Cairo University
Institute of Statistical Studies
and Research (ISSR)

Finite Sample Properties of Double k –Class Estimators in Simultaneous – Equations Model

By

Marwa Ibrahim Mahmoud Sharadah

Submitted to Applied Statistic & Econometric Department,
Institute of Statistical Studies & Research, Cairo University
In Partial Fulfillment of the Requirements for the Degree
Of Master of Statistic

Under the supervision of

Prof. Dr. Ghazal Abdel-Aziz Amer

Professor of Statistics &Econometrics

Institute of Statistical Studies and Research

Cairo University

Prof. Dr. Sayed Mesheal El-sayed

Head: Department of Applied Statistics and Econometrics

Professor of Statistics

Institute of Statistical Studies and Research

Cairo University

2009

To my dear father,

To my dear mother,

To my dear mother in-law

and

To my dear husband

CERTIFICATION

I certify that this work has not been accepted in substance for any academic degree and is not being concurrently submitted in candidature for any other degree.

Any portions of this dissertation for which I am indebted to other sources are mentioned and explicit references are given.

The student

APPROVAL SHEET

Finite Sample Properties of the Double k –Class Estimators in Simultaneous – Equations Model

By

Marwa Ibrahim Mahmoud Sharadah

This dissertation is submitted as a requirement for the Degree of Master of Statistics to the Institute of Statistical Studies & Research, Cairo University. This dissertation has been approved by

Prof: Ghazal Abdel-Aziz Amer …………

Prof: Sayed Mesheal El-sayed …………

Prof: Amany Mousa Mohamed …………

Prof: Abd Allah Abdel-Galy ………….

ACKNWLEDGMENT

I am very grateful to my professor Dr. Ghazal Abdel Aziz Amer who suggested the topic involved in this thesis. I am also grateful to him for his assistance, his great advices and for supporting me with encouragement and enthusiasm.

I would like to express my deep thanks to professor Dr. Sayed Mesheal El-Sayed for his guidance, creative comments and continuous encouragement which were all of great help in achieving this thesis.

I wish to offer my deep thanks also to all the staff members of the Institute of Statistical Studies and Research for their continuous help and cooperation throughout the preparation of this thesis.

Finally, I would like to express my sincere appreciation and thanks to my father, my mother, my mother in-law and my husband. Their support and encouragement were and still the light which make me go on.

GLOSSARY OF ABREVATIONS USED IN THE THESIS

OLS ⇨ Ordinary Least Square

ILS ⇨ Indirect Least Square

2SLS ⇨ Two-Stage Least Square

LIML ⇨ Limited Information Maximum Likelihood

pdf ⇨ Probability Density Function

Table of Contents

Chapter One

Introduction

Most conceptual frameworks for understanding economic processes and institutions reflect the vision that in economic the existence of interrelationship everything depends on everything else. This idea translates into the realization that economic data that are a product of the existing economic system must then be described as a system of simultaneous relations among the random economic variables and that these relations involve current and lagged values of some of the variables. This led Marschak (1950) to comment that “economic data are generated by systems of relation that are in general stochastic, dynamic and simultaneous ". He further noted that, although these properties of the data give rise to many unsolved problems in statistical inference, they constitute the basic ingredients underlying economic theory, and quantitative knowledge of them is needed for economic practice.

In these situations there are two way cause-and-effect relationships between the endogenous and predetermined variables, which make the distinction between them of dubious value. It is better to lump together a set of variables that can be determined simultaneously by the remaining set of variables precisely what is done in simultaneous - equation models . In such models there is more than one equation -one for each of the endogenous variables and unlike the single-equation models, in the simultaneous- equation models one may not estimate the parameters of a single equation without taking into account information provided by other equations in the system. If one applied the classical OLS method to estimate the parameters of this single equation , these estimators would be bias , inconsistent and they do not converge to their true population values no matter how large the sample size . This refers to the correlation between the disturbances and some of endogenous variables which appear as predetermined variables.

(1.1) The statistical model

To introduce the idea of simultaneity, we must pay special attention to how the economic variables are classified. Basically, each equation in the model may contain the following types of variables:

i) Endogenous variables have outcome values determined through the joint interaction with other variables within the system.

ii) Predetermined variables that affect the outcome of the endogenous variable, but their values are determined outside the system and they contain the following types of variables:

1] Current and lagged exogenous variables.

