Thailand Statistician
April 2026; 24(2): 429-442
http://statassoc.or.th
Contributed paper

Kullback Information Criterion for a Simultaneous Equations Model

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

Received: 15 October 2024
Revised: 26 August 2025
Accepted: 29 August 2025

Abstract

This paper introduces a new model selection criterion, SKIC(MLE), for simultaneous equations modelling based on the Kullback information criterion (KIC) using a maximum likelihood estimator (MLE). A comprehensive comparative simulation study demonstrates that when dealing with small sample sizes, SKIC(MLE) outperforms SKIC, a criterion proposed by Keerativibool and Jitthavech (2015). The results indicate that the proposed criterion, SKIC(MLE), has the potential to improve the accuracy of model selection significantly and exhibits higher observed L_2 efficiency than SKIC in such scenarios. The superiority of SKIC(MLE) in small sample sizes is attributed to the fact that the penalty term of SKIC increases exponentially as the number of parameters increases, causing the SKIC value to be higher than SKIC(MLE). As a result, SKIC is more likely to select underfitting models or few parameters than SKIC(MLE). However, this issue is not prevalent in medium to large sample sizes, so SKIC outperforms SKIC(MLE). This research has significant practical implications, potentially revolutionizing simultaneous equations modelling in cases with small sample sizes.

Keywords: model selection criterion, selection of model order, L_2 efficiency, maximum likelihood estimator (MLE).

1. Introduction

In the simultaneous equations model (SEM), variables influence each other in two-way flows, making it a method for modelling economic data. Endogenous explanatory variables in SEM are important because they become stochastic and are correlated with the error terms of the equation in which they appear as explanatory variables (Greene 2008). SEM is a widely used statistical technique in economic science for modelling the simultaneous relationship between variables (Perez-Sanchez et al., 2021). Fitting the parameters of SEM to an observed data set poses two main tasks: determining the model order and carefully estimating these parameters. The crucial part of this fitting problem is determining the model order. Such determination is often enabled by using a model selection criterion, where one only needs to evaluate two simple terms that balance the fit to the data and the model's complexity. This balance guarantees that the selected model is accurate and simple enough (Zhang et al. 2023). Model selection criteria have become important in selecting appropriate model order for constructing univariate, multivariate regression equations or systems of equations, such as SEM, since

not all available predictor variables are appropriate for explaining the variation in the response variable and predicting the new response outcome. Several model selection criteria are needed in both univariate and multivariate regression analysis. For instance, the Akaike information criterion (AIC) was developed to act as an asymptotically unbiased estimator (UE) of a variant of Kullback-Leibler's (KL) directed divergence between the true and fitted candidate models (Akaike, 1974). On the other hand, the Bayesian information criterion (BIC) was often preferred over AIC by practitioners due to its Bayesian justification or its tendency to select more parsimonious models than AIC (Schwarz 1978). A corrected version of the AIC (AICc) was designed to be an exact UE of a KL-directed divergence variant and extended to a small sample (Hurvich and Tsai 1989). Kullback information criterion (KIC) (Cavanaugh 1999) and the corrected version of the KIC (KICc) (Cavanaugh 2004) were the criteria constructed to target a KL-symmetric divergence. This divergence was an alternative to directed divergence, obtained by summating the two directed divergences, arguably more sensitive than either of their components. Pham information criterion (PIC) was designed to consider a larger penalty when adding too many coefficients (or estimated parameters) in the model from too small a sample in the presence of too much noise, in addition to minimizing the sum of squares error (Pham 2019). A new method for regression model selection uses the p-value for the t-test and F-test for the variable's selection criteria and the adjusted R-squared or AIC for the best model selection criterion (Niyonzima 2020). And the model selection criteria with goodness of fit tests for practical applications (Rossi et al. 2020).

In previous studies, different model selection criteria were used for univariate and multivariate regression models, while fewer studies focused on model selection criteria for the SEM. The first development of the SEM selection criterion is a system of Akaike information criterion (SAIC) (Keerativibool et al. 2011). Later, a solution to the issue of the SAIC selecting models with an excessive number of independent variables or overfitting was developed. This solution is called a system of Kullback information criterion (SKIC) (Keerativibool and Jitthavech 2015). However, SKIC tends to give overestimations when sample size is small. This problem could lead to a model with too few parameters, resulting in a lower percentage of correct model selection from SKIC. Our research aims to establish a novel criterion using the maximum likelihood estimator (MLE) to address this specific issue, potentially improving model selection in statistical analysis. We conducted a comprehensive simulation study to compare its efficiency. This study compared SKIC(MLE) with SKIC under three distinct sample sizes: small samples T = 15, 20 and 25, medium samples T = 30 and 50, and large samples T = 100 and 500. The efficiency of both criteria was assessed using a corresponding measure, which calculates the percentage of correct model selections. Especially for small to medium sample cases, we utilize an effective measure known as observed L_2 efficiency. This measure is helpful when the criterion needs to meet the correct model standards. The formula for observed L_2 efficiency is outlined in the simulation step. The rest of this paper is structured as follows: Section 2 introduces the SEM with structural and reduced forms, including contemporaneously correlated error vectors. Subsequently, a new criterion for selection in SEM, referred to as SKIC(MLE), is proposed. A simulation study and results from 1,000 samples of the SEM are presented in Section 3 to verify the efficiency of SKIC(MLE) and SKIC (Keerativibool and Jitthavech 2015). Lastly, Section 4 serves as a summary and outlines areas for future study.

2. Methodology

Equation (1) shows the structural form of the SEM with M equations (M external explanatory variables or dependent variables), denoted by $y_{t1}, y_{t2}, \ldots, y_{tM}$, and K predefined variables or

independent variables, denoted by $x_{t1}, x_{t2}, \ldots, x_{tK}$. The first independent variable of each equation, x_{t1}, is assigned 1. The parameters γ and β represent the coefficients of the dependent and independent variables, respectively. Each equation's error terms, denoted by $u_{t1}, u_{t2}, \ldots, u_{tM}$, are related and have a mean equal to 0.

$$\begin{aligned}
y_{t1} &= \gamma_{21}y_{t2} + \gamma_{31}y_{t3} + \ldots + \gamma_{M-1,1}y_{t,M-1} + \gamma_{M1}y_{tM} + \beta_{11}x_{t1} + \beta_{21}x_{t2} + \ldots + \beta_{K1}x_{tK} + u_{t1} \\
y_{t2} &= \gamma_{12}y_{t1} + \gamma_{32}y_{t3} + \ldots + \gamma_{M-1,2}y_{t,M-1} + \gamma_{M2}y_{tM} + \beta_{12}x_{t1} + \beta_{22}x_{t2} + \ldots + \beta_{K2}x_{tK} + u_{t2} \\
&\vdots \\
y_{tM} &= \gamma_{1M}y_{t1} + \gamma_{2M}y_{t2} + \gamma_{3M}y_{t3} + \ldots + \gamma_{M-1,M}y_{t,M-1} + \beta_{1M}x_{t1} + \beta_{2M}x_{t2} + \ldots + \beta_{KM}x_{tK} + u_{tM}. \quad (1)
\end{aligned}$$

Equation (1) illustrates that the SEM's structural form can be reorganized into matrix form, as demonstrated in (2).

$$\mathbf{Y\Gamma} + \mathbf{XB} = \mathbf{U}, \quad (2)$$

where

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

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

From the structural form in (2), it can be expressed in reduced form as (3).

