Thailand Statistician
July 2012; 10(2): 163-181
http://statassoc.or.th
Contributed paper

New Criteria for Selection in Simultaneous Equations Model

Warangkhana Keerativibool

Department of Mathematics and Statistics, Faculty of Science, Thaksin University, Phattalung 93110 Thailand, and Centre of Excellence in Mathematics, CHE, Si Ayutthaya Rd., Bangkok 10400, Thailand.

E-mail: warang27@gmail.com

Received: 22 September 2011

Accepted: 27July 2012

Abstract

When the errors of statistical models are not independent, such as in the existence of the autocorrelation (AR) and/or moving average (MA) problems, the values of the standard model selection criteria are not correct and hence may affect the acquisition of the true model. This paper attempts to modify the Bayesian information criterion (BIC) in order to select the most appropriate simultaneous equations model (SEM). The first criterion, a system of simultaneous equation BIC (SBIC), is constructed after correcting the second-order autocorrelation, AR(2), problem. The second criterion is the adjusted BIC when the AR(2) problem is ignored. If there is no AR(2) problem in the errors, SBIC reduces to BIC. Using an extensive simulation study, SBIC and BIC are compared with SAIC and AIC, the measures of model selection in SEM that were introduced by Keerativibool et al. (2011). From the simulation study we conclude that SBIC convincingly outperformed the other criteria and the rest of the criteria can be ordered according to their performance by BIC, SAIC, and AIC.

Keywords: Bayesian information criterion (BIC), model selection criteria, second-order autocorrelation [AR(2)], simultaneous equations model (SEM), system of simultaneous equation BIC (SBIC).

1. Introduction

In the application of statistics, the statistical modeling is considered a major task of study. Three statistical processes to guide a model, which has the parsimony, goodness-of-fit, and generalizability properties, are the hypothesis testing of parameters, variable selection algorithms, and model selection criterion. The model selection criterion is a popular tool for selecting the best model. The first model selection criterion to gain widespread acceptance was the Akaike information criterion, AIC [1-6]. Other criteria were subsequently introduced and studied such as, Bayesian information criterion, BIC [7-9], Hannan and Quinn criterion, HQ [10-11], and Kullback information criterion, KIC [12-19]. AIC and BIC are two well-known measures, although AIC remains arguably the most widely used of the model selection criterion, BIC is a popular competitor. In fact, BIC is often preferred over AIC by practitioners who find appeal in either its Bayesian justification or its tendency to choose more parsimonious models than AIC. Neath and Cavanaugh [8]; Cavanaugh [12]; Giombini and Szroeter [20] concluded that AIC was an asymptotically efficient criterion, then in the large sample, AIC chose the model with minimum mean squared error (MSE) whereas BIC was a consistent criterion and could identify the correct model asymptotically with probability one. As a result, when the generating model is a finite order and is represented in the collection of candidate families under consideration model, the efficient criterion as AIC is an inconsistent criterion and tends, asymptotically, to overestimate the dimension of the parameter vector for the model.

Unfortunately, all of the standard model selection criteria are stated above cannot be used in a SEM when the autocorrelation (AR) and/or moving average (MA) problems occurred, except SAIC in [6] can be used in the SEM when there is the AR(2) problem. Keerativibool and Keerativibool et al. [21-25] concluded that the AR and MA problems made the overestimated of the errors whether the models were regression or SEM. Consequently, the values of all model selection criteria are incorrect. The AR and MA problems are usually found in time-series and panel data. The economic time-series and panel cross sectional data often display a memory in that variation around the regression function is not independent from one period to the next. The seasonally adjusted price and quantity series published by government agencies are examples. With this motivation, this study has three objectives as follows. Firstly, a GLS transformation matrix proposed in Keerativibool's paper [24] is used to correct the AR(2) problem. Secondly, a system of simultaneous equations BIC, called SBIC, is proposed to select the best SEM. SBIC is considered after the AR(2) problem are corrected, but the contemporaneously correlated errors still being considered. Also, BIC introduced by Schwarz [7] is slightly adjusted in order to use in a SEM when the AR(2) problem is ignored. The last objective, the performance of proposed model selection criteria, SBIC and BIC, are compared with SAIC and AIC, the measures of model selection proposed by Keerativibool et al. [6].

The remainder of this study is organized as follows. In Section 2, we summarize the main characteristics of the model to consider this study, including a GLS transformation matrix to correct the AR(2) problem. Derivations of the model selection criteria, called SBIC and BIC, follow in Section 3. Section 4 demonstrates the steps to construct the SEM when the errors are AR(2) and contemporaneously schemes, the steps to transform the errors of SEM to be independent, the steps of model selection, and shows all results of the simulation study. Finally, Section 5 is the conclusion, discussion, and further study.

2. A simultaneous equations model (SEM) and a GLS transformation matrix to correct the AR(2) problem

The structural and reduced-forms of the SEM [26] may be represented, respectively, as follows:

$$\mathbf{Y\Gamma} + \mathbf{XB} = \mathbf{U} \text{ and } \mathbf{Y} = \mathbf{X\Pi} + \mathbf{V}, \tag{1}$$

where $\mathbf{Y}$ is a $T \times M$ matrix of observations, $\mathbf{X}$ is a $T \times K$ design matrix of full-column rank, $\mathbf{\Gamma}$ is an $M \times M$ nonsingular matrix of coefficients of endogenous variables, $\mathbf{B}$ is a $K \times M$ matrix of coefficients of predetermined variables, $\mathbf{\Pi} = -\mathbf{B\Gamma}^{-1}$ is a $K \times M$ matrix of unknown parameters, $\mathbf{U}$ and $\mathbf{V} = \mathbf{U\Gamma}^{-1}$ are the $T \times M$ matrices of AR(2) and contemporaneously correlated errors. The t^{th} observation vector of reduced-form model in (1) is

$$\mathbf{y}_t = \mathbf{\Pi}^{\mathsf{T}}\mathbf{x}_t + \mathbf{v}_t, \tag{2}$$

where

$$\mathbf{v}_t = \rho_1\mathbf{v}_{t-1} + \rho_2\mathbf{v}_{t-2} + \boldsymbol{\varepsilon}_t \; ; \; t = 1, 2, \ldots, T, \tag{3}$$

the l-periods back error vector $\mathbf{v}_{t-l}$ is called the l^{th} lag of error vector $\mathbf{v}_t$, the autoregressive parameters ρ_1 and ρ_2 of the model must satisfy the stationary conditions [27], and $\boldsymbol{\varepsilon}_t$ is a multivariate normal vector given zero mean vector and contemporaneous covariance matrix $\mathbf{\Sigma}$, which is assumed nonsingular and positive symmetric definite,

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1M} & \sigma_{2M} & \cdots & \sigma_{MM} \end{bmatrix}.$$

That is, the error vector $\boldsymbol{\varepsilon}_t$ should be $\boldsymbol{\varepsilon}_t \sim N_M\left(\mathbf{0}, \mathbf{\Sigma}\right)$. (4)

Combine every observation vectors of the model in (2) to a stack model as follows:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_T \end{bmatrix} = \begin{bmatrix} \mathbf{\Pi}^{\mathsf{T}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Pi}^{\mathsf{T}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Pi}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_T \end{bmatrix} + \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_T \end{bmatrix}$$

$$\mathbf{y} = \tilde{\mathbf{\Pi}}^{\mathsf{T}}\mathbf{x} + \mathbf{v}. \tag{5}$$

