Article

A New Computational Method for Estimating Simultaneous Equations Models Using Entropy as a Parameter Criteria

Belén Pérez-Sánchez [1,*] **, Martín González** [2] **, Carmen Perea** [2] **and Jose J. López-Espín** [2]

1 Department of Statistics, Mathematics and Informatics. Miguel Hernández University, 03202 Elche, Spain
2 Center of Operations Research, Miguel Hernández University, 03202 Elche, Spain;
mgonzaleze@umh.es (M.G.); perea@umh.es (C.P.); jlopez@umh.es (J.J.L.-E.)
* Correspondence: m.perezs@umh.es

Abstract: Simultaneous Equations Models (SEM) is a statistical technique widely used in economic science to model the simultaneity relationship between variables. In the past years, this technique has also been used in other fields such as psychology or medicine. Thus, the development of new estimating methods is an important line of research. In fact, if we want to apply the SEM to medical problems with the main goal being to obtain the best approximation between the parameters of model and their estimations. This paper shows a computational study between different methods for estimating simultaneous equations models as well as a new method which allows the estimation of those parameters based on the optimization of the Bayesian Method of Moments and minimizing the Akaike Information Criteria. In addition, an entropy measure has been calculated as a parameter criteria to compare the estimation methods studied. The comparison between those methods is performed through an experimental study using randomly generated models. The experimental study compares the estimations obtained by the different methods as well as the efficiency when comparing solutions by Akaike Information Criteria and Entropy Measure. The study shows that the proposed estimation method offered better approximations and the entropy measured results more efficiently than the rest.

Keywords: simultaneous equations models; bayesian method of moments; markov chain monte carlo; akaike information criteria; entropy; computational statistics

1. Introduction

Simultaneous Equations Models (SEM) [1] is statistical model formed by a set of regression equations that reflect the simultaneity between the set of dependent and independent variables of the model. SEM is used when there is a bidirectional influence relationship between both types of variables. The estimation of the coefficients of a SEM can be made by methods based on either the classical statistical approach or the Bayesian approach.

Regarding classic inference, the estimation of a SEM can be made by limited information and full information methods. Limited information methods estimate each of the equations of the structural form [1] without making use of the information contained in the detailed specification of the rest of the model, only considering both the endogenous and exogenous variables that are included in this equation. Ordinary Least Squares (OLS), Indirect Least Squares (ILS), and the Two Stage Least Squares (2SLS) are examples of limited information methods [1]. Full information methods consider joint estimation of the whole model in the structural form. These methods require the specification all equations, and all of them have to be identified. In general, they are more asymptotically efficient than the others since they incorporate all the information of the system, but, with the drawback that if any equation is incorrectly specified, estimates that are inconsistent with the other equations may be generated. Examples of these kinds of methods are Full Information Maximum Likelihood (FIML) or Three Stage Least Squares (3SLS) [1].

On the other hand, the Bayesian inference which is based on a given set of data, does not use sampling assumptions, but introduces a high degree of complexity due to the prior specification of the distribution, as well as the obtaining of the posterior distribution. Some estimation techniques in the Bayesian approach are the Bayesian Method of Moments (BMOM) [2] or the methods used by Chao and Phillips [3], Geweke [4], and Kleibergen and Van Dijk [5]. The recent development of the Markov Chain Monte Carlo method has been key in making the computation of large models that require integrations over hundreds or even thousands of unknown parameters possible. The Metropolis–Hastings algorithm and the Gibbs Sampling [6] are examples of them.

Regarding model selection, the literature is limited to a comparison between Bayesian and classical estimators, concluding that Bayesian methods perform better in the case of a small sample [2].

Applications of SEMs can be found mainly in the economic framework, although there are some applications of SEMs in other fields. For instance, in tax research, the effects of fiscal decentralization on regional income inequality in Indonesia have been studied using provincial-level data over the period 2001–2014 [7], or finding a SEM that relates employment to mental health [8], others have studied the impact of foreign trade on energy efficiency in China's textile industry [9] or biomass energy consumption, economic growth, and carbon emission in West Africa [10]. And other studies have concentrated on comparing the results obtained by SEMs versus linear regression modeling of complex phenotypes [11], a model for knowing the forecast demand on the facilities provided airside at airports [12] or a SEM for modeling prescriptions in primary care [13]. The Bayesian SEM approach has been used in agricultural science [14], or for studying the impact of product information on third-party websites on the feedback mechanism between internal word-of-mouth and retail sales on Download.com and Amazon.com (accessed on 1 March 2020) [15]. It can be useful for analyzing the interdependence of a television program viewership between spouses [16], for exploring peer effects in casino gambling behavior [17], for modeling the interaction between people's health risk perception and betel chewing habits in Taiwan [18], and for studying the effects of repetitive iodine thyroid blocking on the development of the foetal brain and thyroid in rats [19].

