)].

(6) Fuller (1977) modified LIML estimators (Fuller1 and Fuller4)

$$K_1 = K_2 = \lambda_* - \alpha/(T - k), \text{ for } \alpha = 1, 4$$

where

$$\lambda_* = \min_{\beta} \frac{(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'Q_Z(y_1 - Y_2\beta)}$$

and it is computed using the LIML estimate.

(7) JIVE

(8) Posterior mode and median from Geweke (1996) approach using Gibbs Sampling. The values of the hyperparameters are chosen to be $\tau^2 = 0.01$, $\nu = m(m + 1)/2$, $S = 0.01I_m$.[11]

(9) Mode and median of the marginal density of β based on classical LIML from Gibbs sampling (LIML-GS). LIML-GS is a by-product of the "Gibbs

[10] Tsurumi (1990) used $\omega = 0.75$ for Zellner's extended MELO (ZEM) in his experiments. BMOM and ZEF are almost identical in our context.

[11] We found that the median-bias and disperson of the posterior density of β from the Geweke (1996) approach increase as τ^2 gets larger. Although one might suspect that the convergence of the Gibbs sampler could be slow with smaller values of τ^2, our convergence diagnostics did not confirm this concern.

within M-H" algorithm for the CP approach since the likelihood function in (3) is used as the candidate-generating density to explore the CP posterior.

(10) Posterior mode and median from CP approach using "Gibbs within M-H" algorithm

(11) Posterior mode and median from KVD approach using "Gibbs within M-H" algorithm

For the recent Bayesian approaches and LIML-GS, we report both (posterior) mode and median to show possible asymmetry in the marginal densities of β. Any preference for one over the other will depend on the researcher's loss function. We obtain 16 estimates for each generated data set. The data are generated from the model,

$$y_1 = Y_2\beta + u,$$

$$Y_2 = Z_2\pi + V_2, \tag{34}$$

where y_1, Y_2 are $T \times 1$ such that $m = 2$, and $Z_2 : T \times k_2$. We further specify $\beta = 1$ and

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{35}$$

for $|\rho| = 0.20$, 0.60, and 0.95.[12] Z_2 is simulated from a $N(0, I_{k_2} \otimes I_T)$ distribution and (u, V_2) from a $N(0, \Sigma \otimes I_T)$ distribution. A constant term is added in each equation, i.e., Z_1 is a $T \times 1$ vector of 1's.

The simulation results are reported in Table 1 through Table 13. Tables 1 to 12 are for cases with $\rho > 0$, each table reporting results for one speci-

[12]We do not report cases with $|\rho| = 0.99$ or 1. As pointed out by Maddala and Jeong (1992), when the instruments are weak and $|\rho|$ is very close to one, the exact finite sample distribution of IV estimator is bimodal. Our experiments show that the marginal posterior density of β from the recent Bayesian approaches exhibits a similar pattern.

fication. Tables 13 summarizes the results for cases with $\rho < 0$ for BMOM and KVD for whom negative ρ made a surprising difference. As mentioned before, we focus on the estimates of the structural parameter β. Specifically, we analyze the sensitivity of the various estimates of β with respect to the strength of the instrumental variables Z, the degree of overidentification $(k_2 - m + 1)$, the degree of endogeneity (ρ), and the sample size (T). Also, we will examine whether the performance of an estimator is symmetric with respect to the sign of parameter ρ, an issue generally overlooked in the literature.[13]

Note that the strength of the instrumental variables for the included endogenous variable Y_2 is measured in terms of the adjusted R^2 by regressing Y_2 on $Z = (Z_1, Z_2)$. In the data generating process, we controlled $\overline{R}^2$ to be within $\pm 2.5\%$ of the specified value to reduce unnecessary variation. We did not experiment with extremely small $\overline{R}^2$ (say, 0.01 or less). In these cases the mean values of all estimators approached the point of concentration ω_{12}/ω_{22}, which is equal to $(\beta + \rho)$ for our data generating process (DGP).

For each specification, the number of replications is 400. The number of burn-ins (nburn_GS and nburn_MH), and subsequent number of iterations (n) determined at the convergence diagnosis step are reported in the footnotes

[13]Denote $\Omega = \begin{bmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{bmatrix}$. Using $\Sigma = C'\Omega C$, we have $\sigma_{11} = w_{11} - 2\beta w_{12} + \beta^2 w_{22}$, $\sigma_{12} = w_{12} - \beta w_{22}$, and $\sigma_{22} = w_{22}$. Letting $\rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$, the second relationship may be rewritten as:

$$\beta - \frac{w_{12}}{w_{22}} = -\rho\sqrt{\frac{\sigma_{11}}{w_{22}}}.$$

If Σ is normalized as in (35) with $\sigma_{11} = w_{22} = 1$, then $w_{12} = \beta + \rho$. Therefore, in our context, given $\beta = 1$, the sign and magnitude of ρ (or w_{12}) has a special significance.

to each table. The average acceptance rate and its standard deviation (in parentheses) across replications for each M-H routine are reported as well. To evaluate alternative estimators, we computed mean, standard deviation (Std), root of mean squared errors (RMSE), and mean absolute deviation (MAD) over repeated experiments for all the estimators considered.[14] Since LIML, posterior densities for CP and KVD, as well as 2SLS in the just-identified case do not have finite moments of positive order in finite samples, one should interpret the computed mean, standard deviation and RMSE across replications for these estimators with caution. In this sense, the MAD across replications is a preferred measure to consider.

We will first look at cases reported in Tables 1 to 12 with $\rho > 0$. In Table 1, we consider a case (T = 50, ρ = 0.60, k_2= 4) with moderately strong instruments ($\overline{R}^2$ = 0.40). It is found that with reasonably strong instruments all estimators designed for simultaneous equations perform reasonably well. As expected, OLS is seriously biased. BMOM has a slight edge over others in terms of RMSE and MAD. For all recent Bayesian approaches and LIML-GS the medians perform a little better than modes, and CP over KVD, in terms of bias, RMSE and MAD. Notice that the classical LIML estimates are different from LIML-GS (mode or median). As noted by Drèze (1976), from a Bayesian viewpoint, LIML produces an estimate of β conditionally on the over-identifying restrictions, the modal values of all the remaining parameters, and a uniform prior. In other words, the concentrated likelihood function of β after concentrating out (i.e., maximizing

[14]Medians were also calculated. Since they were very close to the corresponding means in all our experiments, we did not report them in this paper.

with respect to) other reduced-form and nuisance parameters is a conditional density. However, LIML-GS is a marginal density with all other parameters being integrated out. Due to possible asymmetry in the distribution of the nuisance parameters, the modal/median values of LIML-GS may not coincide with classical LIML estimates. In all our experiments, we find that the median-unbiasedness property of (conditional) LIML does not carry over to the marginal LIML (i.e., LIML-GS); however, the former generally has a much larger standard deviation than the latter. In a way, LIML-GS brings the classical LIML estimator close to its Bayesian counterpart for the purpose comparison.