The GLS transformation matrix $\mathbf{P} \otimes \mathbf{I}_{\mathrm{M}}$ proposed in Keerativibool's paper [24] to correct the AR(2) problem in the error vector $\mathbf{v}$ in (5) is expressed as

$$\mathbf{P} \otimes \mathbf{I}_{\mathrm{M}} = \begin{bmatrix} \mathbf{I}_{\mathrm{M}} & -\rho_1\mathbf{I}_{\mathrm{M}} & -\rho_2\mathbf{I}_{\mathrm{M}} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\mathrm{M}} & -\rho_1\mathbf{I}_{\mathrm{M}} & -\rho_2\mathbf{I}_{\mathrm{M}} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{\mathrm{M}} & -\rho_1\mathbf{I}_{\mathrm{M}} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{\mathrm{M}} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_{\mathrm{M}} & -\rho_1\mathbf{I}_{\mathrm{M}} & -\rho_2\mathbf{I}_{\mathrm{M}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \sqrt{1-\rho_2^2}\,\mathbf{I}_{\mathrm{M}} & -\rho_1\sqrt{\dfrac{1+\rho_2}{1-\rho_2}}\,\mathbf{I}_{\mathrm{M}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \sqrt{\dfrac{(1+\rho_2)\left((1-\rho_2)^2-\rho_1^2\right)}{1-\rho_2}}\,\mathbf{I}_{\mathrm{M}} \end{bmatrix}. \tag{6}$$

The special case of the SEM in (5) is the case of the number of equations equal to one (M = 1), then the SEM can be reduced to the multiple linear regression model and the GLS transformation in (6) can be rewritten as

$$\mathbf{P}=\begin{bmatrix} 1 & -\rho_1 & -\rho_2 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -\rho_1 & -\rho_2 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & -\rho_1 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -\rho_1 & -\rho_2 \\ 0 & 0 & 0 & 0 & \dots & 0 & \sqrt{1-\rho_2^2} & -\rho_1\sqrt{\dfrac{1+\rho_2}{1-\rho_2}} \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & \sqrt{\dfrac{(1+\rho_2)\left((1-\rho_2)^2-\rho_1^2\right)}{1-\rho_2}} \end{bmatrix}. \tag{7}$$

3. Derivations of the proposed model selection criteria

Based on AIC [1-2] and BIC [7], the penalized likelihoods which are the negative log likelihoods plus their penalty term have been proposed as follows:

$$\text{AIC} = -2\log L\left(\hat{\boldsymbol{\theta}}\middle|\mathbf{y}\right) + 2K\ , \tag{8}$$

$$\text{BIC} = -2\log L\left(\hat{\boldsymbol{\theta}}\middle|\mathbf{y}\right) + K\log(T)\ , \tag{9}$$

where $\hat{\boldsymbol{\theta}}$ is the estimator of the parameter vector $\boldsymbol{\theta}$, $L\left(\hat{\boldsymbol{\theta}}\middle|\mathbf{y}\right)$ is the likelihood function corresponding to the candidate model, K represents the dimension of the parameter vector for the model, and T represents the sample size.

The criteria in (8) and (9) are not yet available in the SEM when there is the AR(2) problem, adjustment of the penalty terms are required before. Keerativibool et al. [6] have constructed a system of simultaneous equations AIC, called SAIC, in order to select the best SEM when occurred the AR(2) problem as

$$\text{SAIC} = T\log\left|\hat{\boldsymbol{\Sigma}}\right| + M\left(K+M+5\right), \tag{10}$$

and also constructed an adjusted AIC when the AR(2) problem is ignored as

$$\text{AIC} = T\log\left|\hat{\boldsymbol{\Sigma}}_1\right| + M\left(K+M+1\right), \tag{11}$$

where $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_1$ are the estimated contemporaneous covariance matrices $\boldsymbol{\Sigma}$, which $\hat{\boldsymbol{\Sigma}}_1$ still exists the AR(2) problem.

In this study, we propose two new criteria for selecting the best SEM, that are a system of simultaneous equations BIC, called SBIC, and an adjusted BIC as follows in Theorem 1 and 2, respectively.

Theorem 1. When the AR(2) problem is corrected, a system of simultaneous equations BIC, called SBIC, is defined to be

$$\text{SBIC} = T\log\left|\hat{\boldsymbol{\Sigma}}\right| + KM\left(\log(T)-1\right) + \frac{M(M+5)}{2}\log(T), \tag{12}$$

where $\hat{\boldsymbol{\Sigma}}$ is the unbiased estimator of the contemporaneous covariance matrix $\boldsymbol{\Sigma}$, T is the sample size, M is the number of equations in a SEM, and K is the number of independent variables in each equation.

Theorem 2. When the AR(2) problem is ignored, a Bayesian information criterion for SEM is defined to be

$$\text{BIC} = T\log\left|\hat{\boldsymbol{\Sigma}}_1\right| + KM\left(\log(T)-1\right) + \frac{M(M+1)}{2}\log(T), \tag{13}$$

where $\hat{\boldsymbol{\Sigma}}_1$ is the unbiased estimator of the contemporaneous covariance matrix $\boldsymbol{\Sigma}$ which still includes the AR(2) problem, T is the sample size, M is the number of equations in a SEM, and K is the number of independent variables in each equation.

4. Simulation study

In this simulation study, we use the SAS programming version 9.1 to generate one thousand iterations of a system of three simultaneous equations with four relevant independent variables. Each equation composes of one hundred observations. The steps of simulation and all results are displayed as follows.

1. Generate 100,000 vectors of the 3x1 multivariate normal $\boldsymbol{\varepsilon}_t$ in (4) by the IML procedure, given zero mean vector and contemporaneous covariance matrix $\boldsymbol{\Sigma}$,

$$\boldsymbol{\varepsilon}_t \sim N_3\left(\mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 0.49 & 0.392 & 0.504 \\ 0.392 & 0.64 & 0.648 \\ 0.504 & 0.648 & 0.81 \end{bmatrix}\right);\ t = 1, 2, \ldots, 100{,}000.$$

2. Construct the 3x1 AR(2) and contemporaneously correlated error vectors $\mathbf{v}_t$ in (3), using the multivariate normal vectors $\boldsymbol{\varepsilon}_t$ in Step 1, where the first-two error vectors $\mathbf{v}_t$ are arbitrarily given as

$$\mathbf{v}_{-1} = \begin{bmatrix} 3 & 5 & 7 \end{bmatrix},\ \mathbf{v}_0 = \begin{bmatrix} 4 & 6 & 8 \end{bmatrix},$$

and the first-two autoregressive parameters are arbitrarily given as $\rho_1 = 0.6$, $\rho_2 = -0.5$. Therefore, we have

$$\mathbf{v}_t = 0.6\mathbf{v}_{t-1} - 0.5\mathbf{v}_{t-2} + \boldsymbol{\varepsilon}_t \ ; \ t = 1, 2, \ldots, 100{,}000.$$

Split the series of error vectors $\mathbf{v}_t$ in sequence to preserve the autocorrelation relationship into 1,000 samples, each of which consists of 100 vectors. Then, estimate the autoregressive parameters, test the AR(2) properties, and test the multivariate normality for the residuals by the ARIMA and MODEL procedures, respectively.

3. Generate 100,000 observations of six series of independent variables x_{t2}, x_{t3}, x_{t4}, x_{t5}, x_{t6}, and x_{t7} by the UNIFORM function where the relevant independent variables x_{t2}, x_{t3}, and x_{t4} are $x_{t2} \sim U(1,7)$, $x_{t3} \sim U(2,9)$, $x_{t4} \sim U(3,6)$, and irrelevant independent variables x_{t5}, x_{t6}, and x_{t7} are $x_{t5} \sim U(4,12)$, $x_{t6} \sim U(5,10)$, $x_{t7} \sim U(6,8)$. In this study, x_{t1} is given as a constant which equal to one. Again, split the series of independent variables in sequence into 1,000 samples, each of which consists of 100 observations. Then, test the multicollinearity problem for the series of independent variables.

