Sampling Distribution of Econometric Estimators in Simultaneous Equations Models

Kianré Boniface Eouanzoui

A Thesis

in

The Department

of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Sciences at

Concordia University

Montréal, Québec, Canada

May 1990

ISBN 0-315-64674-8

ABSTRACT

Sampling Distribution of Econometric

Estimators in Simultaneous Equations Models

K.B. Eouanzoui

As field of applications of simultaneous equations, a complete section is devoted to survey of macroeconometric models mainly applied to the U.S. intra and post world wars economy. Then a comprehensive study of the simultaneous equations model is carried out. A survey of the methods of estimation is also included for ready reference.

A discussion of reduced form and variance covariance matrix is also considered. This is followed by a survey of the developments on the distributions of the econometric estimators in large simultaneous equations model during these last fifteen years.

ACKNOWLEDGEMENTS

Although it does not contain any mathematical formula, this section is the hardest to write. I am grateful to a number of nice people who helped me in one form or another to draw up this thesis. I am particularly indebted to Professor T.D. Dwivedi who, promptly, accepted to offer me the opportunity to taste the stimulating but also disciplining and exhausting savor of research duty. All along the process of writing this thesis, his talent of pedagogue, his wide experience in supervising theses as well as his patience have been, more than once, submitted to stern proofs. In addition to statistics theory and methods, he subtly taught to me human relations. Thank you Professor.

I am also very appreciative of the technical and financial assistance by Dr. F. Wang, Dr. E. Cohen, and Dr. J. Garrido. Without this assistance I would have given up this bold duty. One more time, I am indebted to all those who helped me and whose names cannot be enumerated with respect to the thesis regulations.

Contents

Chapter 1 INTRODUCTION

1.1 Simultaneous equations..... 1

1.2 The simultaneous equation model..... 14

1.3 The reduced form..... 15

1.4 A priori restrictions on structural parameters..... 17

Chapter 2 METHODS OF ESTIMATION

2.1 Introduction..... 21

2.2 Single equation of the complete structural system..... 22

2.3 Ordinary least squares and least squares bias..... 25

2.4 Indirect least squares (ILS) estimator of parameters of a just-identified structural equation..... 27

2.5 The two-stage least-squares (2SLS) estimator of parameters in a structural equation..... 30

2.6 Three-stage least squares..... 35

2.7 Limited information maximum likelihood..... 38

2.8 Full information maximum likelihood..... 51

2.9 Instrumental variable (IV) estimator..... 54

2.10 Families of general k-class, h-class, and double k-class estimators..... 56

Chapter 3 SURVEY OF SOME RECENT DEVELOPMENTS (1973 - 1988) ON THE DISTRIBUTION OF ECONOMETRIC ESTIMATORS

3.1 Introduction..... 59

3.2 Developments on the exact distribution theory..... 60

3.3 Survey of developments on the approximation to the distribution of econometric estimators in simultaneous..... 99

BIBLIOGRAPHY..... 124

Chapter 1

INTRODUCTION

1.1 SIMULTANEOUS EQUATIONS

Simultaneous-equations result from the interdependence of economic phenomena. The concept was first introduced by Tinbergen while he was working for the League of Nations. A simultaneous- equations model consists of several stochastic and deterministic equations. Traditionally we classify these equations into the following four categories:

i)behavioral equations: represent the behavior of some group of economic subjects like the demand and supply equations, the aggregate consumption and investment equations, the price and wage functions, etc...

ii)technological or technical relations: arise due to the given technology of a certain firm or an industry like the production function relating output and inputs.

iii)institutional relations: arise due to various governmental regulations like tax functions, and

iv)definitional relations or identities: arise due to specific definitions of economic variables for example, the total wage bill is equal to the wage rate multiplied by the total number of people employed. Generally speaking, relations in i), ii) and iii) will be stochastic whereas iv) will be deterministic.

Since the equations of a simultaneous-equations model are supposed to describe the structure of an economy, they are called structural equations, and the parameters of the equations system are called structural parameters. Accordingly, the disturbances in these equations are called structural disturbances.

We have two sets of variables in the structural equations system, viz., endogenous and exogenous variables. The endogenous variables are those which are "to be explained" by the equations system and exogenous variables are supposed to be determined from outside the system. However, these variables are measured at different points of time, and, therefore, for statistical purposes, we distinguish between jointly dependent and predetermined variables. The current endogenous variables are called jointly dependent and the lagged endogenous variables along with current and lagged exogenous variables are grouped together in the category of predetermined variables.