When several estimate models are available, it is necessary to have a selection parameter criteria. There are a lot of useful information parameter criteria for comparing SEMs, such as Akaike Information Criteria (AIC) [20,21], its corrected version (AICc) [22], Schwarz Information Criteria (SIC) [23], Bayesian Information Criteria (BIC) [24], Hannan and Quinn (HQ) [25], and Model Selection Criterion based on Kullback–Leibler's Symmetric Divergence [26].

Entropy was initially introduced in thermodynamics, where it was used to provide the basis for the second law of thermodynamics. Subsequently, mechanical statistics provided a connection between the macroscopic properties of entropy and the states of the system, and from a mathematical point of view, are non-negative functions defined in probability distributions with multiple applications such as using information theory for measuring a system stability [27]. Applications of entropy have been used in the fields of finance, [28], environmental and water engineering [29], urban systems [30], and applications for customer satisfaction surveys [31].

In our paper, a new method for the SEM estimation is developed and compared with other methods through the AIC and an entropy measure developed by Amigó [32], which allows us to select the estimation method with the highest homogeneity in the estimation errors.

The organization of the paper is as follows: In Section 2, the model is set up and several classical and Bayesian methods for estimating SEM are briefly reviewed. Section 3 describes the proposed estimation method. In Section 4, the entropy is shown and a new version is obtained as information criteria for selecting the method of estimation with minimum error. The experimental design and the results are shown in Section 5, and, finally, the conclusions and future lines of work are presented in Section 6.

2. Definition of the Model and Methods for Estimating a SEM Problem

2.1. Definition of the Model

Consider m interdependent or endogenous variables which depend on k independent or exogenous variables. Suppose that each endogenous variable can be expressed as a linear combination of the other endogenous variables, the exogenous ones, and white noise that represents stochastic interference. Thus, a SEM as a matrix form is [1]:

$$
\begin{aligned}
y_1 &= B_{1,2}y_2 + B_{1,3}y_3 + \cdots + B_{1,m}y_m + \Gamma_{1,1}x_1 + \cdots + \Gamma_{1,k}x_k + u_1 \\
y_2 &= B_{2,1}y_1 + B_{2,3}y_3 + \cdots + B_{2,m}y_m + \Gamma_{2,1}x_1 + \cdots + \Gamma_{2,k}x_k + u_2 \\
&\ \vdots \\
y_m &= B_{m,1}y_1 + B_{m,2}y_2 + \cdots + B_{m,m-1}y_{m-1} + \Gamma_{k,1}x_1 + \cdots + \Gamma_{m,k}x_k + u_m.
\end{aligned}
\tag{1}
$$

The equations can be represented in matrix form as:

$$
YB^T + X\Gamma^T + U = 0
\tag{2}
$$

where $B \in \mathbb{R}^{m \times m}$ and $\Gamma \in \mathbb{R}^{m \times k}$ are matrices of coefficients, $Y \in \mathbb{R}^{n \times m}$ is the matrix of endogenous variables, $X \in \mathbb{R}^{n \times k}$ is the matrix of exogenous variables, and $U \in \mathbb{R}^{n \times m}$ is the matrix of white noise variables, being n the sample size. Some coefficients of B and Γ are zero, and are known a priori. The number of endogenous and exogenous variables in the ith equation of (1) is denoted by m_i and k_i. An equation is identified if the number of variables (endogenous and exogenous) in the equation is lower than or equal to $k + 1$, that is $m_i - 1 \le k - k_i$ (order condition (1)). When $m_i - 1 = k - k_i$, the equation is exactly identified and when $m_i - 1 < k - k_i$ it is over-identified. Only identified equations can be solved. Solving the model is equivalent to obtaining a estimation of B and Γ in (2) from a representative sample of the model (a set of values of the data variables X and Y), in order to explain a well-known matrix equation that represents the relationship between both sets of variables.

