Thailand Statistician
April 2025; 23(2): 368-376
http://statassoc.or.th
Contributed paper

Akaike Information Criterion for a Simultaneous Equations Model using Unbiased Estimator

Warangkhana Riansut
Department of Mathematics and Statistics, Faculty of Science and Digital Innovation,
Thaksin University, Phatthalung, Thailand.
Corresponding author; e-mail: warang27@gmail.com

Received: 13 March 2024
Revised: 25 June 2024
Accepted: 9 July 2024

Abstract

This article introduces a novel approach in the fields of econometrics and statistics. A new criterion for selection in a simultaneous equation model (SEM) based on the Akaike information criterion using an unbiased estimator (UE) called SAIC(UE). This innovative method replaces the traditional maximum likelihood estimator (MLE) with UE, offering a fresh perspective on model selection. A comparison of the effectiveness of SAIC(UE) with SAIC (Keerativibool et al. 2011) by the extensive simulation study and the observed L_2 performance found that SAIC(UE) is the best criterion because it had the maximum percentage of correct model selections, the maximum average of the observed L_2 performance, and the minimum standard deviation of the observed L_2 performance in all situations of the simulation study. SAIC has a more negative bias than SAIC(UE) because the contemporaneous covariance matrix of error terms obtained from MLE is smaller than that obtained from UE or underestimated. This reason causes the correct order of the model from SAIC to be less than SAIC(UE).

Keywords: Model selection criterion, performance, simultaneous equations model, unbiased estimator, maximum likelihood estimator.

1. Introduction

Model selection criteria have become important in selecting appropriate variables for constructing a regression equation or system of equations, such as simultaneous equation models (SEM) since not all available predictor variables are appropriate for explaining the variation in the response variable as well as predicting the new response outcome. There are now many model selection criteria in regression analysis, for example, Akaike information criterion, AIC (Akaike 1974), Bayesian information criterion, BIC (Schwarz 1978), Hannan and Quinn criterion, HQ (Hannan and Quinn 1979), modified version of AIC (AICc) (Hurvich and Tsai 1989), Kullback information criterion, KIC (Cavanaugh 1999), modified version of KIC (KICc) (Cavanaugh 2004), Pham information criterion, PIC (Pham 2019), a new method for regression model selection uses the p-value for t-test and F-test for the variables selection criteria and the adjusted R-squared or AIC for the best model selection criterion (Niyonzima 2020), and upgrading model selection criteria with goodness of fit tests for

practical applications (Rossi et al. 2020). AIC is the most commonly used and well-known criterion among all the discussed criteria. It was found that several model selection criteria were used for regression models, while there are still few studies and developments of model selection criteria used for the SEM. SEM is a statistical technique widely used in economic science to model the simultaneity relationship between variables (Perez-Sanchez et al. 2021). External explanatory variables will become random and be related to the error terms of the equations appearing as explanatory variables. An example of SEM was a study of planting costs and benefits gained from selling rubber in Pa Phayom District, Phatthalung Province and Thung Song District, Nakhon Si Thammarat Province. It was found that the cost variable, which was the dependent variable in the first equation, was the independent variable of the benefits gained equation in the next equation (Keerativibool 2014). Keerativibool et al. (2011) proposed a system of simultaneous equations AIC (SAIC) by developing a model selection criterion using the maximum likelihood estimator (MLE). Many studies (Neath and Cavanaugh 1997, Cavanaugh 1999, Giombini and Szroeter 2007) have found that AIC was an asymptotically efficient criterion. Then, AIC selected the model with the minimum mean square error (MSE) in a large sample. Thus, when the generation model had a finite order and was presented among the candidate families under the consideration model, the efficient criterion as AIC was, therefore, an inconsistent criterion and tends, asymptotically, to overestimate the dimensionality of the parameter vectors for the model. Therefore, McQuarrie et al. (1997) replaced the MLE of variance of errors $\left(\tilde{\sigma}^2\right)$ with an unbiased estimator (UE) $\left(\hat{\sigma}^2\right)$ in AICc, yields AICu had a greater penalty for overfitting, especially as the sample size increased, where

$$\text{AICc} = \log\left(\tilde{\sigma}^2\right) + \frac{2\left(p+1\right)}{n-p-2}, \tag{1}$$

$$\text{AICu} = \log\left(\hat{\sigma}^2\right) + \frac{2\left(p+1\right)}{n-p-2}, \tag{2}$$

$$\tilde{\sigma}^2 = \frac{1}{n}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right),$$

$$\hat{\sigma}^2 = \frac{1}{n-p}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right) = \frac{n\tilde{\sigma}^2}{n-p},$$