The earliest and still one the most important application of simultaneous-equations techniques is to macroeconometric models [61]. The latest are reputed to involve more equations and variables by including certain factors not treated explicitly in prototype model, which focuses on national income variables like prices, wages, interest rates, employment, and unemployment.

The first macroeconometric model was the Tinbergen model of U.S. business cycles in the periode 1919-1932 [87]. This model was quite influential in three respects. First, it influenced futur research by developing a quantitative approach to the subject of business-cycle analysis. Second, it frostered the further development and use of econometrics. Third, it was partly responsible for the later work on

problems of estimation of a simultaneous-equations system.

The major macroeconometric models are:

a)The Brookings model (1960) [29] & [39]: the largest and most ambitious macroeconometric model of the U.S. economy. More than thirty economists at various universities and research organizations collaborated in the development of the model, with individual specialists working in particular sectors. Brooking model is a highly disaggregated quaterly model, involving, in the "standard" version, 176 endogenous and 89 exogenous variables. It has been used both for structural analysis of cycles and for growth and policy evaluation. The model was divided into various interacting blocks, and two-stage least squares was then used to make consistent estimates of the individual equations within each major block of equations. The entire model was then reestimated to take account of the interactions among the blocks, and the resulting estimated model was the one used for policy simulation experiments. Interconnections among the blocks of the Brookings model are:

A							Fixed Business Investment and Exports
B	C						Other Final Demand
D	E	F					Sector Outputs
G		H	I				Employment and Hours
	J		K	L			Labor Supply, Unemployment
		M	N	O	P		Wages, Prices, and Profits
Q	R				S	T	Interest, Money, and Other Factor shares

The most important outcome of Brookings model was, without doubt, its role in integrating various sectors of the economy, methodologies, and data into a single unified framework and its influence in these respects on later models, estimation approaches, and data banks.

b)The Chase Econometrics model [35]: a large-scale quaterly model used for short term forecasting. The model was used to estimate new passenger car sales. The variables are: new passenger car sales, personal disposable income, transfer payments, ratio of nonwage personal income to wages and salaries, credit-rationing variable, ratio of the average monthly payment for new cars to the consumer price index, unemployment rate, dummy variable for auto strikes, and the stock of new cars. One of the important features of the model is the explicit inclusion of credit rationing and capacity utilization.

c)The D R I model [32]: developed by Data Ressources, Inc. is one of the largest models of the U.S economy. It is a highly disaggregated model that was influenced by the brookings model, the Wharton model, and other earlier models. The 1976 DRI model includes 718 endogenous variables and 170 exogenous variables.

d)The Duesenberry-Eckstein-Fromm model [28]: a quaterly model of the U.S.A economy in recession, emphasizing tax and transfer payments. It is not a direct descendant of the Klein-Goldberger model; however it was influenced by the latter.

e)The Fair model [36]: a small short-run quaterly forecasting econometric model consisting of 14 stochastic equations and 5 identities. The model explicitly allows for disequilibrium in the housing sector and makes use of the concept of "excess labor" to explain employment.

f)The M P S model [75]: is the public version of an econometric model of the U.S. economy developed by the Federal Reserve Board, M I T, and the University of Pennsylvania (FMP). While the FMP model is used for forecasting and policy evaluation by the Federal Reserve System, the MPS model (M I T, Pensylvania, and the Social Science Research Council) is a large-scale quaterly econometric model involving over 100 equations. Its main focus is in estimating the impacts of alternative monetary policies. The model includes six major blocks of equations. Listing of the version 4.1 of FMP model, dated April 15, 1969. is

	Stochastic equations	Nonstochastic equations	Total equations
final demand	24	20	44
distribution of income	5	21	26
tax and transfert	12	9	21
labor market	3	10	13
prices	10	22	32
financial sector	21	14	35
Total	75	96	171

g) The Klein-Goldberger model [44]: is a "medium-size" econometric model of the U.S. economy for the period 1929-1952, excluding the war years 1942-1945. It consists of twenty equations, of which 15 are stochastic and 5 are identities. It contains 34 variables, of which 20 are endogenous and 14 are exogenous.

