E C O N O M I C S   B U L L E T I N

The Optimal Prediction Simultaneous Equations Selection

Alexander Gorobets
Sevastopol National Technical University

Abstract

This paper presents a method for selection of the optimal simultaneous equation system from a set of nested models under the condition of a small sample. The purpose of selection is to identify a model with the best prognostic possibilities. Multivariate AIC, BIC and AICC are used as the selection criteria. The selection properties of this method are investigated by Monte−Carlo simulations. They show that the structural form of system can outperform its reduced form for making predictions.

Author is grateful to Erasmus Research Institute of Management for given opportunity to conduct this research. Special gratitude is given to Tim Conley and anonymous reviewers for many helpful discussions and comments on the manuscript.
Citation: Gorobets, Alexander, (2005) "The Optimal Prediction Simultaneous Equations Selection." *Economics Bulletin*, Vol. 3, No. 36 pp. 1−8
Submitted: May 6, 2005. **Accepted:** July 26, 2005.
URL: http://www.economicsbulletin.com/2005/volume3/EB−05C30009A.pdf

1. Introduction

The main contribution to model selection in econometrics is devoted to single equation regression models and many criteria for this purpose have been presented, e.g. Akaike Information Criterion (AIC), corrected version of AIC (AICC), Schwarz' BIC, Bozdogan' ICOMP, etc. (Akaike, 1973; Shi and Tsai, 1998; Bozdogan and Haughton, 1998). Some of them, particularly AIC, BIC and AICC have been modified for selecting multivariate regression models (Bedrick and Tsai, 1994; Fujikoshi and Satoh, 1997). Bedrick and Tsai (1994) showed that AICC for multiresponse models is unbiased for the expected Kullback-Leibler information and provides better model choices than other criteria, including AIC, in small samples. However the problem of simultaneous equation model (SEM) selection has been explored insufficiently. Therefore the objective of this paper is to apply multivariate AIC, BIC and AICC for selection of a SEM with the best prognostic possibility from the given models set in the case of a small sample. The efficiency of the proposed method is investigated by Monte-Carlo simulations. Our special point of interest is to test a hypothesis that estimated structural equations can outperform the reduced form equations for making predictions depending on the specification of SEM and statistical data.

Potential applications of this method include building and analyzing models of economic processes in the countries with transition economics, for example, in the New Independent States, which are characterized by a short period of reforms. In this case, only relatively simple SEMs can be used for prediction of interrelated (endogenous) macroeconomic indicators, e.g. gross domestic product and aggregate consumption.

The order of the presentation is as follows: Section 2 defines the problem of simultaneous equation model selection. Section 3 proposes the method of selection. Section 4 presents an illustrative example and properties of the method. Finally, Section 5 gives the conclusions.

2. The model

The model to be considered is a system of m simultaneous equations

$$y_{it} = \eta_i(\mathbf{x}_t, \mathbf{y}_t, \boldsymbol{\alpha}_i) + u_{it}, \quad i = 1,2 ..., m, \; t = 1,2 ...,n, \tag{1}$$

where y_{it} is a scalar endogenous variable, $\boldsymbol{\eta}' = (\eta_1, \eta_2, ..., \eta_m)$ is the true but an unknown m-vector of models, $\mathbf{x}'_t = (x_{1t}, x_{2t}, ..., x_{kt})$ is a k-vector of exogenous variables, $\mathbf{y}'_t = (y_{1t}, y_{2t}, ..., y_{mt})$ is a m-vector of endogenous variables, $\boldsymbol{\alpha}'_i = (\alpha_{1i}, \alpha_{2i}, ..., \alpha_{Pi})$ is a p_i vector of unknown parameters in a i-th structural equation, and $\mathbf{u}'_t = (u_{1t}, u_{2t}, ..., u_{mt})$ is a m-vector of independent normally distributed random disturbances with zero mean and a covariance matrix $\boldsymbol{\Sigma}_u$, n is the total number of observations. There is usually some prior information about the regions of possible values for variables: $\mathbf{X} \in \mathbf{W}_1$ and $\mathbf{Y} \in \mathbf{W}_2$, where $\mathbf{W}_1$ and $\mathbf{W}_2$ are sets of possible values for the matrices $\mathbf{X}$ and $\mathbf{Y}$. The objective of this research is to identify the model of simultaneous equation system (1), which has the optimal prediction quality on the basis of n observations over matrices $\mathbf{X}$ and $\mathbf{Y}$ under the condition of the small sample. In this case, the order of possible models is limited and relatively simple models can be used. It is necessary to develop a selection method that reflects the trade-off between forecast accuracy and model parsimony.