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y},$$

$\mathbf{y}$ is an $n\times 1$ dependent random vector of observations, $\mathbf{X}$ is an $n\times p$ matrix of independent variables with full-column rank, $\boldsymbol{\beta}$ is a $p\times 1$ parameter vector of regression coefficients, n is the sample size, and p is the number of parameters in the approximate model, including the constant. In our rigorous research, we have replaced the MLE of the contemporaneous covariance matrix of error terms with UE into the SAIC formula (Keeratibool et al. 2011), creating the SAIC(UE). We then conducted a comprehensive simulation study to compare the performance of SAIC(UE) with SAIC under three sample sizes: small samples T = 15, 20 and 25, medium samples T = 30 and 50, and large samples T = 100 and 500. The performance of both criteria was verified by the corresponding measure, which is the measure of the percentage of correct model selections. Especially for small to medium sample cases, we use an effective measure, observed L_2 performance. This measure is useful when the criterion is not achieved according to the correct model. The observed L_2 performance formula is shown in Section 2. The remainder of this article is organized as follows. Section 2 introduces the SEM with structural and reduced forms, including contemporaneously correlated error vectors. Then, a new criterion for selection in SEM called SAIC(UE) is proposed. A simulation study and results of

1,000 samples of the SEM to verify the performance of SAIC and SAIC(UE) are presented in Section 3. Finally, Section 4 is considered as a summary and future study.

2. Methodology

The structural form of the SEM with M equations (M external explanatory variables), denoted by $y_{t1}, y_{t2}, \ldots, y_{tM}$ and K predefined variables, denoted by $x_{t1}, x_{t2}, \ldots, x_{tK}$ can be written as,

$$\mathbf{Y\Gamma} + \mathbf{XB} = \mathbf{U}, \tag{3}$$

where

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1M} \\ y_{21} & y_{22} & \cdots & y_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ y_{T1} & y_{T2} & \cdots & y_{TM} \end{bmatrix}, \ \mathbf{\Gamma} = \begin{bmatrix} 1 & -\gamma_{12} & \cdots & -\gamma_{1M} \\ -\gamma_{21} & 1 & \cdots & -\gamma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma_{M1} & -\gamma_{M2} & \cdots & 1 \end{bmatrix},$$

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T1} & x_{T2} & \cdots & x_{TK} \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} -\beta_{11} & -\beta_{12} & \cdots & -\beta_{1M} \\ -\beta_{21} & -\beta_{22} & \cdots & -\beta_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ -\beta_{K1} & -\beta_{K2} & \cdots & -\beta_{KM} \end{bmatrix}, \ \mathbf{U} = \begin{bmatrix} \mathrm{u}_{11} & \mathrm{u}_{12} & \cdots & \mathrm{u}_{1M} \\ \mathrm{u}_{21} & \mathrm{u}_{22} & \cdots & \mathrm{u}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{u}_{T1} & \mathrm{u}_{T2} & \cdots & \mathrm{u}_{TM} \end{bmatrix}.$$

The reduced form of the model described by the model structure in (3) can be written as

