Songklanakarin J. Sci. Technol.
32 (5), 519-525, Sep. - Oct. 2010

Songklanakarin Journal
of Science and Technology
http://www.sjst.psu.ac.th

Original Article

A goodness-of-fit measure for a system-of-equations model

Jirawan Jitthavech*

School of Applied Statistics
National Institute of Development Administration, Bang Kapi, Bangkok, 10240 Thailand.

Received 5 October 2009; Accepted 8 August 2010

Abstract

Two statistics, SR^2 and adjusted SR^2, are developed to measure the goodness–of–fit for a system-of-equations model based on a new definition of the norm of a square matrix with positive diagonal elements, and by this overcome the short-comings of the McElroy R^2 and the system weighted R^2. The proposed measures are tested with the simulation data and the simulation results show that the proposed measures outperform the McElroy R^2 and give similar results as the AIC and BIC criteria. The proposed measures are fairly constant when the irrelevant variables are eliminated but the sharp change in the measure is obviously visible when one relevant variable is eliminated from the model. The movement of the McElroy R^2 statistic is very slow comparing with the other four statistics and the sharp change in the measure is not visible when one of relevant variables is eliminated.

Keywords: system R-square, adjusted system R-square, goodness–of–fit, a system-of-equations model, contemporaneous correlation.

1. Introduction

Variable selection criteria and procedures in the single–equation model are widely discussed in the literature (e.g. Draper and Smith, 1998; Stock and Watson, 2003; Montgomery *et al.*, 2006). The concepts of R^2 and adjusted R^2 are widely used as the measures of goodness–of–fit in the single–equation model. McElroy (1977) extends the concept of R^2 to measure the goodness–of–fit in a system-of-equations model:

$$R^2 = 1 - \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\Omega}^{-1}\hat{\boldsymbol{\varepsilon}}}{\mathbf{y}'\hat{\Omega}^{-1}\mathbf{y}} \quad (1)$$

where $\hat{\boldsymbol{\varepsilon}}$ is the $nm \times 1$ error vector, $\mathbf{y}$ is the $nm \times 1$ dependent variable vector, n is the number of observations, m is the number of equations in the model, $\hat{\Omega}^{-1} = S^{-1} \otimes I$, and is the estimated $m \times m$ cross–equation covariance matrix of errors. The system weighted R^2 for the 3SLS, IT3SLS, SUR, ITSUR, and FIML methods in the SAS PROC SYSLIN (SAS Inc., 2003) is defined as

$$R^2 = \frac{\mathbf{y}'WR\left(X'X\right)^{-1}R'W\mathbf{y}}{\mathbf{y}'W\mathbf{y}} \quad (2)$$

where $\mathbf{y}$ is the dependent variable vector, R is the vector of regressor set, W is the weighted matrix defined as $W = S^{-1} \otimes Z\left(Z'Z\right)^{-1}Z'$, Z is the matrix of instrument set, and $X'X$ is the matrix of $R'WR$. The system weighted R^2 reduces to the McElroy R^2 if the matrix of the instrument set is the identity matrix.

McElroy R^2 in (1) or system weighted R^2 in (2) is not popular for variable selection in the system-of-equations model since both measures are based on the comparison reference depending on the selected variables. The denominator in McElroy R^2 or system weighted R^2 is not a constant for a set of data because it is the total sum of squares weighted by the inverse matrix of the estimated cross–equation covariance matrix of errors, depending on the variables in the model. Furthermore, if the mean of errors generated in