4. Construct 1,000 samples of the 3x1 dependent vectors $\mathbf{y}_t$ in (2), using the relevant independent variables in Step 3 and the AR(2) error vectors in Step 2, as the following form where the parameters of the model ($\mathbf{\Pi}$) are arbitrarily given,

$$\begin{bmatrix} y_{t1} \\ y_{t2} \\ y_{t3} \end{bmatrix} = \begin{bmatrix} 20 & 5 & 7 & 12 \\ 12 & 4 & 11 & 20 \\ 15 & 6 & 9 & 16 \end{bmatrix} \begin{bmatrix} 1 \\ x_{t2} \\ x_{t3} \\ x_{t4} \end{bmatrix} + \begin{bmatrix} v_{t1} \\ v_{t2} \\ v_{t3} \end{bmatrix}; \ t = 1, 2, \ldots, 100.$$

For each sample, we combine all one hundred observations of $\mathbf{y}_t$, $\mathbf{x}_t$, and $\mathbf{v}_t$ in Steps 4, 3, and 2, respectively, as a stack model in (5).

5. Construct the estimate of GLS transformation matrix $\mathbf{P} \otimes \mathbf{I}_{\mathrm{M}}$ in (6) to eliminate the AR(2) problem of the model in Step 4.

6. Estimate the parameters π of the models before and after transformation in Steps 4 and 5, respectively, by the MODEL procedure. For the model after transformation, examine the AR(2) problem in the residuals and the normality of the residuals by the ARIMA and MODEL procedures, respectively. The tests confirm that the residuals of all 1,000 transformed samples are independent.

7. Compare the values of estimated parameters π and compare their standard deviations of the models before and after transformation by the TTEST procedure. The results in Tables 1 and 2 show that, the averages of all estimated parameters from the

models before and after transformation in 1,000 samples are insignificantly different. Whereas, the averages of all standard deviations of estimated parameters from the model before transformation are greater than the model after transformation at the 5% level of significance.

Table 1. Summary statistics of the estimated parameters from the models before and after transformation in 1,000 samples with the t and p values of the tests.

Equations	True Parameters	Estimated Parameters		Statistics				t-test	p-value
				Average	S.D.	Max	Min		
1st	$\pi_{11}=20$	$\hat{\pi}_{11}$	before	19.999	0.584	21.843	18.419	-0.42	0.6761
			after	19.990	0.372	21.153	18.603		
	$\pi_{21}=5$	$\hat{\pi}_{21}$	before	5.001	0.051	5.180	4.824	0.35	0.7242
			after	5.002	0.033	5.108	4.907		
	$\pi_{31}=7$	$\hat{\pi}_{31}$	before	6.999	0.045	7.141	6.848	0.72	0.4746
			after	7.000	0.028	7.097	6.923		
	$\pi_{41}=12$	$\hat{\pi}_{41}$	before	11.999	0.110	12.348	11.697	0.08	0.9331
			after	12.000	0.067	12.271	11.772		
2nd	$\pi_{12}=12$	$\hat{\pi}_{12}$	before	11.985	0.651	14.198	10.028	0.07	0.9445
			after	11.986	0.420	13.144	10.414		
	$\pi_{22}=4$	$\hat{\pi}_{22}$	before	4.002	0.057	4.217	3.815	0.23	0.8209
			after	4.002	0.038	4.123	3.891		
	$\pi_{32}=11$	$\hat{\pi}_{32}$	before	10.999	0.049	11.158	10.805	0.75	0.4544
			after	11.001	0.032	11.111	10.905		
	$\pi_{42}=20$	$\hat{\pi}_{42}$	before	20.002	0.123	20.393	19.626	-0.52	0.6002
			after	20.000	0.077	20.319	19.731		
3rd	$\pi_{13}=15$	$\hat{\pi}_{13}$	before	15.006	0.752	17.629	12.750	-0.67	0.5048
			after	14.987	0.480	16.360	13.198		
	$\pi_{23}=6$	$\hat{\pi}_{23}$	before	6.003	0.064	6.237	5.806	-0.07	0.9460
			after	6.002	0.043	6.145	5.884		
	$\pi_{33}=9$	$\hat{\pi}_{33}$	before	8.999	0.058	9.176	8.742	0.95	0.3419
			after	9.001	0.037	9.142	8.868		
	$\pi_{43}=16$	$\hat{\pi}_{43}$	before	15.997	0.138	16.467	15.591	0.37	0.7087
			after	15.999	0.087	16.309	15.704		

Table 2. Summary statistics of the standard deviations of estimated parameters from the models before and after transformation in 1,000 samples with the t and p values of the tests.

Equations	Standard Deviations of Estimated Parameters		Statistics				t-test	p-value
			Average	S.D.	Max	Min		
1st	$sd(\hat{\pi}_{11})$	before	0.572	0.062	0.789	0.357	-87.12	< 0.0001
		after	0.367	0.041	0.509	0.258		
	$sd(\hat{\pi}_{21})$	before	0.052	0.005	0.073	0.037	-95.08	< 0.0001
		after	0.033	0.004	0.047	0.023		
	$sd(\hat{\pi}_{31})$	before	0.045	0.005	0.063	0.030	-91.22	< 0.0001
		after	0.028	0.003	0.039	0.018		
	$sd(\hat{\pi}_{41})$	before	0.104	0.011	0.145	0.072	-90.89	< 0.0001
		after	0.066	0.007	0.096	0.049		
2nd	$sd(\hat{\pi}_{12})$	before	0.649	0.074	0.923	0.447	-82.73	< 0.0001
		after	0.418	0.049	0.589	0.273		
	$sd(\hat{\pi}_{22})$	before	0.059	0.006	0.080	0.038	-89.36	< 0.0001
		after	0.037	0.004	0.052	0.025		
	$sd(\hat{\pi}_{32})$	before	0.051	0.006	0.070	0.033	-86.27	< 0.0001
		after	0.032	0.004	0.044	0.022		
	$sd(\hat{\pi}_{42})$	before	0.118	0.012	0.158	0.083	-88.84	< 0.0001
		after	0.075	0.009	0.110	0.051		
3rd	$sd(\hat{\pi}_{13})$	before	0.729	0.081	1.065	0.475	-84.00	< 0.0001
		after	0.470	0.054	0.652	0.315		
	$sd(\hat{\pi}_{23})$	before	0.066	0.007	0.090	0.048	-89.64	< 0.0001
		after	0.042	0.005	0.061	0.029		
	$sd(\hat{\pi}_{33})$	before	0.057	0.006	0.081	0.038	-86.21	< 0.0001
		after	0.036	0.004	0.051	0.024		
	$sd(\hat{\pi}_{43})$	before	0.132	0.014	0.185	0.096	-89.36	< 0.0001
		after	0.085	0.010	0.127	0.059		

8. Calculate the sum of squares errors (SSE) of the models before and after transformation, using the corresponding residuals of the model in Step 6,

$$SSE(before) = \hat{\mathbf{v}}^{\mathsf{T}}\hat{\mathbf{v}} \text{ and } SSE(after) = \hat{\boldsymbol{\varepsilon}}^{\mathsf{T}}\hat{\boldsymbol{\varepsilon}}.$$

Compare the SSE of both models by the TTEST procedure. The results in Table 3 and Figure 1 show that, the average of SSE from the model before transformation in 1,000 samples is greater than the model after transformation at the 5% level of significance. The relative efficient of the SSE's variances before and after transformation is equal to 22.68%, which means that the SSE's variance of the transformed model is less than the one before transformation about 4 times. These results implied that, the SSE values of SEM with AR(2) errors are overestimate.

Table 3. Summary statistics of the SSE from the models before and after transformation in 1,000 samples with the t and p values of the tests.