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{E}, \quad (3)$$

where $\mathbf{\Pi} = -\mathbf{B\Gamma}^{-1}$ represents a $K \times M$ matrix of unknown parameters and $\mathbf{E} = \mathbf{U\Gamma}^{-1}$ represents a $T \times M$ matrix of contemporaneously correlated errors. Considering all the M equations, (3) can be represented as a nested form of the individual equations as (4).

$$\mathbf{y} = \mathbf{X}^{*}\boldsymbol{\pi} + \boldsymbol{\varepsilon}, \quad (4)$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \vdots \\ \mathbf{y_M} \end{bmatrix}, \ \mathbf{X}^{*} = \begin{bmatrix} \mathbf{X} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X} \end{bmatrix}, \ \boldsymbol{\pi} = \begin{bmatrix} \boldsymbol{\pi_1} \\ \boldsymbol{\pi_2} \\ \vdots \\ \boldsymbol{\pi_M} \end{bmatrix}, \ \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon_1} \\ \boldsymbol{\varepsilon_2} \\ \vdots \\ \boldsymbol{\varepsilon_M} \end{bmatrix}.$$

That is, $\mathbf{y}$ represents a $TM \times 1$ vector of dependent variables consisting of M vectors $\mathbf{y}_j$ of size $T \times 1$, $\mathbf{X}^{*}$ represents a $TM \times KM$ diagonal matrix consisting of M identical matrices $\mathbf{X}$ of size $T \times K$, $\boldsymbol{\pi}$ represents a $KM \times 1$ vector of unknown parameters consisting of M vectors $\boldsymbol{\pi}_j$ of size $K \times 1$, $\boldsymbol{\varepsilon}$ represents a $TM \times 1$ vector of error terms consisting of M vectors $\boldsymbol{\varepsilon}_j$ of size $T \times 1$, and ε_{tj} represents an error term that is normally distributed, independent, has a mean equal to 0 and a variance equal to σ_{jj}, $\varepsilon_{tj} \sim N(0, \sigma_{jj})$. Accordingly, we express the error terms in each system as (5)

$$\boldsymbol{\varepsilon}_{\mathbf{t}} = \begin{bmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \\ \vdots \\ \varepsilon_{tM} \end{bmatrix} \sim \mathrm{N}_{\mathrm{M}}\left(\mathbf{0}, \boldsymbol{\Sigma}\right), \tag{5}$$

where $\boldsymbol{\Sigma}$ represents an $\mathrm{M} \times \mathrm{M}$ contemporaneous covariance matrix of error terms which is nonsingular and positive symmetric definite. To establish model selection criterion, we utilized the multivariate normal density (Johnson and Wichern 1998) to build the joint normal density of these M error terms, $\boldsymbol{\varepsilon} = \mathbf{y} - \tilde{\mathbf{X}}\boldsymbol{\pi}$, as (6).

$$\mathrm{L}\left(\boldsymbol{\theta} \middle| \mathbf{y}\right) = \left(2\pi\right)^{-\frac{\mathrm{TM}}{2}} \left| \boldsymbol{\Sigma} \otimes \mathbf{I}_{\mathbf{T}} \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\mathbf{y} - \mathbf{X}^{*}\boldsymbol{\pi}\right)' \left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_{\mathbf{T}}\right)\left(\mathbf{y} - \mathbf{X}^{*}\boldsymbol{\pi}\right) \right), \tag{6}$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\pi}, \boldsymbol{\Sigma}\right)$ represents the parameters of the model and $\otimes$ represents Kronecker Product. Logarithm of (6) by substituting values of $\log\left| \boldsymbol{\Sigma} \otimes \mathbf{I}_{\mathbf{T}} \right|$ with $\mathrm{T}\log\left| \boldsymbol{\Sigma} \right|$, we have $\log \mathrm{L}\left(\boldsymbol{\theta} \middle| \mathbf{y}\right)$ as in (7).

$$\log \mathrm{L}\left(\boldsymbol{\theta} \middle| \mathbf{y}\right) = -\frac{\mathrm{TM}}{2}\log\left(2\pi\right) - \frac{\mathrm{T}}{2}\log\left| \boldsymbol{\Sigma} \right| - \frac{1}{2}\left(\mathbf{y} - \mathbf{X}^{*}\boldsymbol{\pi}\right)' \left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_{\mathbf{T}}\right)\left(\mathbf{y} - \mathbf{X}^{*}\boldsymbol{\pi}\right). \tag{7}$$

The minus twice of the log-likelihood function in (7) can be written as (8).

$$-2\log \mathrm{L}\left(\boldsymbol{\theta} \middle| \mathbf{y}\right) = \mathrm{TM}\log\left(2\pi\right) + \mathrm{T}\log\left| \boldsymbol{\Sigma} \right| + \left(\mathbf{y} - \mathbf{X}^{*}\boldsymbol{\pi}\right)' \left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_{\mathbf{T}}\right)\left(\mathbf{y} - \mathbf{X}^{*}\boldsymbol{\pi}\right). \tag{8}$$

Before considering the KL-symmetric divergence to distinguish between the candidate and true models, we define (4) to represent the candidate model and (9) below represents the true model, where it is assumed to be correctly specified or overfitted by all the candidate models. The true parameter $\boldsymbol{\pi}_0$ has $\mathrm{K}_0\mathrm{M}$ nonzero entries with $0 < \mathrm{K}_0\mathrm{M} \le \mathrm{KM}$ and the rest of $\left(\mathrm{K} - \mathrm{K}_0\right)\mathrm{M}$ entries are equal to zero.

$$\mathbf{y} = \mathbf{X}^{*}\boldsymbol{\pi}_0 + \boldsymbol{\varepsilon}_0. \tag{9}$$

The KL-symmetric divergence defined by (10).

$$2\mathrm{J}\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}\right) = \mathrm{d}\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}\right) - \mathrm{d}\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0\right) + \mathrm{d}\left(\boldsymbol{\theta}, \boldsymbol{\theta}_0\right) - \mathrm{d}\left(\boldsymbol{\theta}, \boldsymbol{\theta}\right), \tag{10}$$

where $\mathrm{d}\left(\boldsymbol{\theta}_{\mathrm{i}}, \boldsymbol{\theta}_{\mathrm{j}}\right) = \mathrm{E}_{\boldsymbol{\theta}_{\mathrm{i}}}\left\{ -2\log \mathrm{L}\left(\boldsymbol{\theta}_{\mathrm{j}} \middle| \mathbf{y}\right) \right\}$. Given a set of MLE $\tilde{\boldsymbol{\theta}} = \left(\tilde{\boldsymbol{\pi}}, \tilde{\boldsymbol{\Sigma}}\right)$, we have the estimate of the KL-symmetric divergence as (11).