* Corresponding author.
Email address: jirawan@as.nida.ac.th

the parameter estimation from each equation is equal to zero and the uncorrected form of degrees of freedom is used, Lorchirachoonkul and Jitthavech (2008) have shown that the weighted error sum of squares $\hat{\boldsymbol{\varepsilon}}'S^{-1}\hat{\boldsymbol{\varepsilon}}$ is equal to the total error degrees of freedom, nm. If the corrected error degrees of freedom are $EDF_{ij} = n - p_i$, i = 1, 2,..., m; j = 1, 2,..., m then $\hat{\boldsymbol{\varepsilon}}'S^{-1}\hat{\boldsymbol{\varepsilon}}$ is equal to $nm - \sum_{i=1}^{m} p_i$, where p_i is the number of parameters in equation i in the model. If the degrees of freedom correction are $EDF_{ij} = n - \max(p_i, p_j)$, i = 1, 2, … m ; j = 1, 2, …, m, then $\hat{\boldsymbol{\varepsilon}}'S^{-1}\hat{\boldsymbol{\varepsilon}} = nm - \sum_{i=1}^{m} p_i + \frac{1}{|S|}\sum_{i=1}^{m}\sum_{j=1}^{i}(p_i - p_j)S_{ij}C_{ij}$, where S_{ij} and C_{ij} are the elements in i^{th} row and j^{th} column of matrix S and the co-factor ij of matrix S, respectively. Therefore, from the first normal condition, it may be concluded that the numerator in the second term in McElroy R^2 in (1) is equal to a fixed value depending on the number of parameters in the model. Consequently, the values of McElroy R^2 in the different models will slightly change when the numbers of regressors are equal since the dependent variable sum of squares is usually much greater than the error variance. A comparison between the models should be done carefully since the weighted total sum of squares varies from model to model. Moreover, since the measures in Eq. 1 and 2 are both based on the weighted error sum of squares and the weighted total sum of squares, their values can possibly be increased by simply adding more regressors. These overall fit measures also obscure the variation in fit across equations (Greene, 2003).

The information AIC and BIC statistics for the model selection in a multivariate model (Bedrick *et al.*, 1994; Gorobets, 2005) are given by

$$AIC = n\ln|\hat{\Sigma}_\varepsilon| + 2\sum_{i=1}^{m} p_i + m(m+1) \tag{3}$$

$$BIC = n\ln|\hat{\Sigma}_\varepsilon| + \ln(n)\left(\sum_{i=1}^{m} p_i + 0.5m(m+1)\right) \tag{4}$$

The AIC for model selection in a model described by a system of equations is developed for the generalized estimating equations (GEE) model by using the quasi–likelihood function (Pan, 2001) and for the state–space model under the normality assumption (Bengtsson and Cavanaugh, 2006). The GEE model extends the generalized linear model for the correlated observations (Hedeker and Gibbons, 2006) but does not take into account the contemporaneous correlation. The main criticism of the information-type criteria is that the minimum value of the criterion is unknown and we cannot determine how the selected model fits the unknown true model.

The paper is organized as follows: In the next section, a statistic different from the McElroy R^2 and the system weighted R^2 is proposed to measure the goodness–of–fit of a system-of-equations model in which the contemporaneous correlation exists. A simulation is conducted to test the proposed measure and the results are compared with the results using the McElroy R^2, AIC, and BIC criteria.

2. Variable selection in a system-of-equations model

Based on our theoretical and/or empirical knowledge we may define an initial specification of the *good* model. But it is unknown whether the specification is an overspecification or underspecification of the *true* model. However, one usually has a tendency to include too many variables in the initial specification causing the overspecification problem in order that the estimate of the covariance matrix with the corrected degrees of freedom is an unbiased estimate of the true covariance matrix (Fujikoshi and Satoh, 1997). Selecting the appropriate variables from an overspecification set of qualified regressors is, therefore, a critical step in model building.

Consider a system-of-equations model, which can be written as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \vdots \\ \mathbf{y}_m \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 & \cdots & 0 \\ 0 & X_2 & 0 & \cdots & 0 \\ 0 & 0 & X_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & X_m \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \\ \vdots \\ \boldsymbol{\beta}_m \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \boldsymbol{\varepsilon}_3 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{bmatrix}, \tag{5}$$

where y_i is the $n \times 1$ dependent variable vector in equation i, X_i is the $n \times p_i$ matrix of independent variables including a constant unit vector 1 in equation i, $\boldsymbol{\beta}_i$ is the $p_i \times 1$ parameter vector in equation i, $\boldsymbol{\varepsilon}_i$ is the $n \times 1$ random error $IN(0, \sigma_i^2)$ vector in equation i, n is the number of observations in each equation, and p_i is the number of parameters including a constant term in equation i. The error covariance matrix given by $\Sigma_\varepsilon(p)$ is not a diagonal matrix if the equations are correlated. The total number of parameters in the system-of-equations model (5) is equal to

$$p = \sum_{i=1}^{m} p_i.$$

Given X_i, i = 1, 2, …, m, the expectation of the random disturbance is equal to zero.

