No.448

A Bayesian Test of Independence of Stochastic Regressors in a Simultaneous Equation Model with Autoregressive Residuals

by

Shiba Tsunemasa & Hiroki Turumi

November 1990

A Bayesian Test of Independence of Stochastic Regressors in a Simultaneous Equation Model with Autoregressive Residuals

Tsunemasa Shiba
University of Tsukuba
and
Hiroki Tsurumi
Rutgers University

October 1990

1. Introduction

Several procedures are available for the estimation of linear simultaneous equation models with autoregressive errors [Sargan (1959, 1961, 1973), Amemiya (1966), Fair (1970, 1972), Dhrymes, Berner and Cummins (1974), Hatanaka (1976), and Hendry (1971, 1976), Bowden and Turkington (1984) among others]. Since these procedures are based on the assumptions that (1) stochastic regressors are correlated with the error terms and that (2) the errors are serially correlated, for most empirical work there is a need to test these assumptions. Godfrey (1976, 1988) discusses testing for serial correlation in simultaneous equation models. Sargan (1973) discusses misspecification tests in the context of instrumental variable procedures.

In this paper we propose a Bayesian test of highest posterior density credible set (HPDCS) test of independence of stochastic regressors in the presence of the serially correlated errors. We can easily test independence of stochastic regressors *and* serially independent errors, but for empirical work, the basic concern may be whether or not one can use a familiar single equation method such as the Cochrane-Orcutt procedure in the presence of the serially correlated error. Accordingly, we shall focus our attention on testing whether or not the equation of interest can be treated as the classical regression with the serially correlated errors. If one is interested in testing any other hypothesis, our test statistics can be modified to serve the purpose. For simplicity, we assume a vector autoregressive process of order 1, VAR(1), for the errors, but the test statistics can be extended to the VAR(p) error processes, assuming that p is known.

The organization of the paper is as follows. In section 2 we derive test statistics and in section 3 we present results of sampling experiments, and we apply the test statistics to Klein's model I. Concluding remarks are given in section 4.

II. Test Statistics

Let the simultaneous equations model be given by

$$Y\Gamma + XB = U \tag{1}$$
$$U = U_{-1}R + E \tag{2}$$

where equation (1) describes the m structural equations and equation (2) specifies the VAR(1) process for the errors. The notations are defined in Appendix A. Let us express the structural equation of interest as

$$y_1 = Y_1\gamma_1 + X_1\beta_1 + u_1 \tag{3}$$

where y_1 is an $(n \times 1)$ vector of observations on the dependent variable; Y_1 is an $(n \times m_1)$ matrix of observations on stochastic regressors; X_1 is a $(n \times k_1)$ matrix of exogenous variables included in the equation; u_1 is the $(n \times 1)$ vector of the structural error terms; γ_1 is a $(m_1 \times 1)$ vector and β_1 is a $(k_1 \times 1)$ vector of structural regression coefficients.

As shown in Appendix A from equations (1) and (2) we may derive

$$\begin{aligned} y_1 - r_{11}y_{1,-1} &= (Y_1 - r_{11}Y_{1,-1})\gamma_1 + (X_1 - r_{11}X_{1,-1})\beta_1 \\ &\quad + Y_{-1}\Gamma_2 R_{21} + X_{-1}B_2R_{21} + \epsilon_1 \end{aligned} \tag{4}$$

Let Y_{t1} be the t-th row of Y_1, and ϵ_{t1} be the t-th element of ϵ_1. From equation (4) it is clear if

$$\mathrm{Cov}(Y_{t1}, \epsilon_{t1}) = 0 \quad \text{and} \quad R_{21} = 0$$

then equation (4) can be treated as a single equation classical regression model with the AR(1) error. Hence, the hypothesis of interest for most empirical work may be put as

$$\mathrm{H}_1 : \begin{bmatrix} \mathrm{Cov}(Y_{t1}, \epsilon_{t1}) \\ R_{21} \end{bmatrix} = 0 \quad \text{versus} \quad \mathrm{K}_1 : \begin{bmatrix} \mathrm{Cov}(Y_{t1}, \epsilon_{t1}) \\ R_{21} \end{bmatrix} \neq 0 \tag{5}$$

If we are certain that the stochastic regressors Y_{t1} are correlated with ϵ_{t1}, then we may test

$$\mathrm{H}_2 : \begin{bmatrix} \mathrm{Cov}(Y_{t1}, \epsilon_{t1}) \neq 0 \\ R_{21} = 0 \end{bmatrix} \quad \text{versus} \quad \mathrm{K}_2 : \begin{bmatrix} \mathrm{Cov}(Y_{t1}, \epsilon_{t1}) \neq 0 \\ R_{21} \neq 0 \end{bmatrix} \tag{6}$$

Assuming that $\mathrm{Cov}(Y_{t1}, \epsilon_{t1}) \neq 0$ and $R_{21} = 0$, Sargan (1961) and Amemiya (1966) derived a limited information maximum likelihood estimator and modified Sargan's two-stage least squares estimator, respectively. For some empirical work one may be interested in testing H_2.

In the following, we shall focus our attention on deriving test statistics to test H_1 versus K_1. The test statistics to test H_2 versus K_2 can be similarly derived. To test the hypothesis H_1 versus K_1, let us focus on the system

$$\begin{aligned} y_1 - r_{11} y_{1,-1} = (Y_1 - r_{11} Y_{1,-1})\gamma_1 + (X_1 - r_{11} X_{1,-1})\beta_1 \\ + Y_{-1}\Gamma_2 R_{21} + X_{-1} B_2 R_{21} + \epsilon_1 \\ = (Y_1 - r_{11} Y_{1,-1})\gamma_1 + (X_1 - r_{11} X_{1,-1})\beta_1 + U_{2,-1} R_{21} + \epsilon_1 \quad (7) \\ Y_1 = X\Pi_2 + Y_{-1}\Upsilon_2 + X_{-1}\Phi_2 + V_1 \quad (8) \end{aligned}$$

where $U_{2,-1} = Y_{-1}\Gamma_2 + X_{-1}B_2$. From equations (7) and (8) we derive in Appendix A the highest posterior density credible set (HPDCS) [or highest posterior density region (HPDR)] test:

$$\mathrm{F}_{m-1+m_1, n-2m_1-k_1-m+1} = \frac{(SSR_c - SSR_u)/(m - 1 + m_1)}{SSR_u/(n - 2m_1 - k_1 - m + 1)} \tag{9}$$

where SSR_u is the unconstrained sum of squared residuals under K_1 and SSR_c is the constrained sum of squared residuals under H_1.

The test statistic (9) is conditioned on $r_{11}, \Pi_2, \Upsilon_2, \Phi_2, \Gamma_2$ and B_2 and we need to estimate them. The estimates of Π_2, Υ_2, and Φ_2 are obtained from equation (8) as the posterior means or the maxium likelihood estimators (MLE's). There are two ways to estimate Γ_2 and B_2, one under the null and the other under the alternative hypothesis.

The first way of obtaining estimates of Γ_2 and B_2 is to estimate structural equations in the system under the null hypothesis. From an equation just below equation (26) of Appendix A, we have

$$\begin{aligned} y_1 - r_{11} y_{1,-1} &= (Y_1 - r_{11} Y_{1,-1})\gamma_1 + (X_1 - r_{11} X_{1,-1})\beta_1 \\ &\quad + Y_{-1}\Gamma_2 R_{21} + X_{-1} B_2 R_{21} + V_1 \delta + \epsilon_1' \end{aligned} \tag{10}$$

and under the null hypothesis H_1, equation (10) becomes

$$y_1 - r_{11}y_{1,-1} = (Y_1 - r_{11}Y_{1,-1})\gamma_1 + (X_1 - r_{11}X_{1,-1})\beta_1 + \epsilon_1' \qquad (11)$$

We obtain estimates of γ_1 and β_1, $\hat{\gamma}_1$ and $\hat{\beta}_1$. The SSR_c is the sum of squared residuals from equation (11). As for the SSR_u, we use the estimates $\hat{\gamma}_i$ and $\hat{\beta}_i$ from equation (11) for the i-th structural equation and form $\widehat{\Gamma}_2$ and $\widehat{B}_2$ and obtain the SSR_u from

$$\begin{aligned} y_1 - r_{11}y_{1,-1} &= (Y_1 - r_{11}Y_{1,-1})\gamma_1 + (X_1 - r_{11}X_{1,-1})\beta_1 \\ &\quad + \widehat{U}_{2,-1}R_{21} + \widehat{V}_1\delta + \epsilon_1' \end{aligned} \qquad (12)$$

where $\widehat{U}_{2,-1} = Y_{-1}\widehat{\Gamma}_2 + X_{-1}\widehat{B}_2$ and $\widehat{V}_1 = Y_1 - X\widehat{\Pi}_2 - Y_{-1}\widehat{\Upsilon}_2 - X_{-1}\widehat{\Phi}_2$. Let us call the F test (9) using this procedure as the FTH.