Statistics	**SSE**	
	before transformation	**after transformation**
Average	296.5040	184.7150
S.D.	50.5266	24.1581
Max	487.4125	285.4189
Min	160.6269	110.9055
t-test	-63.12	
p-value	< 0.0001	
Relative efficient of the SSE_{before} with SSE_{after} $= \frac{Var(SSE_{after})}{Var(SSE_{before})} = 0.2286 \approx \frac{1}{4}$		

9. Calculate SAIC and SBIC in (10) and (12), respectively, for 1,000 transformed samples of 100 observations, using the estimated contemporaneous covariance matrix of the model after transformation, $\hat{\boldsymbol{\Sigma}}$, from the MODEL procedure in Step 6. For each sample, we use SAIC and SBIC to determine which potential independent variables, x_{t2} until x_{t7}, should be included in the model by the criteria of minimum-SAIC and minimum-SBIC. Therefore, the candidate models to consider in this study are equal to 2^6 = 64 models.

10. Calculate AIC and BIC in (11) and (13), respectively, for 1,000 samples of 100 observations, using the estimated contemporaneous covariance matrix of the model before transformation, $\hat{\boldsymbol{\Sigma}}_1$, from the MODEL procedure in Step 6. As in Step 9, for each sample we use AIC and BIC to determine which potential independent variables, x_{t2} until x_{t7}, should be included in the model by considering among 64 candidate models and using the criteria of minimum-AIC and minimum-BIC.

11. The results of Steps 9 and 10 in Table 4 can be concluded that, SBIC convincingly outperformed the other criteria. It correctly chooses the true model 99.0% of the time, compared to a 67.1% correctly selection rate for SAIC. The correctly selection rates of SAIC and SBIC are more than AIC and BIC, respectively, because under the correct specification of the true independent variables, the SSE of the model before transformation tends to overestimate as shown in Table 3 and Figure 1. Comparison of SAIC, SBIC, AIC, and BIC in Table 5 and Figure 2 found that, on the average of 1,000

samples, the model with relevant independent variables x_{t2}, x_{t3}, and x_{t4}, including a constant x_{t1} in the candidate no. 23, has the minimum values of all model selection criteria because it is the generating model.

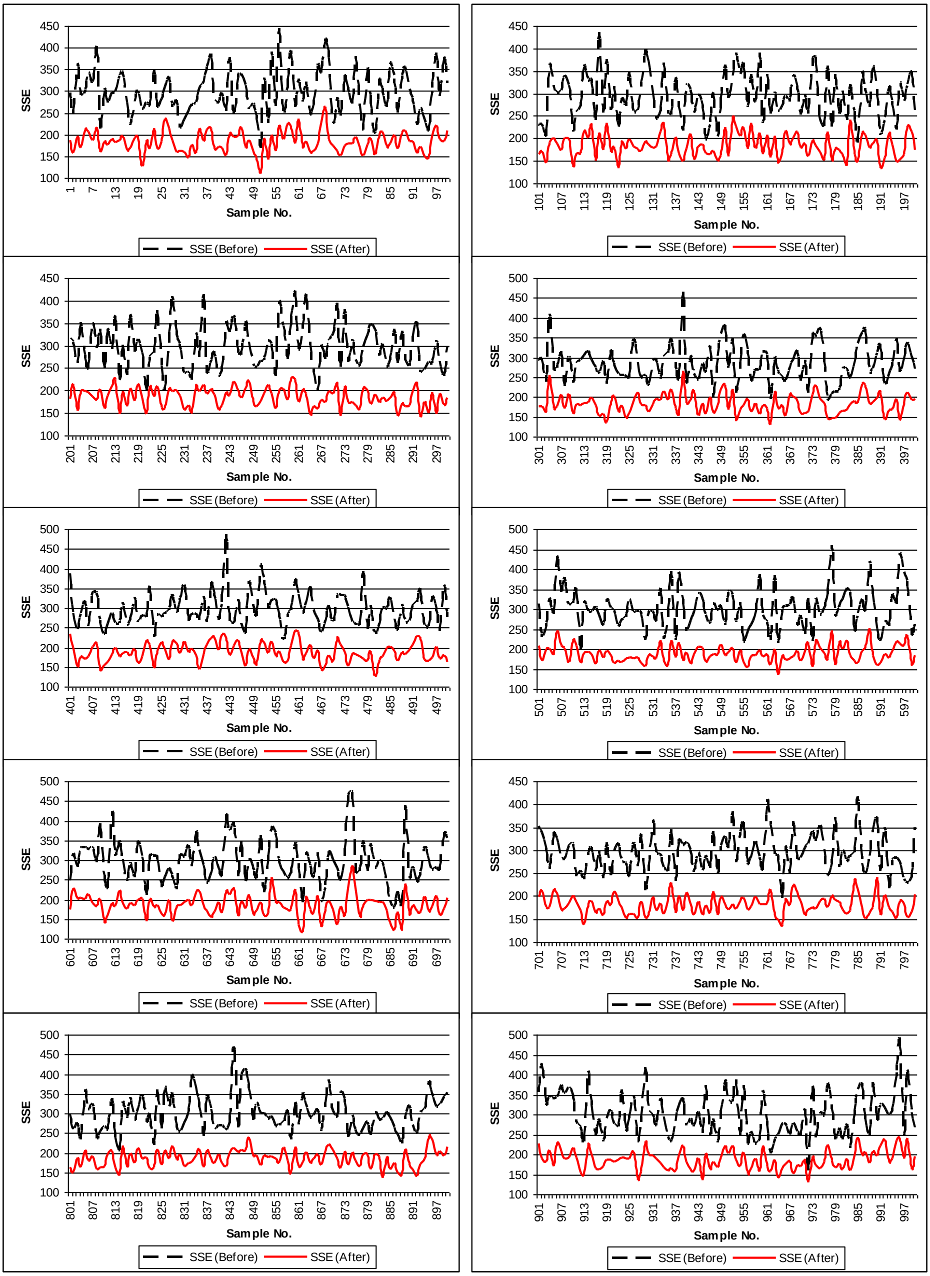

Figure 1. SSE of 1,000 samples from the models before and after transformation.

Table 4. Frequency of order selected by SAIC, SBIC, AIC, and BIC in 1,000 samples.

The number of independent variables in the model	SAIC	SBIC	AIC	BIC
None or only a constant x_{t1}	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3 or True model: x_{t1}, x_{t2}, x_{t3}, and x_{t4}	671	990	639	987
4	294	10	319	13
5	30	0	37	0
6	5	0	5	0

Table 5. Summary statistics of the model selection criteria in 64 candidate models.