$$E(\boldsymbol{\varepsilon}(p)|X_i, i = 1, 2, \ldots, m) = \mathbf{0}.$$

The dependent variables and the random disturbances in Eq. 5 are rearranged in the matrix form as

$$\mathbf{y} = (\mathbf{y}_1\ \mathbf{y}_2\ \ldots\ \mathbf{y}_m),$$

$$\boldsymbol{\varepsilon}(p) = [\boldsymbol{\varepsilon}_1(p_1)\ \boldsymbol{\varepsilon}_2(p_2)\ \cdots \boldsymbol{\varepsilon}_m(p_m)].$$

The error sum of squares generated in the parameter estimation can be written as

$$\begin{aligned}\hat{\boldsymbol{\varepsilon}}(p)'\boldsymbol{\varepsilon}(p) &= [\mathbf{y}-\hat{\mathbf{y}}(p)]'[\mathbf{y}-\hat{\mathbf{y}}(p)] \\ &= [\mathbf{y}-E(\mathbf{y})]'[\mathbf{y}-E(\mathbf{y})]-[\hat{\mathbf{y}}(p)-E(\mathbf{y})]'[\mathbf{y}-E(\mathbf{y})] \\ &\quad -[(\mathbf{y}-E(\mathbf{y})]'[\hat{\mathbf{y}}(p)-E(\mathbf{y})]+[\hat{\mathbf{y}}(p)-E(\mathbf{y})]'[\hat{\mathbf{y}}(p)-E(\mathbf{y})].\end{aligned} \tag{6}$$

Consider the second term in the RHS of Eq. 6 which can be rewritten as

$$\begin{aligned}&[\hat{\mathbf{y}}(p)-E(\mathbf{y})]'[\mathbf{y}-E(\mathbf{y})] \\ &= [\hat{\mathbf{y}}(p)-E(\mathbf{y})]'\hat{\boldsymbol{\varepsilon}}(p)+[\hat{\mathbf{y}}(p)-E(\mathbf{y})]'[\hat{\mathbf{y}}(p)-E(\mathbf{y})] \\ &= \hat{\mathbf{y}}'(p)\hat{\boldsymbol{\varepsilon}}(p)-E(\mathbf{y})'\hat{\boldsymbol{\varepsilon}}(p)+[\hat{\mathbf{y}}(p)-E(\mathbf{y})]'[\hat{\mathbf{y}}(p)-E(\mathbf{y})].\end{aligned}$$

The diagonal elements of the first matrix in the RHS are equal to zero since the least squares estimate $\hat{\mathbf{y}}(p)$ and $\hat{\boldsymbol{\varepsilon}}(p)$ are known to be orthogonal (Johnson and Wichern, 1998), and the second matrix is equal to zero since the sum of residuals is equal to zero. Therefore, it can be concluded that the diagonal elements in the matrices in Eq. 6 are all positive. Therefore, the norm of a square matrix A with positive diagonal elements, $\|A\|$, can be defined as trace(A), which satisfies the four basic properties of the matrix norm (Leon, 2002):

a) $\|A\| = \text{trace}(A) > 0$ and $\|A\| = 0$ if and only if $A = 0$.

b) $\|cA\| = |c|\,\|A\| = |c|\,\text{trace}(A)$ where c is a scalar.

c) $\|A + B\| = \text{trace}(A + B) = \text{trace}(A) + \text{trace}(B) = \|A\| + \|B\|$, where B is also a matrix of same dimension as A and its diagonal elements are positive.

d) $\|A'\| = \|A\|$

The matrix norm as defined above is different from the trace norm, which is equal to the sum of the absolute value of the singular values of the matrix, which is the upper bound of the Frobenius norm (Srebro, 2004). The matrix norm of A in the above definition equal to trace(A) can be easily shown to be equal to the sum of eigenvalues of A.