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{E}, \tag{4}$$

where $\mathbf{\Pi} = -\mathbf{B\Gamma}^{-1}$ is a $K \times M$ matrix of unknown parameters and $\mathbf{E} = \mathbf{U\Gamma}^{-1}$ is a $T \times M$ matrix of contemporaneously correlated errors. For all M equations, the models in (4) can be expressed as a cascaded model of the individual equations $\mathbf{y_j} = \mathbf{X\pi_j} + \boldsymbol{\varepsilon_j}$, $j = 1, 2, \ldots, M$ as follows

$$\mathbf{y} = \tilde{\mathbf{X}}\boldsymbol{\pi} + \boldsymbol{\varepsilon}, \tag{5}$$

where $\mathbf{y}$ is a $TM \times 1$ observation vector consisting of M $(T \times 1)$ $\mathbf{y_j}$ vectors, $\tilde{\mathbf{X}}$ is a $TM \times KM$ diagonal matrix of rank KM consisting of $M(T \times K)$ identical $\mathbf{X}$ matrices, $\boldsymbol{\pi}$ is a $KM \times 1$ unknown parameter vector consisting of M $(K \times 1)$ $\boldsymbol{\pi_j}$ vectors, $\boldsymbol{\varepsilon}$ is a $TM \times 1$ contemporaneously correlated error vector consisting of $M(T \times 1)$ $\boldsymbol{\varepsilon_j}$ vectors, and ε_{tj} is independent identically distributed random variable that corresponds to

$$\varepsilon_{tj} \sim N(0, \sigma_{jj}),$$

and

$$\boldsymbol{\varepsilon_t} = \begin{bmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \\ \vdots \\ \varepsilon_{tM} \end{bmatrix} \sim N_M(\mathbf{0}, \mathbf{\Sigma}), \tag{6}$$

where $\mathbf{\Sigma}$ is the $M \times M$ contemporaneous covariance matrix of error terms which is nonsingular and positive symmetric definite.

The observed L_2 performance is defined as (McQuarrie et al. 1997)

$$\text{Observed } L_2 \text{ performance } = \frac{\min_{1 \le K \le K^*} L_2(K)}{L_2(K_s)}, \tag{7}$$

where

$$L_2(K) = \frac{1}{T}(\boldsymbol{\pi} - \hat{\boldsymbol{\pi}})' \tilde{\mathbf{X}}' \left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_{\mathbf{T}}\right) \tilde{\mathbf{X}} (\boldsymbol{\pi} - \hat{\boldsymbol{\pi}})$$

is the observed L_2 distance between the true and the candidate model, K^* is the class of all possible candidate models, K is the rank of the fitted candidate model, and K_s is the model selected by a specific model selection criterion. The closer the selected model is to the true model, the higher the performance. Therefore, the best model selection criterion selects a model that performs well even in small sample sizes.

To create the model selection criterion, we use the multivariate normal density (Johnson and Wichern 1998) to obtain the joint normal density of these M error terms $\left(\boldsymbol{\varepsilon} = \mathbf{y} - \tilde{\mathbf{X}}\boldsymbol{\pi}\right)$,

$$L\left(\mathbf{y} - \tilde{\mathbf{X}}\boldsymbol{\pi},\, \boldsymbol{\pi},\, \boldsymbol{\Sigma} \otimes \mathbf{I}_{\mathbf{T}}\right) = (2\pi)^{-\frac{TM}{2}} \left|\boldsymbol{\Sigma} \otimes \mathbf{I}_{\mathbf{T}}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\mathbf{y} - \tilde{\mathbf{X}}\boldsymbol{\pi}\right)' \left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_{\mathbf{T}}\right)\left(\mathbf{y} - \tilde{\mathbf{X}}\boldsymbol{\pi}\right)\right). \quad (8)$$

Logarithmic replacement of the determinant of the contemporaneous covariance matrix of the error terms in (8), i.e., $\log\left|\boldsymbol{\Sigma} \otimes \mathbf{I}_{\mathbf{T}}\right| = T \log\left|\boldsymbol{\Sigma}\right|$, we have

$$\log L\left(\mathbf{y} - \tilde{\mathbf{X}}\boldsymbol{\pi},\, \boldsymbol{\pi},\, \boldsymbol{\Sigma} \otimes \mathbf{I}_{\mathbf{T}}\right) = -\frac{TM}{2}\log(2\pi) - \frac{T}{2}\log\left|\boldsymbol{\Sigma}\right| - \frac{1}{2}\left(\mathbf{y} - \tilde{\mathbf{X}}\boldsymbol{\pi}\right)' \left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_{\mathbf{T}}\right)\left(\mathbf{y} - \tilde{\mathbf{X}}\boldsymbol{\pi}\right). \quad (9)$$