The second way of obtaining Γ_2 and B_2 is to use equation (10) by replacing V_1 by $\widehat{V}_1$ and estimate γ_1 and β_1, $\tilde{\gamma}_1$ and $\tilde{\beta}_1$. From the estimates, $\tilde{\gamma}_i$ and $\tilde{\beta}_i$ fo the i-th structural equation , we make $\tilde{\Gamma}_2$ and $\tilde{B}_2$ and obtain the SSR_u from

$$\begin{aligned} y_1 - r_{11}y_{1,-1} &= (Y_1 - r_{11}Y_{1,-1})\gamma_1 + (X_1 - r_{11}X_{1,-1})\beta_1 \\ &\quad \tilde{U}_{2,-1}R_{21} + \widehat{V}_1\delta + \epsilon_1' \end{aligned} \qquad (13)$$

where $\tilde{U}_{2,-1} = Y_{-1}\tilde{\Gamma}_2 + X_{-1}\tilde{B}_2$. Let us call the F test (9) using this procedure as the FTK.

If the structural errors are not autocorrelated, *i.e.* $R = 0$, then the FTH and FTK are identical and they reduce to the Wu-Hausman test statistic [Wu (1973) and Hausman (1978), and Nakamura and Nakamura (1981)]. The F statistics, FTH and FTK, are asympotically distributed as $\chi^2_{m-1+m_1}/(m-1+m_1)$ under the null hypothesis, H_1.

III. Sampling Experiments

As for the designs of the sampling experiments, we modify the model in Tsurumi (1990). for the AR(1) errors. The design of experiments are explained in Appendix B in detail. The model consists of three structural equations and we focus on the first equation which has one stochastic regressor ($m_1 = 1$) and four exogenous variables ($k_1 = 4$):

$$y_{t1} = \gamma_{12}y_{t2} + \beta_{11}x_{t1} + \beta_{13}x_{t3} + \beta_{15}x_{t5} + \beta_{17}x_{t7} + u_{t1}$$
$$\gamma_{12} = .222,\ \beta_{11} = 6.2,\ \beta_{13} = .7,\ \beta_{15} = .96\ \beta_{17} = .06\ .$$

As shown in Appendix B the performances of the F test statistics depend, among others, on R^2 (the coefficient of determination of the reduced form for y_{t2}) and on multicollinearity among the exogenous variables. Hence we control sampling experiments for R^2 as well as for multicollinearity. As for the matrix of the autoregressive coefficients, r_{ij} of R in equation (2), we use four cases which are given in Table 1.

Table 1: The Matrices of the Autoregressive Coefficients R Used in the Sampling Experiments

$$R1 = \begin{bmatrix} .8 & 0 & 0 \\ 0 & .8 & 0 \\ 0 & 0 & .4 \end{bmatrix}, \quad R2 = \begin{bmatrix} .2 & 0 & 0 \\ 0 & .8 & 0 \\ 0 & 0 & .4 \end{bmatrix}, \quad R3 = \begin{bmatrix} .2 & 0 & 0 \\ 0 & .2 & 0 \\ 0 & 0 & .4 \end{bmatrix}$$

$$R4 = \begin{bmatrix} .2 & .2 & 0 \\ 0 & .2 & 0 \\ 0 & 0 & .4 \end{bmatrix}$$

$R1$, $R2$ and $R3$ are all diagonal matrices. $R4$ has one nonzero off-diagonal element, $r_{12} = 0.2$.

The results of the sampling experiments are given in Tables 2–6. For Tables 2–6, we made sampling experiments for (1) low R^2 ($R^2 = .3$) and high R^2 ($R^2 = .9$); (2) sample sizes of $n = 40$, and $n = 100$, and (3) with and without multicollinearity. The values of ρ^2 (the squared correlation coefficient between y_{t2} and ϵ_{t1}) are set at 0, .3, .5, .7, and .9. The figures in parentheses are the empirical powers of the tests that are adjusted so that the empirical sizes of the tests become the 5% significance level when the null hypothesis, H_1 is true (*i.e.* $\rho^2 = 0$ and $R_{21} = 0$). For each combination of R^2, ρ^2, and n, the number of replications is 1000. In all cases we estimated r_{11} by the grid method to obtain the value of r_{11} that minimizes the sum of squared residuals. Table 6 present the sampling results for $n = 500$, $R^2 = .3$, with multicollinearity and for the matrices of autocorrelation coefficients of R1 and R2.

The results of the sampling experiments may be summarized as follows:

(1) The sizes of the tests are larger for FTK than for FTH. In most cases the sizes of the tests are reasonably close to the nominal level of 5%.

(2) Given the R matrix and R^2, multicollinearity reduces the powers of the tests.

(3) Given the R matrix and sample size, the higher is R^2, the larger are the powers of the tests.

(4) As the sample size increases from $n = 40$ to $n = 100$, the powers of the tests increase.

(5) The cases of $r_{11} = r_{22}$ yield low powers of the tests than the cases of $r_{11} \neq r_{22}$, where r_{ii} is the ith diagonal element of R. For the cases of the low values of r_{ii} ($r_{11} = r_{22} = .2$), the powers of the tests are quite low when multicollinearity exists, and R^2 is low. However, if the 1-2 element of R, r_{12}, is not zero (*i.e.* $R4$ matrix) the powers are better than those for the $R3$ matrix.

(6) From Table 6 we see that as the sample size increases to $n = 500$ the powers of the tests increase even for the case of multicollinearity with $R^2 = .3$, and $r_{11} = r_{22}$.

In summary the sampling experiments show that both the FTH and FTK have sizes of the tests close to the nominal size of 5% and that the powers of the tests are sensitive especialy to the presence of multicollinearity. The sensitivity of the powers of the tests to multicollinearity, R^2, and sample size can be explained by the noncentrality parameter of the F statistics. Conditionally on $S = (\widehat{U}_{2,-1}, \widehat{V}_1)$ and on r_{11}, the F statistics (FTH and FTK) have the noncentrality parameter under the alternative hypothesis:

$$\eta = \xi' S' M_E S \xi / \sigma_{11 \cdot 2}$$

where $\xi = (R_{21}', \delta_1)'$, $E = (Y_1 - r_{11} Y_{1,-1}, X_1 - r_{11} X_{1,-1})$, and $M_E = I - E(E'E)^{-1}E'$. The noncentrality parameter, η, is proportionately related to the sample size, R^2, ρ^2, and multicollinearity. Table 7 gives how η is affected by n, R^2, ρ^2, and multicollinearity.

Table 2: Empirical Sizes and Powers of the FTH and FTK; R1 Matrix

		Without Multicollinearity							
R^2	ρ^2	$n = 40$				$n = 100$			
		FTH		FTK		FTH		FTK	
0.3	0.0	5.5	(5.0)	7.5	(5.0)	4.8	(5.0)	5.9	(5.0)
	0.3	24.8	(24.3)	28.6	(22.6)	91.8	(92.2)	92.5	(90.5)
	0.5	45.2	(44.2)	48.6	(43.8)	99.3	(99.3)	99.3	(99.2)
	0.7	65.1	(64.5)	68.8	(63.4)	100	(100)	100	(100)
	0.9	81.9	(81.4)	84.5	(82.2)	100	(100)	100	(100)
0.9	0.0	5.2	(5.0)	7.5	(5.0)	5.7	(5.0)	5.9	(5.0)
	0.3	62.1	(61.1)	68.0	(60.5)	99.8	(99.7)	99.8	(99.6)
	0.5	92.0	(91.3)	94.0	(90.0)	100	(100)	100	(100)
	0.7	99.1	(99.1)	99.2	(98.8)	100	(100)	100	(100)
	0.9	100	(100)	100	(100)	100	(100)	100	(100)
		With Multicollinearity							
0.3	0.0	4.0	(5.0)	8.4	(5.0)	3.3	(5.0)	5.9	(5.0)
	0.3	4.9	(5.6)	8.1	(4.8)	4.0	(5.7)	6.4	(5.2)
	0.5	5.0	(5.7)	8.6	(5.5)	4.9	(7.1)	8.4	(7.4)
	0.7	4.2	(5.0)	8.6	(4.5)	5.8	(8.5)	11.1	(9.5)
	0.9	5.0	(5.9)	11.2	(7.6)	11.9	(15.8)	24.5	(22.0)
0.9	0.0	4.5	(5.0)	7.7	(5.0)	4.8	(5.0)	6.3	(5.0)
	0.3	6.8	(7.7)	10.2	(6.0)	33.8	(34.1)	38.1	(34.8)
	0.5	12.0	(12.7)	14.5	(10.3)	67.2	(67.3)	70.6	(68.1)
	0.7	20.3	(21.9)	25.4	(18.3)	92.8	(92.9)	94.7	(93.6)
	0.9	44.0	(46.0)	54.9	(45.6)	99.3	(99.3)	99.5	(99.5)