2.2. Methods for Estimating an SEM Problem

There are different techniques for estimating SEM parameters. In the classical approach, examples of these are 3SLS, 2SLS, OLS, K-class estimators [33], etc., and in the Bayesian approach, MCMC algorithms, and several conjugate distributions are some examples. In this section, a brief review of those estimators that have been used in our work is shown.

2.2.1. Two Stage Least Squares (2SLS)

2SLS is the most common estimation method for a SEM [1], developed independently by Theil (1953) [34] and Basmann (1957) [35]. This method is called two stage, because in its mathematical expression, Ordinary Least Square (OLS) is applied twice. In the first step, new variables, called proxy, are calculated by Least Squares using the exogenous variables of the model as independent variables, and in the second step, the endogenous variables are substituted by the *proxys*, and then Least Squares is applied once more. Both, the 2SLS and OLS method are single K-class estimators [33], expressed in (3) when $K_1 = K_2 = 0$ (OLS), and when $K_1 = K_2 = 1$ (2SLS).

2.2.2. Bayesian Method of Moments (*BMOM*)

This method, proposed by Zellner in 1998 [2], applies the principle of maximum entropy and generates optimal estimation evaluated by double K-class estimators shown in Table 1. When there is not enough information available to obtain the likelihood function, allowing for data analysis without specifying a probability function and sampling assumption.

Considering, for example, the first structural equation, $y_1 = Y_1\beta_1 + X_1\gamma_1 + u_1$, being $Y_1 \in \mathbb{R}^{n \times m}$ and $X_1 \in \mathbb{R}^{n \times k}$ the matrix of endogenous and exogenous variables and

$u_1 \in \mathbb{R}^{n \times 1}$, the white noise vector. The parameters $\delta_1 = (\beta_1 \; \gamma_1)' \in \mathbb{R}^{1 \times (m+k)}$, are estimated by BMOM, that minimize the loss functions are given by:

$$\widehat{\delta}_1(K_1, K_2) = \begin{bmatrix} Y_1'Y_1 - K_1\widehat{V}_1'\widehat{V}_1 & Y_1'Z_1 \\ Z_1'Y_1 & Z_1'Z_1 \end{bmatrix}^{-1} \begin{bmatrix} \left(Y_1 - K_2\widehat{V}_1\right)' y_1 \\ Z_1'y_1 \end{bmatrix} \tag{3}$$

with $Z_1 = (X\Pi_1 \; X_1)'$, where Π_1 is the coefficient matrix of the reduced form equation for Y_1, being $Y_1 = X_1\Pi_1 + V_1$. Those coefficients can be calculated by LS obtaining $\widehat{\Pi}_1 = (X'X)^{-1}X'Y_1$ and $\widehat{V}_1 = Y_1 - X\widehat{\Pi}_1$.

Table 1 shows the K_1 and K_2 parameters for two loss functions using the BMOM approach, being n and k, the sample size and the number of exogenous variables:

Table 1. K1 and K2 parameters proposed to minimize loss function. Bayesian Method of Moments (BMOM).

	Loss Function	BMOM Approach
Goodness of fit	$L_g = (y_1 - Z_1\widehat{\delta}_1)'(y_1 - Z_1\widehat{\delta}_1)$	$K_1 = 1 - k/(n-k), K_2 = 1$
Precision of estimation	$L_p = (\delta_1 - \widehat{\delta}_1)'Z_1'Z_1(\delta_1 - \widehat{\delta}_1)$	$K_1 = K_2 = 1 - k/(n-k)$

2.2.3. Bayesian Approach in Two Stages ($Bayes_{2S}$)

In Bayesian inference, a pragmatic solution for choosing the prior distribution is to select a member of the distribution family so that it remains in the same family as the prior distribution. If the prior is conjugate, the posterior distribution after the first observation belongs, by definition, to the same type and is used as the new prior distribution in the next observation. Incorporating this second observation, the new posterior distribution, also belongs to the class of conjugation. This sequential process only updates the value of parameters of the distribution [6]. In this work, the $Bayes_{2S}$ method uses a Normal-Inverse Gamma prior to obtaining the exact analytic expressions for the posterior distribution of the structural B and Γ coefficients of the SEM. Basically, it has been applied in two stages as the 2SLS method, used in the two-steps Bayesian Least Squares instead of Ordinary Least Squares.