3. Selection method

The method of selection of the optimal prediction simultaneous equation model consists of the following main stages:

1. The special case in which the possible models are nested as in polynomial regression models or moving-average models for time series is considered. Let the nested set of models be denoted by

$$\eta_{il}(\mathbf{X},\mathbf{Y},\boldsymbol{\alpha}_{il}) \in \mathbf{S}_l, \qquad l = 1, 2,..., q, \tag{2}$$

where $\boldsymbol{\alpha}_{il}$ is a vector of parameters in a i-th structural equation of class l and $\mathbf{S}_1 \subset \mathbf{S}_2 \subset ... \subset \mathbf{S}_q$, $\mathbf{S}_l$ - being the set of all possible models for class l.
For models which are linear in the parameters, model (2) can be rewritten as

$$\eta_{il}(\mathbf{X},\mathbf{Y},\boldsymbol{\alpha}_{il}) = \mathrm{f}'_{il}(\mathbf{X},\mathbf{Y})\boldsymbol{\alpha}_{il}, \tag{3}$$

where $\mathrm{f}_{il}(\mathbf{X},\mathbf{Y})$ is a vector of known functions in a i-th structural equation of class l.

2. The models from every l-th class are tested for identifiability by special conditions such as order condition and rank condition (Fisher, 1966; Brown, 1983).
3. The parameters of each simultaneous equation system from the given models set are estimated by the consistent, full information maximum likelihood method (FIML) (Amemiya, 1986).
4. Multivariate AIC, BIC and AICC are used for selecting the optimal simultaneous equation system from the given models set:

$$\mathrm{AIC} = n\ln\left|\widehat{\boldsymbol{\Sigma}}_u\right| + 2\sum_{i=1}^{m} p_i + m(m+1), \tag{4}$$

$$\mathrm{BIC} = n\ln\left|\widehat{\boldsymbol{\Sigma}}_u\right| + \ln(n)\left(\sum_{i=1}^{m} p_i + 0.5m(m+1)\right), \tag{5}$$

$$\mathrm{AICC} = n\ln\left|\widehat{\boldsymbol{\Sigma}}_u\right| + 2d(mp + 0.5m(m+1)), \tag{6}$$

where $d=n/(n-(m+p+1))$.

Therefore AICC is applicable to the structural form of a SEM only if $p_i = p$ for $i = 1,...m$.
It should be noted that the selection properties of AICC derived for a multivariate regression can be directly generalized to a SEM (Bedrick and Tsai, 1994).

5. The selection properties of AIC, BIC and AICC are explored by Monte-Carlo simulations for a particular experimental situation.
6. The average of the mean squared error of prediction (AMSEP) (Herzberg and Tsukanov, 1985) is used for evaluation of method efficiency and for comparison of the selection criteria:

$$R = \sum_{l=1}^{q} v_l L_l, \tag{7}$$

where v_l is the probability of selection of the model l by a particular criterion, L_l is the loss function for the model l, which is defined as

$$L_l = \left|\boldsymbol{\Sigma}_{el}\right|, \tag{8}$$

where

$$\boldsymbol{\Sigma}_{el} = \begin{pmatrix} \sigma^2_{1l} & cov_{12l} & \cdots & cov_{1ml} \\ cov_{21l} & \sigma^2_{2l} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ cov_{m1l} & \cdots & \cdots & \sigma^2_{ml} \end{pmatrix} \tag{9}$$

is the mean squared errors of prediction (MSEP) matrix for the model l.
where $\sigma_{il}^2 = E\{e_{il}^2\}$, $cov_{ijl} = E\{e_{il}e_{jl}\}$ $(i, j=1..m)$;

$$e_{il}^2 = \frac{1}{n_p}\sum_{t=1}^{n_p}(y_{it} - \eta_{il}(\mathbf{x}_t, \mathbf{y}_t, \hat{\boldsymbol{\alpha}}_{il}))^2 \tag{10}$$

is the mean squared error of prediction for the i-th equation of the model l, n_p is the total number of prediction points.