Given a vector of MLE $\tilde{\boldsymbol{\theta}} = \left(\tilde{\boldsymbol{\pi}},\, \tilde{\boldsymbol{\Sigma}}\right)$, the minus twice of the log-likelihood function in (9) can be written as

$$-2\log L\left(\mathbf{y},\, \tilde{\mathbf{X}},\, \tilde{\boldsymbol{\pi}},\, \tilde{\boldsymbol{\Sigma}}\right) = TM\log(2\pi) + T\log\left|\tilde{\boldsymbol{\Sigma}}\right| + \left(\mathbf{y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\pi}}\right)' \left(\tilde{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_{\mathbf{T}}\right)\left(\mathbf{y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\pi}}\right). \quad (10)$$

Consider the last term in (10)

$$\begin{aligned}
\left(\mathbf{y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\pi}}\right)' \left(\tilde{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_{\mathbf{T}}\right)\left(\mathbf{y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\pi}}\right) &= tr\left[\left(\tilde{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_{\mathbf{T}}\right)\left(\mathbf{y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\pi}}\right)\left(\mathbf{y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\pi}}\right)'\right] \\
&= tr\left[\tilde{\boldsymbol{\Sigma}}^{-1}\left(\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\Pi}}\right)\left(\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\Pi}}\right)'\right] \\
&= T\, tr\left(\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\Sigma}}\right) = T\, tr\left(\mathbf{I}_{\mathbf{M}}\right) = TM. \quad (11)
\end{aligned}$$

From (11), the minus twice of the log-likelihood function in (10) becomes

$$-2\log L\left(\mathbf{y},\, \tilde{\mathbf{X}},\, \tilde{\boldsymbol{\pi}},\, \tilde{\boldsymbol{\Sigma}}\right) = TM\left[\log(2\pi) + 1\right] + T\log\left|\tilde{\boldsymbol{\Sigma}}\right|. \quad (12)$$

The first term in (12), $TM\left[\log(2\pi) + 1\right]$ is a constant for a given system of M simultaneous equations, T observations and can be omitted in any benchmark. From AIC $= -2\log L\left(\tilde{\boldsymbol{\theta}} \middle| \mathbf{y}\right) + 2(K+1)$ (Akaike 1974), we replace the parameters "$K+1$" with a parameters in an SEM, i.e., KM parameters for $\boldsymbol{\pi}$ and $M(M+1)/2$ parameters for $\boldsymbol{\Sigma}$. Therefore, when adjusting the value by scaling the sample size (T), the Akaike information criterion for an SEM or SAIC (Keerativibool et al. 2011) is

$$\text{SAIC} = \log\left|\tilde{\boldsymbol{\Sigma}}\right| + \frac{2}{T}\left(KM + M(M+1)/2\right) = \log\left|\tilde{\boldsymbol{\Sigma}}\right| + M(2K + M + 1)/T,$$

(13)

where $\tilde{\boldsymbol{\Sigma}} = \frac{1}{T}\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\Pi}}\right)' \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\Pi}}\right)$.

For this article, we use the knowledge from McQuarrie et al. (1997) to create a new criterion for selecting the SEM by substituting UE of $\boldsymbol{\Sigma}$ into (13) to get SAIC(UE) as follows

$$\text{SAIC(UE)} = \log\left|\hat{\boldsymbol{\Sigma}}\right| + M\left(2K + M + 1\right)/T, \tag{14}$$

where $\hat{\boldsymbol{\Sigma}} = \frac{1}{T-K}\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{\Pi}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{\Pi}}\right)$.