20 endogenous variables	14 exogenous variables
5 income	government expenditure
consumption	4 direct taxes
gross private unvestment	indirect tax
depreciation	5 population & labor force
imports	hours worked
corporate saving	excess reserves
corporate surplus	import price
capital stock	
2 liquid assets	
3 prices	
2 interest rates	

The Klein-Goldberger model has extremely influenced the construction of most of the later models.

h)The Klein Interwar Model (or Klein Model I) [43]: a "small" model of less than ten stochastic equations developed by Lawrence R. Klein to analyze the U.S. economy during the period between world wars I and II, 1921-1941. It has been used to study policy pursued during the depression year. Variables of the Klein model I are:

6 endogenous variables	4 exogenous variables
Y = output	G = governm. nonwage expend.
C = consumption	W_G = public wages
I = investment	T = taxes
W_p = private wages	t = time
Π = profits	
K = capital stock (at year end)	

The model:

. aggregate consumption in year t.

$$(1)\ C_t = \alpha_0 + \alpha_1(W_{pt} + W_{Gt}) + \alpha_2\Pi_t + \alpha_3\Pi_{t-1} + u_{1t}$$

where u_{1t} is structural disturbance.

. Investment function:

$$(2)\ I_t = \beta_0 + \beta_1\Pi_t + \beta_2\Pi_{t-1} + \beta_3K_{t-1} + u_{2t}$$

. private wages:

$$(3)\ W_{pt} = \gamma_0 + \gamma_1(Y + T - W_G)_t + \gamma_2(Y + T - W_G)_{t-1} + \gamma_3(t-1931) + u_3(t).$$

. national income identity:

$$(4)\ Y_t = C_t + I_t + G_t$$

.total profits of the private sector

$$(5)\ \Pi_t = Y_t - W_{pt} + T_t$$

.net investment

$$(6)\ K_t = K_{t-1} + I_t$$

The Klein interwar model is said to be complete.

i)The Liu model [47]: an exploratory model of effective demand in the postwar economy, starting with the first quarter of 1947. It has a monetary sector, involving five liquid assets, and five interest rates. Exogenous variables are excess reserves relative to required reserves and the discount rate.

j)The Liu-Hwa model [48]: first major monthly econometric model; it was influenced by the Liu model and intended to analyze forecast and policy. It includes 12 policy instruments, the values of which, together with lagged endogenous and exogenous variables, generate monthly forecasts of GNP and its components.

k)The Michigan Quaterly Econometric Model (MQEM) [42]: direct descendant of a small quarterly model developed in the late 1960's at the Council of Economic Advisers for use in forecasting. It was influenced by the Suits model. MQEM is a medium-size nonlinear model designed for short-term prediction. Its variables are: wages and prices, productivity and employment, expenditures, income flows, interest rates, and output composition.

l)The Morishima-Saito model [60]: is a model of long-term growth. It has been used for structural analysis and for policy evaluation, particularly for the study of the relative effectiveness of monetary

and fiscal policy. Variables of the Morishima-Saito model of U.S. economy over the period 1902-1952 (excluding 1941-1945) are:

9 endogenous variables	6 exogenous variables
Y = net national output (national income)	I = investment (gross)
C = consumption	B = trade balance
D = capital consumption allowances	M = cash balance
K = capital stock (at year end)	N = population 15 and over
L = employment	t = time
p = price level (1929:1)	u = dummy variable (0 before 1941, 1 after 1946)
w = wage rate	
r = corporate bond yield	
h = hours worked per person per y.	

The model:

.consumption function:

$$(1)\ \mathrm{Log}(C/N)_t = \alpha_1 \mathrm{Log}(Y/N)_t + \alpha_2 \mathrm{Log}(M/pN)_t + \alpha_3 \mathrm{Log}(C/N)_{t-1} + \alpha_0 + e(t)$$

.liquidity preference function:

$$(2)\ \mathrm{Log}(M/p)_t = \beta_0 + \beta_1(\beta_2 \mathrm{Log}Y_t + \beta_3 \mathrm{Log}(M_{t-1}/p)) + \beta_4) + \beta_5 \mathrm{Log}r_t + \beta_6 u + e(t).$$

.production function:

$$(3)\ \mathrm{Log}Y_t = \gamma_0 + \gamma_1 \mathrm{Log}(hL)_t + \gamma_2 \mathrm{Log}K_{t-1} + \gamma_3 t + \gamma_4 u + e(t).$$

.relative-share equation:

$$(4)\ (wL/pY)_t = a_0 \text{ or } Log(wL/pY)_t = a_0 + e(t).$$

.wage-determination equation:

$$(5)\quad Log(w/h)_t = b_0 + b_1 Log(w/h)_{t-1} + b_2(Log(p/p_{t-1})^{1/5} + Log(p/p_{t-2})^{4/5}) + b_3 Log(L/.57N) + e(t).$$