Candidate models	Independent variables in the model	SAIC		SBIC		AIC		BIC	
		Average	S.D.	Average	S.D.	Average	S.D.	Average	S.D.
1	x_{t1} or constant equal to one	980.58	30.91	1019.66	30.91	900.45	23.41	923.89	23.41
2	x_{t1} and x_{t2}	501.16	34.72	548.05	34.72	481.59	27.71	512.86	27.71
3	x_{t1} and x_{t3}	842.43	29.14	889.33	29.14	788.62	23.86	819.88	23.86
4	x_{t1} and x_{t4}	888.29	29.71	935.19	29.71	833.50	24.21	864.76	24.21
5	x_{t1} and x_{t5}	981.45	31.16	1028.34	31.16	903.48	23.61	934.74	23.61
6	x_{t1} and x_{t6}	981.61	30.89	1028.50	30.89	903.31	23.57	934.57	23.57
7	x_{t1} and x_{t7}	985.69	31.00	1032.59	31.00	903.32	23.57	934.58	23.57
8	x_{t1}, x_{t2} and x_{t3}	302.12	26.46	356.83	26.46	338.24	28.61	377.31	28.61
9	x_{t1}, x_{t2} and x_{t4}	352.10	26.92	406.81	26.92	387.16	28.73	426.24	28.73
10	x_{t1}, x_{t2} and x_{t5}	493.23	29.04	547.94	29.04	484.59	27.98	523.67	27.98
11	x_{t1}, x_{t2} and x_{t6}	493.23	28.84	547.93	28.84	484.35	27.80	523.43	27.80
12	x_{t1}, x_{t2} and x_{t7}	496.66	29.04	551.37	29.04	484.49	27.92	523.56	27.92
13	x_{t1}, x_{t3} and x_{t4}	180.83	27.39	235.54	27.39	215.97	29.61	255.05	29.61
14	x_{t1}, x_{t3} and x_{t5}	843.70	29.36	898.41	29.36	791.63	24.09	830.71	24.09
15	x_{t1}, x_{t3} and x_{t6}	843.73	29.23	898.44	29.23	791.50	24.01	830.58	24.01
16	x_{t1}, x_{t3} and x_{t7}	847.41	29.23	902.12	29.23	791.55	24.03	830.63	24.03
17	x_{t1}, x_{t4} and x_{t5}	889.52	29.82	944.23	29.82	836.58	24.40	875.66	24.40
18	x_{t1}, x_{t4} and x_{t6}	889.79	29.73	944.50	29.73	836.46	24.37	875.54	24.37
19	x_{t1}, x_{t4} and x_{t7}	893.43	29.86	948.14	29.86	836.37	24.36	875.45	24.36
20	x_{t1}, x_{t5} and x_{t6}	982.46	31.10	1037.17	31.10	906.32	23.75	945.40	23.75
21	x_{t1}, x_{t5} and x_{t7}	986.55	31.29	1041.26	31.29	906.36	23.78	945.43	23.78
22	x_{t1}, x_{t6} and x_{t7}	986.73	30.98	1041.44	30.98	906.18	23.75	945.26	23.75
23	**x_{t1}, x_{t2}, x_{t3} and x_{t4}**	**-378.08**	**25.71**	**-315.55**	**25.71**	**-253.86**	**32.79**	**-206.97**	**32.79**
24	x_{t1}, x_{t2}, x_{t3} and x_{t5}	304.22	26.47	366.74	26.47	341.18	28.87	388.07	28.87
25	x_{t1}, x_{t2}, x_{t3} and x_{t6}	304.12	26.57	366.64	26.57	341.03	28.79	387.93	28.79
26	x_{t1}, x_{t2}, x_{t3} and x_{t7}	306.63	26.72	369.16	26.72	341.16	28.81	388.06	28.81
27	x_{t1}, x_{t2}, x_{t4} and x_{t5}	354.20	26.90	416.72	26.90	390.19	28.92	437.08	28.92
28	x_{t1}, x_{t2}, x_{t4} and x_{t6}	354.29	26.92	416.81	26.92	390.02	28.90	436.91	28.90
29	x_{t1}, x_{t2}, x_{t4} and x_{t7}	356.80	27.07	419.32	27.07	390.04	28.84	436.93	28.84
30	x_{t1}, x_{t2}, x_{t5} and x_{t6}	494.72	28.92	557.24	28.92	487.33	28.07	534.22	28.07
31	x_{t1}, x_{t2}, x_{t5} and x_{t7}	498.14	29.11	560.66	29.11	487.47	28.20	534.37	28.20
32	x_{t1}, x_{t2}, x_{t6} and x_{t7}	498.17	28.89	560.69	28.89	487.24	28.02	534.13	28.02
33	x_{t1}, x_{t3}, x_{t4} and x_{t5}	182.98	27.55	245.50	27.55	219.01	29.85	265.90	29.85
34	x_{t1}, x_{t3}, x_{t4} and x_{t6}	183.02	27.52	245.55	27.52	218.86	29.82	265.75	29.82
35	x_{t1}, x_{t3}, x_{t4} and x_{t7}	185.35	27.53	247.87	27.53	218.95	29.77	265.84	29.77
36	x_{t1}, x_{t3}, x_{t5} and x_{t6}	844.98	29.44	907.50	29.44	794.49	24.22	841.38	24.22
37	x_{t1}, x_{t3}, x_{t5} and x_{t7}	848.72	29.51	911.25	29.51	794.57	24.27	841.46	24.27
38	x_{t1}, x_{t3}, x_{t6} and x_{t7}	848.74	29.41	911.27	29.41	794.44	24.20	841.33	24.20
39	x_{t1}, x_{t4}, x_{t5} and x_{t6}	891.02	29.83	953.54	29.83	839.53	24.54	886.42	24.54
40	x_{t1}, x_{t4}, x_{t5} and x_{t7}	894.66	29.93	957.18	29.93	839.44	24.57	886.33	24.57
41	x_{t1}, x_{t4}, x_{t6} and x_{t7}	894.89	29.81	957.42	29.81	839.33	24.56	886.22	24.56
42	x_{t1}, x_{t5}, x_{t6} and x_{t7}	987.56	31.13	1050.09	31.13	909.19	23.94	956.08	23.94
43	x_{t1}, x_{t2}, x_{t3} x_{t4} and x_{t5}	-375.08	25.71	-304.74	25.71	-250.90	33.03	-196.19	33.03
44	x_{t1}, x_{t2}, x_{t3}, x_{t4} and x_{t6}	-375.19	25.87	-304.85	25.87	-251.10	32.93	-196.39	32.93
45	x_{t1}, x_{t2}, x_{t3}, x_{t4} and x_{t7}	-375.20	25.93	-304.86	25.93	-250.90	32.97	-196.19	32.97
46	x_{t1}, x_{t2}, x_{t3}, x_{t5} and x_{t6}	306.17	26.62	376.51	26.62	343.95	29.05	398.66	29.05
47	x_{t1}, x_{t2}, x_{t3}, x_{t5} and x_{t7}	308.75	26.75	379.09	26.75	344.11	29.06	398.82	29.06
48	x_{t1}, x_{t2}, x_{t3}, x_{t6} and x_{t7}	308.64	26.83	378.98	26.83	343.95	28.99	398.66	28.99
49	x_{t1}, x_{t2}, x_{t4}, x_{t5} and x_{t6}	356.37	26.91	426.71	26.91	393.03	29.09	447.74	29.09
50	x_{t1}, x_{t2}, x_{t4}, x_{t5} and x_{t7}	358.91	27.08	429.25	27.08	393.05	29.02	447.76	29.02
51	x_{t1}, x_{t2}, x_{t4}, x_{t6} and x_{t7}	359.00	27.12	429.34	27.12	392.88	29.03	447.59	29.03
52	x_{t1}, x_{t2}, x_{t5}, x_{t6} and x_{t7}	499.69	28.99	570.03	28.99	490.21	28.30	544.92	28.30
53	x_{t1}, x_{t3}, x_{t4}, x_{t5} and x_{t6}	185.17	27.69	255.51	27.69	221.89	30.07	276.60	30.07
54	x_{t1}, x_{t3}, x_{t4}, x_{t5} and x_{t7}	187.46	27.70	257.80	27.70	221.97	30.01	276.68	30.01
55	x_{t1}, x_{t3}, x_{t4}, x_{t6} and x_{t7}	187.53	27.66	257.87	27.66	221.83	30.00	276.54	30.00
56	x_{t1}, x_{t3}, x_{t5}, x_{t6} and x_{t7}	849.93	29.61	920.27	29.61	797.43	24.41	852.13	24.41

Table 5. (Continued)