3. Simulation Study

The true model of SEM to be considered in this article is a system of three equations (M = 3),

$$\begin{aligned} y_{t1} &= 1 + 2x_{t2} + 3x_{t3} + 4x_{t4} + \varepsilon_{t1} \\ y_{t2} &= 1 - 0.5x_{t2} - 5x_{t3} - 1.5x_{t4} + \varepsilon_{t2} \\ y_{t3} &= 1 + x_{t2} + x_{t3} + x_{t4} + \varepsilon_{t3}, \end{aligned} \tag{15}$$

where there are three sample sizes: small samples T = 15, 20 and 25, medium samples T = 30 and 50, and large samples T = 100 and 500 observations. All simulation procedures and results are as follows.

3.1 Generate 1,000,000 vectors of the 3×1 multivariate normal $\boldsymbol{\varepsilon}_t$ in (6) with the average vector being zero. The correlation coefficients of error between the equations are

$$\rho_{12} = 0.9,\ \rho_{13} = 0.7,\ \rho_{23} = 0.8,$$

and the variances-covariances of the errors are

$$\sigma_{11} = 0.9^2 = 0.81,\ \sigma_{22} = 0.8^2 = 0.64,\ \sigma_{33} = 0.7^2 = 0.49,$$
$$\sigma_{12} = \rho_{12}\sqrt{\sigma_{11}\sigma_{22}} = 0.648,\ \sigma_{13} = \rho_{13}\sqrt{\sigma_{11}\sigma_{33}} = 0.441,\ \sigma_{23} = \rho_{23}\sqrt{\sigma_{22}\sigma_{33}} = 0.448,$$

then, the form to generate $\boldsymbol{\varepsilon}_t$ in (6) is represented by

$$\boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \\ \varepsilon_{t3} \end{bmatrix} \sim N_3\left(\mathbf{0},\ \boldsymbol{\Sigma} = \begin{bmatrix} 0.81 & 0.648 & 0.441 \\ 0.648 & 0.64 & 0.448 \\ 0.441 & 0.448 & 0.49 \end{bmatrix}\right).$$

3.2 Divide the error set obtained in 3.1 into 1,000 samples, each containing three sample sizes: small samples T = 15, 20 and 25, medium samples T = 30 and 50, and large samples T = 100 and 500 observations.

3.3 Generate the independent variables x_{t2} until $x_{t,10}$ with approximately 1,000,000 observations as normal random variables with zero mean and variance equal to one. The relevant independent variables are x_{t2}, x_{t3}, and x_{t4} and irrelevant independent variables are x_{t5} until $x_{t,10}$.

3.4 Divide the independent variables x_{t2} until $x_{t,10}$ obtained in 3.3 into 1,000 samples, each containing three sample sizes: small samples T = 15, 20 and 25, medium samples T = 30 and 50, and large samples T = 100 and 500 observations.

3.5 Test the multicollinearity problem for the independent variables set in 3.4 and discard samples that do not pass the test. Only 1,000 samples were kept for further study.

3.6 Create the dependent variables described in (15) by using the relevant independent variables x_{t2}, x_{t3}, and x_{t4} obtained in 3.5 and the errors obtained in 3.2.

3.7 Adhere to a systematic and structured approach using the nested model assumption to construct the candidate models. This process follows a step-by-step addition of columns of independent variables, starting from columns 1 to *K*, which define the design matrix for the candidate model with dimension K. For a comprehensive sample size of 1,000, we calculate SAIC in (13) and SAIC(UE) in (14). The candidate model with the lowest value of the model selection criterion is identified as the best model. To illustrate, consider the example of calculating the values of SAIC in (13) and SAIC(UE) in (14) for T = 15, M = 3, K = 2, we have

$$\text{SAIC} = \log\left|\tilde{\boldsymbol{\Sigma}}\right| + M\left(2K+M+1\right)/T = \log\left(16.97685721\right) + 3\left(2\left(2\right)+3+1\right)/15 = 4.4319,$$

$$\text{SAIC(UE)} = \log\left|\hat{\boldsymbol{\Sigma}}\right| + M\left(2K+M+1\right)/T = 15\log\left(26.0796054\right) + 3\left(2\left(2\right)+3+1\right)/15 = 4.8612.$$

Table 1 details the performance comparison of the percentage of correct model selections from SAIC in (13) and SAIC(UE) in (14) for 1,000 samples.