.hours-worked equation:

$$(6)\ Logh_t = C_0 + C_1 Log(w_{t-1}/p_{t-1}h_{t-1}) + C_2 Log(L_t/.57N_t) + e(t).$$

.depreciation equation:

$$(7)\ LogD_t = d_0 + d_1 LogK_{t-1} + e(t).$$

identities:

$$(8)\ Y_t = C_t + I_t - D_t + B_t$$

$$(9)\ K_t = K_{t-1} + I_t - D_t$$

m)The OBE/BEA model [46]: quaterly model influenced by the Wharton model and developed by the Bureau of Economic Analysis (formely the Office of Business Economics) in the U.S. Department of Commerce. It includes three sectors: the output market for components of GNP; the labor market, for hours, wages rate, labor force, and labor income; and

prices, for price deflators for GNP components and the wage rate. The model is used for short-term forecasting and policy evaluation in the Department of Commerce and other government agencies, including the Council of Economic Advisers.

n)**The suits model [84]:** an expanded version of Klein-Goldberger model where variables of the model are replaced by first differences. It influenced the MQEM model.

o)**The Tinberger model [87]:** the first macroeconometric model of U.S business cycles in the period 1919-1932. It includes 50 equations of which 32 are stochastic and contains 50 endogenous and 14 exogenous variables.

p)**The Wharton model [34]:** is a "medium-size" macroeconometric model of the U.S. economy. It is quaterly model, involving variables and data defined over a three-month period, particularly useful for analyzing and forecasting short-term macroeconomic phenomena -especially the national income components and unemployment-, involves a greater degree of disaggregation, a better treatment of accounting identities, and a better integration of the monetary sector than the previous models. The original version of the model contains 118 variables, of which 76 are endogenous and 42 are exogenous. It consists of 76 equations, of which 47 are stochastic and 29 are identities.

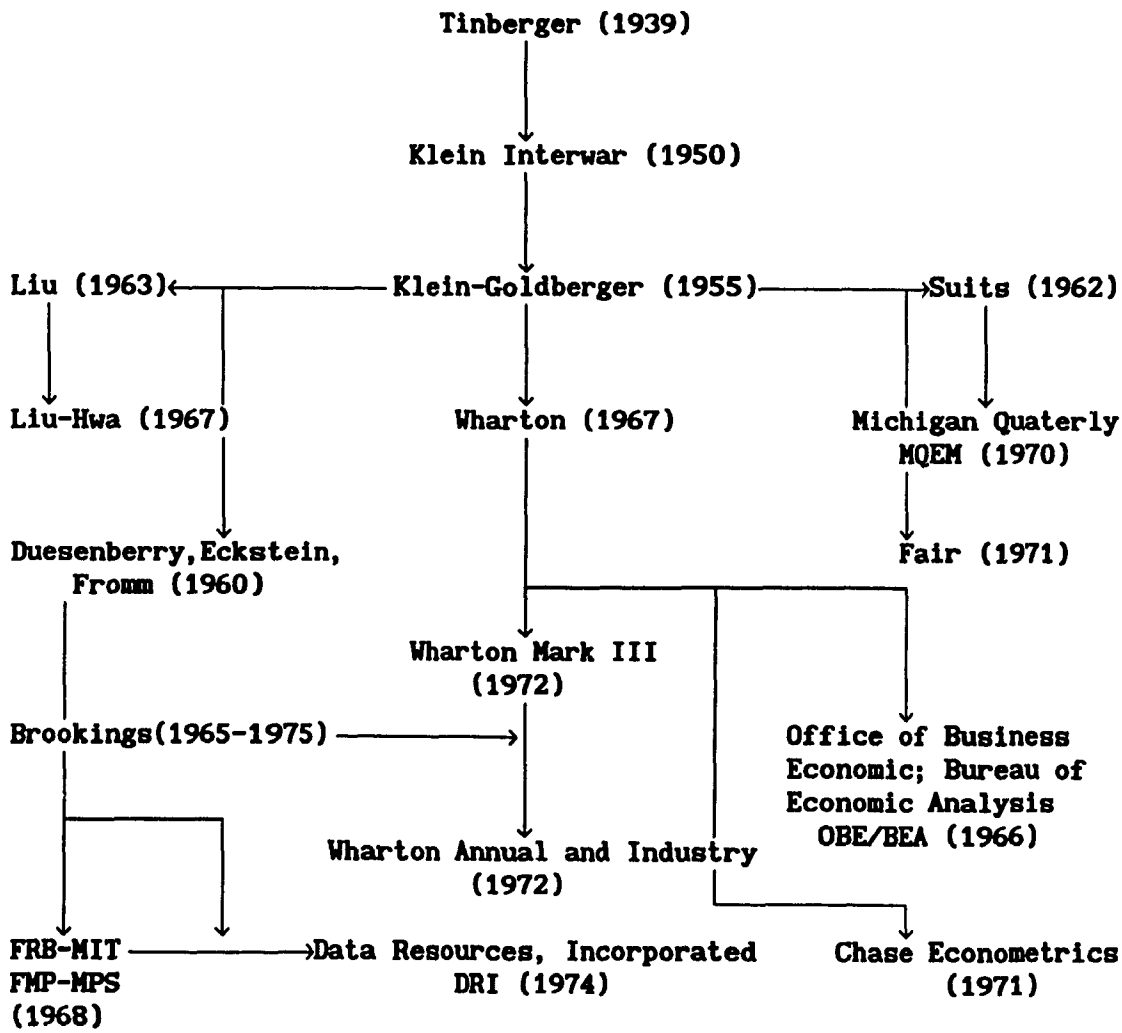
76 endogenous variables	42 exogenous variables
5 output	2 output
2 sales	income
4 income	consumption anticipat.
5 consumption	farm inventories
5 fixed investment	2 investment anticipation
4 depreciation	depreciation
exports	2 government purchases
3 imports	interest payments
2 corporate profits	2 social security contrib.
dividends	housing starts
retained earnings	population
cash flow	5 labor force
inventory valuation adjustment	2 wage bill
rent and interest payments	7 prices
3 taxes	discount rate
transfert payments	net free reserves
labor force	time
hours worked	6 dummy variables
wage bill	productivity trend
unemployment rate	index of world rate
capital stocks	statistical discrepancy
inventories	
unfilled orders	
index of capacity utilization	
10 prices	
2 wage rates	
2 interest rates	