2.2.4. Markov Chain Monte Carlo (MCMC)

In the Bayesian approach, it is essential to select prior distribution, however, there are situations where this selection is somewhat difficult due to the absence of previous model information. The MCMC methodology provides a wide scope for statistical modeling, and is widely used to summarize complicated posterior distributions in econometrics models. In particular, Bayesian methods need to integrate the posterior distribution of model parameters for that reason and MCMC draws samples from these posterior distributions. There are many ways for constructing these chains like Gibbs sampler or special cases of the general framework of Metropolis and Hastings [6]. In particular, in this work, Gibbs Sampling has been used for the simulation of posterior distribution, calculating the average to estimate the model parameters.

3. The Proposed Estimation Method: Optimized BMOM Method ($Bmom_{OPT}$)

A variation of BMOM (Section 2.2.2) is proposed obtaining K_1 and K_2 by the optimization of different parameter criteria, instead of setting them by the proposed values in (1). Concretely, the optimization of K_1 and K_2 parameters that minimize the AIC it is proposed, which is a quality measure of statistical models [20] based on sample fit to estimate the likelihood of a model. Thus, given a collection of model-based estimates for data, AIC obtains the quality of each model with respect to other models, providing a way for a model selection.

The expression for this measure is:

$$AIC = n \ln|\hat{\Sigma}_e| + 2 \sum_{i=1}^{m} (m_i + k_i - 1) + m(m + 1) \tag{4}$$

where n is the sample size, m is the number of equations, m_i and k_i are the number of endogenous and exogenous variables in i-equation and $\hat{\Sigma}_e$, the variance-covariance matrix of the errors $e_j = Y_j - \hat{Y}_j$, $j=1, \ldots$, m.

For the experimental study, a large number of SEMs models have been randomly generated through a model generator tool, called real models. Thus, new errors, denoted by e_j^r, have been obtained as the difference between the values generated by the real models, Y_j^r, and the estimated values by each method described in the previous section, $\hat{Y}_j$. These errors are substituted in (4), and, a new measure, denoted AIC_{real} is calculated. This value is a measure of the error and we propose it as a reference parameter, which could be an indicator of the goodness of the estimated model. It can only be calculated if the real coefficients are known, that is, in an experimental study. To reach the minimum AIC_{real} value, an algorithm based on the Quasi-Newton method is used, so that, in each iteration the algorithm looks for an approximation of the inverse of the Hessian matrix. So, the algorithm guarantees approximation at every step of the process.

4. Entropy as an Information Parameter Criteria

Another motivation in this work is to obtain an alternative measure to AIC which is closely related to the BIC. In this section, it presents a parameter criteria of a SEM quality based on entropy, $H(e)$. Although entropy was introduced by Clausius [36] in thermodynamics to measure the amount of energy in a system that cannot produce work, this concept appears in many contexts (statistical mechanics, information theory, etc.) as disorder, uncertainty, randomness, complexity, etc. Claude Shannon [37] in 1948 built his theory of information and communication, being generalized by some authors as Tsallis [38]. The expression for Shannon entropy [39], $H_S(x)$, is:

$$H_S(x) = - \sum_{i=1}^{n} p(x) \log p(x). \tag{5}$$

On the other hand, a new entropy measure, developed by Amigó [32] as a variation of generalized entropy, has been developed to allow having small $p(x)$ values in the distribution. The expression for this entropy, $H_A(p)$, is:

$$H_A(p) = \prod_{i=1}^{n} \left(2 - (p_i)^{p_i}\right). \tag{6}$$

The model SEM has m equations, the same number as endogenous variables. Then, applying this entropy to each equation in the estimated model obtains:

$$H_j(e) = \prod_{i=1}^{n} \left(2 - (p_{ij})^{p_{ij}}\right) \quad j = 1, 2, \ldots m \tag{7}$$

where n is the sample size and m the number of endogenous variables, and the p_{ij} values for each endogenous variable have been obtained as follows:

$$p_{ij} = \frac{e_{ij}}{\sum_{i=i}^{n} e_{ij}} \quad j = 1, 2, \ldots m, \quad \text{where} \quad e_{ij} = Y_{ij} - \hat{Y}_{ij} \tag{8}$$

where p_{ij} is the error mass in each endogenous variable and has been calculated from the error matrix as the difference between endogenous variables and its estimation through

each method. Finally, for each method and estimated model, the average of logarithms of (7) is calculated:

$$H(e) = \frac{\sum_{j=1}^{m} \ln H_j(e)}{m}. \tag{9}$$

The minimum value of $H(e)$ is reached when e_{ij} are homogeneous, so, the $\widehat{Y}_{ij}$ values are more well-balanced.