Based on such definition of the matrix norm the norms of matrices in Eq. 6 can be written as

$$\|\hat{\boldsymbol{\varepsilon}}'(p)\hat{\boldsymbol{\varepsilon}}(p)\| = \sum_{i=1}^{m} SSE_i(p_i)$$

$$\|[\mathbf{y}-E(\mathbf{y})]'[\mathbf{y}-E(\mathbf{y})]\| = \sum_{i=1}^{m} SST_i$$

$$\|[\hat{\mathbf{y}}(p)-E(\mathbf{y})]'[\mathbf{y}-E(\mathbf{y})]\| = \|[\hat{\mathbf{y}}(p)-E(\mathbf{y})]'[\hat{\mathbf{y}}(p)-E(\mathbf{y})]\|$$

$$\|[\hat{\mathbf{y}}(p)-E(\mathbf{y})]'[\hat{\mathbf{y}}(p)-E(\mathbf{y})]\| = \sum_{i=1}^{m} SSR_i.$$

Replacing the matrix norms in Eq. 6 and the rearranging terms yield the coefficient of multiple determination in a system-of-equations model

$$SR^2 = 1 - \frac{\sum_{i=1}^{m} SSE_i(p_i)}{\sum_{i=1}^{m} SST_i} \tag{7}$$

$$= \frac{\sum_{i=1}^{m} SSR_i}{\sum_{i=1}^{m} SST_i}. \tag{8}$$

If $\hat{\mathbf{y}}(p)$ approaches $\mathbf{y}$ the denominator in Eq. 8 the regression sum of squares in the system-of-equations model will approach the total sum of squares. Consequently, SR^2 approaches 1. It can be seen that the properties of R^2 statistic in a single-equation model can be extended to a system-of-equations by the proposed SR^2. It can be seen from Eq.7 that maximizing SR^2 is equivalent to minimizing the sum of error sum of squares. Following the same arguments against R^2 as in the single-equation model we define the adjusted SR^2 in the system-of-equations model to take the complexity of the model into account as

$$\text{adjusted } SR^2 = 1 - \frac{\sum_{i=1}^{m} SSE_i(p_i)/(nm - p)}{\sum_{i=1}^{m} SST_i/[m(n-1)]}, \tag{9}$$

where p is the total number of parameters in the system-of-equations model, m is the number of equations, and n is the number of observations in each equation. Again, maximizing the adjusted SR^2 is equivalent to minimizing the mean of error sum of squares.

3. Simulation

A set of simulation data is generated to compare the proposed SR^2 and adjusted SR^2 statistics as measures of goodness-of-fit with the well-known information AIC and BIC criteria and McElroy R^2 in the system-of-equations model. The system of equations used to simulate the data consists of two equations and four exogenous variables with six parameters:

$$\begin{aligned} y_1 &= 5 + a_{11}x_{11} + a_{12}x_{12} + a_{13}x_{13} + \varepsilon_1 \\ y_2 &= 4 + a_{21}x_{21} + a_{22}x_{22} + a_{23}x_{23} + \varepsilon_2 \end{aligned} \tag{10}$$

where x_{11} U(5, 30), $x_{12} = x_{21}$ U(3, 20), $x_{13} = x_{22} \sim$ U(15, 35), $x_{23} \sim$ U(10, 25), $a_{11} = 2$, $a_{12} = 5$, $a_{13} = 6$, $a_{21} = 3$, $a_{22} = 4$, $a_{23} = 7$, ε_1 and ε_2 are bivariate normal with $\sigma_1 = 0.15$ and $\sigma_2 = 0.25$, respectively, and the contemporaneous correlation = 0.4.

In the simulation study the system of equations is initially specified by adding another two irrelevant variables in each equation as

$$\begin{aligned} y_1 &= 5 + a_{11}x_{11} + a_{12}x_{12} + a_{13}x_{13} + a_{14}x_{14} + a_{15}x_{15} + \varepsilon_1 \\ y_2 &= 4 + a_{21}x_{21} + a_{22}x_{22} + a_{23}x_{23} + a_{24}x_{24} + a_{25}x_{25} + \varepsilon_2 \end{aligned} \tag{11}$$

In this simulation, the irrelevant variables $x_{14} = x_{24}$ and $x_{15} = x_{25}$ are generated from U(15,115) and U(3,63), respectively, and a_{14}, a_{15}, a_{24} and a_{25} are expected to be equal to 0. The data set in the simulation consists of 5,000 records; each consists of twelve variables, y_1, y_2, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{21}, x_{22}, x_{23}, x_{24}, and x_{25}. The set of simulation data is randomly sampled to create 20 datasets of 50 observations to test the effectiveness of the proposed SR^2 and adjusted SR^2 statistics for variable selection.