$$\begin{aligned} 2\mathrm{J}\left(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}\right) \; &= \mathrm{TM}\log\left(2\pi\right) + \mathrm{T}\log\left| \tilde{\boldsymbol{\Sigma}} \right| + \mathrm{E}_{\boldsymbol{\theta}_0}\left\{ \left(\mathbf{y} - \mathbf{X}^{*}\tilde{\boldsymbol{\pi}}\right)' \left(\tilde{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_{\mathbf{T}}\right)\left(\mathbf{y} - \mathbf{X}^{*}\tilde{\boldsymbol{\pi}}\right) \right\} \\ &\quad - \mathrm{TM}\log\left(2\pi\right) - \mathrm{T}\log\left| \boldsymbol{\Sigma}_0 \right| - \mathrm{E}_{\boldsymbol{\theta}_0}\left\{ \left(\mathbf{y} - \mathbf{X}^{*}\boldsymbol{\pi}_0\right)' \left(\boldsymbol{\Sigma}_0^{-1} \otimes \mathbf{I}_{\mathbf{T}}\right)\left(\mathbf{y} - \mathbf{X}^{*}\boldsymbol{\pi}_0\right) \right\} \\ &\quad + \mathrm{TM}\log\left(2\pi\right) + \mathrm{T}\log\left| \boldsymbol{\Sigma}_0 \right| + \mathrm{E}_{\tilde{\boldsymbol{\theta}}}\left\{ \left(\mathbf{y} - \mathbf{X}^{*}\boldsymbol{\pi}_0\right)' \left(\boldsymbol{\Sigma}_0^{-1} \otimes \mathbf{I}_{\mathbf{T}}\right)\left(\mathbf{y} - \mathbf{X}^{*}\boldsymbol{\pi}_0\right) \right\} \\ &\quad - \mathrm{TM}\log\left(2\pi\right) - \mathrm{T}\log\left| \tilde{\boldsymbol{\Sigma}} \right| - \mathrm{E}_{\tilde{\boldsymbol{\theta}}}\left\{ \left(\mathbf{y} - \mathbf{X}^{*}\tilde{\boldsymbol{\pi}}\right)' \left(\tilde{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_{\mathbf{T}}\right)\left(\mathbf{y} - \mathbf{X}^{*}\tilde{\boldsymbol{\pi}}\right) \right\}. \end{aligned} \tag{11}$$

The second line in (11) or the term $\mathrm{d}\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0\right)$ can be disregarded as it is constant and independent of $\tilde{\boldsymbol{\theta}}$. The estimate of the KL-symmetric divergence becomes (12).

$$2\mathrm{J}\left(\boldsymbol{\theta}_0,\tilde{\boldsymbol{\theta}}\right) = \mathrm{TM}\log\left(2\pi\right)+\mathrm{T}\log\left|\tilde{\boldsymbol{\Sigma}}\right|+\left(\boldsymbol{\pi}_0-\tilde{\boldsymbol{\pi}}\right)'\mathbf{X}^{*\prime}\left(\tilde{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{I}_{\mathrm{T}}\right)\mathbf{X}^{*}\left(\boldsymbol{\pi}_0-\tilde{\boldsymbol{\pi}}\right)+\mathrm{T}\,\mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_0\right)$$
$$+\mathrm{TM}\log\left(2\pi\right)+\mathrm{T}\log\left|\boldsymbol{\Sigma}_0\right|+\left(\tilde{\boldsymbol{\pi}}-\boldsymbol{\pi}_0\right)'\mathbf{X}^{*\prime}\left(\boldsymbol{\Sigma}_0^{-1}\otimes\mathbf{I}_{\mathrm{T}}\right)\mathbf{X}^{*}\left(\tilde{\boldsymbol{\pi}}-\boldsymbol{\pi}_0\right)+\mathrm{T}\,\mathrm{tr}\left(\boldsymbol{\Sigma}_0^{-1}\tilde{\boldsymbol{\Sigma}}\right)$$
$$-\mathrm{TM}\log\left(2\pi\right)-\mathrm{T}\log\left|\tilde{\boldsymbol{\Sigma}}\right|-\mathrm{TM}. \qquad (12)$$

The expected value of the KL-symmetric measure estimates in (12) can be written as (13).

$$\mathrm{E}_{\boldsymbol{\theta}_0}\left\{2\mathrm{J}\left(\boldsymbol{\theta}_0,\tilde{\boldsymbol{\theta}}\right)\right\} = \mathrm{TM}\left[\log\left(2\pi\right)+1\right]+\mathrm{T}\,\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\log\left|\tilde{\boldsymbol{\Sigma}}\right|\right\}+\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\left(\boldsymbol{\pi}_0-\tilde{\boldsymbol{\pi}}\right)'\mathbf{X}^{*\prime}\left(\tilde{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{I}_{\mathrm{T}}\right)\mathbf{X}^{*}\left(\boldsymbol{\pi}_0-\tilde{\boldsymbol{\pi}}\right)\right\}$$
$$+\mathrm{T}\,\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_0\right)\right\}+\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\left(\tilde{\boldsymbol{\pi}}-\boldsymbol{\pi}_0\right)'\mathbf{X}^{*\prime}\left(\boldsymbol{\Sigma}_0^{-1}\otimes\mathbf{I}_{\mathrm{T}}\right)\mathbf{X}^{*}\left(\tilde{\boldsymbol{\pi}}-\boldsymbol{\pi}_0\right)\right\}$$
$$+\mathrm{T}\,\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\mathrm{tr}\left(\boldsymbol{\Sigma}_0^{-1}\tilde{\boldsymbol{\Sigma}}\right)\right\}-\mathrm{T}\,\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\log\left(\left|\tilde{\boldsymbol{\Sigma}}\right|/\left|\boldsymbol{\Sigma}_0\right|\right)\right\}-2\mathrm{TM}. \qquad (13)$$