2] Lagged endogenous variables placed in the same type as the exogenous variables since for the current period the observed values are predetermined.

iii) Random disturbances.

The equations of the system are called structural equations, and the corresponding parameters are called structural parameters .The system of equations is complete if the number of independent equations are equal to the number of endogenous variables.

With these definition and concepts let the statistical model which consists of G linear stochastic equation relating G endogenous variables by the (Tx1) vectors $y_1,y_2,\ldots,y_G$; L predetermined variables by the (Tx1) vectors $z_1,z_2,\ldots,z_L$ and G random disturbances by the (Tx1) vectors $u_1,u_2,\ldots,u_G$.

In matrix notation the linear statistical model may by written compactly as:

$$\underset{(TxG)}{\mathbf{Y}}\ \underset{(GxG)}{\mathbf{B}} + \underset{(TxL)}{\mathbf{Z}}\ \underset{(LxG)}{\mathbf{\Gamma}} = \underset{(TxG)}{\mathbf{U}} \tag{1-1}$$

The following assumptions will be made throughout this thesis:

Assumption 1 :(Completeness and linearity)

The T values taken by the G jointly dependent variables are the realizations of an **(**TxG**)** random matrix **Y** which satisfies (1-1), where **Z** is an **(**TxL**)** matrix of predetermined variables observations; **U** is an **(**TxG**)** matrix of random disturbances, and **B** and $\mathbf{\Gamma}$ are unknown parameters matrices of order **(**GxG**)** and **(**LxG**)** respectively, **B** being nonsingular (i.e) the reduced form exists.

Assumption 2 :(Independence and identical distribution of the rows of the disturbance matrix)

The T rows of the **(**TxG**)** disturbance matrix $\mathbf{U}$ are independent random drawings from an G-dimensional population with zero mean vector and unknown but finite covariance matrix $\mathbf{\Sigma}$ and also they are identically distributed.

i.e $E(u_i) = 0$ for $i = 1, 2, ..., G$

then $E(\mathbf{U}) = \mathbf{0}$ (1-2)

and $E(u_i u_j') = \sigma_{ij} \mathbf{I}_T$ for $i, j = 1, 2, ..., G$

then the variance matrix can be written as

$$V(\mathbf{U}) = E(\mathbf{UU'}) = \begin{pmatrix} \sigma_{11}\mathbf{I} & \sigma_{12}\mathbf{I} & \cdots & \sigma_{1G}\mathbf{I} \\ \sigma_{21}\mathbf{I} & \sigma_{22}\mathbf{I} & \cdots & \sigma_{2G}\mathbf{I} \\ \vdots & & \ddots & \vdots \\ \sigma_{G1}\mathbf{I} & \sigma_{G2}\mathbf{I} & \cdots & \sigma_{GG}\mathbf{I} \end{pmatrix}$$

$$V(\mathbf{U}) = E(\mathbf{UU'}) = \mathbf{\Sigma} \otimes \mathbf{I} \tag{1-3}$$

where the unknown variance covariance matrix Σ is an (GxG) symmetric and positive semidefinite matrix and $\otimes$ is the kronecker product.

Assumption 3: (Purely exogenous case)

The (TxL) matrix $\mathbf{Z}$ of the values taken by the predetermined variables has rank L **(**L<T**)** and consists of nonstochastic elements. The second part of this assumption requires that there be no lagged endogenous variables in the complete system.

From the structural form (1-1) and under the assumption that **B** is nonsingular matrix one can derive the reduced form that expresses the endogenous variables solely in terms of predetermined variables and the random disturbances. The structural form represented by (1-1) may be written in the following equivalent form:

$$\mathbf{Y}\mathbf{B}\mathbf{B}^{-1} + \mathbf{Z}\Gamma\mathbf{B}^{-1} = \mathbf{U}\mathbf{B}^{-1}$$

$$\underset{(T\times G)}{\mathbf{Y}} = \underset{(T\times L)}{\mathbf{Z}}\ \underset{(L\times G)}{\Pi} + \underset{(T\times G)}{\mathbf{V}} \qquad (1\text{-}4)$$

where: $\Pi = -\Gamma\mathbf{B}^{-1}$ (1-5)