It is interesting to note that across all our tables, the difference between LIML-GS and CP can only be attributed to the importance of Jeffreys prior. Compared to LIML-GS, typically CP has a smaller bias, but slightly larger standard deviation, even though the differences are very small. In some cases, however, the use of Jeffreys prior reduces the bias in CP quite substantially. For example, in Table 4 with T = 50 and a high degree of overidentification, the bias is reduced from 0.36 to 0.25.

A simple case when the structural model is just identified ($k_2 = 1$) is reported in Table 2. For this case it is well known that classical LIML coincides with 2SLS. KVD approach does not accommodate the case of just-

identification since (15) requires $k_2 > (m-1)$.[15] In this case, we find that CP-Mode produces results closer to LIML-GS-Mode than to LIML. CP (1998) showed that for a two-equation just-identified SEM in orthonormal canonical form, the posterior density of β with Jeffreys prior has precisely the same functional form as the density of the finite sample distribution of the corresponding LIML estimator as obtained by Mariano and McDonald (1979). Our simulation results show that the assumption of orthonormal canonical form is crucial for their exact correspondence, which cannot be extended to a general SEM.[16] In general, the Bayesian marginal density is not the same as the classical conditional density. Interestingly, JIVE is considerably more biased and has larger standard deviation than 2SLS. Also, CP-Median and LIML-GS-Median perform significantly worse than their modes. This is because in an exactly identified model with weak instruments the probability of local non-identification is substantial, and the resulting non-standard marginal density exhibits a very high variance. The same result holds true for Geweke-Median, but to a lesser extent. Thus, for exactly identified SEMs

[15] When $k_2 = (m-1)$, a diffuse prior in (20) for the linear model implies that the prior for the parameters of the LISEM (4) is

$$p(\beta, \pi_1, \Pi_1, \Pi_2, \Omega) \propto |\Omega|^{-(k+m+1)/2} |\Pi_2|,$$

and the prior for the parameters of the LISEM (1) and (2) is

$$p(\beta, \gamma, \Pi_1, \Pi_2, \Sigma) \propto |\Sigma|^{-(k+m+1)/2} |\Pi_2|,$$

which is identical to the Jeffreys prior; see also expressions (22) and (42) in CP.

[16] Note that the relationship between the standardized parameter vector and the original parameter vector involves the nuisance parameters, cf. Phillips (1983). However, when a SEM is in orthonormal canonical form (i.e., the exogenous regressors are orthonormal and the disturbance covariance matrix Ω is an identity matrix), both the density of random parameter β from the CP approach and the probability density of the classical LIML estimator for β are conditional on these information.

with very weak instruments, mode of the marginal density is a more dependable measure of β. We should point out that in all other cases in this study the medians generally turned out to be more preferable than the modes in terms of bias, RMSE, and MAD (see Tables 11 and 12, for instance).

Results reported in Tables 3 through 12 consider cases with general over-identification and weak instruments. As noted in the literature, OLS and 2SLS are median-biased in the direction of the correlation coefficient ρ, and the bias in 2SLS grows with the degree of over-identification, and decreases as sample size increases. Results in Tables 3 through 10 confirm these results.. Since MELO is a single K-class estimator with $0 < K < 1$, it is always between OLS and 2SLS estimates. The bias in MELO shows the same pattern as that of 2SLS. With moderate simultaneity, the median-bias in 2SLS can be as large as about 40% of the true value (see Table 8). We note that MELO, LIML-GS-Mode, and KVD-Mode or KVD-Median are also median-biased in the direction of ρ. But the bias in JIVE is consistently in the opposite direction of ρ. Classical LIML is remarkably median-unbiased when the instrumental variables are not very weak, which is well documented in the literature. We find that LIML is median-biased in the direction of ρ when the instruments are very weak (Table 8), which is consistent with the finding in Staiger and Stock (1998) using local-to-zero asymptotic theory. Even in this situation, the bias of LIML is much smaller than that of any other estimator, except BMOM.

The MAD of OLS is very close to its bias (i.e., relatively small Std) across all cases and it implies that OLS method is robust in the sense that it does not suffer from heavy tails or outlying estimates, see Zellner (1998). In this

sense, MELO and BMOM are all robust with relatively small standard deviations across replications. However, OLS exhibits large bias in the presence of simultaneity and is not so appealing. It is known that for a degree of overidentification strictly less than 7, 2SLS would have a smaller asymptotic mean squared error (AMSE) than LIML, cf. Mariano (1982). In cases with weak instruments the situation gets more complicated in finite samples. In our experiments, LIML has larger RMSE and MAD than 2SLS except in Tables 11 and 12 where ρ was 0.95. Note that the degree of over-identification is 8 in Tables 4, 6, 8 and 10.

Among classical estimators, JIVE turns out to be least appealing. Monte Carlo simulations in Angrist, Imbens and Krueger (1999) showed that JIVE has slight median bias in the opposite direction of ρ (but less than 2SLS) and have heavier tails than LIML. Our Table 6 is comparable to panel 2 of their Table I, and the results are also similar. Our other experiments show that JIVE may also have large absolute bias (larger than LIML) in the case with weak instruments, sometimes even greater than 2SLS (see Table 2). Generally, JIVE has slightly less bias than 2SLS, but this gain is overshadowed by enlarged standard deviation such that in finite samples it has no advantage over 2SLS in terms of MAD and RMSE. We also find that JIVE has greater RMSE and MAD than LIML. Blomquist and Dahlberg (1999) experimented with much larger sample sizes than ours. Comparing our Table 4 with Table 6 and with an unreported simulation with a sample size of 500, we found that the relative gain in JIVE is more than other estimators as sample size increases, even though its relative low standing remains valid.

Fuller's modified LIML estimators are included because Fuller1 is de-

signed to minimize the median-bias, and Fuller4 to minimize the mean-squared error. It seems that this conclusion is also problematic in the presence of weak instruments. Between the two, Fuller1 has smaller median-bias, and Fuller4 has smaller standard deviations across replications. However, in terms of RMSE or MAD, Fuller4 shows no advantage over Fuller1 in most of the cases.