Notes: (1) FTH = the F-statistic using parameters estimates under H_1;
FTK = the F-statistic using parameter estimates under K_1.
(2) R^2 = the coefficient of determination in the reduced form equation for y_{t2}.
(3) ρ^2 = the squared correlation coefficient between y_{t2} and ϵ_{t1}.
(4) Figures in parentheses are empirical powers adjusted to make the sizes of the tests equal to 5% when H_1 is true.
(5) The number of replications is 1000 for each combination of R^2, ρ^2, and n.

Table 3: Empirical Sizes and Powers of the FTH and FTK; R2 Matrix

R^2	ρ^2	$n = 40$				$n = 100$			
		FTH		FTK		FTH		FTK	
		Without Multicollinearity							
0.3	0.0	4.0	(5.0)	5.4	(5.0)	4.4	(5.0)	5.4	(5.0)
	0.3	51.1	(54.7)	55.4	(54.6)	97.8	(97.9)	97.8	(97.8)
	0.5	87.9	(89.7)	90.2	(89.6)	100	(100)	100	(100)
	0.7	99.2	(99.5)	99.3	(99.3)	100	(100)	100	(100)
	0.9	100	(100)	100	(100)	100	(100)	100	(100)
0.9	0.0	4.5	(5.0)	8.4	(5.0)	5.2	(5.0)	5.5	(5.0)
	0.3	66.1	(67.1)	74.2	(66.1)	99.7	(99.7)	99.7	(99.7)
	0.5	92.7	(93.4)	96.3	(92.7)	100	(100)	100	(100)
	0.7	99.5	(99.5)	99.6	(99.5)	100	(100)	100	(100)
	0.9	100	(100)	100	(100)	100	(100)	100	(100)
		With Multicollinearity							
0.3	0.0	3.6	(5.0)	5.8	(5.0)	2.8	(5.0)	4.7	(5.0)
	0.3	51.5	(56.6)	61.4	(59.9)	96.1	(97.3)	97.1	(97.3)
	0.5	87.5	(89.7)	93.3	(92.8)	100	(100)	100	(100)
	0.7	98.3	(98.5)	99.8	(99.8)	100	(100)	100	(100)
	0.9	98.1	(98.5)	100	(100)	100	(100)	100	(100)
0.9	0.0	2.9	(5.0)	4.6	(5.0)	3.3	(5.0)	4.4	(5.0)
	0.3	51.7	(60.0)	58.7	(59.4)	96.2	(97.1)	96.5	(96.6)
	0.5	87.4	(90.6)	91.2	(91.4)	100	(100)	100	(100)
	0.7	99.2	(99.7)	99.5	(99.7)	100	(100)	100	(100)
	0.9	99.9	(99.9)	100	(100)	100	(100)	100	(100)

Notes: See the footnotes under Table 2.

Table 4: Empirical Sizes and Powers of the FTH and FTK; R3 Matrix

R^2	ρ^2	$n=40$ FTH		$n=40$ FTK		$n=100$ FTH		$n=100$ FTK	
		Without Multicollinearity							
0.3	0.0	2.9	(5.0)	5.1	(5.0)	4.3	(5.0)	5.2	(5.0)
	0.3	10.5	(14.5)	13.3	(13.2)	52.2	(56.8)	53.1	(52.2)
	0.5	18.3	(23.7)	22.0	(21.8)	87.2	(88.7)	87.4	(87.2)
	0.7	32.6	(39.1)	37.0	(36.8)	98.2	(98.5)	98.5	(98.5)
	0.9	62.0	(67.8)	67.6	(67.5)	100	(100)	100	(100)
0.9	0.0	3.6	(5.0)	4.8	(5.0)	5.9	(5.0)	6.7	(5.0)
	0.3	58.7	(63.1)	63.2	(63.7)	99.6	(99.4)	99.7	(99.4)
	0.5	88.3	(90.7)	90.3	(90.5)	100	(100)	100	(100)
	0.7	98.1	(98.1)	99.5	(99.5)	100	(100)	100	(100)
	0.9	99.8	(99.9)	100	(100)	100	(100)	100	(100)
		With Multicollinearity							
0.3	0.0	3.3	(5.0)	5.9	(5.0)	3.0	(5.0)	5.6	(5.0)
	0.3	3.5	(5.2)	6.1	(4.9)	3.6	(5.2)	6.1	(5.4)
	0.5	2.9	(4.0)	5.3	(4.6)	3.5	(5.2)	6.1	(5.4)
	0.7	2.8	(4.2)	5.5	(4.4)	3.5	(5.2)	6.8	(5.8)
	0.9	2.2	(3.2)	6.2	(5.3)	3.6	(6.2)	9.5	(7.7)
0.9	0.0	3.2	(5.0)	6.0	(5.0)	3.4	(5.0)	6.3	(5.0)
	0.3	4.7	(6.4)	6.2	(5.1)	7.0	(10.1)	9.9	(8.2)
	0.5	4.5	(6.5)	6.0	(5.1)	12.9	(17.4)	16.8	(14.4)
	0.7	5.4	(7.9)	7.0	(6.5)	26.7	(31.1)	32.3	(29.1)
	0.9	10.3	(13.7)	15.2	(13.6)	60.5	(65.6)	68.4	(65.4)

Notes: See the footnotes under Table 2.

Table 5: Empirical Sizes and Powers of the FTH and FTK; R4 Matrix

		Without Multicollinearity							
R^2	ρ^2	$n = 40$				$n = 100$			
		FTH		FTK		FTH		FTK	
0.3	0.0	2.8	(5.0)	4.1	(5.0)	4.4	(5.0)	5.1	(5.0)
	0.3	10.0	(15.3)	14.5	(15.7)	57.0	(62.1)	61.2	(60.8)
	0.5	21.7	(28.6)	26.3	(28.3)	91.9	(93.6)	93.1	(92.8)
	0.7	41.7	(51.7)	48.9	(51.3)	99.6	(99.7)	99.6	(99.6)
	0.9	80.6	(85.4)	85.2	(86.3)	100	(100)	100	(100)
0.9	0.0	4.1	(5.0)	5.3	(5.0)	5.7	(5.0)	6.1	(5.0)
	0.3	58.5	(61.3)	61.6	(61.3)	99.5	(99.4)	99.5	(99.3)
	0.5	89.4	(90.4)	90.4	(90.4)	100	(100)	100	(100)
	0.7	99.1	(99.1)	99.5	(99.5)	100	(100)	100	(100)
	0.9	100	(100)	100	(100)	100	(100)	100	(100)
		With Multicollinearity							
0.3	0.0	3.4	(5.0)	5.9	(5.0)	3.1	(5.0)	6.1	(5.0)
	0.3	4.5	(5.9)	8.1	(6.7)	5.2	(8.5)	10.1	(9.0)
	0.5	6.4	(9.7)	14.0	(12.6)	18.6	(26.2)	29.5	(27.2)
	0.7	14.9	(21.6)	30.2	(27.4)	58.1	(66.1)	72.5	(69.9)
	0.9	58.6	(64.8)	80.0	(77.9)	99.2	(99.4)	100	(100)
0.9	0.0	3.3	(5.0)	5.9	(5.0)	3.4	(5.0)	6.3	(5.0)
	0.3	5.1	(7.2)	7.3	(6.7)	9.8	(14.6)	13.2	(10.7)
	0.5	8.9	(12.3)	11.6	(11.1)	33.2	(40.7)	37.6	(32.4)
	0.7	19.7	(24.6)	24.0	(22.8)	75.9	(81.0)	79.3	(75.8)
	0.9	58.1	(64.5)	64.4	(62.6)	99.2	(99.4)	99.5	(99.4)

Notes: See the footnotes under Table 2.