7. The preferential criterion is used for selection of the optimal prediction model.

4. Simulation results

It is necessary to test the efficiency of the proposed method for optimal model selection. Because of the complexity of analytical exploration the analysis was carried out by the method of statistical trials (Monte-Carlo simulations). Computing was done on the IBM PC Pentium 3 by the tools of the MatLab program. For simplicity consider only simultaneous equation models that are linear in the parameters and endogenous variables. The correct model consists of two equations with one varying parameter, α, and the remaining parameters are given:

$$\begin{cases} y_1 = \eta_1(x_1, y_2, \alpha) + u_1 = x_1 + \alpha x_1^2 - 2y_2 + u_1 \\ y_2 = \eta_2(x_2, y_1, \alpha) + u_2 = x_2 + \alpha x_2^2 - y_1 + u_2 \end{cases}. \tag{11}$$

The system (11) is identified, because both order condition and rank condition are satisfied for every equation of the system (Fisher, 1966). The reduced form of the structural model (11) is

$$\begin{cases} y_1(x_1, x_2, \alpha) = -x_1 - \alpha x_1^2 + 2x_2 + 2\alpha x_2^2 + v_1 \\ y_2(x_1, x_2, \alpha) = x_1 + \alpha x_1^2 - x_2 - \alpha x_2^2 + v_2 \end{cases}, \tag{12}$$

where v_1, v_2 - are the normally distributed random disturbances with zero mean and the covariance matrix $\mathbf{\Omega}_v$.

In practice, the system (11) can roughly describe a supply-demand model for the meat market of Ukraine, where y_1 (quantity) and y_2 (price) are endogenous (interdependent) variables and x_1 and x_2 are exogenous variables. In this case we are particularly interested in using the structural equations for prediction of next period's quantity given external information on next period's price and current data due to special governmental policy of setting price limits. And in turn, forecasting next period's real market price given next period's estimate of quantity (internal and external supply) is important task for the local community due to sharp economic conditions. The data generating process is based on the real economic, social and political circumstances in many countries of the former Soviet Union, e.g. mismanagement of correct statistical data collection, significant proportion and variability of the shadow economy (x_1), high instability of institutional factor (x_2) and a short period of reforms (13 years). Therefore the design matrix $\mathbf{X}$=(x_1, x_2) is a fixed matrix of independent, identically distributed normal random variables with mean zero, variance one and n=13. The random disturbances were simulated by the generator of random numbers built in the computer program with mean zero and $\sigma_{u1}^2 = \sigma_{u2}^2 = 1.0$; parameter α varied from 0.05 to 5 (adjusted to the specific set of models). The selection of both structural and reduced models was simulated by the criteria AIC, BIC and AICC for every realization of the experimental data (t=1,..,n–1) from the correct models (11) and (12). The parameters were estimated by FIML method and the experiment was repeated 5000 times. The mean squared error of prediction is calculated in the last point of design (out of the estimation period) in order to verify the prognostic efficiency of this method. The selection was made from the following nested classes of models η:

1) linear (underfitted), i.e. the quadratic terms were excluded from the initial system (11);

2) quadratic (correct), i.e. the structure of the initial system remained the same;

3) cubic (overfitted), i.e. the exogenous variables in a third power were added to each equation of the initial system.

The results of simulations are averaged across random samples and presented by figures as functions of α. Figure 1 shows the loss functions for all models and AMSEP by AIC and AICC for the reduced form of a SEM.