Because all the estimators except OLS are consistent and their asymptotic distributions are also the same, results in Tables 3 through 6 confirm that their bias and dispersion decrease as sample size increases. But if the instruments are very weak (see Tables 7 and 8), their bias and dispersion may remain significant, a point emphasized forcefully by Zellner (1998). However, when the endogeneity is not strong (see Tables 9 and 10), their bias and dispersion may not be a big concern for some of the estimators.

Across all cases, we find that the bias in BMOM is small if ρ is not too small and the structural equation (1) is overidentified. As sample size increases or degree of over-identification rises, the observed bias in BMOM decreases. The most striking feature of BMOM is that it exhibits the smallest MAD and Std when ρ is not too small. MELO shows slightly smaller MAD and Std than BMOM if ρ is small (see Tables 9-10). In cases with very weak instruments and high degree of over-identification, the MAD of BMOM is only one-fourth of that of other estimators (see Table 8). These are in accordance with Tsurumi (1990)'s finding that in many cases, ZEM has the least relative mean absolute deviation. Meanwhile, if ρ is very small and the structural equation is overidentified, the bias in BMOM can be large; 2SLS, LIML-GS, Geweke, and CP perform remarkably well in these situations.

Next, we examine in more detail the recent Bayesian approaches. Overall, the median bias resulting from these approaches exhibits the same pattern as the bias of 2SLS, it increases with the degree of over-identification, and decreases as sample size rises. The Geweke (1996) approach used a shrinkage prior but its performance is comparable with LIML-GS and CP. The median-bias from PMOD-Geweke is the same or slightly less than that of LIML-GS-Mode, and the bias from Geweke-Median is always slightly less than that of LIML-GS-Median. Similar relationships are observed for MADs. These reflects the impact of the (informative) shrinkage prior on the posterior density.

For each specification, the acceptance rate in the M-H algorithm using CP approach is stable while that using KVD approach shows huge variation across replications. The acceptance rate for CP is generally above 40% except when sample size is small and the degree of overidentification is high. This shows that the posterior of CP is largely dominated by the likelihood function (3) and the Jeffreys prior generally carries little information. Second, in terms of the computed standard deviations (Stds) of the estimates across replications, CP-Mode has larger dispersion than LIML-GS-Mode, and CP-Median has larger dispersion than LIML-GS-Median. These also shed light on the notion that Jeffreys prior is less informative than a uniform prior. However, between the Jeffreys prior (13) used by CP and the implied prior (21) resulting from diffuse/Jeffreys prior on a linear model used by KVD, it is not clear which one is less informative.

As for the KVD (1998) approach, we observe that it performs as well as any other estimator if the instruments are not weak (see Table 1). But

when the instruments are weak, and ρ is positive, KVD shows more bias and higher MAD than those from CP. In Tables 4 with T = 50 and high degree of overidentification, KVD performs as bad as OLS.

Next we consider cases with negative ρ, and the results are summarized in Table 13. We replicate each case in Tables 1 - 12 with the same specification except ρ being negative. Since the performance of all estimators except BMOM and KVD were basically the same with respect to the sign of ρ, we only report results on these two in Table 13. We find that when ρ changes sign, the bias of BMOM does not change sign and even increases in magnitude. Also note that the computed Stds for BMOM when $\rho < 0$ are close to the respective ones when $\rho > 0$. Therefore, for cases with $\rho < 0$, BMOM has large RMSEs/MADs and loses its attraction. Note that BMOM is the same as the double K-class estimator (DKC) with K values fixed. This asymmetry in the performance in DKC is not well recognized in the literature. However, the observed asymmetry in its bias with respect to ρ in our experiments is readily explained by examining an expression for the mean of double K-class estimator (DKC) in Dwivedi and Srivastava (1984, Theorem 1). We can express $\widehat{\delta}_{DKC}$ as:

$$\widehat{\delta}_{DKC} = \widehat{\delta}_{K_1} + \begin{bmatrix} Y_2'Y_2 - K_1\widehat{V}_2'\widehat{V}_2 & Y_2'Z_1 \\ Z_1'Y_2 & Z_1'Z_1 \end{bmatrix}^{-1} \begin{bmatrix} (K_1 - K_2)\widehat{V}_2' y_1 \\ 0 \end{bmatrix}, \qquad (36)$$

where $\widehat{\delta}_{K_1}$ is a single K-class estimator with characterizing scalar K_1. When $Z_1'Z_2 = 0$, which is satisfied in our experimental specifications, a double K-class estimator of β may be written as

$$\widehat{\beta}_{DKC} = \widehat{\beta}_{K_1} + (K_1 - K_2)\frac{Y_2'Q_Z y_1}{Y_2'\Delta Y_2}$$

where $\Delta = (1 - K_1)Q_{z_1} + K_1 P_{Z_2}$. Observe that for $0 < K_1 < 1$, $\widehat{\beta}_{K_1}$ is biased in the direction of ρ, as noted in Mariano (1982). Note also that $Y_2'\Delta Y_2 > 0$, and $Y_2'Q_Z y_1$ provides an estimate of w_{12}. Although Dwivedi and Srivastava (1984) explored the dominance of double K-class over K-class using the exact MSE criterion, their guidelines for the selection of K_2 for a given K_1 are not entirely valid, because the conditions were derived from a small Monte Carlo simulation with cases with positive w_{12} and negative ρ only. Since $K_1 < K_2$ for BMOM, when ρ and w_{12} have the opposite sign, the second term in $\widehat{\beta}_{DKC}$ will be of the same sign as the bias of $\widehat{\beta}_{K_1}$, therefore $\widehat{\beta}_{DKC}$ (hence BMOM) will exhibit large bias. Otherwise, when $\rho w_{12} > 0$, the bias is mitigated. Based on our simulation results, we found that the sign of ρ has no effect on the standard deviation of BMOM. This finding shows that the greater RMSE of BMOM when $\rho w_{12} < 0$ is due to the aggravated bias. For the specification corresponding to table 4 in Table 13 (i.e., T = 50, ρ = -0.60, $K_2 = 4$, $\overline{R}^2 =$ 0.10), we find that for given $K_1 = 0.947$, RMSE is minimized if K_2 is chosen to be 0.829, which is much less than K_1, and also less than $K_2 = 0.987$ used in BMOM.

In tables 3 - 12 we found that KVD with $\rho > 0$ performs very poorly, often with substantial bias and relatively high RMSE and MAD. CP uniformly dominates KVD in these cases. However, with $\rho < 0$ the picture turns around remarkably well in favor of KVD. As we see in Table 13, across all cases the bias tends to be negative and relatively small. With other parameter values being the same, KVD with $\rho < 0$ has significantly less RMSE and MAD than cases when $\rho > 0$, and performs unequivocally the best among all estimators when endogeneity is strong. However, since this

observed asymmetry is essentially a finite sample problem with KVD, the improved performance when $\rho < 0$ becomes less significant when the sample size increases from 50 to 100. With $\rho < 0$ the overall performance of KVD is very comparable to that of CP, if not slightly better in some cases.