Table 6: Empirical Sizes and Powers of the FTH and FTK; R1 and R2 Matrices with Multicollinearity, $n = 500$, $R^2 = 0.3$

		Multicollinearity, $n = 500$							
R^2	ρ^2	R1 Matrix				R2 Matrix			
		FTH		FTK		FTH		FTK	
	0.0	3.0	(5.0)	5.0	(5.0)	2.4	(5.0)	4.6	(5.0)
	0.3	8.8	(12.1)	11.6	(11.6)	100	(100)	100	(100)
0.3	0.5	91.4	(94.3)	93.5	(93.5)	100	(100)	100	(100)
	0.7	99.1	(99.4)	99.5	(99.5)	100	(100)	100	(100)
	0.9	99.8	(99.9)	100	(100)	100	(100)	100	(100)

Notes: See footnotes below Table 2.

From Table 7 it is clear that the noncentrality parameter, η, is relatively insensitive to the choice of the VAR(1) matrix, R in equation (2) when multicollinearity is absent. This explains why the powers of the tests are similar across the different values of the VAR(1) matrices when multicollinearity does not exist. When it does, however, the noncentrality parameter, η, is sensitive to the choice of the VAR(1) matrix. With the use of $R2$ matrix, the nocentrality parameter η under the presence of multicollinearty is close to that under the absence of of multicollinearity, whereas with the $R1$ matrix, the noncentrality parameter is reduced drastically under the presence of multicollinearity. The values of the noncentrality parameters for the $R3$ and $R4$ matrices are close to those for the $R1$ matrix.

The F statistics, FTH and FTK, are derived under the assumption that the structural error process is VAR(1). Suppose that the true error process is VAR(2):

$$U = U_{-1}R_1 + U_2R_2 + E \qquad (14)$$

with

$$R_1 = \begin{bmatrix} .8 & 0 & 0 \\ 0 & .8 & 0 \\ 0 & 0 & .4 \end{bmatrix} \quad \text{and} \quad R_2 = \begin{bmatrix} .3 & 0 & 0 \\ 0 & .3 & 0 \\ 0 & 0 & .5 \end{bmatrix}$$

Table 7: Effects of Multicollinearity, R^2, ρ^2 and Sample Size on the Average of Noncentrality Parameter [$R1$ and $R2$ Matrices for the VAR(1) Error Process]

			without multi-collinearity		with multi-collinearity	
R^2	n	ρ^2	$R1$	$R2$	$R1$	$R2$
		0.0	0	0	0	0
		0.3	3.60	4.10	1.94	3.61
.3	40	0.5	6.80	8.41	3.94	7.46
		0.7	13.83	18.07	8.17	16.72
		0.0	0	0	0	0
		0.3	5.53	5.52	2.20	3.67
.9	40	0.5	11.25	11.37	4.75	7.68
		0.7	23.16	23.34	10.35	17.15
		0.9	80.39	82.41	33.17	58.94
		0.0	0	0	0	0
		0.3	12.57	15.07	2.74	12.55
.3	100	0.5	25.38	31.63	5.48	27.02
		0.7	55.02	65.50	11.11	56.00
		0.9	163.29	229.66	41.22	201.76
		0.0	0	0	0	0
		0.3	20.17	20.14	5.83	13.08
.9	100	0.5	42.27	42.22	9.85	27.74
		0.7	83.87	83.51	21.93	57.64
		0.9	295.19	297.13	83.98	208.28

Notes: R^2 = the coefficient of determination of the reduced form equation for Y_1.
ρ^2 = the squared correlation between Y_{t1} and ϵ_{t1}.
$\eta = \xi' S' M_E S\xi/\sigma_{11\cdot 2}$ is the noncentrality parameter. The value of η is the average of 1000 replications.
$R1$ = VAR(1) matrix in Table 1.
$R2$ = VAR(1) matrix in Table 1.

but we use the FTH and FTK to test H_1 versus K_1. The sampling experiments for this set-up are presented in Table 8, and the results show that the FTH and FTK are fairly robust in terms of the sizes and powers under the VAR(2) process.

One may wonder if a test statistic that is derived assuming that $R = 0$ is robust when a VAR(1) error process exists. Among many test statistics that are derived under $R = 0$, let us choose the Wu-Hausman test (WHT) [Wu (1973),and Hausman (1978)], since as we mentioned earlier the FTH and FTK collapse to the WHT when $R = 0$. Table 9 report the results of sampling experiments for $R1$ matrix in Table 1. The numbers under FTH and FTK are the same as those in Table 2. The results show that when multicollinearity is absent, the sizes of the WHT are much higher than the nominal 5%, indicating that the WHT is more likely to reject the null hypothesis of independence of stochastic regressors when the null hypothesis is true. When multicollinearity is present, the powers of all the three statistics are equally poor.

Let us apply the test statistics to Klein's mode I. Klein's model I has been used frequently as an example of simultaneous estimation procedures [Zellner and Theil (1962), Theil (1971), and Pindyck and Rubinfeld (1981), among others.] There are three behavioral equations in the model: consumption, investment, and labor demand equations. The FTH and FTK are given in Table 10 for each of the three quations. Table 10 shows that the null hypothesis H_1 is rejected for the consumption and investment equations but it cannot be rejected for the labor demand equation, which may be treated as the single equation with an AR(1) error process.

IV. Concluding Remarks

In this paper we derived test statistics for testing whether or not the structural equation of interest can be regarded as the classical regression with the serially correlated error, and we conducted sampling experiments to see how the test statistics perform. We find that the sizes of the tests are reasonable, and the powers of the tests are sensitive to the existence of multicollinearity, the value of R^2, and to the values of the matrix of autocorrelation coefficients, R. When R is close to 0 (*i.e.* case of R3) and

Table 8: Empirical Sizes and Powers of the FTH and FTK; VAR(2) Error

		Without Multicollinearity							
R^2	ρ^2	$n = 40$				$n = 100$			
		FTH		FTK		FTH		FTK	
0.3	0.0	5.3	(5.0)	6.4	(5.0)	4.9	(5.0)	5.6	(5.0)
	0.3	66.2	(65.9)	69.3	(64.3)	100	(100)	100	(100)
	0.5	92.5	(92.5)	94.0	(91.6)	100	(100)	100	(100)
	0.7	99.0	(99.0)	99.0	(98.7)	100	(100)	100	(100)
	0.9	99.6	(99.6)	99.6	(99.6)	100	(100)	100	(100)
0.9	0.0	5.2	(5.0)	6.3	(5.0)	4.9	(5.0)	5.6	(5.0)
	0.3	67.5	(67.1)	70.6	(66.7)	100	(200)	100	(100)
	0.5	94.2	(94.1)	95.1	(93.8)	100	(100)	100	(100)
	0.7	99.4	(99.5)	99.4	(99.1)	100	(100)	100	(100)
	0.9	100	(100)	100	(100)	100	(100)	100	(100)
		With Multicollinearity							
0.3	0.0	5.3	(5.0)	7.6	(5.0)	5.7	(5.0)	6.3	(5.0)
	0.3	51.2	(50.2)	55.9	(47.6)	99.8	(99.7)	99.8	(99.7)
	0.5	77.7	(77.0)	80.5	(76.1)	99.8	(99.8)	100	(99.9)
	0.7	88.7	(88.4)	90.2	(88.2)	99.9	(99.9)	99.9	(99.9)
	0.9	91.6	(91.6)	92.9	(91.9)	99.9	(99.9)	99.9	(99.9)
0.9	0.0	5.4	(5.0)	7.4	(5.0)	5.7	(5.0)	6.3	(5.0)
	0.3	62.6	(61.7)	67.1	(61.2)	100	(100)	100	(100)
	0.5	90.5	(90.2)	92.2	(89.9)	100	(100)	100	(100)
	0.7	98.0	(98.0)	98.5	(97.9)	100	(100)	100	(100)
	0.9	99.9	(98.9)	99.1	(98.9)	100	(100)	100	(100)

Notes: See the footnotes under Table 2.
The VAR(2) process $U = U_{-1}R_1 + U_{-2}R_2 + E$ is generated with $R_1 = \text{Diag}(.8, .8, .4)$ and $R_2 = \text{Diag}(.3, .3\ .5)$.