3.8 To provide confidence about the process, we further validated the performance of the model selection criteria by measuring the L_2 performance observed in (7), detailed in Table 2.

Table 1 Percentage of correct model selections

Sample sizes	T	Criteria	K								
			2	3	4*	5	6	7	8	9	10
Small	15	SAIC	0	0	73.6	7	3.7	1.9	3.1	0.3	10.4
		SAIC(UE)	0	0	**93.7**	4.6	0.4	0.6	0	0	0.7
	20	SAIC	0	0	87.1	7.6	2.7	1.1	0.8	0	0.7
		SAIC(UE)	0	0	**97.1**	2.4	0.5	0	0	0	0
	25	SAIC	0	0	90.8	6.8	1.3	0.8	0.2	0	0.1
		SAIC(UE)	0	0	**97.7**	2.1	0.2	0	0	0	0
Medium	30	SAIC	0	0	92.7	5.7	0.9	0.4	0.3	0	0
		SAIC(UE)	0	0	**97.9**	2	0.1	0	0	0	0
	50	SAIC	0	0	95.2	3.7	1	0.1	0	0	0
		SAIC(UE)	0	0	**99.1**	0.9	0	0	0	0	0
Large	100	SAIC	0	0	95.8	3.4	0.5	0.2	0.1	0	0
		SAIC(UE)	0	0	**99.3**	0.7	0	0	0	0	0
	500	SAIC	0	0	97.6	2	0.3	0.1	0	0	0
		SAIC(UE)	0	0	**99.3**	0.7	0	0	0	0	0

* represents the true model, and bold type indicates the maximum percentage of correct model selections.

Table 2 Average and standard deviation of the observed L_2 performance

Sample sizes	T	Criteria	Average L_2	S.D. L_2
Small	15	SAIC	0.7387	0.3656
		SAIC(UE)	**0.8939**	**0.2173**
	20	SAIC	0.9027	0.2263
		SAIC(UE)	**0.9665**	**0.1252**
	25	SAIC	0.9417	0.1773
		SAIC(UE)	**0.9819**	**0.0951**
Medium	30	SAIC	0.9554	0.1574
		SAIC(UE)	**0.9843**	**0.0963**
	50	SAIC	0.9735	0.1196
		SAIC(UE)	**0.9948**	**0.0534**
Large	100	SAIC	0.9786	0.1058
		SAIC(UE)	**0.9963**	**0.0448**
	500	SAIC	0.9867	0.0866
		SAIC(UE)	**0.9956**	**0.0534**

Bold type indicates the maximum average of the observed L_2 performance and the minimum standard deviation of the observed L_2 performance.

3.9 From the percentage of correct model selections in Table 1, it can be concluded that the performance of SAIC(UE) in (14) convincingly outperforms SAIC in (13) for all sample size levels

(T = 15, 20, 25, 30, 50, 100, 500) because SAIC tends to provide an overfit the model order than SAIC(UE), especially when the sample size is small (T = 15, 20, 25). The observed L_2 performance in Table 2 also confirms that SAIC(UE) has more average observed L_2 performance and lower discrepancies. Therefore, SAIC(UE) is better than SAIC in all situations of the simulation study. Table 3 shows the average and standard deviation of SAIC and SAIC(UE) for 1,000 samples. This table shows that SAIC has a more negative bias than SAIC(UE) since SAIC estimates the contemporaneous covariance matrix of error terms using MLE, while SAIC(UE) estimates it using UE. MLE is smaller than UE because it is divisible by T, while UE is divisible by $T-K$. This problem causes the value of SAIC to be lower than it should be or underestimated, which may be the main reason why the correct number of model sequences is selected less often.