After experimenting with widely different negative and positive values of β and ρ, we found out that the performance of KVD is dependent on the sign of $\beta\rho$, rather than on the sign of ρ. When $\beta\rho > 0$, it performs very unsatisfactorily as documented in Tables 3-12. Kleibergen and Zivot (1998) have recently derived exact analytical expressions for the conditional densities of β given Ω for both the KVD and CP posteriors. They show that the difference between the two is in the Jacobian relating the unrestricted linear multivariate model to the restricted reduced form model. We expect that this additional term may account for the asymmetry in KVD with respect to $\beta\rho$. In our experiments, we found that in finite samples, when $\beta\rho > 0$, the reduced rank restriction using singular value decomposition shifts the marginal posterior for KVD away from the marginal posterior of the linear multivariate model. However, when the sample size gets large, the problem seems to go away.

6 Conclusions

This paper examines the relative merits of some recent developments in the Bayesian and classical analysis of limited information simultaneous equations models in situations where the instruments are very weak. Since the posterior densities and their conditionals in the Bayesian approaches devel-

oped by Chao and Phillips (1998) and Kleibergen and van Dijk (1998) are non-standard, we proposed and implemented a "Gibbs within Metropolis-Hastings" algorithm, which only requires the availability of the conditional densities from the candidate generating density. These conditional densities are used in a Gibbs sampler (GS) to simulate the candidate generating density, whose drawings, after convergence, are then weighted to generate drawings from the target density in a Metropolis-Hastings (M-H) algorithm. We rely on Raftery and Lewis (1992) method to determine the number of burn-ins, and the subsequent number of required iterations in order to ensure convergence. Through a MCMC simulation study, our results provide useful guidelines for empirical practitioners.

The first comforting result is that with reasonably strong instruments (marginal $\overline{R}^2$ in excess of 0.40) all estimators perform equally well in finite samples. In cases with very weak instruments (marginal $\overline{R}^2$ less than 0.10), there is no single estimator that is superior to others in all cases. When endogeneity is weak (ρ less than 0.20), Zellner's MELO does the best. When the endogeneity is relatively strong (ρ in excess of 0.60) and $\rho w_{12} > 0$, BMOM outperforms all other estimators by wide margins. When the endogeneity is strong but $\beta\rho < 0$, KVD approach seems to get very appealing; but, otherwise, its performance is surprisingly very poor. With $\beta\rho > 0$, as the sample size gets larger, the performance of KVD improves rapidly. Fortunately, the Geweke and CP approaches exhibit no such asymmetry and their performances based on bias, RMSE, and MAD are very similar. Based on the medians of marginal posteriors, their performance ranking is consistently a distant second. The record of JIVE is quite disappointing across all our

experiments, and is not recommended in practice. Even though JIVE is slightly less biased than 2SLS in most cases, its standard deviation is considerably higher, particularly in small samples. The most remarkable result in this study is that poor instruments can affect the performance of different estimators differently, depending on the signs and magnitudes of certain key parameters of the model. Given the finding that even in finite samples with very weak instruments BMOM and KVD perform so remarkably well on certain parts of the parameter space, more research is needed to understand the reasons for the asymmetry and find ways to fix the problem.

Appendix

This is to show that $|J(\Phi,(\Pi_2,\beta,\lambda))| \geq |J(\Phi,(\Pi_2,\beta,\lambda))|_{\lambda=0}|$ does not hold. In current notation, we need to show that $|J(\Phi,(\Pi_{21},\theta_2,\beta,\lambda))| \geq |J(\Phi,(\Pi_{21},\theta_2,\beta,\lambda))|_{\lambda=0}|$ fails, where $\theta_2 = \Pi_{22}\Pi_{21}^{-1}$. The fact is that $J(\Phi,(\Pi_{21},\theta_2,\beta,\lambda))|_{\lambda=0}$ and W are not orthogonal, where $W = J(\Phi,(\Pi_{21},\theta_2,\beta,\lambda)) - J(\Phi,(\Pi_{21},\theta_2,\beta,\lambda))|_{\lambda=0}$. Consider a simple case with $m = k_2 = 2$. In this case,

$$\Phi = \begin{pmatrix} 1 \\ \theta_2 \end{pmatrix} \Pi_{21} \begin{pmatrix} \beta & 1 \end{pmatrix} + \begin{pmatrix} -\theta_2 \\ 1 \end{pmatrix} (1+\theta_2^2)^{-\frac{1}{2}} \lambda (1+\beta^2)^{-\frac{1}{2}} \begin{pmatrix} 1 & -\beta \end{pmatrix}.$$

Denote $K = (1+\theta_2^2)^{-\frac{1}{2}}(1+\beta^2)^{-\frac{1}{2}}$. Therefore,

$$J(\Phi,(\Pi_{21},\theta_2,\beta,\lambda))|_{\lambda=0} = \begin{pmatrix} \beta & 0 & \Pi_{21} & K(-\theta_2) \\ \beta\theta_2 & \beta\Pi_{21} & \theta_2\Pi_{21} & K(1) \\ 1 & 0 & 0 & K(\beta\theta_2) \\ \theta_2 & \Pi_{21} & 0 & K(-\beta) \end{pmatrix},$$

$$W = \begin{pmatrix} 0 & \lambda K(1+\theta^2)^{-1}(-1) & \lambda K(1+\beta^2)^{-1}(\beta\theta_2) & 0 \\ 0 & \lambda K(1+\theta^2)^{-1}(-\theta_2) & \lambda K(1+\beta^2)^{-1}(-\beta) & 0 \\ 0 & \lambda K(1+\theta^2)^{-1}(\beta) & \lambda K(1+\beta^2)^{-1}(\theta_2) & 0 \\ 0 & \lambda K(1+\theta^2)^{-1}(\beta\theta_2) & \lambda K(1+\beta^2)^{-1}(-1) & 0 \end{pmatrix}.$$

Easy to check that $(J(\Phi,(\Pi_{21},\theta_2,\beta,\lambda))|_{\lambda=0})W'$ is not a zero matrix but with its third row being 0's. Interestingly,

$$(J(\Phi,(\Pi_{21},\theta_2,\beta,\lambda))|_{\lambda=0})'W = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda K(-\Pi_{21}) & 0 \\ 0 & \lambda K(-\Pi_{21}) & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

References

Angrist, J., G.W. Imbens and A. Krueger (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics* 14, 57-67.