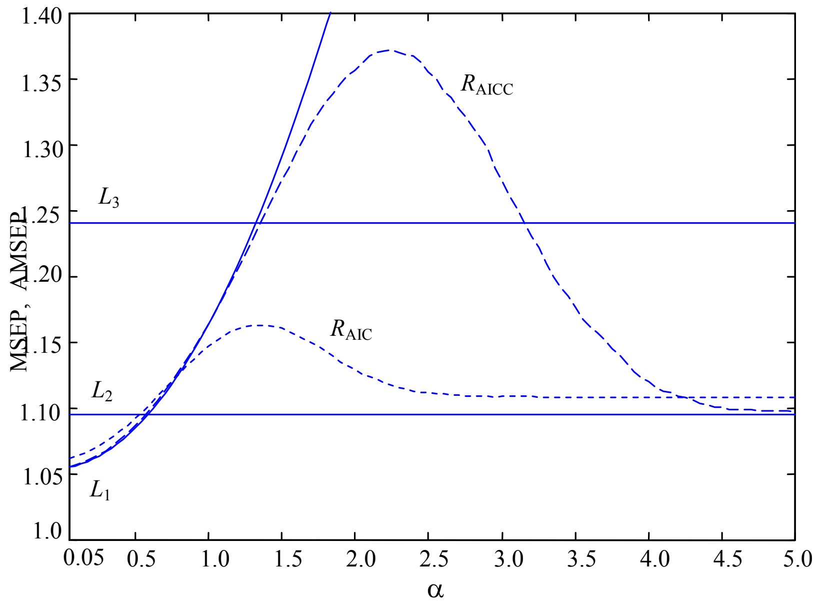

Fig.1. MSEP and AMSEP for the reduced system:
L_1 – MSEP for linear model;
L_2 – MSEP for quadratic model;
L_3 – MSEP for cubic model;
R_{AIC} / R_{AICC} – AMSEP by AIC / AICC.

It can be seen (Fig.1.) that there is a region of the correct model parameter variation where an underfitted model (linear in this case) is better for making predictions than the correct (quadratic) model. The loss function for the overfitted (cubic) model is always bigger than for the correct model and they do not depend on α variation. This has been confirmed by theoretical research (Gorobets, 2005). Because the number of the chosen model by the selection criterion is random, AMSEP is always bigger than the minimal loss function or equal to that in the entire region of parameter variation. Although there is a region of α variation where AMSEP by AICC is bigger than AMSEP by AIC, with increasing parameter only AMSEP by AICC gradually converges to the loss function of the correct model, which verifies the efficiency of criterion AICC.

Figure 2 gives the number of times each model was selected by AIC and AICC for the reduced system. With increasing α AICC consistently identifies the correct model ($AICC_2$), whereas AIC tends to overfit the model (AIC_3). On the contrary under the small values of α AIC performs better than AICC, which tends to underfit the model ($AICC_1$).

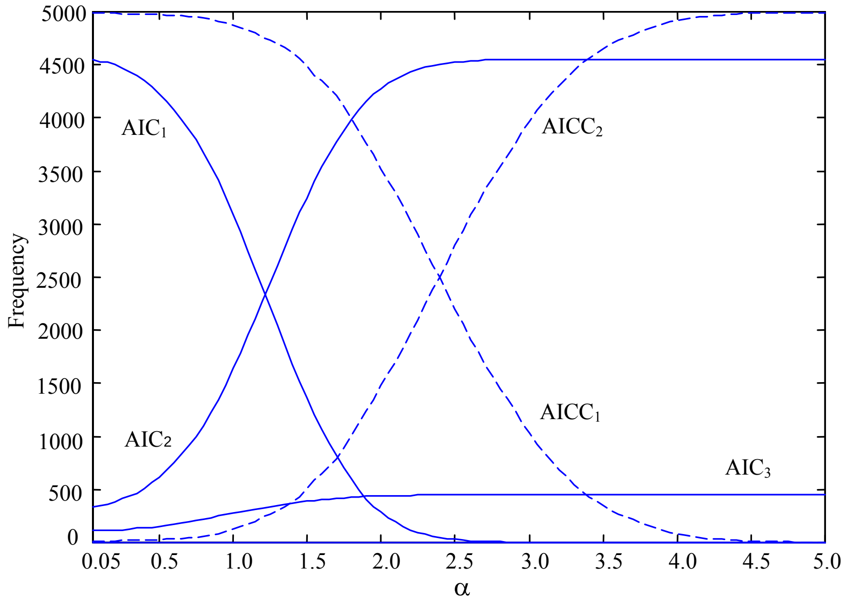

Fig.2. Frequency of models selected by AIC and AICC for the reduced system:
AIC$_1$ /AICC$_1$ – frequency of linear model selected by AIC /AICC;
AIC$_2$ /AICC$_2$ – frequency of quadratic model selected by AIC /AICC;
AIC$_3$ – frequency of cubic model selected by AIC.