Candidate models	Independent variables in the model	SAIC		SBIC		AIC		BIC	
		Average	S.D.	Average	S.D.	Average	S.D.	Average	S.D.
57	x_{t1}, x_{t4}, x_{t5}, x_{t6} and x_{t7}	896.18	29.92	966.52	29.92	842.38	24.74	897.09	24.74
58	x_{t1}, x_{t2}, x_{t3}, x_{t4}, x_{t5} and x_{t6}	-372.21	25.87	-294.06	25.87	-248.14	33.17	-185.62	33.17
59	x_{t1}, x_{t2}, x_{t3}, x_{t4}, x_{t5} and x_{t7}	-372.22	25.94	-294.06	25.94	-247.94	33.20	-185.42	33.20
60	x_{t1}, x_{t2}, x_{t3}, x_{t4}, x_{t6} and x_{t7}	-372.33	26.09	-294.17	26.09	-248.14	33.12	-185.62	33.12
61	x_{t1}, x_{t2}, x_{t3}, x_{t5}, x_{t6} and x_{t7}	310.72	26.90	388.87	26.90	346.87	29.24	409.40	29.24
62	x_{t1}, x_{t2}, x_{t4}, x_{t5}, x_{t6} and x_{t7}	361.08	27.08	439.24	27.08	395.87	29.20	458.40	29.20
63	x_{t1}, x_{t3}, x_{t4}, x_{t5}, x_{t6} and x_{t7}	189.65	27.83	267.81	27.83	224.85	30.25	287.37	30.25
64	x_{t1}, x_{t2}, x_{t3}, x_{t4}, x_{t5}, x_{t6} and x_{t7}	-369.34	26.11	-283.37	26.11	-245.20	33.35	-174.86	33.35

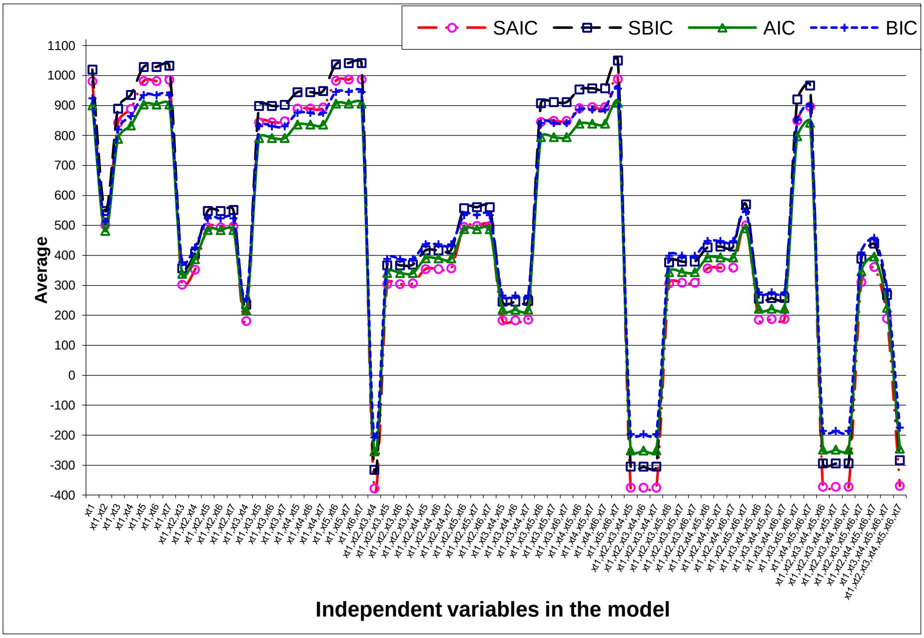

Figure 2. Averages of the model selection criteria in 1,000 samples.

5. Conclusion, discussion, and further study

This study would like to show that, the AR(2) problem makes the overestimated values of the model selection criteria and then affects the performance of selecting the true model. Two model selection criteria are proposed to select the most appropriate SEM. Firstly, SBIC, is proposed after correcting the AR(2) problem in the errors by the GLS transformation. Secondly, the original BIC is slightly adjusted when the AR(2) problem is ignored. If there is no AR(2) problem in the errors, SBIC reduces to BIC. The results of simulation can be concluded as follows. The averages of the estimated parameters from the models before and after transformation are insignificantly different. Whereas, the averages of the standard deviations of estimated parameters from the model before transformation are greater than the model after transformation at the 5%

level of significance. The errors after using the GLS transformation are white noises and that before removing the AR(2) errors; the average of SSE is greater than the one after the AR(2) problem has been corrected, at the 5% level of significance. The relative efficient of the SSE's variances before and after transformation is equal to 22.68%, which means that the SSE's variance of the transformed model is less than the one before transformation about 4 times. These results implied that, the SSE values of SEM with AR(2) errors are overestimate. Comparing the correctly selection performance of the proposed model selection criteria, SBIC and BIC, with the model selection criteria proposed by Keerativibool et al. [6], SAIC and AIC, found that SBIC convincingly outperformed the other criteria and the rest of the criteria can be ordered according to their performance by BIC, SAIC, and AIC. On the average of 1,000 iterations, the SEM with relevant independent variables and constant term has the minimum values of all model selection criteria because it is the generating model.

AIC and BIC are most popular model selection criteria. While AIC is an asymptotically efficient criterion which motivated by minimized MSE, BIC is a consistent criterion which developed from the Bayesian idea that the model with the largest posterior probability should be chosen. Consequently, AIC is inconsistent whereas BIC is consistent and then the new criteria, SBIC and BIC perform favorably than SAIC and AIC. When there is the AR(2) problem in a SEM, AIC and BIC provide the inflated values and then should not be used. Hence, SBIC is a recommended criterion when the AR(2) problem is found.

Nowadays, there is not much the criterion to select the best SEM. Therefore, it should be studied and established the other criteria, such as Kullback information criterion (KIC), the corrected versions of AIC, BIC, and KIC for small sample cases. Including, other schema of the error-generation might also be considered, such as the moving average, autoregressive and moving average schemes instead of only the autoregressive scheme.

Acknowledgements

This article is (partially) supported by the Centre of Excellence in Mathematics, the Commission of Higher Education, Thailand. The author is thankful for the help and support of several people in my organization, Thaksin University.

References

[1] Akaike, H., Information Theory and an Extension of the Maximum Likelihood Principle, In *2nd International Symposium on Information Theory*, B.N. Petrov and F. Csaki, eds. Akademiai Kiado, Budapest, 1973: 267-281.

[2] Akaike, H., A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, 1974; 19: 716-723.

[3] Hurvich, C.M., and Tsai, C.L., Regression and Time Series Model Selection in Small Samples, *Biometrika*, 1989; 76: 297-307.

[4] Bedrick, E.J., and Tsai, C.L., Model Selection for Multivariate Regression in Small Samples, *Biometrics*, 1994; 50: 226-231.

[5] Fujikoshi, Y., and Satoh, K., Modified AIC and C_p in Multivariate Linear Regression, *Biometrika*, 1997; 84: 707-716.

[6] Keerativibool, W., Jitthavech, J., and Lorchirachoonkul, V., Model Selection in a System of Simultaneous Equations Model, *Communications in Statistics-Theory and Methods*, 2011; 40: 373-393.

[7] Schwarz, G., Estimating the Dimension of a Model, *The Annals of Statistics*, 1978; 6: 461-464.

[8] Neath, A., and Cavanaugh, J.E., Regression and Time Series Model Selection using Variants of the Schwarz Information Criterion, *Communications in Statistics-Theory and Methods*, 1997; 26: 559-580.

[9] McQuarrie, A.D.R., A Small-Sample Correction for the Schwarz SIC Model Selection Criterion, *Statistics & Probability Letters*, 1999; 44: 79-86.

[10] Hannan, E.J., and Quinn, B.G., The Determination of the Order of an Autoregression, *Journal of the Royal Statistical Society: Series B (Methodological),* 1979; 41: 190-195.

[11] McQuarrie, A.D.R., and Tsai, C.L., *Regression and Time Series Model Selection*, Singapore, World Scientific, 1998.