4. Conclusions

In this article, we propose the Akaike information criterion for the simultaneous equations model using an unbiased estimator called SAIC(UE) to select the optimal system of the models. Comparison of the performance of selecting the correct order of the model from SAIC(UE) with SAIC proposed by Keerativibool et al. (2011) shows that in all situations with three sample sizes: small samples T = 15, 20 and 25, medium samples T = 30 and 50, and large samples T = 100 and 500 observations, SAIC(UE) outperforms SAIC since there is a higher percentage of correct model selections. SAIC tends to provide an overfit the model order than SAIC(UE). The observed L_2 performance result also confirms that SAIC(UE) has better observed L_2 performance on average and lower discrepancies. Therefore, SAIC(UE) is likely to be better than SAIC in all situations of the simulation study. The mean and standard deviation of SAIC and SAIC(UE) for 1000 samples show that SAIC has a more negative bias than SAIC(UE) or underestimated, which may be the main reason for selecting the correct order of models from SAIC less than SAIC(UE). There are few criteria for selecting SEM, so additional criteria should be studied and determined. In further work, we attempt to construct the other model selection criteria for SEM using an unbiased estimator, such as BIC and KIC and then compare the performance of selecting the correct order of the model with SBIC proposed by Keerativibool (2012) and SKIC proposed by Keerativibool and Jitthavech (2015), respectively.

Table 3 Average and standard deviation of SAIC and SAIC(UE) for 1,000 samples

Sample sizes	T	Criteria	Stat	K								
				2	3	4*	5	6	7	8	9	10
Small	15	SAIC	Ave.	5.330	2.458	**-2.707**	-2.379	-2.058	-1.773	-1.529	0.546	-1.306
			S.D.	0.693	0.769	0.854	0.900	0.957	1.031	1.151	1.330	1.524
		SAIC(UE)	Ave.	5.760	3.127	**-1.776**	-1.162	-0.526	0.113	0.757	3.295	1.989
			S.D.	0.693	0.769	0.854	0.900	0.957	1.031	1.151	1.330	1.524
	20	SAIC	Ave.	5.182	2.243	**-2.978**	-2.705	-2.444	-2.178	-1.923	0.014	-1.426
			S.D.	0.574	0.627	0.683	0.706	0.738	0.769	0.799	0.874	0.893
		SAIC(UE)	Ave.	5.498	2.730	**-2.309**	-1.842	-1.374	-0.886	-0.390	1.808	0.654
			S.D.	0.574	0.627	0.683	0.706	0.738	0.769	0.799	0.874	0.893
	25	SAIC	Ave.	5.095	2.088	**-3.156**	-2.935	-2.711	-2.494	-2.278	-0.430	-1.860
			S.D.	0.528	0.545	0.566	0.593	0.607	0.623	0.645	0.745	0.695
		SAIC(UE)	Ave.	5.345	2.472	**-2.633**	-2.266	-1.887	-1.509	-1.121	0.909	-0.328
			S.D.	0.528	0.545	0.566	0.593	0.607	0.623	0.645	0.745	0.695

* represents the true model and bold type indicates the minimum average of the criterion.

Table 3 (Continued)

Sample sizes	*T*	Criteria	Stat	*K*								
				2	3	4*	5	6	7	8	9	10
Medium	30	SAIC	Ave.	5.019	1.984	**-3.302**	-3.118	-2.927	-2.743	-2.551	-0.730	-2.176
			S.D.	0.459	0.487	0.493	0.501	0.506	0.522	0.540	0.591	0.571
		SAIC(UE)	Ave.	5.226	2.300	**-2.872**	-2.571	-2.258	-1.946	-1.621	0.340	-0.960
			S.D.	0.459	0.487	0.493	0.501	0.506	0.522	0.540	0.591	0.571
	50	SAIC	Ave.	4.886	1.768	**-3.587**	-3.468	-3.353	-3.233	-3.115	-1.414	-2.885
			S.D.	0.362	0.377	0.382	0.387	0.390	0.391	0.396	0.437	0.406
		SAIC(UE)	Ave.	5.008	1.954	**-3.337**	-3.152	-2.969	-2.780	-2.592	-0.818	-2.215
			S.D.	0.362	0.377	0.382	0.387	0.390	0.391	0.396	0.437	0.406
Large	100	SAIC	Ave.	4.778	1.600	**-3.811**	-3.751	-3.691	-3.631	-3.573	-1.957	-3.453
			S.D.	0.256	0.263	0.264	0.266	0.267	0.268	0.269	0.295	0.272
		SAIC(UE)	Ave.	4.839	1.692	**-3.688**	-3.597	-3.506	-3.413	-3.322	-1.674	-3.137
			S.D.	0.256	0.263	0.264	0.266	0.267	0.268	0.269	0.295	0.272
	500	SAIC	Ave.	4.673	1.450	**-4.013**	-4.001	-3.989	-3.977	-3.965	-2.402	-3.941
			S.D.	0.113	0.108	0.110	0.110	0.110	0.110	0.110	0.118	0.110
		SAIC(UE)	Ave.	4.685	1.468	**-3.989**	-3.971	-3.953	-3.935	-3.917	-2.348	-3.880
			S.D.	0.113	0.108	0.110	0.110	0.110	0.110	0.110	0.118	0.110