Billingsley, P. (1986). *Probability and Measure* (Wiley, New York).

Blomquist, S. and M. Dahlberg (1999). Small sample properties of LIML and jackknife IV estimators: experiments with weak instruments. *Journal of Applied Econometrics* 14, 69-88.

Bound, J., D.A. Jaeger, and R.M. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90, 443-450.

Brooks, S. and G.O. Roberts (1999). Assessing convergence of Markov chain Monte Carlo algorithms. *Statistics and Computing* (forthcoming).

Brooks, S. (1996). Quantitative convergence diagnosis for MCMC via CUSUMS. Technical report, University of Bristol.

Buse, A. (1992). The bias of instrumental variable estimators. *Econometrica* 60, 173-180.

Casella, G. and E. George (1992). Explaining the Gibbs sampler. *The American Statistician* 46, 167-174.

Chao, J.C. and P.C.B. Phillips (1998). Posterior distributions in limited information analysis of the simultaneous equations model using the Jeffreys prior. *Journal of Econometrics* 87, 49-86.

Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49, 327-335.

Chib, S. and E. Greenberg (1996). Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory* 12, 409-431.

Cowles, M.K. and B.P. Carlin (1996). Markov chain Monte Carlo convergence diagnosis: a comparative review. *Journal of the American Statistical Association* 91, 883-904.

Drèze, J.H. (1976). Bayesian limited information analysis of the simultaneous equation model. *Econometrica* 44, 1045-1075.

Drèze, J.H. and J.A. Morales (1976). Bayesian full information analysis of simultaneous equations. *Journal of American Statistical Association* 71, 329-354.

Drèze, J.H. and J.-F. Richard (1983). Bayesian analysis of simultaneous equation systems. In: Z. Griliches, and M. Intriligator, eds., *Handbook of Econometrics* (North Holland, Amsterdam).

Dwivedi, T.D, and V.K. Srivastava (1984). Exact finite sample properties of double k-class estimators in simultaneous equations. *Journal of Econometrics* 25, 263-283.

Fuller W.A. (1977). Some properties of a modification of the limited information estimator. *Econometrica* 45, 939-953.

Gao, C. and K. Lahiri (1999). Further consequences of viewing LIML as an iterated Aitken estimator. *Journal of Econometrics* (forthcoming).

Gao, C. and K. Lahiri (2000). MCMC algorithms for two recent Bayesian limited information estimators. *Economics Letters* 66, 121-126.

Geweke, J. (1996). Bayesian reduced rank regression in econometrics. *Journal of Econometrics* 75, 121-146.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrica* 57, 97-109.

Kleibergen, F. (1997). Equality restricted random variables: densities and sampling algorithms. *Econometric Institute Report* 9662/A. Erasmus University Rotterdam.

Kleibergen, F. (1998). Conditional densities in econometrics (1998). Discussion paper.

Kleibergen, F. and H.K. van Dijk (1998). Bayesian simultaneous equation analysis using reduced rank structures. *Econometric Theory* 14, 701-743.

Kleibergen, F. and E. Zivot (1998). Bayesian and classical approaches to instrumental variable regression. *Econometric Institute Report* 9835/A. Erasmus University Rotterdam.

MacEachern, S.N. and L.M. Berliner (1994). Subsampling the Gibbs sampler. *The American Statistician* 48, 188-190.

Maddala, G.S. (1976). Weak priors and sharp posteriors in simultaneous equation models. *Econometrica* 44, 345-351.

Maddala, G.S. and J. Jeong (1992). On the exact small sample distribution of the instrumental variable estimator. *Econometrica* 60, 181-183.

Mariano, R.S. and J.B. McDonald (1979). A note on the distribution functions of LIML and 2SLS structural coefficient in exactly identified case. *Journal of the American Statistical Association* 74, 847-848.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21, 1087-1092.

Pagan, A. (1979). Some consequences of viewing LIML as an iterated Aitken estimator, *Economics Letters* 3, 269-372.

Percy, D.F. (1992). Prediction for seemingly unrelated regressions. *Journal of the Royal Statistical Society B* 54, 243-252.

Poirier, D. (1995). *Intermediate Statistics and Econometrics* (M.I.T. Press, Cambridge, Mass.).

Poirier, D. (1996). Prior beliefs about fit. In: J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds., *Bayesian Statistics 5* (Clarendon Press, Oxford).

Raftery, A.E. and S.M. Lewis (1992). How many iterations in the Gibbs sampler? In: J.M. Bernardo, A.F.M. Smith, A.P. Dawid and J.O. Berger, eds., *Bayesian Statistics 4* (Oxford University Press).

Staiger, D. and J.H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica* 65, 557-586.

Smith, A.F.M. and G.O. Roberts (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B* 55, 3-23.

Tsurumi, H. (1990). Comparing Bayesian and non-Bayesian limited information estimators. In: Geisser, S., Hodges, J.S., Press, S.J., Zellner, A., eds., *Bayesian and Likelihood Methods in Statistics and Econometrics* (North-Holland, Amsterdam).

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics* 22, 1701-1767.

Yu, B. and P. Mykland (1994). Looking at Markov samplers through Cusum path plots: a simple diagnostic idea. Technical Report 413, University of California at Berkeley, Dept. of Statistics.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics* (Wiley, New York).

Zellner, A. (1978). Estimation of functions of population means and regression coefficients: a minimum expected loss (MELO) approach. *Journal of Econometrics* 8, 127-158.

Zellner, A. (1986). Further results on Bayesian minimum expected loss (MELO) estimates and posterior distributions for structural coefficients. In: Slottje, D., eds., *Advances in Econometrics*, Vol. 5, pp. 171-182.

Zellner, A. (1994). Bayesian and Non-Bayesian estimation using balanced loss functions. In: Gupta, S.S., Berger, J.O., eds., *Statistical Decision Theory and Related Topics*, Vol. V. Springer, New York, Chapter 28, pp. 377-390.

Zellner, A. (1998). The finite sample properties of simultaneous equations' estimates and estimators: Bayesian and non-Bayesian approaches. *Journal of Econometrics* 83, 185-212.

Zellner, A., L. Bauwens, and H.K. van Dijk (1988). Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo methods. *Journal of Econometrics* 38, 39-72.