[12] Cavanaugh, J.E., A Large-Sample Model Selection Criterion based on Kullback's Symmetric Divergence, *Statistics & Probability Letters*, 1999; 42: 333-343.

[13] Cavanaugh, J.E., Criteria for Linear Model Selection based on Kullback's Symmetric Divergence, *Australian & New Zealand Journal of Statistics,* 2004; 46: 257-274.

[14] Seghouane, A.K., and Bekara, M., A Small Sample Model Selection Criterion based on Kullback's Symmetric Divergence, *IEEE Transactions on Signal Processing*, 2004; 52: 3314-3323.

[15] Kim, H.J., and Cavanaugh, J.E., Model Selection Criteria based on Kullback Information Measures for Nonlinear Regression, *Journal of Statistical Planning and Inference*, 2005; 134: 332-349.

[16] Seghouane, A.K., Bekara, M., and Fleury, G., A Criterion for Model Selection in the Presence of Incomplete Data based on Kullback's Symmetric Divergence, *Signal Processing*, 2005; 85: 1405-1417.

[17] Hafidi, B., A Small-Sample Criterion based on Kullback's Symmetric Divergence for Vector Autoregressive Modeling, *Statistics & Probability Letters*, 2006; 76: 1647-1654.

[18] Hafidi, B., and Mkhadri, A., A corrected Akaike Criterion based on Kullback's Symmetric Divergence: Applications in Time Series, Multiple and Multivariate Regression, *Computational Statistics & Data Analysis*, 2006; 50: 1524-1550.

[19] Seghouane, A.K., Multivariate Regression Model Selection from Small Samples using Kullback's Symmetric Divergence, *Signal Processing*, 2006; 86: 2074-2084.

[20] Giombini, G., and Szroeter, J., Quasi Akaike and Quasi Schwarz Criteria for Model Selection: A Surprising Consistency Result, *Economics Letters*, 2007; 95: 259-266.

[21] Keerativibool, W., Jitthavech, J., and Lorchirachoonkul, V., Autocorrelation Correction in a Regression Model, *Proceedings: Applied Statistics Conference (ASCONF2009)*, Pattaya, Chon Buri, Thailand, 2009: 519-533.

[22] Keerativibool, W., Jitthavech, J., and Lorchirachoonkul, V., Autocorrelation Correction in a Simultaneous Equations Model, *Proceedings: the 5th International Conference on Mathematics, Statistics and their Applications (ICMSA2009)*, West Sumatra, Indonesia, 2009: 192-197.

[23] Keerativibool, W., Moving Average Correction in a Regression Model, *Thailand Statistician: Journal of Thai Statistical Association*, 2010; 8: 63-80.

[24] Keerativibool, W., Generalized Least Squares Transformation with the Second-Order Autoregressive Error, *Thailand Statistician: Journal of Thai Statistical Association*, 2011; 9: 77-92.

[25] Keerativibool, W., Simultaneous Equations Model Selection using Variants of the Akaike Information Criterion, *Proceedings: the 14th International Conference on Applied Stochastic Models and Data Analysis (ASMDA2011)*, Rome, Italy, 2011: 669-676.

[26] Greene, W., *Econometric Analysis,* 6th ed., Upper Saddle River, New Jersey, Prentice Hall, 2008.

[27] Box, G.E.P., Jenkins, G.M., and Reinsel, G.C., *Time Series Analysis: Forecasting and Control*, 3rd ed., Englewood Cliffs, New Jersey, Prentice Hall, 1994.
[28] Johnson, R.A., and Wichern, D.W., *Applied Multivariate Statistical Analysis*, 4th ed., Englewood Cliffs, New Jersey, Prentice Hall, 1998.

Appendix

Proof of Theorem 1. Since, the reduced-form model in (5) after using the GLS transformation in (6) satisfies the usual assumptions of multivariate model that is the mean of error vector is insignificantly different from zero and does not have the AR(2) problem, but the contemporaneously correlated errors still exist [24],

$$\left(\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}\right)\mathbf{y}=\left(\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}\right)\tilde{\mathbf{\Pi}}^{\mathsf{T}}\mathbf{x}+\boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}=\left(\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}\right)\mathbf{v}$ and $\boldsymbol{\varepsilon}\sim N_{TM}\left(\mathbf{0},\ \left(\mathbf{I}_{\mathrm{T}}\otimes\mathbf{\Sigma}\right)\right)$.

Therefore, we may apply the multivariate normal density [28] to yield the likelihood,

$$L\left(\boldsymbol{\varepsilon},\left(\mathbf{I}_{\mathrm{T}}\otimes\mathbf{\Sigma}\right)\right)=\left(2\pi\right)^{-\frac{TM}{2}}\left|\mathbf{I}_{\mathrm{T}}\otimes\mathbf{\Sigma}\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}^{\mathsf{T}}\left(\mathbf{I}_{\mathrm{T}}\otimes\mathbf{\Sigma}\right)^{-1}\boldsymbol{\varepsilon}\right). \qquad \text{(A.1)}$$

The determinant of contemporaneous covariance $\left(\mathbf{I}_{\mathbf{T}}\otimes\mathbf{\Sigma}\right)$ in (A.1) is equal to

$$\left|\mathbf{I}_{\mathrm{T}}\otimes\mathbf{\Sigma}\right|=\left|\mathbf{I}_{\mathrm{T}}\right|^{M}\times\left|\mathbf{\Sigma}\right|^{T}=\left|\mathbf{\Sigma}\right|^{T}. \qquad \text{(A.2)}$$

Using (A.2), we have the log of likelihood function in (A.1) as follows:

$$\log L\left(\boldsymbol{\varepsilon},\left(\mathbf{I}_{\mathrm{T}}\otimes\mathbf{\Sigma}\right)\right)=-\frac{TM}{2}\log\left(2\pi\right)-\frac{T}{2}\log\left|\mathbf{\Sigma}\right|-\frac{1}{2}\boldsymbol{\varepsilon}^{\mathsf{T}}\left(\mathbf{I}_{\mathrm{T}}\otimes\mathbf{\Sigma}\right)^{-1}\boldsymbol{\varepsilon}. \qquad \text{(A.3)}$$

Given a vector of unbiased estimators $\hat{\boldsymbol{\theta}}=\left(\hat{\boldsymbol{\varepsilon}},\ \hat{\mathbf{\Sigma}}\right)$, the minus twice of the log-likelihood function in (A.3) can be written as

$$-2\log L\left(\hat{\boldsymbol{\varepsilon}},\left(\mathbf{I}_{\mathrm{T}}\otimes\hat{\mathbf{\Sigma}}\right)\right)=TM\log\left(2\pi\right)+T\log\left|\hat{\mathbf{\Sigma}}\right|+\hat{\boldsymbol{\varepsilon}}^{\mathsf{T}}\left(\mathbf{I}_{\mathrm{T}}\otimes\hat{\mathbf{\Sigma}}\right)^{-1}\hat{\boldsymbol{\varepsilon}}. \qquad \text{(A.4)}$$

Consider the last term in (A.4),

$$\hat{\boldsymbol{\varepsilon}}^{\mathsf{T}}\left(\mathbf{I}_{\mathrm{T}}\otimes\hat{\mathbf{\Sigma}}\right)^{-1}\hat{\boldsymbol{\varepsilon}}=\begin{bmatrix}\hat{\boldsymbol{\varepsilon}}_1^{\mathsf{T}} & \hat{\boldsymbol{\varepsilon}}_2^{\mathsf{T}} & \dots & \hat{\boldsymbol{\varepsilon}}_T^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{\Sigma}}^{-1} & \mathbf{0} & \dots & \mathbf{0}\\ \mathbf{0} & \hat{\mathbf{\Sigma}}^{-1} & \dots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \dots & \hat{\mathbf{\Sigma}}^{-1}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\varepsilon}}_1\\ \hat{\boldsymbol{\varepsilon}}_2\\ \vdots\\ \hat{\boldsymbol{\varepsilon}}_T\end{bmatrix}$$

$$= \sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t^{\mathsf{T}} \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\varepsilon}}_t = \sum_{t=1}^{T} \left(\sum_{j=1}^{M} \hat{\varepsilon}_{tj}^2 \hat{\sigma}^{jj} + 2 \sum_{i<j}^{M} \hat{\varepsilon}_{ti} \hat{\varepsilon}_{tj} \hat{\sigma}^{ij} \right), \qquad \text{(A.5)}$$

where $\hat{\sigma}^{ij}$ represents the element of the ith row and jth column of $\hat{\boldsymbol{\Sigma}}^{-1}$.