* represents the true model and bold type indicates the minimum average of the criterion.

Acknowledgements

This research was supported by National Science, Research and Innovation Fund (NSRF), Thaksin University (Research project grant) Fiscal Year 2024.

References

Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr. 1974; 19(6): 716-723.

Cavanaugh JE. A large-sample model selection criterion based on Kullback's symmetric divergence. Stat Prob Lett. 1999; 42(4): 333-343.

Cavanaugh JE. Criteria for linear model selection based on Kullback's symmetric divergence. Aust N Z J Stat. 2004; 46(2): 257-274.

Giombini G, Szroeter J. Quasi Akaike and quasi Schwarz criteria for model selection: a surprising consistency result. Econ Lett. 2007; 95(2): 259-266.

Hannan EJ, Quinn BG. The determination of the order of an autoregression. J R Stat Soc, Ser B, Methodol. 1979; 41(2): 190-195.

Hurvich CM, Tsai CL. Regression and time series model selection in small samples. Biometrika. 1989; 76(2): 297-307.

Johnson RA, Wichern DW. Applied multivariate statistical analysis. 4th ed. Englewood Cliffs, New Jersey: Prentice Hall; 1998.

Keerativibool W, Jitthavech J, Lorchirachoonkul V. Model selection in a system of simultaneous equations model. Commun Stat Theory Methods. 2011; 40(3): 373-393.

Keerativibool W. New criteria for selection in simultaneous equations model. Thailand Statistician J. 2012; 10(2): 163-181.

Keerativibool W. Costs and benefits gained from planting para rubber in the form of system of simultaneous equations: a case study of Pa Phayom district, Phatthalung province and Thung Song district, Nakhon Si Thammarat province. NIDA Development J. 2014; 54(3): 95-124.

Keerativibool W, Jitthavech J. Model selection criterion based on Kullback-Leibler's symmetric divergence for simultaneous equations model. Chiang Mai J Sci. 2015; 42(3): 761-773.

McQuarrie ADR, Shumway RH, Tsai CL. The model selection criterion AICu. Stat. Probabil Lett. 1997; 34(3): 285-292.

Neath AA, Cavanaugh JE. Regression and time series model selection using variants of the Schwarz information criterion. Commun Stat Theory Methods. 1997; 26(3): 559-580.

Niyonzima F. A new method for regression model selection. Global Sci J. 2020; 8(2): 1858-1899.

Perez-Sanchez B, Gonzalez M, Perea C, Lopez-Espin JJ. A new computational method for estimating simultaneous equations models using entropy as a parameter criteria. Mathematics. 2021; 9(700): 1-9.

Pham H. A new criterion for model selection. Mathematics. 2019; 7(12): 1215.

Rossi R, Murari A, Gaudio P, Gelfusa M. Upgrading model selection criteria with goodness of fit tests for practical applications. Entropy. 2020; 22(447): 1-13.

Schwarz G. Estimating the dimension of a model. Ann Stat. 1978; 6(2): 461-464.