5. Experimental Design and Results

5.1. Experimental Design

In the experimental study a large number of SEMs are generated (that is, generating the matrices X, Y, B y Γ of each model) and then are estimated through the methods presented in Sections 2 and 3 and finally, the models are compared to their estimations.

The SEMs have been generated as follows: The values of matrices B and Γ are generated randomly following a Uniform distribution in [0, 10], matrix X a multivariate normal distribution, and finally, matrix Y as ΠX plus a Normal distribution with mean 0 and sigma 0.1. Two functions of the R package have been used: In MCMC, Markov chains have been simulated through the *MCMCregress* function of MCMCpack package without prior information, and, in $Bmom_{OPT}$ the function *optim* has been used to obtain the optimal values K_1 and K_2. In $Bayes_{2S}$, a Normal-Inverse Gamma prior has been used, with an average of 5 and precision of 0.2 as the initial parameters.

For the comparison study, measures based on the generated model parameters and measures based on the estimated parameters have been calculated, in order to have the criteria to find the best estimation method. So, it is possible to have criteria to choose the best estimation method. In the first type, the Euclidean distance between δ and $\hat{\delta}$, denoted by $D_{\delta,\hat{\delta}}$, where $\delta = [B \ \Gamma]$ is the coefficient matrix and $\hat{\delta}$ its estimation, and AIC_{real}. In the second type, AIC and the entropy $H(e)$ have been calculated. And finally, the execution time has been calculated. Table 2 shows the average and the standard deviation of 50 simulations for each measure, model, and method when the number of variables and sample size of SEMs are varying.

Table 2. Average and standard deviation of 50 simulations of $D_{\delta,\hat{\delta}}$, AIC_{real}, AIC, $H(e)$, and execution time in seconds. Markov Chain Monte Carlo (MCMC). Sigma 0.1.

	m	k	n	2SLS	BMOM		$Bmom_{OPT}$	$Bayes_{2S}$ [a]	MCMC [b]
					Goodness of Fit	Precision of Estimation			
$D_{\delta,\hat{\delta}}$	10	20	100	27.670 7.778	40.914 7.538	40.966 7.546	**20.647** 8.826	33.673 12.084	71.410 10.491
	10	40	100	40.927 9.104	58.635 8.039	58.932 8.100	**26.769** **8.209**	56.340 13.721	91.076 4.918
	20	60	100	115.852 8.537	141.029 6.294	140.906 6.250	**94.999** 12.498	146.640 5.904	163.771 6.092
	10	20	400	16.563 11.257	27.508 6.955	27.534 6.974	**10.619** 5.868	22.233 17.026	70.449 7.899
	10	40	400	15.199 4.366	30.538 6.467	30.494 6.446	**7.923** 2.576	26.009 22.856	90.357 5.568
	10	40	1000	7.394 2.830	17.218 5.450	17.229 5.458	**5.130** 1.944	9.233 7.540	95.210 10.424