q)**The Wharton Annual and Industry Forecasting model [73]**: a variant of the Wharton model which provides annual long-term forecasts of up to ten years on an industry basis. It utilizes input-output information,

and explicitly accounts for final demand, input-output, labor requirements, sector wage, sector price, final demand price, income payments, and financial factors.

r)The Wharton Mark III model [56]: first variant of Wharton model. This model contains 201 endogenous and 104 exogenous variables in 67 stochastic equations and 134 identities. It makes extensive use of distributed lag analysis investment and other areas involving 25 policy instruments with considerably more detailed treatment of both monetary variables and fiscal policy variables, particularly tax rates.

The connections between different macroeconometric models of the U.S, and other countries is illustrated by the following "family tree".



1.2 THE SIMULTANEOUS EQUATION MODEL

The structural equations may be linear or nonlinear. Here we consider only linear equations. However, the theoretical analysis of nonlinear equations is not straight forward.

In general, a system of M linear structural equations in M jointly dependent and K predetermined variables may be expressed in algebraic form as follows:

$$\left.\begin{array}{lllll} \gamma_{11}y_1(t) + \ldots + \gamma_{M1}y_M(t) & + \beta_{11}x_1(t) + \ldots & + \beta_{K1}x_K(t) & = u_1(t) \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \gamma_{1M}y_1(t) + \ldots + \gamma_{MM}y_M(t) & + \beta_{1M}x_1(t) + \ldots & + \beta_{KM}x_K(t) & = u_M(t) \end{array}\right\} \quad (1.1)$$

for $t = 1,\ldots,T$. Here y's are jointly dependent variables and x's are predetermined. We have assumed that T observations are available on each of these variables. The structural disturbances in successive equations are represented by $u_1(t),\ldots,u_M(t)$, and γ's and β's are the structural coefficients.

The system (1.1) is said to be <u>complete system</u> if there are as many equations as the number of jointly dependent variables to be explained. In matrix notation (1.1) may be written as:

$$Y_W\Gamma + XB = U \quad (1.2)$$

where

$$Y_W = \begin{pmatrix} y_1(1)\ldots\ldots y_M(1) \\ \vdots \qquad\qquad \vdots \\ y_1'(T)\ldots\ldots y_M'(T) \end{pmatrix} \quad (1.3)$$

and

$$X = \begin{pmatrix} x_1(1) \ldots\ldots x_K(1) \\ \vdots \qquad\qquad \vdots \\ x_1'(T) \ldots\ldots x_K'(T) \end{pmatrix} \tag{1.4}$$

are the matrices of observations on the jointly dependent and predetermined variables, respectively;