Π is the matrix of reduced form parameters

$$\mathbf{V} = \mathbf{U}\mathbf{B}^{-1} \qquad (1\text{-}6)$$

V is the matrix of reduced form disturbances

The system in equation (1-4) is reduced form. Since only the predetermined variables and random disturbances appear on the right side of this equation and since the predetermined variables are assumed to be uncorrelated with the disturbance terms, the OLS method can be applied to estimate the coefficients of the reduced form as following:

$$\hat{\Pi} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \qquad (1\text{-}7)$$

(1.2) <u>The identification problems</u>

Since the reduced form parameters in equation (1-5) can be estimated consistently by using the OLS method, it seems natural to try to estimate the structural parameters $\mathbf{B}$ and Γ . This method is known as indirect least square (ILS), since the reduced form parameters can be estimated by the OLS method and these estimators in (1-7) can be used to obtain the estimators of the structural parameters. One problem with this method is that the ILS method is not always successful in providing unique estimators of the structural parameters. This result occurs because the relation in (1-5) sometimes provides insufficient information for obtaining estimators of the structural parameters.

If the estimators of the structural parameters can be obtained from the estimators of the reduced form parameters, the particular equation is identified. If this cannot be done, then the equation under consideration is unidentified or under identified. An identified equation may be either just identified or over identified. It is just identified if unique estimators of the structural parameters can be obtained .It is said to be overidentified if more than one estimators can be obtained for some of the parameters of the structural equation.

The identification problem asks whether one can obtain unique estimators of the structural parameters from the estimated reduced form parameters. The model as a whole is identified if each equation in it is identified. To secure identification, one may resort to reduced form equations or use an alternative and perhaps less time - consuming method of determining whether or not an equation in simultaneous-equations model is identified.

(1.3) <u>Conditions of identification and methods of estimation</u>

There are two conditions for identifying an equation: an order condition which is just necessary and rank condition which is both necessary and sufficient. To understand the order and rank condition, the following notations are introduced as:

G = number of endogenous variables in the model

L = number of predetermined variables in the model

G_i number of endogenous variables in the i th equation

L_i number of predetermined variables in the i th equation

N_i number of parameters to be estimated in the i th equation

$N_i = G_i + L_i - 1$ where $i = 1, 2, \ldots, G$

(1.3.1) The order condition for identification:

The order condition states that the number of predetermined variables excluded from the equation must not be less than the number of endogenous variables included in that equation less 1, that is

$L - L_i \geq G_i - 1$ or $L \geq N_i$

if $L = N_i$,then the equation is just identified

if $L > N_i$,then the equation is overidentified

if $L < N_i$,then the equation is under identified .

(1.3.2) The rank condition for identification:

The rank condition states that an equation is identified if and only if at least one nonzero determinant of order (G-1)x(G-1) can be constructed from the coefficients of the variables (both endogenous and predetermined) excluded from that particular equation but included in the other equations of the model .

If the rank condition is satisfied, the order condition is satisfied, too, although the converse may not be true. That means the rank conditions necessary and sufficient condition but the order condition is necessary but not sufficient condition for identification. If an equation in a simultaneous- equation model is identified (either just or over), there are several methods to estimate it. These methods fall into two broad categories: single equation methods and system-equation methods. In the sense of econometric theory, the single-equation methods are by far the most popular. A feature of these methods is that one can estimate a single equation in a multiequation model without worrying too much about other equation in the system. In the system methods, on the other hand, such methods are not commonly used for a variety of reasons .First, the computational burden is enormous. Second, the system-equation methods, lead to solutions that are highly nonlinear in the parameters and are therefore often difficult to determine. Third, if there is a specification error in one or more equations of the system, that error is transmitted to the rest of the system. As a result, the system methods become very sensitive to specification errors.

Here we are concerned with the single-equations methods. The ILS method is a single-equation method which provides consistent estimators for the reduced from parameters, but the unique derivation of the parameters of a structural equation from the reduced from parameters is possible only when the structural equation is just identified .Unfortunately, if we actually try to model the underlying data generation process, not all structural equations are just identified. Consequently, specifying models that we believe are consistent with the way in which the economic data were generated leads in many cases to structural equation that are overidentified .In this case the problem is to use all the information contained in the model when carrying out estimation relative to the parameters of the structural