Table 1. T = 50, $\rho = 0.60$, $k_2 = 4$, $R^2 = 0.40$

	Mean	Std	RMSE	MAD
OLS	1.348	0.089	0.359	0.348
2SLS	1.045	0.144	0.151	0.121
MELO	1.115	0.126	0.171	0.144
BMOM	0.967	0.127	0.131	0.102
LIML	0.998	0.152	0.152	0.118
Fuller1	1.015	0.147	0.148	0.116
Fuller4	1.061	0.136	0.149	0.120
JIVE	0.957	0.178	0.183	0.141
Geweke_Mode	1.056	0.140	0.151	0.122
Geweke_Median	1.031	0.143	0.146	0.116
LIML_GS_Mode	1.061	0.139	0.152	0.123
LIML_GS_Median	1.036	0.142	0.146	0.116
CP_Mode	1.046	0.144	0.151	0.121
CP_Median	1.021	0.145	0.147	0.115
KVD_Mode	1.090	0.148	0.173	0.143
KVD_Median	1.079	0.137	0.158	0.130

Notes. Number of replications: 400
Geweke: nburn = 100, n = 2000
CP: nburn_GS = 100, nburn_MH = 100, n = 5000, acceptance rate = 0.482 (0.015)
KVD: nburn_GS = 100, nburn_MH = 100, n = 4000, acceptance rate = 0.215 (0.136)

Table 2. T = 50, $\rho = 0.60$, $k_2 = 1$, $R^2 = 0.10$

	Mean	Std	RMSE	MAD
OLS	1.537	0.111	0.548	0.537
2SLS	1.030	0.345	0.346	0.267
MELO	1.173	0.262	0.314	0.248
BMOM	0.881	0.264	0.290	0.229
LIML	1.030	0.345	0.346	0.267
Fuller1	1.107	0.300	0.319	0.245
Fuller4	1.250	0.219	0.332	0.277
JIVE	0.803	0.491	0.529	0.409
Geweke_Mode	1.089	0.331	0.343	0.265
Geweke_Median	0.907	0.518	0.526	0.358
LIML_GS_Mode	1.091	0.313	0.326	0.255
LIML_GS_Median	0.778	1.386	1.404	0.592
CP_Mode	1.108	0.309	0.327	0.256
CP_Median	0.797	1.383	1.398	0.580
KVD_Mode	n.a.	n.a.	n.a.	n.a.
KVD_Median	n.a.	n.a.	n.a.	n.a.

Notes. Number of replications: 400
Geweke: nburn = 100, n = 3000
CP: nburn_GS = 200, nburn_MH = 200, n = 10000, acceptance rate = 0.551 (0.023)

Table 3. T = 50, $\rho = 0.60$, $k_2 = 4$, $R^2 = 0.10$

	Mean	Std	RMSE	MAD
OLS	1.539	0.111	0.550	0.539
2SLS	1.231	0.279	0.362	0.296
MELO	1.366	0.186	0.411	0.368
BMOM	0.943	0.184	0.193	0.154
LIML	1.043	0.579	0.581	0.386
Fuller1	1.143	0.367	0.394	0.307
Fuller4	1.281	0.244	0.372	0.307
JIVE	0.816	0.568	0.597	0.474
Geweke_Mode	1.244	0.287	0.377	0.309
Geweke_Median	1.204	0.309	0.370	0.300
LIML_GS_Mode	1.260	0.268	0.373	0.308
LIML_GS_Median	1.220	0.298	0.370	0.300
CP_Mode	1.230	0.293	0.372	0.301
CP_Median	1.194	0.315	0.370	0.298
KVD_Mode	1.351	0.384	0.520	0.389
KVD_Median	1.381	0.367	0.529	0.405

Notes. Number of replications: 400
Geweke: nburn = 100, n = 2000
CP: nburn_GS = 100, nburn_MH = 100, n = 10000, acceptance rate = 0.475 (0.010)
KVD: nburn_GS = 100, nburn_MH = 100, n = 3000, acceptance rate = 0.400 (0.217)

Table 4. T = 50, $\rho = 0.60$, $k_2 = 9$, $R^2 = 0.10$

	Mean	Std	RMSE	MAD
OLS	1.535	0.111	0.546	0.535
2SLS	1.363	0.221	0.425	0.371
MELO	1.463	0.139	0.483	0.463
BMOM	0.969	0.132	0.136	0.106
LIML	1.090	0.864	0.869	0.534
Fuller1	1.182	0.479	0.512	0.366
Fuller4	1.302	0.291	0.419	0.333
JIVE	0.706	0.933	0.978	0.728
Geweke_Mode	1.357	0.239	0.430	0.367
Geweke_Median	1.350	0.245	0.427	0.361
LIML_GS_Mode	1.375	0.218	0.328	0.380
LIML_GS_Median	1.367	0.228	0.432	0.374
CP_Mode	1.215	0.629	0.665	0.466
CP_Median	1.255	0.388	0.464	0.346
KVD_Mode	1.550	0.376	0.666	0.556
KVD_Median	1.573	0.322	0.657	0.576

Notes. Number of replications: 400
Geweke: nburn = 100, n = 1000
CP: nburn_GS = 200, nburn_MH = 200, n = 10000, acceptance rate = 0.242 (0.040)
KVD: nburn_GS = 200, nburn_MH = 100, n = 10000, acceptance rate = 0.267 (0.188)

Table 5. T = 100, $\rho = 0.60$, $k_2 = 4$, $R^2 = 0.10$

	Mean	Std	RMSE	MAD
OLS	1.538	0.077	0.543	0.538
2SLS	1.138	0.208	0.250	0.200
MELO	1.257	0.156	0.301	0.264
BMOM	0.954	0.156	0.163	0.127
LIML	1.023	0.280	0.281	0.210
Fuller1	1.069	0.250	0.259	0.197
Fuller4	1.171	0.195	0.259	0.209
JIVE	0.914	0.320	0.331	0.262
Geweke_Mode	1.149	0.215	0.262	0.208
Geweke_Median	1.111	0.228	0.254	0.198
LIML_GS_Mode	1.162	0.205	0.261	0.209
LIML_GS_Median	1.117	0.225	0.254	0.199
CP_Mode	1.155	0.207	0.259	0.206
CP_Median	1.107	0.228	0.252	0.196
KVD_Mode	1.233	0.205	0.310	0.258
KVD_Median	1.215	0.210	0.301	0.243

Notes. Number of replications: 400
Geweke: nburn = 100, n = 2000
CP: nburn_GS = 200, nburn_MH = 200, n = 10000, acceptance rate = 0.616 (0.008)
KVD: nburn_GS = 200, nburn_MH = 100, n = 10000, acceptance rate = 0.312 (0.175)