$$\Gamma = \begin{pmatrix} \gamma_{11} \ldots\ldots \gamma_{1M} \\ \vdots \qquad\qquad \vdots \\ \gamma_{MM}' \quad \ldots \gamma_{MM}' \end{pmatrix} \tag{1.5}$$

and

$$B = \begin{pmatrix} \beta_{11} \ldots\ldots \beta_{1M} \\ \vdots \qquad\qquad \vdots \\ \beta_{K1}' \ldots\ldots \beta_{KM}' \end{pmatrix} \tag{1.6}$$

are the matrices of structural coefficients and

$$U = \begin{pmatrix} u_1(1) \ldots\ldots u_M(1) \\ \vdots \qquad\qquad \vdots \\ u_1'(T) \ldots\ldots u_M'(T) \end{pmatrix} \tag{1.7}$$

is the matrix of structural disturbances.

1.3 THE REDUCED FORM

We may solve the structural equations for jointly dependent variables in terms of predetermined variables and write

$$\left.\begin{array}{l} y_1(t)=\pi_{11}x_1(t)+\ldots+\pi_{K1}x_K(t)+\bar{v}_1(t) \\ \ldots\ldots \quad \ldots\ldots\ldots \qquad \ldots\ldots \quad \ldots \\ y_M(t)=\pi_{1M}x_1(t)+\ldots+\pi_{KM}x_K(t)+\bar{v}_M(t) \end{array}\right\} \tag{1.8}$$

for $t = 1,\ldots,T$. This is the <u>reduced form</u> of the structural equations system. The coefficients π's are called the reduced form coefficients and disturbances $\bar{v}_i(t)$'s, are the reduced form disturbances.

The reduced form can be derived conveniently, in matrix form, by post multiplying both sides of (1.2) by Γ^{-1}, provided Γ is a non-singular matrix. In that case we write the reduced form as:

$$Y = X\Pi + V \tag{1.9}$$

where

$$\Pi_W = \begin{pmatrix} \pi_{11} & \cdots & \pi_{1M} \\ \vdots & & \vdots \\ \pi_{MM} & \cdots & \pi_{MM} \end{pmatrix} = -B\Gamma^{-1} \tag{1.10}$$

is the matrix of reduced form coefficients and

$$\bar{V}_W = \begin{pmatrix} \bar{v}_1(1) & \cdots & \bar{v}_M(1) \\ \vdots & & \vdots \\ \bar{v}_1(T) & \cdots & \bar{v}_M(T) \end{pmatrix} = U\Gamma^{-1} \tag{1.11}$$

is the matrix of reduced form disturbances.

<u>ASSUMPTIONS</u>

We assume that

i) the elements of X are nonstochastic and fixed in repeated samples.

ii) rank of X, $\rho(X)$, $= K < T$

iii) $\lim_{T\to\infty} \frac{1}{T}X'X = \Sigma_{X'X}$ is a positive definite matrix.

Finally, we make following assumptions about the structural disturbances:

iv) the M-dimensional row vectors of U are independently and identically distributed so that

$E(u_i(t)) = 0$ for all $i = 1,\ldots\ldots,M$ and all $t = 1,\ldots\ldots,T$ and

$$E(u_i(t)u_j(t')) = \sigma_{ij} \quad \text{if } t = t'$$
$$= 0 \quad \text{if } t \neq t'$$

or, alternatively,

iv)' the T rows of U are independently distributed according to an M-dimensional normal law with means, variances and covariances as defined in iv).

<u>Remarks</u>: 1) iv)' is required for the maximum likelihood estimation of structural equations and for finite sample analysis of the estimators. 2) iv) is enough for the least squares estimation and for asymptotic analysis of the estimators. In matrix notation we have

$$E(U) = 0 \text{ and } \frac{1}{T}E(U'U) = \Sigma = ((\sigma_{ij})) \tag{1.12}$$

where 0 is a TxM zero matrix and Σ is MxM positive definite covariance matrix of contemporaneous structural disturbances. If iv)' holds, then the T rows of $\bar{V}_W$ are also independently and identically normally distributed; and for both iv) and iv)' we have

$$E\bar{V}_W = 0 \quad \text{and} \quad \frac{1}{T}E(V'V_W) = \Omega = \Gamma'^{-1}\Sigma\Gamma^{-1} \tag{1.13}$$

where Ω is the variance-covariance matrix of the contemporaneous reduced form disturbances, because $\bar{V}_W = U\Gamma^{-1}$ as defined in (1.11)