Disregarding the constant term $\mathrm{TM}\left[\log\left(2\pi\right)+1\right]$ for a given system of M simultaneous equations and T observations, we can now highlight the robustness and reliability of our statistical analysis. Due to $\tilde{\boldsymbol{\pi}}$ and $\mathrm{T}\tilde{\boldsymbol{\Sigma}}$ are asymptotically independent where $\tilde{\boldsymbol{\pi}}$ is asymptotically distributed as a Gaussian distribution with mean vector $\boldsymbol{\pi}$ and covariance matrix $\left[\mathbf{X}^{*\prime}\left(\boldsymbol{\Sigma}_0^{-1}\otimes\mathbf{I}_{\mathrm{T}}\right)\mathbf{X}^{*}\right]^{-1}$ and $\mathrm{T}\tilde{\boldsymbol{\Sigma}}$ is asymptotically distributed as the Wishart distribution with T degrees of freedom, $\mathrm{W}_{\mathrm{KM}}\left(\boldsymbol{\Sigma}_0,\mathrm{T}\right)$ (Anderson 2003), leads us to the result $\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\tilde{\boldsymbol{\Sigma}}^{-1}\right\}=\dfrac{\mathrm{T}}{\mathrm{T}-\mathrm{M}-1}\boldsymbol{\Sigma}_0^{-1}$. We can move forward with computing the values in (13) using the following approach:

$$\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\left(\tilde{\boldsymbol{\pi}}-\boldsymbol{\pi}_0\right)'\mathbf{X}^{*\prime}\left(\boldsymbol{\Sigma}_0^{-1}\otimes\mathbf{I}_{\mathrm{T}}\right)\mathbf{X}^{*}\left(\tilde{\boldsymbol{\pi}}-\boldsymbol{\pi}_0\right)\right\}=\mathrm{KM},$$

$$\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\left(\boldsymbol{\pi}_0-\tilde{\boldsymbol{\pi}}\right)'\mathbf{X}^{*\prime}\left(\tilde{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{I}_{\mathrm{T}}\right)\mathbf{X}^{*}\left(\boldsymbol{\pi}_0-\tilde{\boldsymbol{\pi}}\right)\right\}=\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\mathrm{tr}\left[\left(\tilde{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{I}_{\mathrm{T}}\right)\mathbf{X}^{*}\left(\boldsymbol{\pi}_0-\tilde{\boldsymbol{\pi}}\right)\left(\boldsymbol{\pi}_0-\tilde{\boldsymbol{\pi}}\right)'\mathbf{X}^{*\prime}\right]\right\}$$
$$=\mathrm{tr}\left\{\left[\mathrm{E}_{\boldsymbol{\theta}_0}\left(\tilde{\boldsymbol{\Sigma}}^{-1}\right)\otimes\mathbf{I}_{\mathrm{T}}\right]\mathrm{E}_{\boldsymbol{\theta}_0}\left[\mathbf{X}^{*}\left(\boldsymbol{\pi}_0-\tilde{\boldsymbol{\pi}}\right)\left(\boldsymbol{\pi}_0-\tilde{\boldsymbol{\pi}}\right)'\mathbf{X}^{*\prime}\right]\right\}$$
$$=\frac{\mathrm{T}}{\mathrm{T}-\mathrm{M}-1}\mathrm{tr}\left\{\mathrm{E}_{\boldsymbol{\theta}_0}\left[\left(\boldsymbol{\pi}_0-\tilde{\boldsymbol{\pi}}\right)'\mathbf{X}^{*\prime}\left(\boldsymbol{\Sigma}_0^{-1}\otimes\mathbf{I}_{\mathrm{T}}\right)\mathbf{X}^{*}\left(\boldsymbol{\pi}_0-\tilde{\boldsymbol{\pi}}\right)\right]\right\}=\frac{\mathrm{TKM}}{\mathrm{T}-\mathrm{M}-1},$$

$$\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\mathrm{T}\,\mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_0\right)\right\}=\mathrm{T}\,\mathrm{tr}\left\{\mathrm{E}_{\boldsymbol{\theta}_0}\left(\tilde{\boldsymbol{\Sigma}}^{-1}\right)\boldsymbol{\Sigma}_0\right\}=\mathrm{T}\,\mathrm{tr}\left\{\frac{\mathrm{T}}{\mathrm{T}-\mathrm{M}-1}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_0\right\}=\frac{\mathrm{T}^2\mathrm{M}}{\mathrm{T}-\mathrm{M}-1},$$

$$\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\mathrm{T}\,\mathrm{tr}\left(\boldsymbol{\Sigma}_0^{-1}\tilde{\boldsymbol{\Sigma}}\right)\right\}=\mathrm{tr}\left\{\boldsymbol{\Sigma}_0^{-1}\mathrm{E}_{\boldsymbol{\theta}_0}\left(\mathrm{T}\tilde{\boldsymbol{\Sigma}}\right)\right\}=\mathrm{T}\,\mathrm{tr}\left\{\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_0\right\}=\mathrm{TM},$$

$$\mathrm{T}\,\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\log\left(\left|\tilde{\boldsymbol{\Sigma}}\right|/\left|\boldsymbol{\Sigma}_0\right|\right)\right\}=\mathrm{T}\,\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\log\left|\mathrm{T}\tilde{\boldsymbol{\Sigma}}\right|-\log\left|\boldsymbol{\Sigma}_0\right|-\log\left(\mathrm{T}^{\mathrm{M}}\left|\tilde{\boldsymbol{\Sigma}}\right|\right)+\log\left|\tilde{\boldsymbol{\Sigma}}\right|\right\}$$
$$=\mathrm{T}\,\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\log\left(\left|\mathrm{T}\tilde{\boldsymbol{\Sigma}}\right|/\left|\boldsymbol{\Sigma}_0\right|\right)\right\}-\mathrm{TM}\log\mathrm{T}.$$

Therefore, the expected value in (13) can be more simply expressed in (14).

$$\mathrm{E}_{\boldsymbol{\theta}_0}\left\{2\mathrm{J}\left(\boldsymbol{\theta}_0,\tilde{\boldsymbol{\theta}}\right)\right\}=\mathrm{T}\,\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\log\left|\tilde{\boldsymbol{\Sigma}}\right|\right\}+\frac{\mathrm{TKM}}{\mathrm{T}-\mathrm{M}-1}+\frac{\mathrm{T}^2\mathrm{M}}{\mathrm{T}-\mathrm{M}-1}+\mathrm{KM}+\mathrm{TM}$$
$$-\mathrm{T}\,\mathrm{E}_{\boldsymbol{\theta}_0}\left\{\log\left(\left|\mathrm{T}\tilde{\boldsymbol{\Sigma}}\right|/\left|\boldsymbol{\Sigma}_0\right|\right)\right\}+\mathrm{TM}\log\mathrm{T}-2\mathrm{TM}$$

$$= T E_{\boldsymbol{\theta}_0}\left\{\log\left|\tilde{\boldsymbol{\Sigma}}\right|\right\} - T E_{\boldsymbol{\theta}_0}\left\{\log\left(\left|T\tilde{\boldsymbol{\Sigma}}\right| / \left|\boldsymbol{\Sigma}_0\right|\right)\right\} + TM\log T + \frac{TM(2K+M+1)}{T-M-1} - \frac{KM(M+1)}{T-M-1}. \quad (14)$$