The ten independent variables and the two random disturbances ε_1 and ε_2 are tested to confirm that their distributions in the generated population are consistent with the assumed distributions. From the Pearson correlation coefficient matrix in Table 1 it can be concluded that there is a significant contemporaneous correlation of $\rho = 0.4$ in the system of equations and the pairwise correlations between the x_i's are insignificant. Since we assume that no information on the covariance matrix of the random disturbances is available the parameters are estimated by the SAS Proc Model using the IT3SLS method to correct the contemporaneous correlation.

The variable selection strategy in the simulation follows the concept of the backward elimination procedure to reduce partially the computational demand in the variable selection. The backward elimination algorithm is often less adversely affected by the correlative structure of the regressors than the forward selection (Mantel, 1970). The simulation algorithm can be summarized as follows:

1. Estimate the parameters in the system-of-equations model with all twelve parameters in Eq. 11 referred as a full model. From the simulation results of 20 runs, the averages of the goodness-of-fit measures are used as the estimates of all five statistics as shown in Table 2.

2. Repeat Step 1 ten times with the system-of-equations model consisting of 9 independent variables, by dropping one variable at a time. All statistics except McElroy R^2 suggest dropping x_{25} permanently from the model, while McElroy R^2 suggests dropping x_{14}. The optimal values of statistics at the end of iteration 1 are summarized in Table 3. It can be seen that the values of McElroy R^2 of the reduced models almost remain unchanged within the same simulation run when dropping one variable at a time. The maximum of the average values of McElroy R^2 of the reduced model decreases insignificantly compared with the full model. The optimal values of the averages of the other four statistics of the reduced models are not significantly different from the

Table 1. Pearson correlation coefficients.

	y_{11}	y_2	x_{11}	x_{12}	x_{13}	x_{23}	x_{14}	x_{15}	ε_1	ε_2
y_1	1.0000	0.6081	0.3106	0.5400	0.7610	-0.0163	0.0083	0.0097	0.0084	0.0148
		<.0001	<.0001	<.0001	<.0001	0.2488	0.5595	0.4943	0.5540	0.2968
y_2	0.6081	1.0000	-0.0242	0.3572	0.5557	0.7307	-0.0086	0.0047	0.0132	0.0093
	<.0001		0.0869	<.0001	<.0001	<.0001	0.5416	0.7416	0.3507	0.5128
x_{11}	0.3106	-0.0242	1.0000	-0.0180	0.0029	-0.0225	0.0110	0.0119	-0.0053	-0.0187
	<.0001	0.0869		0.2042	0.8386	0.1125	0.4353	0.3986	0.7094	0.1869
x_{12}	0.5400	0.3572	-0.0180	1.0000	0.0002	-0.0090	-0.0053	0.0105	-0.0151	0.0153
	<.0001	<.0001	0.2042		0.9875	0.5233	0.7063	0.4590	0.2845	0.2789
x_{13}	0.7610	0.5557	0.0029	0.0002	1.0000	-0.0025	0.0089	-0.0009	0.0146	0.0154
	<.0001	<.0001	0.8386	0.9875		0.8594	0.5277	0.9468	0.3028	0.2774
x_{23}	-0.0163	0.7307	-0.0225	-0.0090	-0.0025	1.0000	-0.0158	0.0027	0.0157	-0.0146
	0.2488	<.0001	0.1125	0.5233	0.8594		0.2644	0.8469	0.2672	0.3011
x_{14}	0.0083	-0.0086	0.0110	-0.0053	0.0089	-0.0158	1.0000	0.0061	0.0099	0.0088
	0.5595	0.5416	0.4353	0.7063	0.5277	0.2644		0.6645	0.4828	0.5351
x_{15}	0.0097	0.0047	0.0119	0.0105	-0.0009	0.0027	0.0061	1.0000	-0.0089	-0.0213
	0.4943	0.7416	0.3986	0.4590	0.9468	0.8469	0.6645		0.5302	0.1330
ε_1	0.0084	0.0132	-0.0053	-0.0151	0.0146	0.0157	0.0099	-0.0089	1.0000	0.3934
	0.5540	0.3507	0.7094	0.2845	0.3028	0.2672	0.4828	0.5302		<.0001
ε_2	0.0148	0.0093	-0.0187	0.0153	0.0154	-0.0146	0.0088	-0.0213	0.3934	1.0000
	0.2968	0.5128	0.1869	0.2789	0.2774	0.3011	0.5351	0.1330	<.0001	

Table 2. Estimates of measures of goodness-of-fit of the full model.