Figures 3,4 demonstrate the simulations results for the structural form of a SEM.

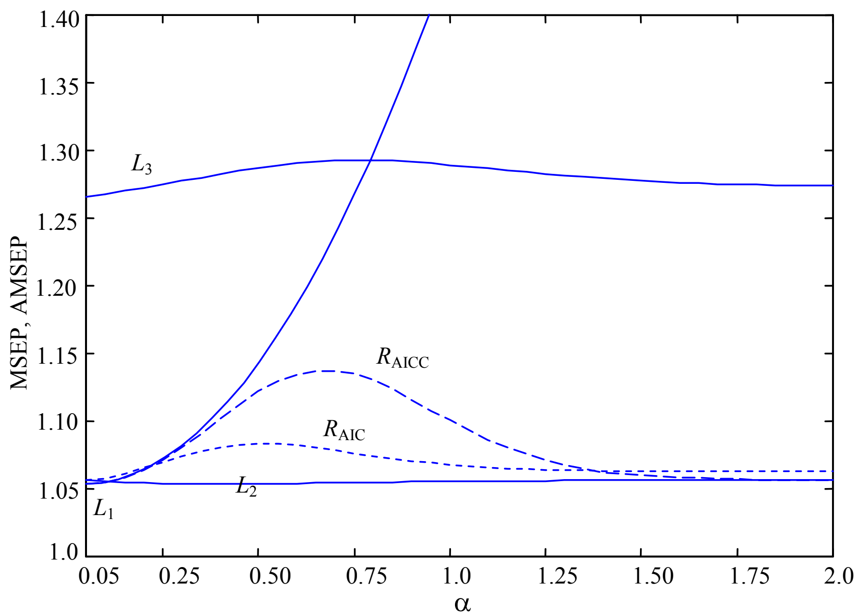

Fig.3. MSEP and AMSEP for the structural system:
L_1 – MSEP for linear model;
L_2 – MSEP for quadratic model;
L_3 – MSEP for cubic model;
R_{AIC} / R_{AICC} – AMSEP by AIC / AICC.