Because $\left|T\tilde{\boldsymbol{\Sigma}}\right| / \left|\boldsymbol{\Sigma}_0\right|$ in (14) is the distribution of a product of independent χ^2 random variables, $\prod_{i=1}^{M}\chi^2_{T-K-M+i}$, then we have $\log\left(\left|T\tilde{\boldsymbol{\Sigma}}\right| / \left|\boldsymbol{\Sigma}_0\right|\right) \sim \sum_{i=1}^{M}\log\chi^2_{T-K-M+i}$. Using the second-order of Taylor series expansions to expand the function of $\log\left(\chi^2_p\right)$ about the mean p, we have

$$\log\left(\chi^2_p\right) \doteq \log(p) + \frac{1}{p}\left(\chi^2_p - p\right) - \frac{1}{2p^2}\left(\chi^2_p - p\right)^2 \text{ and } E\left[\log\left(\chi^2_p\right)\right] \doteq \log(p) - \frac{1}{p}.$$

Then, the term $E_{\boldsymbol{\theta}_0}\left\{\log\left(\left|T\tilde{\boldsymbol{\Sigma}}\right| / \left|\boldsymbol{\Sigma}_0\right|\right)\right\}$ in (14) becomes (15).

$$E_{\boldsymbol{\theta}_0}\left\{\log\left(\left|T\tilde{\boldsymbol{\Sigma}}\right| / \left|\boldsymbol{\Sigma}_0\right|\right)\right\} \doteq \sum_{i=1}^{M}\left[\log(T-K-M+i) - \frac{1}{T-K-M+i}\right]. \quad (15)$$

McQuarrie and Tsai (1998) gave the simplification formulae for any T, K, M and assume $T-K-M$ is much larger than M, therefore the (15) become

$$\sum_{i=1}^{M}\log(T-K-M+i) = M\log\left(T-K-\frac{M-1}{2}\right) = M\log\left(\frac{2T-2K-M+1}{2}\right), \quad (16)$$

and

$$\sum_{i=1}^{M}\frac{1}{T-K-M+i} \doteq \frac{M}{T-K-\frac{M-1}{2}} = \frac{2M}{2T-2K-M+1}. \quad (17)$$

Replacing the results in (15), (16) and (17) into (14), we have the expected KL-symmetric measure as

$$E_{\boldsymbol{\theta}_0}\left\{2J\left(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}\right)\right\} = T E_{\boldsymbol{\theta}_0}\left\{\log\left|\tilde{\boldsymbol{\Sigma}}\right|\right\} + TM\log\left(\frac{2T}{2T-2K-M+1}\right) + \frac{2TM}{2T-2K-M+1} + \frac{TM(2K+M+1)}{T-M-1} - \frac{KM(M+1)}{T-M-1}. \quad (18)$$

We can define a new model selection criterion based on (18) and adjust the value by scaling the sample size (T). This criterion, named SKIC(MLE), is particularly relevant when using maximum likelihood estimation for parameter estimation and is defined by the following formula.

$$\text{SKIC(MLE)} = \log\left|\tilde{\boldsymbol{\Sigma}}\right| + M\log\left(\frac{2T}{2T-2K-M+1}\right) + \frac{2M}{2T-2K-M+1} + \frac{M(2K+M+1)}{T-M-1} - \frac{KM(M+1)}{T(T-M-1)}, \quad (19)$$

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

In this study, we conduct data simulation to compare the model selection efficiency of the newly developed criterion in (19) with the existing criterion, SKIC (Keerativibool and Jitthavech 2015) in (20) below.

$$\text{SKIC} = \log\left|\hat{\boldsymbol{\Sigma}}\right| + \text{M}\log\left(\frac{2\text{T}}{2\text{T}-2\text{K}-\text{M}+1}\right) + \frac{2\text{M}}{2\text{T}-2\text{K}-\text{M}+1} + \frac{\text{M}(2\text{K}+\text{M}+1)}{\text{T}-\text{K}-\text{M}-1}, \tag{20}$$

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

3. Simulation Study

In this paper, we delve into the true model of SEM, which comprises two systems of three equations (M = 3). One system has low parameter values and is challenging to estimate, while the other has high parameter values and is thus easy to estimate. These systems are detailed in Table 1, along with the three sample sizes we will work with: small samples T = 15, 20 and 25, medium samples T = 30 and 50, and large samples T = 100 and 500 observations. The following simulation procedures and results aim to provide a comprehensive understanding of the model selection process.

Table 1 True model of SEM to be considered in this paper

Challenging to estimate parameters	Easily to estimate parameters
$y_{t1} = 1 + 0.9x_{t2} + 0.8x_{t3} + 0.7x_{t4} + \varepsilon_{t1}$	$y_{t1} = 1 + 3x_{t2} + 2x_{t3} + x_{t4} + \varepsilon_{t1}$
$y_{t2} = 1 - 0.25x_{t2} - 0.13x_{t3} - 0.05x_{t4} + \varepsilon_{t2}$	$y_{t2} = 1 - 0.5x_{t2} - 0.25x_{t3} - 0.1x_{t4} + \varepsilon_{t2}$
$y_{t3} = 1 + 0.5x_{t2} + 0.5x_{t3} + 0.5x_{t4} + \varepsilon_{t3}$	$y_{t3} = 1 + x_{t2} + x_{t3} + x_{t4} + \varepsilon_{t3}$

3.1. Generate 1,000,000 vectors of the 3×1 multivariate normal $\boldsymbol{\varepsilon}_t$ in (5) with a mean vector of zero and specified correlation coefficients of errors between the equations are equal to $\rho_{12} = 0.9$, $\rho_{13} = 0.7$, $\rho_{23} = 0.8$, and the variances-covariances of the errors are

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

The form to be generated $\boldsymbol{\varepsilon}_t$ in (5) is represented by:

$$\boldsymbol{\varepsilon}_t = \begin{bmatrix}\varepsilon_{t1}\\ \varepsilon_{t2}\\ \varepsilon_{t3}\end{bmatrix} \sim \text{N}_3\left(\mathbf{0},\ \boldsymbol{\Sigma} = \begin{bmatrix}0.81 & 0.648 & 0.441\\ 0.648 & 0.64 & 0.448\\ 0.441 & 0.448 & 0.49\end{bmatrix}\right).$$

3.2. Partition the errors set obtained in 3.1. into 1,000 samples, with each sample comprising three different sample sizes: small (T = 15, 20, 25), medium (T = 30, 50), and large (T = 100, 500) observations.

3.3. Generate the independent variables x_{t2} until x_{t8} with approximately 1,000,000 observations as normal random variables with zero means and variances equal to one. The relevant independent variables are x_{t2}, x_{t3}, and x_{t4} and irrelevant independent variables are x_{t5}, x_{t6} x_{t7}, and x_{t8}.

3.4. Partition the independent variables x_{t2} until x_{t8} obtained in 3.3. into 1,000 samples, with each sample comprising three different sample sizes: small (T = 15, 20, 25), medium (T = 30, 50), and large (T = 100, 500) observations.