Full model	McElroy R^2	AIC	BIC	SR^2	Adjusted SR^2
Mean	0.9994	403.4638	432.1441	0.9754	0.9726
Standard error	1.59E-05	3.0969	3.0969	0.0013	0.0014

Table 3. Estimates of measures of goodness-of-fit of the reduced models at the end of iteration 1.

Reduced model	Max McElroy R^2	Min AIC	Min BIC	Max SR^2	Max Adjusted SR^2
Mean	0.9994	402.5267	429.2951	0.9752	0.9727
Standard error	1.58E-05	3.0860	3.0860	0.0013	0.0014
Full model-Reduced model	5.36E-06	0.9371	2.8491	-0.0002	-0.0001
Pooled standard error	1.59E-05	3.0914	3.0914	0.0013	0.0014
t-statistic	0.3384	0.3031	0.9216	0.1591	0.0581

Table 4. Estimates of measures of goodness-of-fit of the reduced models at the end of iteration 2.

Reduced model	Max McElroy R^2	Min AIC	Min BIC	Max SR^2	Max Adjusted SR^2
Mean	0.9994	401.7743	426.6306	0.9749	0.9727
Standard error	1.71E-05	3.1153	3.1153	0.0013	0.0014
Full model-Reduced model	1.27E-05	1.6895	5.5135	0.0005	0.0001
Pooled standard error	1.65E-05	3.1061	3.1061	0.0013	0.0014
t-statistic	0.7688	0.5439	1.7751	0.3592	0.0725

Table 5. Estimates of measures of goodness-of-fit of the reduced models at the end of iteration 3.

Reduced model	Max McElroy R^2	Min AIC	Min BIC	Max SR^2	Max Adjusted SR^2
Mean	0.9994	402.0897	425.0339	0.9743	0.9723
Standard error	1.77E-05	2.9451	2.9451	0.0013	0.0014
Full model-Reduced model	2.60E-05	1.3741	7.1102	0.0011	-0.0003
Pooled standard error	1.68E-05	3.0220	3.0220	0.0013	0.0014
t-statistic	1.5467	0.4547	2.3528	0.8610	0.2035

Table 6. Estimates of measures of goodness-of-fit of the reduced models at the end of iteration 4.

Reduced model	Max McElroy R^2	Min AIC	Min BIC	Max SR^2	Max Adjusted SR^2
Mean	0.9994	400.6573	421.6895	0.9741	0.9994
Standard error	1.62E-05	2.9899	2.9899	0.0013	0.0014
Full model-Reduced model	3.26E-05	2.8065	10.4546	0.0013	0.0001
Pooled standard error	1.60E-05	3.0439	3.0439	0.0013	0.0014

full model. The significant level used in the tests in the simulation is 0.05.

3. Repeat Step 1 nine times with the system-of-equations model consisting of 8 independent variables, by dropping x_{25} permanently from the model and dropping another one variable at a time. The optimal values of statistics are summarized in Table 4. At the end of iteration 2, all five goodness-of-fit measures of the reduced models decrease insignificantly. McElroy R^2 still insists to drop x_{14} permanently from the model while the other four statistics suggest dropping x_{24} permanently from the model.

4. Repeat Step 3 eight times by dropping another x_{24} permanently and dropping another one variable at a time. All statistics suggest dropping x_{15} permanently without any significant deterioration in the goodness-of-fit measure of the model. In fact, the significant improvement of goodness-of-fit measure of the reduced model under BIC criterion can be observed. The optimal values of statistics are summarized in Table 5.