Table 9: Empirical Sizes and Powers of the FTH, FTK and WHT; R1 Matrix

		Without Multicollinearity											
R^2	ρ^2	$n = 40$						$n = 100$					
		FTH		FTK		WHT		FTH		FTK		WHT	
0.3	0.0	5.5	(5.0)	7.5	(5.0)	8.5	(5.0)	4.8	(5.0)	5.9	(5.0)	15.4	(5.0)
	0.3	24.8	(24.3)	28.6	(22.6)	35.0	(27.5)	91.8	(92.2)	92.5	(90.5)	74.8	(58.9)
	0.5	45.2	(44.2)	48.6	(43.8)	56.8	(49.4)	99.3	(99.3)	99.3	(99.2)	74.8	(58.9)
	0.7	65.1	(64.5)	68.8	(63.4)	78.7	(74.9)	100	(100)	100	(100)	99.2	(97.9)
	0.9	81.9	(81.4)	84.5	(82.2)	93.0	(91.4)	100	(100)	100	(100)	100	(100)
0.9	0.0	5.2	(5.0)	7.5	(5.0)	24.1	(5.0)	5.7	(5.0)	5.9	(5.0)	32.0	(5.0)
	0.3	62.1	(61.1)	68.0	(60.5)	79.5	(52.1)	99.8	(99.7)	99.8	(99.6)	97.1	(83.4)
	0.5	92.0	(91.3)	94.0	(90.0)	95.1	(82.7)	100	(100)	100	(100)	99.8	(99.1)
	0.7	99.1	(99.1)	99.2	(98.8)	99.1	(96.6)	100	(100)	100	(100)	100	(100)
	0.9	100	(100)	100	(100)	100	(100)	100	(100)	100	(100)	100	(100)
		With Multicollinearity											
0.3	0.0	4.0	(5.0)	8.4	(5.0)	8.1	(5.0)	3.3	(5.0)	5.9	(5.0)	3.3	(5.0)
	0.3	4.9	(5.6)	8.1	(4.8)	6.9	(3.9)	4.0	(5.7)	6.4	(5.2)	4.3	(6.1)
	0.5	5.0	(5.7)	8.6	(5.5)	4.8	(3.5)	4.9	(7.1)	8.4	(7.4)	3.8	(6.6)
	0.7	4.2	(5.0)	8.6	(4.5)	6.7	(4.0)	5.8	(8.5)	11.1	(9.5)	3.2	(6.1)
	0.9	5.0	(5.9)	11.2	(7.6)	8.2	(5.6)	11.9	(15.8)	24.5	(22.0)	7.3	(13.1)
0.9	0.0	4.5	(5.0)	7.7	(5.0)	6.6	(5.0)	4.8	(5.0)	6.3	(5.0)	3.1	(5.0)
	0.3	6.8	(7.7)	10.2	(6.0)	7.8	(6.8)	33.8	(34.1)	38.1	(34.8)	17.7	(24.9)
	0.5	12.0	(12.7)	14.5	(10.3)	11.9	(9.9)	67.2	(67.3)	70.6	(68.1)	36.2	(46.6)
	0.7	20.3	(21.9)	25.4	(18.3)	21.6	(18.7)	92.8	(92.9)	94.7	(93.6)	64.6	(73.3)
	0.9	44.0	(46.0)	54.9	(45.6)	51.0	(47.6)	99.3	(99.3)	99.5	(99.5)	93.9	(95.2)

Notes: See footnotes below Table 2.
WHT = the Wu-Hausman test.

Table 10: The FTH and FTK Applied to Klein's Model I

	FTH	FTK
Consumption equation	5.192*	5.131*
Investment equation	10.127*	8.580*
Labor demand equation	1.590	1.275

Notes * indicates significance at 1% level.

multicollinearity exists, the sizes of the tests are close to the nominal values but the powers are poor except when the correlation between the endogenous variable in the right hand side of the equation of interest and the error term, ρ, is high. But as shown in Tsurumi (1990) when multicollinearity exists, simultaneous equation estimators are not so much better than the ordinary least squares estimator, and thus even if the exogeneity tests lead one to a choice of an inappropriate estimator its cost may not amount to any significance.

Although we assumed that the structural errors follow the AR(1) process, we can extend the test statistics to the AR(p) process as long as p is known. The derivation of the test statistics become cumbersome as the value of p increases. Since for most empirical work one does not know the value of p, an interesting question is whether the test statistics that are derived under the AR(1) process are criterion robust or not. As shown in Table 8, it appears that the FTH and FTK are robust.

Appendix A. Derivation of the System (7) and (8) in the Text

Let the simultaneous equations model be given by

$$Y\Gamma + XB = U \tag{15}$$

$$U = U_{-1}R + E \tag{16}$$

$$y_1 = Y_1\gamma_1 + X_1\beta_1 + u_1 \tag{17}$$

where

$$R \sim m \times m, \quad E = (\epsilon_1, E_2) \sim n \times m$$

$$\text{vec}(E') \sim \text{N}(0, I \otimes \Sigma) \quad \Sigma \sim m \times m, \text{ positive definite}$$

$$
\begin{aligned}
Y_{n\times m} &= (y_1, Y_1, Y_2); \\
& \quad y_1 \sim n\times 1, \quad Y_1 \sim n\times m, \quad Y_2 \sim n\times m_2, \quad m_2 = m - m_1 - 1 \\
X_{n\times k} &= (X_1, X_2); \\
& \quad X_1 \sim n\times k_1, \quad X_2 \sim n\times k_2 \\
U_{n\times m} &= (u_1, U_1, U_2); \\
& \quad u_1 \sim n\times 1, \quad U_1 \sim n\times m_1, \quad U_2 \sim n\times m_2 \\
\Gamma_{m\times m} &= \begin{bmatrix} 1 & \\ -\gamma_1 & \Gamma_2 \\ 0 & \end{bmatrix}, \quad \gamma_1 \sim m_1\times 1, \quad \Gamma_2 \sim m\times(m-1) \\
Y\Gamma &= (y_1, Y_1, Y_2)\begin{bmatrix} 1 & \\ -\gamma_1 & \Gamma_2 \\ 0 & \end{bmatrix} = (y_1 - Y_1\gamma_1, Y\Gamma_2) \\
B_{k\times m} &= \begin{bmatrix} -\beta_1 & \\ & B_2 \\ 0 & \end{bmatrix} \\
B_2 &= \begin{bmatrix} B_{12} \\ B_{22} \end{bmatrix}, \quad B_{12} \sim k_1\times(m-1), \quad B_{22} \sim k_2\times(m-1) \\
XB &= (X_1, X_2)\begin{bmatrix} -\beta_1 & \\ & B_2 \\ 0 & \end{bmatrix} = (-X_1\beta_1, XB_2)
\end{aligned}
$$

The system becomes

$$
\begin{aligned}
Y\Gamma + XB &= U \\
Y_{-1}\Gamma + X_{-1}B &= U_{-1} \quad \text{or} \\
Y_{-1}\Gamma R + X_{-1}BR &= U_{-1}R
\end{aligned}
$$

and thus

$$
Y\Gamma = -XB + Y_{-1}\Gamma R + X_{-1}BR + E \tag{18}
$$

The reduced form is

$$
Y = X\Pi + Y_{-1}\Upsilon + X_{-1}\Phi + V \tag{19}
$$

where $\Pi = -B\Gamma^{-1}$, $\Upsilon = \Gamma R\Gamma^{-1}$, $\Phi = BR\Gamma^{-1}$, and $V = E\Gamma^{-1}$. Let V be partitioned as

$$
V = (v_1, V_1, V_2) \quad v_1 \sim n\times 1, \quad V_1 \sim n\times m_1, \quad V_2 \sim n\times m_2
$$

and post-multiply the reduced form by

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ -\gamma_1 & I_{m_1} & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix}$$

and obtain

$$Y\Lambda = X\Pi\Lambda + Y_{-1}\Upsilon\Lambda + X_{-1}\Phi\Lambda + V\Lambda \tag{20}$$