3.5. Test for multicollinearity issues among the independent variables in 3.4 and eliminate any samples that fail the test. Only 1,000 samples were retained for further analysis.

3.6. Create the dependent variables as described in Table 1 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. Utilize a systematic and structured approach, employing the nested model assumption to form the candidate models. This method involves a thorough and gradual process of adding columns of independent variables, ranging from columns 1 to K, resulting in a design matrix for a candidate model with dimensions K. For a comprehensive sample size of 1,000, we calculate SKIC(MLE) in (19) and also SKIC in (20). The method's effectiveness is demonstrated when we identify the candidate model with the lowest value of the model selection criterion as the best model. For simulation of the first sample when the true model of SEM is challenging to estimate, for the model on the right-hand side of Table 1, with T = 15, M = 3, and K = 2, 3, …, 8, we have the values of SKIC(MLE) and SKIC presented in Table 2. Under the appropriate model selection conditions mentioned above, it is found that under SKIC(MLE), the appropriate model should have an order equal to 4, while under SKIC, the appropriate model should have an order equal to 2. Therefore, SKIC(MLE) selects the correct model because it has the same order as the true model. It is worth noting that this SKIC leads to an incorrect model ranking. This selection is evident from the minimum value of the criterion in the K = 2 case. The rapidly increasing inappropriateness of SKIC values highlights the limitations of current model ranking methods and the urgent necessity for an alternative, such as SKIC(MLE). Tables 3 and 4 detail the efficiency comparison of the percentage of correct model selections from SKIC(MLE) and SKIC for 1,000 samples. The evidence presented in these tables proves that SKIC(MLE) consistently outperforms SKIC when model estimation is challenging for small samples T = 15 and 20, and the model is easy to estimate for T = 15. The main reason is SKIC's tendency to provide faster estimates as K increases. However, for medium to large samples, SKIC performs better than SKIC(MLE).

Table 2 Values of SKIC(MLE) and SKIC from the first sample with T = 15, M = 3

K	SKIC(MLE)	SKIC	K	SKIC(MLE)	SKIC
2	-0.2359	**0.8237**	6	0.5499	7.7551
3	-0.3474	1.563	7	1.1678	12.1537
4	**-1.4647**	1.6268	8	2.1882	19.6019
5	-0.4163	4.3455			

Bold type indicates the minimum value of model selection criterion.

3.8. To increase confidence, we further validated the model selection criteria by measuring observed L_2 efficiency, as shown in (21). The comparison details are shown in Tables 5 and 6. The observed L_2 efficiency is defined as (McQuarrie et al. 1997)

$$\text{Observed } L_2 \text{ efficiency } = \frac{\min_{1\le K\le K^*} L_2(K)}{L_2(K_s)}, \tag{21}$$

where

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

is the observed L_2 distance between the true and the candidate model, K^* represents the class of all possible candidate models, while K represents the rank of the candidate model that has been fitted, and K_s represents the model chosen by a specific model selection criterion. A higher efficiency level is achieved when the selected model closely aligns with the true model. Therefore, the best model

selection criterion will choose a model that performs well even with small sample sizes. Tables 5 and 6 meticulously and comprehensively detail the mean and standard deviation of the observed L_2 efficiency from SKIC(MLE) in (19) and SKIC in (20) for 1,000 samples. They show that SKIC(MLE) is a better criterion than SKIC when model estimation is challenging for small samples with T = 15 and 20. However, the same conclusion holds for T = 15 when the model is easy to estimate. Whereas for medium to large samples, whether the model is easy or challenging to estimate, SKIC performs better than SKIC(MLE).

Table 3 Percentage of correct model selections from the model is challenging to estimate

Sample sizes	T	Criteria	K						
			2	3	4*	5	6	7	8
Small	15	SKIC(MLE)	2.3	1.6	**93.1**	2.6	0.1	0.1	0.2
		SKIC	81.9	5.1	13	0	0	0	0
	20	SKIC(MLE)	0	0.1	**96.8**	2.7	0.4	0	0
		SKIC	7.3	2.6	90.1	0	0	0	0
	25	SKIC(MLE)	0	0	98.2	1.7	0.1	0	0
		SKIC	0.1	0.4	**99.4**	0.1	0	0	0
Medium	30	SKIC(MLE)	0	0	98.7	1.3	0	0	0
		SKIC	0	0	**100**	0	0	0	0
	50	SKIC(MLE)	0	0	99.4	0.6	0	0	0
		SKIC	0	0	**99.9**	0.1	0	0	0
Large	100	SKIC(MLE)	0	0	99.3	0.7	0	0	0
		SKIC	0	0	**100**	0	0	0	0
	500	SKIC(MLE)	0	0	99.3	0.7	0	0	0
		SKIC	0	0	**99.8**	0.2	0	0	0

* Represents the true model; bold type indicates the maximum percentage of correct model selections.

Table 4 Percentage of correct model selections from the model is easy to estimate

Sample sizes	T	Criteria	K						
			2	3	4*	5	6	7	8
Small	15	SKIC(MLE)	0	0	**97**	2.6	0.1	0.1	0.2
		SKIC	4.3	6	89.7	0	0	0	0
	20	SKIC(MLE)	0	0	96.9	2.7	0.4	0	0
		SKIC	0	0	**100**	0	0	0	0
	25	SKIC(MLE)	0	0	98.2	1.7	0.1	0	0
		SKIC	0	0	**99.9**	0.1	0	0	0
Medium	30	SKIC(MLE)	0	0	98.7	1.3	0	0	0
		SKIC	0	0	**100**	0	0	0	0
	50	SKIC(MLE)	0	0	99.4	0.6	0	0	0
		SKIC	0	0	**99.9**	0.1	0	0	0
Large	100	SKIC(MLE)	0	0	99.3	0.7	0	0	0
		SKIC	0	0	**100**	0	0	0	0
	500	SKIC(MLE)	0	0	99.3	0.7	0	0	0
		SKIC	0	0	**99.8**	0.2	0	0	0

* Represents the true model; bold type indicates the maximum percentage of correct model selections.

Table 5 Mean and standard deviation of the observed L_2 efficiency from the model is challenging to estimate

Sample sizes	T	Criteria	$\bar{X}$	s	Sample sizes	T	Criteria	$\bar{X}$	s
Small	15	SKIC(MLE)	**0.8698**	**0.2137**	Medium	30	SKIC(MLE)	0.9904	0.0699
		SKIC	0.7090	0.2157			SKIC	**0.9981**	**0.0168**
	20	SKIC(MLE)	**0.9542**	**0.1403**		50	SKIC(MLE)	0.9962	0.0455
		SKIC	0.9288	0.1667			SKIC	**0.9991**	**0.0176**
	25	SKIC(MLE)	0.9807	0.0928	Large	100	SKIC(MLE)	0.9963	0.0448
		SKIC	**0.9881**	**0.0630**			SKIC	**1.0000**	**0.0005**
						500	SKIC(MLE)	0.9956	0.0534
							SKIC	**0.9986**	**0.0308**

Bold type indicates the highest mean of the observed L_2 efficiency and the lowest standard deviation.