Table 2. *Cont.*

	m	k	n	2SLS	BMOM		$Bmom_{OPT}$	$Bayes_{2S}$ [a]	MCMC [b]
					Goodness of Fit	Precision of Estimation			
AIC_{real}	10	20	100	1361.764 781.683	1895.361 801.349	1896.600 800.971	**1156.465** 792.277	1993.402 831.727	4598.074 395.783
	10	40	100	1915.270 651.931	2345.388 636.213	2351.286 635.582	**1718.000** 626.334	2531.052 715.689	4547.361 286.781
	20	60	100	5941.844 779.559	6458.904 762.773	6456.303 763.063	**5692.839** 793.354	6732.681 777.638	10,172.018 206.247
	10	20	400	4438.187 3271.180	7319.586 3449.337	7323.685 3450.560	**3409.573** 3232.656	7517.898 5184.224	22875.120 1641.293
	10	40	400	4645.919 2358.815	6989.900 2744.681	6983.853 2744.491	**3877.930** 2299.173	7382.701 4943.642	22,057.334 917.641
	10	40	1000	6824.562 8001.673	11,738.582 8294.696	11,742.499 8294.949	**5913.439** 8123.421	9046.596 8850.159	63,247.410 2586.679
AIC	10	20	100	2168.030 854.334	1784.122 808.851	**1783.602** 808.150	2419.415 893.327	2391.887 785.620	4413.688 421.952
	10	40	100	2009.636 639.153	1850.753 626.141	**1850.716** 626.059	2372.355 692.930	2254.874 728.737	4348.852 256.390
	20	60	100	3866.102 1051.585	3647.221 1056.690	**3645.439** 1056.018	4543.780 1054.504	4119.416 1128.524	9856.712 241.454
	10	20	400	15,027.648 3347.331	13,448.372 3161.253	**13,446.459** 3161.039	15,524.231 3459.419	15626.142 3279.198	22,160.759 1516.667
	10	40	400	12,849.587 2699.761	**11,990.079** 2509.827	11,991.659 2509.734	13,479.303 2744.344	13,606.665 2961.619	21,421.728 858.468
	10	40	1000	37,879.770 9621.598	36,438.251 9459.390	**36,437.331** 9459.167	38,479.010 9605.284	38,035.942 9483.024	61,720.600 2138.345
$H(e)$	10	20	100	4.074 0.013	4.081 0.012	4.081 0.012	**4.074** 0.013	4.084 0.018	4.096 0.016
	10	40	100	4.076 0.012	4.080 0.010	4.080 0.010	**4.075** 0.014	4.084 0.011	4.087 0.014
	20	60	100	**4.086** 0.008	4.087 0.009	4.087 0.009	**4.086** 0.009	**4.086** 0.009	4.088 0.009
	10	20	400	**5.579** 0.005	**5.579** 0.005	5.579 0.005	**5.579** 0.005	5.587 0.021	5.609 0.008
	10	40	400	**5.590** 0.005	**5.590** 0.005	5.590 0.005	**5.590** 0.005	5.593 0.014	5.602 0.007
	10	40	1000	6.530 0.003	6.531 0.003	6.531 0.003	**6.530** 0.004	6.532 0.013	6.568 0.004
Time (s)	10	20	100	0.073 0.145	0.732 0.246	0.732 0.246	258.227 98.562	**0.056** 0.017	294.453 17.949
	10	40	100	0.143 0.047	1.022 0.395	1.022 0.395	274.088 345.929	**0.140** 0.050	499.554 68.116
	20	60	100	**0.314** 0.047	2.495 0.395	2.495 0.395	748.504 345.929	0.407 0.145	1435.791 0.145
	10	20	400	0.125 0.031	4.265 0.864	4.265 0.864	2586.186 1005.068	**0.109** 0.030	328.571 18.174
	10	40	400	0.235 0.037	4.533 0.765	4.533 0.765	2281.255 804.186	**0.214** 0.032	507.791 38.694
	10	40	1000	0.426 0.066	21.385 1.595	21.385 1.595	14,534.080 21,689.539	**0.376** 0.107	524.904 26.564

[a] All initial parameters have average 5.0 and precision 0.2. [b] Without prior information. Chain size 10,000.

5.2. Experimental Results

Regarding $D_{\delta,\hat{\delta}}$, the results shows that the $Bmom_{OPT}$ method is better in all estimated models. Nevertheless, the high computational cost used by this method could be an

issue in large problems. In such cases it would be more efficient to use $Bayes2S$, since this method obtains good estimation in less time. AIC_{real} shows similar results than $D_{\delta,\hat{\delta}}$. Regarding estimated measures, AIC does not offer satisfactory results because its values are not in agreement with the Euclidean distance or with AIC_{real}, which is proposed as a reformulation that works well. The entropy $H(e)$ has yielded satisfactory results and it could be considered as a new comparative measurement.