The each term in equation (20) is given by

$$Y\Lambda = (y_1, Y_1, Y_2)\begin{bmatrix} 1 & 0 & 0 \\ -\gamma_1 & I_{m_1} & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix} = (y_1 - Y_1\gamma_1, Y_1, Y_2)$$

$$\begin{aligned} X\Pi\Lambda &= (X_1, X_2)\begin{bmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} \\ \Pi_{21} & \Pi_{22} & \Pi_{23} \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ -\gamma_1 & I_{m_1} & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix} \\ &= (X_1, X_2)\begin{bmatrix} \Pi_{11} - \Pi_{12}\gamma_1 & \Pi_{12} & \Pi_{13} \\ \Pi_{12} - \Pi_{22}\gamma_1 & \Pi_{22} & \Pi_{23} \end{bmatrix} \\ &= [X_1(\Pi_{11} - \Pi_{12}\gamma_1) + X_2(\Pi_{21} - \Pi_{22}\gamma_1), X\Pi_2, X\Pi_3] \end{aligned}$$

where

$$\Pi_2 = \begin{bmatrix} \Pi_{12} \\ \Pi_{22} \end{bmatrix}, \quad \Pi_3 = \begin{bmatrix} \Pi_{13} \\ \Pi_{23} \end{bmatrix}$$

$$\begin{aligned} Y_{-1}\Upsilon\Lambda &= (y_{1,-1}, Y_{1,-1}, Y_{2,-1})\begin{bmatrix} \Upsilon_{11} & \Upsilon_{12} & \Upsilon_{13} \\ \Upsilon_{21} & \Upsilon_{22} & \Upsilon_{23} \\ \Upsilon_{31} & \Upsilon_{32} & \Upsilon_{33} \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ -\gamma_1 & I_{m_1} & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix} \\ &= (y_{1,-1}, Y_{1,-1}, Y_{2,-1})\begin{bmatrix} \Upsilon_{11} - \Upsilon_{21}\gamma_1 & \Upsilon_{12} & \Upsilon_{13} \\ \Upsilon_{21} - \Upsilon_{22}\gamma_1 & \Upsilon_{22} & \Upsilon_{23} \\ \Upsilon_{31} - \Upsilon_{32}\gamma_1 & \Upsilon_{32} & \Upsilon_{33} \end{bmatrix} \\ &= [y_{1,-1}(\Upsilon_{11} - \Upsilon_{12}\gamma_1) + Y_{1,-1}(\Upsilon_{21} - \Upsilon_{22}\gamma_1) + Y_{2,-1}(\Upsilon_{31} - \Upsilon_{32}\gamma_1), \\ &\qquad Y_{-1}\Upsilon_2, Y_{-1}\Upsilon_3] \end{aligned} \tag{21}$$

where

$$\Upsilon_2 = \begin{bmatrix} \Upsilon_{12} \\ \Upsilon_{22} \\ \Upsilon_{32} \end{bmatrix}, \quad \Upsilon_3 = \begin{bmatrix} \Upsilon_{31} \\ \Upsilon_{32} \\ \Upsilon_{33} \end{bmatrix}$$

$$\begin{aligned}
X_{-1}\Phi\Lambda &= (X_{1,-1}, X_{2,-1}) \begin{bmatrix} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & \Phi_{22} & \Phi_{23} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -\gamma_1 & I_{m_1} & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix} \\
&= (X_{1,-1}, X_{2,-1}) \begin{bmatrix} \Phi_{11} - \Phi_{12}\gamma_1 & \Phi_{12} & \Phi_{13} \\ \Phi_{21} - \Phi_{22}\gamma_1 & \Phi_{22} & \Phi_{23} \end{bmatrix} \\
&= [X_{1,-1}(\Phi_{11} - \Phi_{12}\gamma_1) + X_{2,-1}(\Phi_{21} - \Phi_{22}\gamma_1), X_{-1}\Phi_2, X_{-1}\Phi_3]
\end{aligned}$$

where

$$\Phi_2 = \begin{bmatrix} \Phi_{12} \\ \Phi_{22} \end{bmatrix}, \quad \text{and} \quad \Phi_3 = \begin{bmatrix} \Phi_{13} \\ \Phi_{23} \end{bmatrix}$$

and

$$V\Lambda = (v_1, V_1, V_2) \begin{bmatrix} 1 & 0 & 0 \\ -\gamma_1 & I_{m_1} & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix} (v_1 - V_1\gamma_1, V_1, V_2)$$

So the first equation of (20) becomes

$$\begin{aligned}
y_1 - Y_1\gamma_1 = X_1(\Pi_{11} - \Pi_{12}\gamma_1) + X_2(\Pi_{21} - \Pi_{22}\gamma_1) + y_{1,-1}(\Upsilon_{11} - \Upsilon_{12}\gamma_1) \\
+Y_{1,-1}(\Upsilon_{21} - \Upsilon_{22}\gamma_1) + Y_{2,-1}(\Upsilon_{31} - \Upsilon_{32}\gamma_1) + X_{1,-1}(\Phi_{11} - \Phi_{12}\gamma_1) \\
+X_{2,-1}(\Phi_{21} - \Phi_{22}\gamma_1) + v_1 - V_1\gamma_1 \qquad (22)
\end{aligned}$$

and we have $v_1 - V_1\gamma_1 = \epsilon_1$. On the other hand the structural equation is given by equation (18). The first equation in (18) is unscrambled as follows. Let

$$\Gamma = \begin{bmatrix} 1 & \\ -\gamma_1 & \Gamma_2 \\ 0 & \end{bmatrix};$$

$$B = \begin{bmatrix} \beta_1 & \\ & B_2 \\ 0 & \end{bmatrix}, \quad B_2 \sim k \times (m-1)$$

$$R = \begin{bmatrix} r_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix},$$

where $R_{21} \sim (m-1) \times 1$, $R_{12} \sim 1 \times (m-1)$, $R_{22} \sim (m-1) \times (m-1)$. Then

$$Y\Gamma = (y_1, Y_1, Y_2) \begin{bmatrix} 1 & \\ -\gamma_1 & \Gamma_2 \\ 0 & \end{bmatrix} = (y_1 - Y_1\gamma_1, Y\Gamma_2)$$

$$
\begin{aligned}
XB &= (X_1, X_2)\begin{bmatrix} -\beta_1 & \\ & B_2 \\ 0 & \end{bmatrix} = (-X_1\beta_1, XB_2) \\
Y_{-1}\Gamma R &= (y_{1,-1}, Y_{1,-1}, Y_{2,-1})\begin{bmatrix} 1 & \\ -\gamma_1 & \Gamma_2 \\ 0 & \end{bmatrix}\begin{bmatrix} r_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \\
&= [(y_{1,-1} - Y_{1,-1}\gamma_1)r_{11} + Y_{-1}\Gamma_2 R_{21}, (y_{1,-1} - Y_{1,-1}\gamma_1)R_{12} + Y_{-1}\Gamma_2 R_{22}] \\
X_{-1}BR &= (X_{1,-1}, X_{2,-1})\begin{bmatrix} -\beta_1 & \\ & B_2 \\ 0 & \end{bmatrix} R \\
&= (-r_{11}X_{1,-1}\beta_1 + X_{-1}B_2R_{21},\ -X_{1,-1}\beta_1 R_{12} + X_{-1}B_2R_{22})
\end{aligned}
$$

Hence the first equation becomes

$$
\begin{aligned}
y_1 - Y_1\gamma_1 &= X_1\beta_1 + (y_{1,-1} - Y_{1,-1}\gamma_1)r_{11} + Y_{-1}\Gamma_2 R_{21} \\
&\quad -X_{1,-1}\beta_1 r_{11} + X_{-1}B_2R_{21} + \epsilon_1
\end{aligned}
$$

or

$$
\begin{aligned}
y_1 - r_{11}y_{1,-1} &= (Y_1 - r_{11}Y_{1,-1})\gamma_1 + (X_1 - r_{11}X_{1,-1})\beta_1 \\
&\quad +Y_{-1}\Gamma_2 R_{21} + X_{-1}B_2R_{21} + \epsilon_1 \qquad (23)
\end{aligned}
$$

Comparing (23) to (22), we see

(1) $\Pi_{11} - \Pi_{12}\gamma_1 = \beta_1$

(2) $\Pi_{21} - \Pi_{22}\gamma_1 = 0$

(3) $r_{11} + d_1 = \Upsilon_{11} - \Upsilon_{12}\gamma_1$, where d_1 is the first element of $\Gamma_2 R_{21}$.