Table 6 Mean and standard deviation of the observed L_2 efficiency from the model is easy to estimate

Sample sizes	T	Criteria	$\bar{X}$	s	Sample sizes	T	Criteria	$\bar{X}$	s
Small	15	SKIC(MLE)	**0.9089**	**0.1895**	Medium	30	SKIC(MLE)	0.9907	0.0690
		SKIC	0.8848	0.1959			SKIC	**0.9984**	**0.0127**
	20	SKIC(MLE)	0.9645	0.1304		50	SKIC(MLE)	0.9963	0.0453
		SKIC	**0.9849**	**0.0638**			SKIC	**0.9992**	**0.0172**
	25	SKIC(MLE)	0.9833	0.0884	Large	100	SKIC(MLE)	0.9963	0.0448
		SKIC	**0.9935**	**0.0400**			SKIC	**1.0000**	**0.0005**
						500	SKIC(MLE)	0.9956	0.0534
							SKIC	**0.9986**	**0.0308**

Bold type indicates the highest mean of the observed L_2 efficiency and the lowest standard deviation.

4. Conclusions

Our paper introduces the Kullback information criterion for the simultaneous equations model. This criterion uses a maximum likelihood estimator called SKIC(MLE), a precise tool for selecting the best model set. Our method, SKIC(MLE), performs better than the SKIC method proposed by Keerativibool and Jitthavech (2015) when dealing with small sample sizes (T = 15 and 20). It excels in challenging estimation situations, resulting in more accurate model selections and greater observed L_2 efficiency. However, for medium to larger sample sizes, SKIC performs better than SKIC(MLE). The proposed criterion outperforms SKIC in small sample sizes because the penalty term of SKIC grows exponentially as the number of parameters increases, as depicted in Figure 1. This rapid increase in the penalty term of SKIC leads to underfitting the order of the model, as shown in Tables 3 and 4, which is in line with the findings of Keerativibool (2014) and Pham (2019).

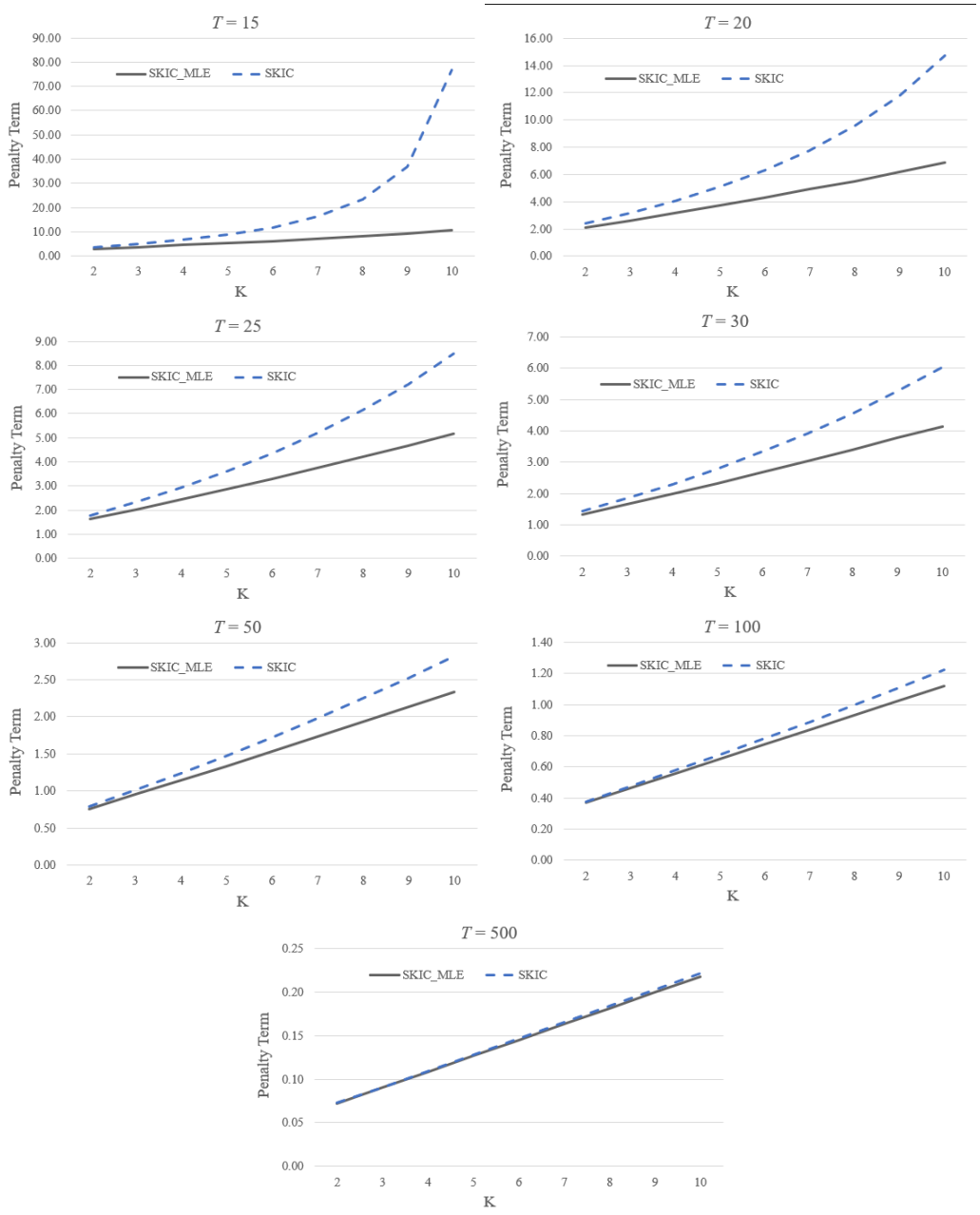


Figure 1 Comparisons of the penalty terms from SKIC(MLE) and SKIC

In Tables 7 and 8, the means and standard deviations of SKIC(MLE) and SKIC values for 1,000 samples indicate that the SKIC value has a higher mean than the SKIC(MLE) value, especially when dealing with small samples. The unrealistically high value of the criterion is likely the main reason for selecting a correct model order from SKIC that is lower than SKIC(MLE). This study greatly benefits the fields of statistics and economics in small sample sizes of simultaneous equations modelling. However, the need for additional criteria in SEM selection is not just important, it's urgent. This urgency underscores the significance of our research. In the future, we plan to investigate the potential

for developing other model selection criteria for SEM. We aim to compare the performance of selecting the correct model order with the criteria proposed in this paper and apply it to real data.