Using the fact that, the unbiased estimators of σ_{jj} and σ_{ij} are, respectively,

$$\hat{\sigma}_{jj} = \frac{1}{T-K} \sum_{t=1}^{T} \hat{\varepsilon}_{tj}^2 \text{ and } \hat{\sigma}_{ij} = \frac{1}{T-K} \sum_{t=1}^{T} \hat{\varepsilon}_{ti} \hat{\varepsilon}_{tj} .$$

Then, (A.5) becomes

$$\hat{\boldsymbol{\varepsilon}}^{\mathsf{T}} \left(\mathbf{I_T} \otimes \hat{\boldsymbol{\Sigma}} \right)^{-1} \hat{\boldsymbol{\varepsilon}} = (T-K) \left(\sum_{j=1}^{M} \hat{\sigma}_{jj} \hat{\sigma}^{jj} + 2 \sum_{i<j}^{M} \hat{\sigma}_{ij} \hat{\sigma}^{ij} \right)$$

$$= (T-K) \Big(\left(\hat{\sigma}_{11}\hat{\sigma}^{11} + \hat{\sigma}_{22}\hat{\sigma}^{22} + \ldots + \hat{\sigma}_{MM}\hat{\sigma}^{MM} \right) + 2 \big(\hat{\sigma}_{12}\hat{\sigma}^{12} + \hat{\sigma}_{13}\hat{\sigma}^{13} + \ldots + \hat{\sigma}_{1M}\hat{\sigma}^{1M}$$

$$+ \hat{\sigma}_{23}\hat{\sigma}^{23} + \hat{\sigma}_{24}\hat{\sigma}^{24} + \ldots + \hat{\sigma}_{2M}\hat{\sigma}^{2M} + \ldots + \hat{\sigma}_{(M-1),M}\hat{\sigma}^{(M-1),M} \big) \Big)$$

$$= (T-K) \Big(\left(\hat{\sigma}_{11}\hat{\sigma}^{11} + \hat{\sigma}_{12}\hat{\sigma}^{12} + \ldots + \hat{\sigma}_{1M}\hat{\sigma}^{1M} \right) + \left(\hat{\sigma}_{12}\hat{\sigma}^{12} + \hat{\sigma}_{22}\hat{\sigma}^{22} + \ldots + \hat{\sigma}_{2M}\hat{\sigma}^{2M} \right)$$

$$+ \ldots + \left(\hat{\sigma}_{1M}\hat{\sigma}^{1M} + \hat{\sigma}_{2M}\hat{\sigma}^{2M} + \ldots + \hat{\sigma}_{MM}\hat{\sigma}^{MM} \right) \Big). \qquad \text{(A.6)}$$

Since the estimated contemporaneous covariance matrix, $\hat{\boldsymbol{\Sigma}}$, is symmetric, we have $\hat{\sigma}_{ij} = \hat{\sigma}_{ji}$, including $\hat{\sigma}^{ij} = \hat{\sigma}^{ji}$. Then,

$$\hat{\boldsymbol{\varepsilon}}^{\mathsf{T}} \left(\mathbf{I}_T \otimes \hat{\boldsymbol{\Sigma}} \right)^{-1} \hat{\boldsymbol{\varepsilon}} = (T-K) \left(\sum_{j=1}^{M} \hat{\sigma}_{1j} \hat{\sigma}^{j1} + \sum_{j=1}^{M} \hat{\sigma}_{2j} \hat{\sigma}^{j2} + \ldots + \sum_{j=1}^{M} \hat{\sigma}_{Mj} \hat{\sigma}^{jM} \right). \qquad \text{(A.7)}$$

Due to the elements $\sum_{j=1}^{M} \hat{\sigma}_{ij} \hat{\sigma}^{ji}$; $i = 1, 2, \ldots, M$ in (A.7) are the elements in main diagonal of the matrix $\hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\Sigma}}^{-1} = \mathbf{I}_M$. Therefore, (A.7) is equivalent to

$$\hat{\boldsymbol{\varepsilon}}^{\mathsf{T}} \left(\mathbf{I}_T \otimes \hat{\boldsymbol{\Sigma}} \right)^{-1} \hat{\boldsymbol{\varepsilon}} = (T-K) M . \qquad \text{(A.8)}$$

Replacing (A.8) into (A.4) yields the minus twice log-likelihood function as

$$-2 \log L \left(\hat{\boldsymbol{\varepsilon}}, \left(\mathbf{I}_T \otimes \hat{\boldsymbol{\Sigma}} \right) \right) = TM \left(\log(2\pi) + 1 \right) - KM + T \log \left| \hat{\boldsymbol{\Sigma}} \right| . \qquad \text{(A.9)}$$

Dropping the first term in (A.9), $TM\left(\log(2\pi)+1\right)$, which has no effect on the minimum-SBIC, and use the concept of BIC in (9) to construct a system of simultaneous equations BIC, called SBIC,

$$\mathrm{SBIC} = T\log\left|\hat{\boldsymbol{\Sigma}}\right| + \left(KM + \frac{M(M+1)}{2} + 2M\right)\log(T) - KM\ , \tag{A.10}$$

where $\left(KM + \frac{M(M+1)}{2} + 2M\right)\log(T)$ is the bias adjustment, KM represents the number of independent variable in a SEM, $\frac{M(M+1)}{2}$ represents the number of parameters in contemporaneous covariance matrix $\boldsymbol{\Sigma}$, and $2M$ represents the number of autoregressive parameters in a SEM. Combining the last two terms in (A.10),

$$\left(KM + \frac{M(M+1)}{2} + 2M\right)\log(T) - KM = KM\left(\log(T) - 1\right) + \frac{M(M+5)}{2}\log(T)\ . \tag{A.11}$$

Replacing (A.11) into (A.10) yields SBIC in (12).

Proof of Theorem 2. From the minus twice log-likelihood function proposed in [6],

$$-2\log L\left(\hat{\boldsymbol{\Pi}},\ \hat{\boldsymbol{\Sigma}}_1\right) = TM\left(\log(2\pi) + 1\right) + T\log\left|\hat{\boldsymbol{\Sigma}}_1\right| - KM\ . \tag{A.12}$$

Dropping the first term in (A.12), $TM\left(\log(2\pi)+1\right)$, which has no effect on the minimum-BIC, and then use the concept of BIC in (9) to construct an adjusted BIC in order to use in a SEM,

$$\mathrm{BIC} = T\log\left|\hat{\boldsymbol{\Sigma}}_1\right| + \left(KM + \frac{M(M+1)}{2}\right)\log(T) - KM\ , \tag{A.13}$$

where $\left(KM + \frac{M(M+1)}{2}\right)\log(T)$ is the bias adjustment as in (A.10), except the number of autoregressive parameters $2M$. Combine the last two terms in (A.13),

$$\left(KM + \frac{M(M+1)}{2}\right)\log(T) - KM = KM\left(\log(T) - 1\right) + \frac{M(M+1)}{2}\log(T)\ . \tag{A.14}$$

Replacing (A.14) into (A.13) yields BIC in (13).