In all methods, when the complexity of model increases, $D_{\delta,\hat{\delta}}$ increases, being the opposite when the sample size increases. Both BMOMs methods (Goodness of fit and Precision of estimation) show similar results, having a small difference between them. MCMC provides estimations with the biggest average of $D_{\delta,\hat{\delta}}$, in which no trend is appreciated when the number of variables and sample size vary. The minimum average of AIC_{real} is reached by $Bmom_{OPT}$, with MCMC being the worst method. Regarding the entropy, the minimum average value has been obtained by $Bmom_{OPT}$, except in the smallest case, where $Bayes_{2S}$ obtains the minimum value with a small difference. $Bayes_{2S}$ and 2SLS require less execution time in all cases, with $Bmom_{OPT}$ performing the worst.

6. Conclusions and Future Work

In this paper, the estimation of simultaneous equations models was studied through the comparison of models, carried out through an experimental study using randomly generated models. A new estimation method was proposed, $Bmom_{OPT}$, based on the optimization of some parameter of the Bayesian Method of Moments and minimizing the Akaike Information Criteria. The computational study showed that the proposed method was the best one regarding the minimum $D_{\delta,\hat{\delta}}$ and entropy. The study also showed that the AIC parameter presented deficiencies for selecting the estimation method with a minimum $D_{\delta,\hat{\delta}}$ value and minimum AIC_{real}.

The AIC is one of the most used parameter criteria to compare different methods of estimation. Nevertheless, in this study, the results showed that using entropy instead of AIC in the evaluation of the methods provides values according with the quality of the estimation (similarity with the real value).

In future, the study of information criteria parameters and their application in SEM problems, the study of use other criteria for optimization, as well as studying how to reduce the execution costs can be considered.

References

1. Gujarati, D.; Porter, D. *Econometría*, 5th ed.; México, D.F., Ed.; McGraw-Hill Interamericana Editores SA: Ciudad de México, México, 2004.