Table 7 Mean and standard deviation of SKIC(MLE) and SKIC for 1,000 samples from the model is challenging to estimate

Sample sizes	T	Criteria	Stat.	K						
				2	3	4*	5	6	7	8
Small	15	SKIC(MLE)	$\bar{X}$	0.702	0.597	**-0.608**	0.142	0.930	1.742	2.593
			s	0.727	0.785	0.854	0.900	0.957	1.031	1.151
		SKIC	$\bar{X}$	**1.761**	2.508	2.483	4.904	8.135	12.728	20.006
			s	0.727	0.785	0.854	0.900	0.957	1.031	1.151
	20	SKIC(MLE)	$\bar{X}$	0.089	-0.229	**-1.613**	-1.094	-0.550	0.008	0.578
			s	0.603	0.635	0.659	0.694	0.722	0.753	0.794
		SKIC	$\bar{X}$	0.694	0.804	**-0.044**	1.150	2.545	4.188	6.160
			s	0.603	0.635	0.659	0.694	0.722	0.753	0.794
	25	SKIC(MLE)	$\bar{X}$	-0.230	-0.652	**-2.154**	-1.748	-1.330	-0.909	-0.478
			s	0.532	0.545	0.566	0.593	0.607	0.623	0.645
		SKIC	$\bar{X}$	0.186	0.038	**-1.136**	-0.340	0.545	1.522	2.620
			s	0.532	0.545	0.566	0.593	0.607	0.623	0.645
Medium	30	SKIC(MLE)	$\bar{X}$	-0.431	-0.930	**-2.497**	-2.165	-1.821	-1.482	-1.139
			s	0.477	0.489	0.499	0.511	0.518	0.531	0.553
		SKIC	$\bar{X}$	-0.117	-0.417	**-1.754**	-1.157	-0.506	0.188	0.940
			s	0.477	0.489	0.499	0.511	0.518	0.531	0.553
	50	SKIC(MLE)	$\bar{X}$	-0.829	-1.450	**-3.163**	-2.970	-2.778	-2.581	-2.384
			s	0.362	0.374	0.382	0.387	0.390	0.391	0.396
		SKIC	$\bar{X}$	-0.672	-1.204	**-2.817**	-2.516	-2.207	-1.881	-1.545
			s	0.362	0.374	0.382	0.387	0.390	0.391	0.396
Large	100	SKIC(MLE)	$\bar{X}$	-1.090	-1.809	**-3.615**	-3.522	-3.429	-3.335	-3.242
			s	0.259	0.261	0.264	0.266	0.267	0.268	0.269
		SKIC	$\bar{X}$	-1.022	-1.704	**-3.471**	-3.338	-3.202	-3.064	-2.925
			s	0.259	0.261	0.264	0.266	0.267	0.268	0.269
	500	SKIC(MLE)	$\bar{X}$	-1.296	-2.092	**-3.976**	-3.958	-3.940	-3.922	-3.904
			s	0.107	0.108	0.110	0.110	0.110	0.110	0.110
		SKIC	$\bar{X}$	-1.283	-2.073	**-3.951**	-3.927	-3.902	-3.878	-3.853
			s	0.107	0.108	0.110	0.110	0.110	0.110	0.110

* Represents the true model and bold type indicates the lowest mean of the criterion.

Table 8 Mean and standard deviation of SKIC(MLE) and SKIC for 1,000 samples from the model can be easily estimated

Sample sizes	T	Criteria	Stat.	K						
				2	3	4*	5	6	7	8
Small	15	SKIC(MLE)	$\bar{X}$	2.663	1.506	**-0.608**	0.142	0.930	1.742	2.593
			s	0.721	0.789	0.854	0.900	0.957	1.031	1.151
		SKIC	$\bar{X}$	3.723	3.416	**2.483**	4.904	8.135	12.728	20.006
			s	0.721	0.789	0.854	0.900	0.957	1.031	1.151
	20	SKIC(MLE)	$\bar{X}$	2.042	0.681	**-1.613**	-1.094	-0.550	0.008	0.578
			s	0.592	0.634	0.659	0.694	0.722	0.753	0.794
		SKIC	$\bar{X}$	2.647	1.713	**-0.044**	1.150	2.545	4.188	6.160
			s	0.592	0.634	0.659	0.694	0.722	0.753	0.794
	25	SKIC(MLE)	$\bar{X}$	1.724	0.255	**-2.154**	-1.748	-1.330	-0.909	-0.478
			s	0.520	0.547	0.566	0.593	0.607	0.623	0.645
		SKIC	$\bar{X}$	2.140	0.945	**-1.136**	-0.340	0.545	1.522	2.620
			s	0.520	0.547	0.566	0.593	0.607	0.623	0.645
Medium	30	SKIC(MLE)	$\bar{X}$	1.527	-0.024	**-2.497**	-2.165	-1.821	-1.482	-1.139
			s	0.467	0.490	0.499	0.511	0.518	0.531	0.553
		SKIC	$\bar{X}$	1.841	0.489	**-1.754**	-1.157	-0.506	0.188	0.940
			s	0.467	0.490	0.499	0.511	0.518	0.531	0.553
	50	SKIC(MLE)	$\bar{X}$	1.132	-0.543	**-3.163**	-2.970	-2.778	-2.581	-2.384
			s	0.360	0.375	0.382	0.387	0.390	0.391	0.396
		SKIC	$\bar{X}$	1.288	-0.297	**-2.817**	-2.516	-2.207	-1.881	-1.545
			s	0.360	0.375	0.382	0.387	0.390	0.391	0.396
Large	100	SKIC(MLE)	$\bar{X}$	0.865	-0.903	**-3.615**	-3.522	-3.429	-3.335	-3.242
			s	0.259	0.260	0.264	0.266	0.267	0.268	0.269
		SKIC	$\bar{X}$	0.934	-0.798	**-3.471**	-3.338	-3.202	-3.064	-2.925
			s	0.259	0.260	0.264	0.266	0.267	0.268	0.269
	500	SKIC(MLE)	$\bar{X}$	0.662	-1.186	**-3.976**	-3.958	-3.940	-3.922	-3.904
			s	0.109	0.108	0.110	0.110	0.110	0.110	0.110
		SKIC	$\bar{X}$	0.674	-1.167	**-3.951**	-3.927	-3.902	-3.878	-3.853
			s	0.109	0.108	0.110	0.110	0.110	0.110	0.110

* Represents the true model and bold type indicates the lowest mean 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.

Anderson TW. An introduction to multivariate statistical analysis. 3rd ed. New Jersey: John Wiley & Sons; 2003.

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.

Greene W. Econometric analysis. 6th ed. New Jersey: Prentice-Hall; 2008.

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. 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. Study on the penalty functions of model selection criteria. Thail Stat. 2014; 12(2): 161-178.

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 AD, Tsai CL. Regression and time series model selection. Singapore: World Scientific Pub Co Inc; 1998.

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

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, https://doi.org/10.3390/math7121215

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.

Zhang J, Yang Y, Ding J. Information criteria for model selection. WIREs Comp Stats. 2023; 15(5): 1-27.