Table 6. T = 100, $\rho = 0.60$, $k_2 = 9$, $R^2 = 0.10$

	Mean	Std	RMSE	MAD
OLS	1.542	0.078	0.548	0.542
2SLS	1.258	0.197	0.325	0.274
MELO	1.376	0.134	0.399	0.376
BMOM	0.972	0.132	0.135	0.110
LIML	1.003	0.437	0.437	0.291
Fuller1	1.071	0.311	0.319	0.243
Fuller4	1.180	0.233	0.294	0.232
JIVE	0.927	0.408	0.414	0.333
Geweke_Mode	1.253	0.201	0.323	0.269
Geweke_Median	1.238	0.206	0.315	0.261
LIML_GS_Mode	1.265	0.196	0.330	0.278
LIML_GS_Median	1.247	0.202	0.319	0.266
CP_Mode	1.196	0.264	0.329	0.266
CP_Median	1.192	0.232	0.301	0.240
KVD_Mode	1.371	0.278	0.464	0.382
KVD_Median	1.395	0.269	0.478	0.397

Notes. Number of replications: 400
Geweke: nburn = 100, n = 1000
CP: nburn_GS = 200, nburn_MH = 200, n = 6000, acceptance rate = 0.434 (0.029)
KVD: nburn_GS = 200, nburn_MH = 200, n = 10000, acceptance rate = 0.210 (0.179)

Table 7. T = 100, $\rho = 0.60$, $k_2 = 4$, $R^2 = 0.05$

	Mean	Std	RMSE	MAD
OLS	1.565	0.080	0.571	0.565
2SLS	1.254	0.282	0.380	0.309
MELO	1.376	0.184	0.419	0.379
BMOM	0.953	0.183	0.189	0.150
LIML	1.052	0.584	0.586	0.392
Fuller1	1.158	0.377	0.409	0.307
Fuller4	1.296	0.244	0.384	0.317
JIVE	0.833	0.638	0.659	0.527
Geweke_Mode	1.264	0.285	0.388	0.314
Geweke_Median	1.224	0.316	0.387	0.305
LIML_GS_Mode	1.274	0.283	0.394	0.320
LIML_GS_Median	1.232	0.310	0.387	0.306
CP_Mode	1.263	0.295	0.395	0.318
CP_Median	1.223	0.316	0.387	0.304
KVD_Mode	1.388	0.389	0.549	0.418
KVD_Median	1.394	0.315	0.504	0.414

Notes. Number of replications: 400
Geweke: nburn = 100, n = 2000
CP: nburn_GS = 100, nburn_MH = 100, n = 4000, acceptance rate = 0.611 (0.009)
KVD: nburn_GS = 200, nburn_MH = 200, n = 8000, acceptance rate = 0.442 (0.224)

Table 8. T = 100, $\rho = 0.60$, $k_2 = 9$, $R^2 = 0.05$

	Mean	Std	RMSE	MAD
OLS	1.574	0.076	0.579	0.574
2SLS	1.386	0.219	0.444	0.394
MELO	1.478	0.131	0.496	0.478
BMOM	0.979	0.129	0.131	0.105
LIML	1.139	0.882	0.893	0.545
Fuller1	1.224	0.477	0.527	0.389
Fuller4	1.335	0.280	0.437	0.358
JIVE	0.844	0.823	0.838	0.663
Geweke_Mode	1.385	0.243	0.455	0.395
Geweke_Median	1.380	0.246	0.453	0.390
LIML_GS_Mode	1.397	0.230	0.459	0.404
LIML_GS_Median	1.387	0.236	0.453	0.396
CP_Mode	1.338	0.465	0.575	0.433
CP_Median	1.337	0.311	0.459	0.376
KVD_Mode	1.584	0.462	0.745	0.592
KVD_Median	1.608	0.368	0.711	0.610

Notes. Number of replications: 400
Geweke: nburn = 100, n = 2000
CP: nburn_GS = 200, nburn_MH = 200, n = 10000, acceptance rate = 0.433 (0.035)
KVD: nburn_GS = 200, nburn_MH = 200, n = 10000, acceptance rate = 0.371 (0.221)

Table 9. T = 100, $\rho = 0.20$, $k_2 = 4$, $R^2 = 0.10$

	Mean	Std	RMSE	MAD
OLS	1.172	0.090	0.194	0.174
2SLS	1.046	0.253	0.257	0.206
MELO	1.083	0.189	0.206	0.164
BMOM	0.859	0.190	0.237	0.195
LIML	1.017	0.333	0.333	0.260
Fuller1	1.029	0.298	0.299	0.236
Fuller4	1.059	0.235	0.242	0.192
JIVE	0.957	0.417	0.419	0.340
Geweke_Mode	1.053	0.251	0.257	0.200
Geweke_Median	1.041	0.267	0.270	0.214
LIML_GS_Mode	1.058	0.244	0.251	0.197
LIML_GS_Median	1.044	0.265	0.269	0.212
CP_Mode	1.054	0.255	0.261	0.205
CP_Median	1.040	0.271	0.274	0.218
KVD_Mode	1.131	0.368	0.391	0.237
KVD_Median	1.161	0.328	0.365	0.245

Notes. Number of replications: 400
Geweke: nburn = 100, n = 1000
CP: nburn_GS = 100, nburn_MH = 100, n = 5000, acceptance rate = 0.615 (0.011)
KVD: nburn_GS = 100, nburn_MH = 100, n = 1000, acceptance rate = 0.548 (0.200)

Table 10. T = 100, $\rho = 0.20$, $k_2 = 9$, $R^2 = 0.10$

	Mean	Std	RMSE	MAD
OLS	1.179	0.096	0.203	0.181
2SLS	1.085	0.214	0.230	0.182
MELO	1.124	0.146	0.192	0.154
BMOM	0.823	0.143	0.228	0.193
LIML	0.992	0.397	0.397	0.301
Fuller1	1.015	0.347	0.347	0.270
Fuller4	1.055	0.267	0.273	0.216
JIVE	0.991	0.481	0.481	0.390
Geweke_Mode	1.084	0.218	0.234	0.184
Geweke_Median	1.079	0.223	0.237	0.187
LIML_GS_Mode	1.087	0.212	0.229	0.181
LIML_GS_Median	1.082	0.218	0.233	0.185
CP_Mode	1.054	0.308	0.313	0.223
CP_Median	1.063	0.254	0.262	0.207
KVD_Mode	1.249	0.234	0.342	0.283
KVD_Median	1.286	0.235	0.370	0.308

Notes. Number of replications: 400
Geweke: nburn = 100, n = 1000
CP: nburn_GS = 100, nburn_MH = 200, n = 5000, acceptance rate = 0.456 (0.023)
KVD: nburn_GS = 100, nburn_MH = 100, n = 5000, acceptance rate = 0.413 (0.202)