5. Repeat Step 4 seven times by dropping another x_{15} permanently and dropping another one variable at a time. All statistics suggest dropping x_{14} permanently from the model. But at the end of iteration 4 the value of McElroy R^2 decreases significantly as shown in Table 6. It implies that the variable x_{14} cannot be excluded from the model under McElroy R^2 criterion without significant deterioration in the goodness-of-fit measure. The variable elimination under McElroy R^2 criterion is to be terminated. But AIC, SR^2, and adjusted SR^2 do not change significantly from the corresponding values in the full model and BIC shows the signifi-

Table 7. Estimates of measures of goodness-of-fit of the reduced models at the end of iteration 5.

Reduced model	Max McElroy R^2	Min AIC	Min BIC	Max SR^2	Max Adjusted SR^2
Mean	0.9992	475.9044	495.0246	0.9175	0.9130
Standard error	3.16E-05	2.4525	2.4525	0.0039	0.0041
Full model-Reduced model	0.0003	-72.4406	-62.8805	0.0579	0.0596
Pooled standard error	2.50E-05	2.7933	2.7933	0.0029	0.0031
t-statistic	10.4150	-25.9333	-22.5108	19.8357	19.2492

cant improvement in the goodness-of-fit measure. The variable elimination procedure will continue under the remaining four criteria.

6. Repeat Step 5 six times by dropping x_{25}, x_{24}, x_{15}, and x_{14} permanently and dropping one of the remaining variables at a time. It should be noted that all irrelevant variables are eliminated and the reduced models in this iteration encompass only relevant variables in the system-of-equations model. Any further elimination of independent variables is expected to deteriorate the goodness-of-fit if the measure is effective. The simulation result shows that all remaining statistics indicate that any additional elimination of independent variables will deteriorate the goodness-of-fit of the model significantly as shown in Table 7. However, in order to observe the change of McElroy R^2 when one of the relevant variables is dropped the maximum value of McElroy R^2 is also included in Table 7.

Under McElroy R^2 criterion we can eliminate only three out of four irrelevant variables but retain all relevant variables. The McElroy R^2 statistic does not illustrate the sharp change when a significant change in value occurs. The AIC, BIC, SR^2, and adjusted SR^2 can eliminate all irrelevant variables and select only all relevant variables and the statistics do illustrate the sharp change as shown in Figures 1 and 2. This sharp angle is necessary to terminate the algorithm in practice when we have only one sample. Additionally, the

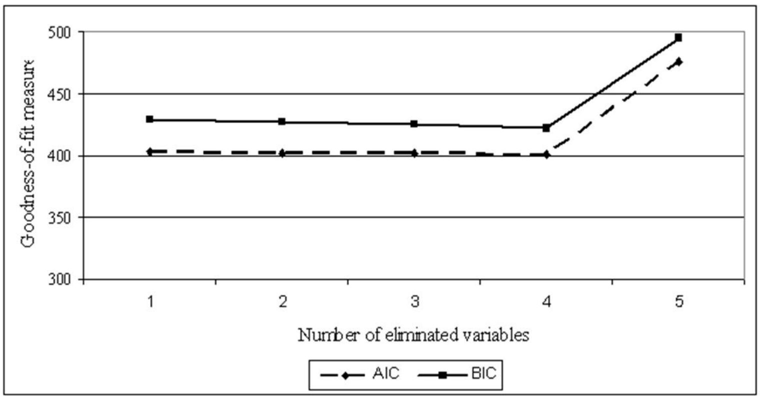

Figure 1. Minimum values of AIC and BIC in each iteration.

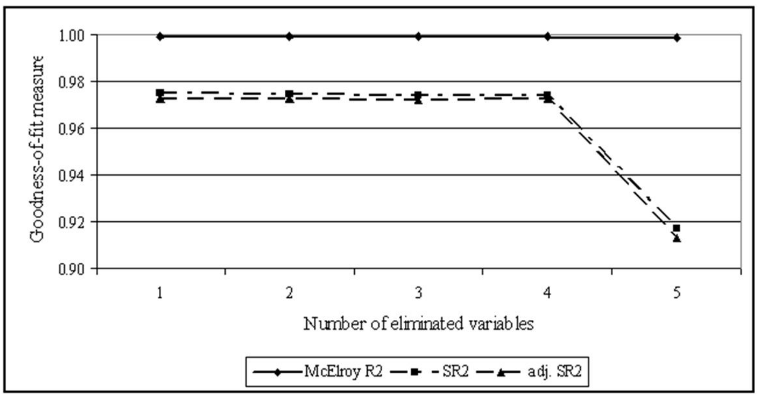

Figure 2. Maximum values of McElroy R^2, SR^2, and adjusted SR^2 in each iteration.

value of McElroy R^2 statistic is quite misleading since it is equal to 0.9995 in the full model, and the maximum value of McElroy R^2 in the reduced model, which drops all irrelevant variables and one of the relevant variables, is equal to 0.9992.