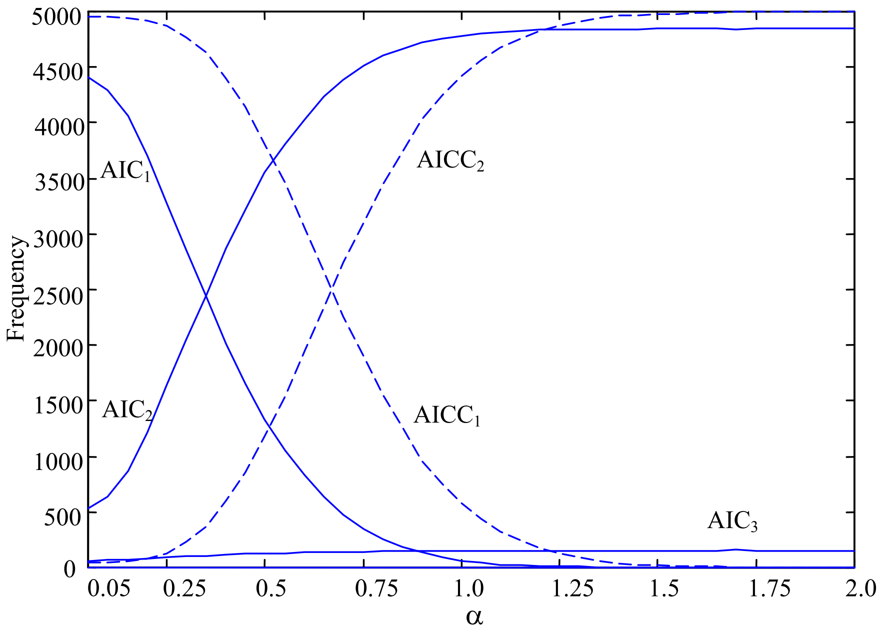

Fig.4. Frequency of models selected by AIC and AICC for the structural system:
AIC_1 /$AICC_1$ – frequency of linear model selected by AIC /AICC;
AIC_2 /$AICC_2$ – frequency of quadratic model selected by AIC /AICC;
AIC_3 – frequency of cubic model selected by AIC.

A difference from the reduced system is that MSEP for all models of the structural system depends on parameter variation. Comparison between figures 1 and 3 illustrates that the structural system can be better for forecasting than the reduced system. The selection properties of criteria AIC and AICC for the structural system are similar to that for the reduced system, but they converge to the correct model faster than for the reduced system. As for criterion BIC, which is not shown on these figures to simplify presentation, it performs between AIC and AICC for both the reduced and the structural systems. Since analogous results were obtained for other values of the SEM parameters and designs we can expect its robustness for the given set of models and error distribution.

On the basis of the simulation results, the following selection properties of proposed method can be formulated:

1. The method allows the selection of the optimal prediction simultaneous equation system from the given models set;
2. For the purposes of the minimization of the prediction error and the speed of convergence to the correct model the structural form of a SEM can be preferred to the reduced system;
3. The criteria efficiency depends on the region of the correct model parameter variation.

5. Conclusions

In this paper a new method for selecting the optimal prediction simultaneous equation system was presented and a first round of computer simulations was carried out to illustrate the performance of the method. The novelty of this method is that the structural form of SEM is identified by criteria for making (better) predictions, whereas traditionally only the reduced form is used for prediction. Therefore in practice it can be recommended to estimate both the reduced and structural forms of SEM for making predictions of real economic indicators. The analysis of the method was done for a particular experimental situation, i.e. for a specific set of models and random disturbances. It is necessary to carry out further investigation of the performance of the method in two complementary ways. First, to increase the generality of our conclusions we should conduct a large number of experiments with various feasible sets of models and error distributions. Second, analytical derivations of the proposed method would be valuable to confirm the simulation results and justify the method. This task was partly resolved by Gorobets (2005), where analytical expressions of the mean squared error of prediction matrices were derived for biased, true and overfitted models of the reduced form of SEM, but they still remain unknown for the structural form of SEM. Furthermore unbiased criterion for selecting a SEM with different number of parameters in each structural equation should be developed.

References

Akaike, H. (1973) "Information theory and an extension of the maximum likelihood principle" in Proc. 2nd Int. Symp. on Information Theory by B.N. Petrov, F. Csaki, Eds., Akademiai Kiado: Budapest, 267 – 281.

Amemiya, T. (1986) Advanced econometrics, TJ Press Ltd*:* Oxford.

Bedrick, E.J., and C.-L. Tsai (1994) "Model selection for multivariate regression in small samples" Biometrics 50, 226-231.

Bozdogan, H., and D. Haughton (1998) "Informational complexity criteria for regression models" Computational Statistics & Data Analysis 28, 51-76.

Brown, B.W. (1983) "The identification problem in systems nonlinear in the variables" Econometrica 51, 175-196.

Fisher, F.M. (1966) The Identification Problem in Econometrics, McGraw-Hill: New York.

Fujikoshi, Y., and K. Satoh (1997) "Modified AIC and C_p in multivariate linear regression" Biometrika 84 (3), 707-716.

Gorobets, A. (2005) "The error of prediction for a simultaneous equation model" Economics Bulletin 17 (3), 1-7.

Herzberg, A.M., and A.V.Tsukanov (1985)"The Monte-Carlo Comparison of Two criteria for the Selection of Models" Journal of Statistical Computation and Simulation 22, 113-126.

Shi, P., and C.-L.Tsai (1998) "A note on the unification of the Akaike information criterion" J.R. Statist. Soc. B. 60 (3), 551-558.