Table 11. T = 50, $\rho = 0.95$, $k_2 = 4$, $R^2 = 0.10$

	Mean	Std	RMSE	MAD
OLS	1.846	0.052	0.848	0.846
2SLS	1.359	0.180	0.402	0.363
MELO	1.572	0.118	0.584	0.572
BMOM	1.057	0.118	0.131	0.102
LIML	0.988	0.404	0.404	0.255
Fuller1	1.169	0.196	0.259	0.221
Fuller4	1.417	0.120	0.434	0.417
JIVE	0.637	0.611	0.711	0.478
Geweke_Mode	1.347	0.302	0.460	0.358
Geweke_Median	1.277	0.377	0.468	0.305
LIML_GS_Mode	1.338	0.155	0.372	0.345
LIML_GS_Median	1.252	0.194	0.318	0.281
CP_Mode	1.314	0.162	0.353	0.325
CP_Median	1.234	0.194	0.304	0.266
KVD_Mode	1.411	0.379	0.559	0.428
KVD_Median	1.462	0.463	0.654	0.514

Notes. Number of replications: 400
Geweke: nburn = 100, n = 3000
CP: nburn_GS = 200, nburn_MH = 200, n = 10000, acceptance rate = 0.476 (0.010)
KVD: nburn_GS = 200, nburn_MH = 200, n = 10000, acceptance rate = 0.036 (0.038)

Table 12. T = 100, $\rho = 0.95$, $k_2 = 4$, $R^2 = 0.10$

	Mean	Std	RMSE	MAD
OLS	1.850	0.033	0.851	0.850
2SLS	1.230	0.126	0.262	0.234
MELO	1.414	0.094	0.425	0.414
BMOM	1.044	0.095	0.105	0.082
LIML	1.025	0.170	0.172	0.132
Fuller1	1.095	0.142	0.171	0.143
Fuller4	1.264	0.099	0.282	0.265
JIVE	0.873	0.199	0.236	0.191
Geweke_Mode	1.216	0.117	0.246	0.223
Geweke_Median	1.150	0.127	0.197	0.172
LIML_GS_Mode	1.227	0.118	0.256	0.235
LIML_GS_Median	1.158	0.128	0.203	0.180
CP_Mode	1.221	0.116	0.250	0.228
CP_Median	1.154	0.127	0.200	0.176
KVD_Mode	1.258	0.207	0.331	0.280
KVD_Median	1.252	0.294	0.387	0.260

Notes. Number of replications: 400
Geweke: nburn = 100, n = 3000
CP: nburn_GS = 200, nburn_MH = 200, n = 10000, acceptance rate = 0.626 (0.007)
KVD: nburn_GS = 200, nburn_MH = 200, n = 10000, acceptance rate = 0.022 (0.022)

Table 13. Performance of BMOM and KVD when $\rho < 0$

	Mean	Std	RMSE	MAD	Remarks
$T = 50, \rho = -0.60, k_2 = 4, R^2 = 0.40$					
BMOM	0.852	0.129	0.196	0.165	Compare Table 1.
KVD_Mode	0.971	0.150	0.152	0.119	Acceptance rate for
KVD_Median	0.999	0.153	0.152	0.119	KVD: 0.713 (0.130)
$T = 50, \rho = -0.60, k_2 = 4, R^2 = 0.10$					
BMOM	0.551	0.191	0.488	0.453	Compare Table 3.
KVD_Mode	0.851	0.327	0.359	0.271	Acceptance rate for
KVD_Median	0.934	0.341	0.347	0.267	KVD: 0.680 (0.133)
$T = 50, \rho = -0.60, k_2 = 9, R^2 = 0.10$					
BMOM	0.420	0.136	0.600	0.580	Compare Table 4.
KVD_Mode	0.857	0.367	0.393	0.296	Acceptance rate for
KVD_Median	0.927	0.399	0.406	0.291	KVD: 0.482 (0.155)
$T = 100, \rho = -0.60, k_2 = 4, R^2 = 0.10$					
BMOM	0.676	0.160	0.362	0.326	Compare Table 5.
KVD_Mode	0.901	0.213	0.235	0.186	Acceptance rate for
KVD_Median	0.964	0.237	0.239	0.190	KVD: 0.772 (0.110)
$T = 100, \rho = -0.60, k_2 = 9, R^2 = 0.10$					
BMOM	0.531	0.129	0.486	0.469	Compare Table 6.
KVD_Mode	0.903	0.240	0.258	0.200	Acceptance rate for
KVD_Median	0.952	0.247	0.252	0.198	KVD: 0.614 (0.138)
$T = 100, \rho = -0.60, k_2 = 4, R^2 = 0.05$					
BMOM	0.514	0.181	0.519	0.486	Compare Table 7.
KVD_Mode	0.813	0.306	0.358	0.285	Acceptance rate for
KVD_Median	0.908	0.362	0.373	0.287	KVD: 0.720 (0.128)
$T = 100, \rho = -0.60, k_2 = 9, R^2 = 0.05$					
BMOM	0.407	0.131	0.608	0.593	Compare Table 8.
KVD_Mode	0.848	0.424	0.450	0.312	Acceptance rate for
KVD_Median	0.907	0.349	0.361	0.275	KVD: 0.585 (0.144)
$T = 100, \rho = -0.20, k_2 = 4, R^2 = 0.10$					
BMOM	0.753	0.195	0.314	0.266	Compare Table 9.
KVD_Mode	1.002	0.267	0.267	0.208	Acceptance rate for
KVD_Median	1.037	0.291	0.293	0.218	KVD: 0.699 (0.162)
$T = 100, \rho = -0.20, k_2 = 9, R^2 = 0.10$					
BMOM	0.673	0.159	0.364	0.328	Compare Table 10.
KVD_Mode	1.093	0.318	0.331	0.233	Acceptance rate for
KVD_Median	1.129	0.279	0.307	0.241	KVD: 0.553 (0.181)
$T = 50, \rho = -0.95, k_2 = 4, R^2 = 0.10$					
BMOM	0.427	0.120	0.585	0.573	Compare Table 11.
KVD_Mode	0.737	0.244	0.359	0.312	Acceptance rate for
KVD_Median	0.836	0.246	0.295	0.239	KVD: 0.173 (0.112)
$T = 100, \rho = -0.95, k_2 = 4, R^2 = 0.10$					
BMOM	0.589	0.097	0.422	0.411	Compare Table 12.
KVD_Mode	0.815	0.155	0.241	0.209	Acceptance rate for
KVD_Median	0.889	0.153	0.189	0.156	KVD: 0.179 (0.103)

Note. Number of replications: 500