(4) $-\gamma_1 r_{11} + D_2 = \Upsilon_{21} - \Upsilon_{22}\gamma_1$, where D_2 is the $2, \cdots, (m_1+1)$ elements of $\Gamma_2 R_{21}$.

(5) $D_3 = \Upsilon_{31} - \Upsilon_{32}\gamma_1$, where D_3, or the last m_2 elements of $\Gamma_2 R_{21}$.

(6) $-\beta_1 r_{11} + (I_{k_1}, 0)B_2R_{21} = \Phi_{11} - \Phi_{12}\gamma_1$

(7) $(0, I_{k_2})B_2R_{21} = \Phi_{21} - \Phi_{22}\gamma_1$

And from equation (20) we get

$$
Y_1 = X\Pi_2 + Y_{-1}\Upsilon_2 + X_{-1}\Phi_2 + V_1 \qquad (24)
$$

Equations (23) and (24) are jointly normal and the likelihood fuction is given by

$$\ell(\gamma_1, \beta_1, r_{11}, R_{21}, \Pi_2, \Upsilon_2, \Phi_2, \Omega \mid \text{data}) \propto \mid \Omega \mid^{-n/2} \exp\left\{-\frac{1}{2}\text{tr}\left[\Omega^{-1}(W - Z\Theta)'(W - Z\Theta)\right]\right\} \quad (25)$$

where $W = (y_1 - r_{11}y_{1,-1},\ Y_1)$, $Z = (Y_1 - r_{11}Y_{1,-1},\ X_1 - r_{11}X_{1,-1},\ X,\ Y_{-1},\ X_{-1})$;

$$\Theta = \begin{bmatrix} \gamma_1 & 0 \\ \beta_1 & 0 \\ 0 & \Pi_2 \\ \Gamma_2 R_{21} & \Upsilon_2 \\ B_2 R_{21} & \Phi_2 \end{bmatrix}$$

and Ω is the covariance matrix

$$\Omega = \begin{bmatrix} \sigma_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} = \begin{bmatrix} \text{var}(\epsilon_{t1}) & \text{Cov}(\epsilon_t, Y_{t1}) \\ \text{Cov}(Y_{t1}, \epsilon_{t1}) & \text{Var}(Y_{t1}) \end{bmatrix}$$

Using the identity

$$\Omega^{-1} = \begin{bmatrix} 1 & 0 \\ -\Omega_{22}^{-1}\Omega_{21} & I \end{bmatrix} \begin{bmatrix} \sigma_{11\cdot 2}^{-1} & 0 \\ 0 & \Omega_{22}^{-1} \end{bmatrix} \begin{bmatrix} 1 & -\Omega_{12}\Omega_{22}^{-1} \\ 0 & I \end{bmatrix}$$

where $\sigma_{11\cdot 2} = \sigma_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$, we see that equation (25) becomes

$$\ell(\gamma_1, \beta_1, r_{11}, R_{21}, \Pi_2, \Upsilon_2, \Phi_2, \sigma_{11\cdot 2}, \delta, \Omega_{22} \mid \text{data}) \propto \sigma_{11\cdot 2}^{-n/2} \mid \Omega_{22} \mid^{-n/2} \exp\left[-\frac{1}{2\sigma_{11\cdot 2}}Q(y_1)\right] \exp\left\{-\frac{1}{2}\text{tr}\left[\Omega_{22}^{-1}Q(Y_1)\right]\right\} \quad (26)$$

where $\delta = \Omega_{22}^{-1}\Omega_{21}$, and

$$Q(y_1) = a'a, \qquad Q(Y_1) = V_1'V_1$$

$$\begin{aligned} a &= y_1 - r_{11}y_{1,-1} - (Y_1 - r_{11}Y_{1,-1})\gamma_1 - (X_1 - r_{11}X_{1,-1})\beta_1 \\ &\quad -Y_{-1}\Gamma_2 R_{21} - X_{-1}B_{21}R_{21} - V_1\delta \\ &= y_1 - r_{11}y_{1,-1} - (Y_1 - r_{11}Y_{1,-1})\gamma_1 - (X_1 - r_{11}X_{1,-1})\beta_1 \\ &\quad -U_{2,-1}R_{21} - V_1\delta \\ V_1 &= Y_1 - X\Pi_2 - Y_{-1}\Upsilon_2 - X_{-1}\Phi_2 \end{aligned}$$

where $U_{2,-1} = Y_{-1}\Gamma_2 + X_{-1}B_2$.

Let the prior probability density function (pdf) be given by

$$p(\gamma_1, \beta_1, r_{11}, R_{21}, \Pi_2, \Upsilon_2, \Phi_2, \sigma_{11\cdot2}\Omega_{22}) \propto \sigma_{11\cdot2}^{-1} \mid \Omega_{22} \mid^{-1}$$

then after integrating out Ω_{22}, we derive the conditional posterior pdf given $r_{11}, \Pi_{22}, \Upsilon_2, \Phi_2, \Gamma_2, B_2$

$$\begin{aligned} & p(\gamma_1, \beta_1, r_{11}, R_{21}, \sigma_{11\cdot2}, \delta \mid r_{11}, \Pi_2, \Upsilon_2, \Phi_2, \Gamma_2, B_2, \text{ data}) \\ \propto \; & \sigma_{11\cdot2}^{-n/2-1} \exp\left[-\frac{1}{2\sigma_{11\cdot2}} Q(y_1)\right] \end{aligned} \tag{27}$$

Integrating out $\sigma_{11\cdot2}, \gamma_1$, and β_1, we obtain the conditional posterior pdf for $\xi = (R_{21}', \delta')'$:

$$p(\xi \mid r_{11}, \Pi_2, \Upsilon_2, \Phi_2, \Gamma_2, B_2, \text{ data}) \propto \left[1 + \frac{Q(\xi)}{SSR_u}\right]^{-(\nu+m_1+m-1)/2} \tag{28}$$

where $\nu = n - 2m_1 - k_1 - (m-1)$ and

$$Q(\xi) = (\xi - \hat{\xi})' S' M_E S (\xi - \hat{\xi})$$

$S = [U_{2,-1}, V_2]$, $E = (Y_1 - r_{11}Y_{1,-1},\ X_1 - r_{11}X_{1,-1})$, $M_E = [I - E(E'E)^{-1}E']$, $\hat{\xi} = (S'M_ES)^{-1}S'M_E(y_1 - r_{11}y_{1,-1})$, $SSR_u = (y_1 - r_{11}y_{1,-1})'M_P(y_1 - r_{11}y_{1,-1})$, $M_P = I - P(P'P)^{-1}P'$ and $P = [E, S]$. The quantity

$$\frac{Q(\xi)/(m-1+m_1)}{SSR_u/(n-2m_1-k_1-m+1)} \tag{29}$$

is distributed *a posteriori* as F with $(m-1+m_1,\ n-2m_1-k_1-m+1)$ degrees of freedom [*c.f.* Box and Tiao (1973, p.117)]. We may employ the highest posterior density credible set (HPDCS) [Berger (1988,p.140] or the highest posterior density region (HPDR) [Box and Tiao (1973, p.125)] to test the hypotheses H_1 versus K_1 which are changed to

$$H_1' : \begin{bmatrix} R_{21} \\ \delta \end{bmatrix} \quad \text{versus} \quad K_1' : \begin{bmatrix} R_{21} \\ \delta \end{bmatrix} \neq 0 \tag{30}$$

Since $\delta = \Omega_{22}^{-1}\Omega_{21}$, testing $\text{Cov}(Y_{t1}, \epsilon_{t1}) = \Omega_{21} = 0$ is the same as testing $\delta = 0$. Rewriting (29) we have the test statistic for $\xi = 0$:

$$F_{m-1+m_1, n-2m_1-k_1-m+1} = \frac{SSR_c - SSR_u)/(m-1+m_1)}{SSR_u/(n-2m_1-k_1-m+1)} \tag{31}$$

since $Q(0) = S'M_E S = SSR_c - SSR_u$. Equation (31) is equation (13) in the text.