4. Conclusions

The norm of a matrix A, whose diagonal elements are positive, is defined as trace(A). Based on such definition, the SR^2 and adjusted SR^2 statistics are developed as measures of goodness-of-fit for a system-of-equations model. Maximizing SR^2 and adjusted SR^2 are shown to be equivalent to minimizing the sum of error sum of squares and the mean of error sum of squares, respectively. From the simulation results it can be concluded that the characteristics of the proposed SR^2 and adjusted SR^2 statistics are similar to the well-known information AIC and BIC criteria in variable selection for a system-of-equations model. The sharp change in the values of the statistics, except McElroy R^2, is clearly visible when any relevant variable is eliminated from the model. The sharp change is necessary in order to terminate the elimination of independent variables from the model when only one sample is available. Therefore the proposed SR^2 and adjusted SR^2 statistics can solve the underfit and overfit problem in practice. The advantage of the proposed SR^2 and adjusted SR^2 statistics over the information criteria is that the values of the proposed statistics can indicate how close the model is to the true model since the maximum values of the proposed SR^2 and adjusted SR^2 statistics are known equal to 1. The analyst knows from this property how close the system-of-equations model fits the data in the case of the single-equation model.

Acknowledgments

The author wishes to thank Dr. Vichit Lorchirachoonkul, School of Applied Statistics, National Institute of Development Administration, for his encouragement and suggestions throughout this research.

References

Bengtsson, T. and Cavanaugh, J.E. 2006. An Improved Akaike Information Criterion for State–Space Model Selection, Computational Statistics & Data Analysis. 50, 2635-2654.

Draper, N.R. and Smith, H. 1998. Applied Regression Analysis. 3rd ed., John Wiley, New York, U.S.A.

Fujikoshi, Y. and Satoh, K. 1997. Modified AIC and Cp in Multivariate Linear Regression. Biometrika. 84(3), 707-716.

Gorobetz, A. 2005. The Optimal Prediction Simultaneous Equations Selection. Economic Bulletin. 3, 1-8.

Greene, W.H. 2003. Econometric Analysis. 5th ed., Prentice Hall, New Jersey, U.S.A.

Johnson, R.A. and Wichern, D.W. 1998. Applied Multivariate Statistical Analysis. 4th ed., Prentice Hall, New Jersey, U.S.A.

Leon, S.J. 2002. Linear Algebra with Applications. 6th ed., Prentice Hall, New Jersey, U.S.A.

Lorchirachoonkul, V. and Jitthavech, J. 2008. A Variable Selection Criterion for a General VARDL(p,q) Model. 13th Asian Pacific Decision Sciences Institute Conference, July 1-5, 2008, University of Griffith, Brisbane, Australia.

Mantel, N. 1970. Why Stepdown Procedures in Variable Selection. Technometrics. 12(3), 621-625.

McElroy, M.B. 1977. Goodness of Fit for Seemingly Unrelated Regressions: Glahn's and Harper's. Journal of Econometrics. 6, 381-387.

Montgomery, D.C., Peck, E.A. and Vining, G.G. 2006. Introduction to Linear Regression Analysis. 4th ed., John Wiley, New York, U.S.A.

Pan, W. 2001 Akaike's Information Criterion in Generalized Estimating Equations. Biometrics. 57, 120-125.

SAS Institute Inc. 2003. SAS/ETS™ Users'Guide, Version 9.1. SAS Institute, Cary, N.C., U.S.A.

Srebo, N. 2004. Learning with Matrix Factorization. Ph.D. Thesis, Massachusetts Institute of Technology, U.S.A.

Stock J.H. and Watson M.W. 2003. Introduction to Econometrics. Addison Wesley, Reading, Massachusetts, U.S.A.