2. Zellner, A. The finite sample properties of simultaneous equations' estimates and estimators Bayesian and non-Bayesian approaches. *J. Econom.* **1998**, *83*, 185–212. [CrossRef]
3. Chao, J.C.; Phillips, P.C. Jeffreys prior analysis of the Simultaneous Equations Model in the case with n+1 endogenous variables. *J. Econom.* **2002**, *111*, 251–283. [CrossRef]
4. Geweke, J. Bayesian Reduced Rank Regression in Econometrics. *J. Econ. Trics* **1996**, *75*, 121–146. [CrossRef]
5. Kleibergen, F.; Dijk, H.V. Bayesian simultaneous equation analysis using reduced rank structures. *Econom. Theory* **1998**, *14*, 701–743. [CrossRef]
6. Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.; Vehtari, A.; Rubin, D.B. *Bayesian Data Analysis*; Chapman and Hall: New York, NY, USA, 2015.
7. Siburian, M.E. Fiscal decentralization and regional income inequality: Evidence from Indonesia. *Appl. Econ. Lett.* **2019**, 1–4. [CrossRef]
8. Steele, F.; French, R.; Bartley, M. Adjusting for Selection Bias in Longitudinal Analysis Using Simultaneous Equations Modeling. The Relationship Between Employment Transitions and Mental Health. *Epidemiology* **2013**, *24*, 703–711. [CrossRef]
9. Zhao, H.; Lin, B. Impact of foreign trade on energy efficiency in China's textile industry. *J. Clean. Prod.* **2019**, *245*, 118878. [CrossRef]
10. Adewuyi, A.O.; Awodumi, O.B. Biomass energy consumption, economic growth and carbon emi-ssions: Fresh evidence from West Africa using a simultaneous equation model. *Energy* **2017**, *119*, 453–471. [CrossRef]
11. King, T. Using simultaneous equation modelling for defining complex phenotypes. *BMC Genet.* **2003**, *4*, S10. [CrossRef] [PubMed]
12. Pitfield, D.; Caves, R.; Quddus, M. Airline strategies for aircraft size and airline frequency with changing demand and competition: A simultaneous-equations approach for traffic on the north Atlantic. *J. Air Transp. Manag.* **2009**, *16*, 151–158. [CrossRef]
13. Olmeda, N.G.; Martinez, I.B. Application of simultaneous equation models to temporary disability prescriptions in primary healthcare centres. *Int. J. Comput. Math.* **2014**, *91*, 252–260. [CrossRef]
14. Strathe, A.; Jørgensen, H.; Kebreab, E.; Danfær, A. Bayesian simultaneous equation models for the analysis of energy intake and partitioning in growing pigs. *J. Agric. Sci.* **2012**, *150*, 764–774. [CrossRef]
15. Zhou, W.; Duan, W. An empirical study of how third-party websites influence the feedback mechanism between online Word-of-Mouth and retail sales. *Decis. Support Syst.* **2015**, *76*, 14–23. [CrossRef]
16. Yang, S.; Narayan, V.; Assael, H. Estimating the Interdependence of Television Program Viewership Between Spouses: A Bayesian Simultaneous Equation Model. *Mark. Sci.* **2006**, *25*, 336–349. [CrossRef]
17. Park, H.; Manchanda, P. When Harry Bet with Sally: An Empirical Analysis of Multiple Peer Effects in Casino Gambling Behavior. *Mark. Sci.* **2015**, *2*, 179–194. [CrossRef]
18. Chen, C.; Chang, K.; Lin, L.; Lee, J. Health risk perception and betel chewing behavior. The evidence from Taiwan. *Addict. Behav.* **2013**, *38*, 2714–2717. [CrossRef]
19. Cohen, D.P.; Benadjaoud, M.A.; Lestaevel, P.; Lebsir, D.; Benderitter, M.; Souidi, M. Effects of repetitive Iodine Thyroid Blocking on the Development of the Foetal Brain and Thyroid in rats: A Systems Biology approach. *bioRxiv* **2019**. . [CrossRef]
20. Akaike, H. *Information Theory and an Extension of the Maximum Likelihood Principle*; Springer: New York, NY, USA, 1998.
21. Keerativibool, W. New Criteria for Selection in Simultaneous Equations Model. *Thail. Stat.* **2012**, *10*, 163–181.
22. Hurvich, C.; Tsai, C. Regression and time series model selection in small samples. *Biometrika* **1989**, *76*, 297–397. [CrossRef]
23. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
24. Findley, D.F. Counterexamples to parsimony and BIC. *Ann. Inst. Stat. Math.* **1991**, *43*, 505–514. [CrossRef]
25. Hannan, E.J.; Quinn, B.G. The Determination of the Order of an Autoregression. *J. R. Stat. Soc. Ser. B (Methodol.)* **1979**, *41*, 190–195. [CrossRef]
26. Keerativibool, W.; Jitthavech, J. Model Selection Criterion Based on Kullback-Leibler's Symmetric Divergence for Simultaneous Equations Model. *Chiang Mai J. Sci.* **2015**, *42*, 761–773.
27. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 1991.
28. Zhou, R.; Cai, R.; Tong, G. Applications of Entropy in Finance: A Review. *Entropy* **2013**, *15*, 4909–4931. [CrossRef]
29. Cui, H.; Sivakumar, B.; Singh, V. Entropy Applications in Environmental and Water Engineering. *Entropy* **2019**, *20*, 598. [CrossRef] [PubMed]
30. Purvis, B.; Mao, Y.; Robinson, D. Entropy and its Application to Urban Systems. *Entropy* **2019**, *21*, 56. [CrossRef]
31. Oruç, Ö.E.; Kuruoglu, E.; Gündüz, A. Entropy Applications for Customer Satisfaction Survey in Information Theory. *Front. Sci.* **2011**, *1*, 1–4. [CrossRef]
32. Amigó, J.M.; Balogh, S.G.; Hernández, S. A Brief Review of Generalized Entropies. *Entropy* **2018**, *20*, 813. [CrossRef]
33. Qayyum, Z.; Hasan, S.S. K-Class estimators-a Review. *Int. J. Math. Trends Technol.* **2017**, *50*, 104–107. [CrossRef]
34. Theil, H. *Repeated Least Squares Applied to Complete Equation Systems*; Central Planning Bureau: The Hague, The Netherlands, 1953.
35. Basmann, R.L. A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation. *Econometrica* **1957**, *25*, 77–83. [CrossRef]
36. Clausius, R. *The Mechanical Theory of Heat*; Macmillan: New York, NY, USA, 1879.
37. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
38. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [CrossRef]
39. Lombardi, O.; Holik, F.; Vanni, L. What is Shannon information? *Synthese* **2016**, *193*, 1983–2012. [CrossRef]