1.4 A PRIORI RESTRICTIONS ON STRUCTURAL PARAMETERS

Before attempting to estimate any structural equation, we must

ensure that the equation is identifiable. That means, we must be able to distinguish between the structural equations. If both the demand and the supply equations, for example, are linear (or have the same functional form) and the same variables have been included in both, one cannot distinguish between them. In fact, in that case several linear combinations of the two will have the same form and variables, as the originally postulated demand and supply equations. If we are not able to distinguish between them there is no question of any statistical estimation of the coefficients of these equations. As we have noted in the previous section, each structural equation represents a certain economic hypothesis. Then it is natural that not all variables will be included in every equation. Let us illustrate this point with the help of Klein's six-equation model for the U.S economy, 1921-1941. The model consists of a consumption function

$$C_t = \alpha_0 + \alpha_1(W_{pt} + W_{Gt}) + \alpha_2\Pi_t + \alpha_3\Pi_{t-1} + u_1(t) \qquad (1.14)$$

which relates the aggregate consumption in year t, C_t, with the aggregate profits in year t, Π_t, the aggregate profits in year t-1, Π_{t-1}, the total wages paid by the private sector in year t, W_{pt}, the total wages paid by the private sector in year t-1, $W_{p(t-1)}$, and the structural disturbance $u_1(t)$. The second equation of the model is an investment function

$$I_t = \beta_0 + \beta_1\Pi_t + \beta_2\Pi_{t-1} + \beta_3K_{t-1} + u_{2t} \qquad (1.15)$$

which relates net investment (I_t) with Π_t and Π_{t-1} as defined above,

the capital stock in year t-1, K_{t-1}, and the structural disturbance term, u_{2t}. Next, there is a demand for labor equation

$$W_{pt} = \gamma_0 + \gamma_1(Y + T - W_G)_t + \gamma_2(Y + T - W_G)_{t-1} + \gamma_3(t-1931) + u_3(t) \tag{1.16}$$

relating W_{pt} (aggregate wages paid by the private sector in year t) with $(Y + T - W_G)_t$ its lagged value and the trend variable $\gamma_3(t-1931)$, where Y_t is the national income in year t, and T_t is "total taxes paid in year t".

Finally, the three identities of the model are

$$Y_t = C_t + I_t + G_t \tag{1.17}$$

$$\Pi_t = Y_t - W_{pt} + T_t \tag{1.18}$$

$$K_t = K_{t-1} + I_t \tag{1.19}$$

where G_t is the aggregate government expenditure.

In this model C_t, Π_t, W_{pt}, I_t, Y_t, and K_t are current endogenous or jointly dependent variables, and Π_{t-1}, K_{t-1}, Y_{t-1} are lagged endogenous. Further, W_{Gt}, T_t, G_t and t are current exogenous and $W_{G(t-1)}$, T_{t-1}, are lagged exogenous. There is also a dummy variable (associated with the intercept term in each equation) which assumes the value "1" always. Accordingly, M = 6 and K = 10. Since the number of structural equations in the model is the same as the number of jointly dependent variables, Klein model I is said to be a complete system. We

note that according to the specification of the consumption function, Π_t, Π_{t-1}, W_{pt}, and W_{Gt} are directly related with C_t while I_t, T_t, and G_t do not affect C_t directly. Structural equation being an economic hypothesis, we impose zero restrictions on coefficients of several variables which are not of any direct relevance in that equation.

Chapter 2

METHODS OF ESTIMATION

2.1 INTRODUCTION

There are mainly three alternative approaches to estimating the simultaneous-equations systems. They are: the naive approach, the limited-information approach, and the full-information approach. These approaches differ in the amount of information u ilized in the estimation process.

The naive approach to estimating parameters of a system of simultaneous equations is that of ordinary least squares. This approach applies least squares to each equation of the model separately, ignoring the distinction between explanatory exogenous and included endogenous variables. It also ignores all information available concerning variables not included in the equation. In general, naive approach leads to biased and inconsistent estimators.

The limited information approach estimates one equation at a time. Unlike ordinary least squares, it utilizes all identifying restrictions pertaining to the equation. The information required is limited, however, to the variables included in or excluded from the equation being estimated. The limited information approach includes several specific estimators, such as indirect least squares (ILS) and the two-stage least squares (2SLS) and k-class estimators.

The full information approach estimates the structural equations simultaneously in "one fell swoop", utilizing prior restrictions on parameters of the entire equation system. This approach includes two specific estimators, of which three-stage least squares (3SLS) and full information maximum likelihood (FIML).