Appendix B: The Design of the Sampling Experiments

We set up a linear simultaneous equations system consisting of three structural equations for our direct Monte Carlo method of sampling experiments. The model is a modification of the model used in Tsurumi (1990).

$$Y\Gamma = XB + U \tag{32}$$

where

$$\Gamma = \begin{pmatrix} 1.0 & -.267 & -.087 \\ -.222 & 1.0 & 0 \\ 0 & -.048 & 1.0 \end{pmatrix} \quad B = \begin{pmatrix} .62 & 4.4 & 4.0 \\ 0 & .74 & 0 \\ .7 & 0 & .53 \\ 0 & 0 & .11 \\ .96 & .13 & 0 \\ 0 & 0 & .56 \\ .06 & 0 & 0 \end{pmatrix}$$

We shall use the first equation of interest:

$$\begin{aligned} y_{t1} &= \gamma_{12} y_{t2} + \beta_{11} x_{t1} + \beta_{13} x_{t3} + \beta_{15} x_{t5} + \beta_{17} x_{t7} + u_{t1} \\ \gamma_{12} &= .222,\ \beta_{11} = 6.2,\ \beta_{13} = .7,\ \beta_{15} = .96\ \beta_{17} = .06\ . \end{aligned} \tag{33}$$

The variance-covariance matrix of the row of E, $Cov(\epsilon_t)$, is specified as

$$Cov(\epsilon_t) = 36.0 \begin{pmatrix} 1.0 & \rho_\omega & .25\rho_\omega \\ \rho_\omega & 1.0 & 0 \\ .25\rho_\omega & 0 & 1.0 \end{pmatrix} \tag{34}$$

where the values of the parameter, ρ_ω, are chosen so that the correlation between y_{t2} and ϵ_{t1}, ρ_{12} will be controlled. The exogenous variables, other than the constants, $x_{t2}, \cdots, x_{t7}$, are drawn from uniform distribution over the interval $[0, \mathrm{a}]$, where a is a scalar whose value will be specified so that R^2 (the coefficient of determination of the reduced form for y_{t2}) is controlled. The combination of the values of ρ_ω, a, and R^2 are given in Table 11. The value of a is affected by the sample size, R^2, and multicollinearity whereas the values of ρ_ω are only influenced by the degrees of simultaneity, ρ^2.

Table 11: Combinations of R^2, Sample Size, ρ_ω, and a Used for Sampling Experiments

Without Multicollinearity					With Multicollinearity				
R^2	n	ρ^2_{12}	ρ_ω	a	R^2	n	ρ^2_{12}	ρ_ω	a
.9	100	0	-.25	83.39	.9	100	0	-.25	45.04
		.3	-.71	72.64			.3	-.71	39.49
		.5	-.823	69.43			.5	-.823	37.74
		.7	-.90	67.02			.7	-.90	36.39
		.9	-.963	65.00			.9	-.936	35.18
.9	20	0	-.25	111.14	.9	20	0	-.25	53.70
		.3	-.71	94.49			.3	-.71	47.08
		.5	-.823	89.76			.5	-.823	45.07
		.7	-.90	86.39			.7	-.90	43.58
		.9	-.963	83.96			.9	-.963	42.41
.35	100	0	-.25	21.57	.35	100	0	-.25	12.49
		.3	-.71	19.51			.3	-.71	11.59
		.5	-.823	18.76			.5	-.823	11.14
		.7	-.90	18.14			.7	-.90	10.72
		.9	-.963	17.55			.9	-.963	10.24
.35	20	0	-.25	37.82	.35	20	0	-.25	19.05
		.3	-.71	34.46			.3	-.71	19.23
		.5	-.823	32.39			.5	-.823	18.57
		.7	-.90	30.35			.7	-.90	17.72
		.9	-.963	28.03			.9	-.963	16.54

Notes: R^2= the coefficient of determination of the reduced form equation for Y_1.
n= sample size.
ρ^2= the multiple correlation coefficient between y_{t2} and ϵ_{t1}.
ρ_ω= parameter in $Cov(\epsilon_t)$ in (27).
a= upper limit of the uniform distribution, $[0, a]$ from which the exogenous variables, x_{ti}, are drawn.

As the structure of the correlation among the exogenous variables, we use the following correlation matrices, one for almost nonexistence of multicollinearity and the other for high multicollinearity:

Correlation Matrix of X,Corr(X)										
	No Multicollinearity					Multicollinearity				
	x_3	x_4	x_5	x_6	x_7	x_3	x_4	x_5	x_6	x_7
x_2	-.15	-.11	-.18	-.37	-.18	.99	.93	.87	.69	.60
x_3		-.18	-.01	-.27	-.03		.93	.88	.73	.61
x_4			-.15	-.37	.02			.87	.67	.64
x_5				-.09	.05				.75	.69
x_6					.56					.85
	Det(Corr(X))=.3074					Det(Corr(X)) =.0000393				

The determinant of the correlation matrix for the case of no multicollinearity is .3074, indicating a low degree of multicollinearity among x_{ti}'s, whereas that for the case of multicollinearity is .0000393, which shows a high degree of multicollinearity.

References:

Amemiya, T. (1966) "Specification analysis in the estimation of parameters of a simultaneous equation model with autoregressive residuals," *Econometrica*, 34, 283–306.

Berger, J.O. (1985) *Statistical decision theory and Bayesian analysis*, 2nd edition, Springer-Verlag, New York.

Box, G.E.P. and G.C. Tiao (1973) *Bayesian inference in statistical analysis*, Addison-Wesley, Reading, Massachusetts.

Bowden, R.J. and D.A. Turkington (1984) *Instrumental variables*, Cambridge University Press, Cambdridge.

Dhrymes, P.J., R. Berner, and D. Cummins (1974) "A comparison of some limited information estimators for dynamic simultaneous equations models with autocorrelated errors," *Econometrica*, 42, 311–332.

Fair, R.C. (1970) "The estimation of simultaneous equations models with lagged endogenous variables and first order serially correlated errors," *Econometrica,* 38, 507–516.

Fair, R.C. (1972) "Efficient estimation of simultaneous equations with autoregressive errors by instrumental variables," *Review of Economics and Statistics,* 54, 444–449.

Godfrey, L.G. (1976) "Testing for serial correlation in dynamic simultaneous equation models," *Econometrica,* 44, 1077–1084.

Godfrey, L.G. (1988) *Misspecification tests in econometrics,* Cambridge University Press, Cambridge, England.

Hatanaka, M. (1974) "Several efficient two-step estimators for the dynamic simultaneous equations model with autoregressive disturbances," *Journal of Econometrics,* 4, 189–204.

Hausman, J. (1978) "Specification tests in econometrics," *Econometrica,* 46, 1251–1271.

Hendry, D. F. (1971) "Maximum likelihood estimation of systems of simultaneous regression equations with errors generated by a vector autoregressive process" *International Economic Review,* 12, 257–272.

Hendry, D. F. (1976) "The structure of simultaneous equations estimators," *Journal of Econometrics,* 4, 51–88.

Nakamura A. and M. Nakamura (1981) "On the relationships between several specification tests presented by Durbin, Wu and Hausman," *Econometrica,* 49, 1583–1588.

Pindyck, R. and D.L. Rubinfeld (1981) *Econometric models and economic forecasts,* 2nd edition, McGraw-Hill, New York.

Sargan, J.D. (1959) "The estimation of relationships with autocorrelated residuals by the use of instrumental variables," *Journal of the Royal Statistical Society,* Series B, 21, 91–105.

Sargan, J.D. (1966) "The maximum likelihood estimation of economic relationships with autoregressive residuals," *Econometrica,* 29, 414–426.

Sargan, J.D. (1973) "Testing for misspecification after estimating using instrumental variables," LSE working paper, also in E. Massoumi (ed.) *Contribution to econometrics: John Dennis Sargan* Vol. 1, 1988, Cambridge University Press, Cambridge, 213–235.

Theil, H. (1971) *Principles of Econometrics*, John Wiley and Sons, New York.

Tsurumi, H. (1990) "Comparing Bayesian and non-Bayesian limited information estimators," Geisser, S., J. Hodges, S.J. Press and A. Zellner (eds.) *Bayesian and likelihood methods in statistics and econometrics, essays in honor of George A. Barnard* North-Holland, Amsterdam, 179–207.

Wu, D. M. (1973) "Alternative tests of independence between stochastic regressors and disturbances," *Econometrica*, 41,733–750.

Zellner, A. and H. Theil (1962) "Three-stage least squares: simultaneous estimation of simultaneous relations," *Econometrica*, 30. 54–78.