It is hard to say immediately that one approach will be invariably better than the other in all cases. The full information method provides more efficient estimates. However, these estimates are sensitive to errors of specification; while limited information estimator is computationally and analytically more convenient.

A new direction to estimating structural parameters in simultaneous-equations econometric models was proposed by Herman O.Wold (1965 and 1969) and Ernest J. Mosback and Herman O. Wold (1970).-See [17] and references within-. They introduced the "fixed point" (FP) method of estimation. One of the difficulties with the FP method is that the fixed point estimator does not always exist and it may not be unique. The FP method is not considered in this work.

2.2 SINGLE EQUATION OF THE COMPLETE STRUCTURAL SYSTEM

Suppose $m + 1 < M$ jointly dependent and $k_1 < K$ predetermined variables enter the equation with non-zero coefficients; and further, the structural coefficients have been normalized by dividing the entire equation by the coefficient of one of the dependent variables. Then, using a priori restrictions, the structural equation, in terms of normalized coefficients, may be expressed as

$$y(t) = \sum_{i=1}^{m} \gamma_i y_i(t) + \sum_{j=1}^{k_1} \beta_j x_j(t) + u(t) \tag{2.1}$$

where the jointly dependent variable with "unit" coefficient is put on the left-hand side and all other jointly dependent and predetermined variables are transferred to the right. The coefficients γ_i's and β_j's are ratios of the original structural coefficients. In matrix notation, (2.1) may be written as :

$$\underset{T\times 1}{y} = \underset{T\times 1}{Y\gamma} + \underset{K_1\times 1}{X_1\beta} + \underset{T\times 1}{U} \tag{2.2}$$

where y is the column vector of jointly dependent variables with "unit" coefficient, and Y and X_1 are matrices of all other jointly dependent and predetermined variables included in the right hand side of the equation. γ and β are the coefficient vectors; and U is the vector of structural disturbances.

Now consider one just identified equation. The Y matrix can be partitioned into

$$Y_w = (y \mid Y_1 \mid Y_2) \tag{2.3}$$

corresponding to the one dependent endogenous variable y, the m explanatory endogenous variables Y_1 and the M-m-1 excluded jointly dependent variables Y_2. The X matrix can be similarly partitioned into

$$X = (X_1 \mid X_2) \tag{2.4}$$

corresponding to the K_1 included exogenous variables X_1 and the $K-K_1$ excluded predetermined variables X_2. The reduced form

$$Y_W = X\Pi_W + \bar{V}_W \tag{2.5}$$

of the complete structural system can be rewritten as

$$(y \mid Y_1 \mid Y_2) = (X_1 \mid X_2)\left(\begin{array}{c|c|c} \pi^* & \Pi^* & \Pi_2^* \\ \hline \pi & \Pi & \Pi_2 \end{array}\right) + (\bar{v} \mid \bar{V}_1 \mid \bar{V}_2) \tag{2.6}$$

where

$$K_2 = K - K_1 \tag{2.7}$$

and the stochastic disturbance terms have been partitionned in the same way as Y_W, corresponding to the dependent, explanatory, and excluded endogenous variables. The matrix of reduced-form coefficients Π_W has been partitioned here into six submatrices in order to carry out the matrix multiplication.

Suppose that (2.2) happens to be the first equation of the complete structural equation system. Then, only the first columns of Γ and B are involved. Using the normalization, zero restrictions, and the partitioned matrix, Π, in (2.6) we have

$$\left(\begin{array}{c|c|c} \pi^* & \Pi^* & \Pi_2^* \\ \hline \pi & \Pi & \Pi_2 \end{array}\right)\begin{pmatrix} 1 \\ -\gamma \\ 0_{M-m-1} \end{pmatrix} = -\begin{pmatrix} -\beta \\ \\ 0_{K_2} \end{pmatrix} \tag{2.8}$$

where

$$\begin{pmatrix} 1 \\ -\gamma \\ 0_{M-m-1} \end{pmatrix} \text{ and } \begin{pmatrix} -\beta \\ \\ 0_{K_2} \end{pmatrix}$$

are first columns of Γ and B respectively. 0_{M-m-1} and 0_{k_2} are columns vectors of M-m-1 and K_2 (= $K-K_1$) "zero" elements, respectively. Writing out the resulting two sets of equation, we obtain

$$\pi^* - \Pi^*\gamma = \beta \tag{2.9}$$

$$\pi - \Pi\gamma = 0_{K_2} \tag{2.10}$$

and it follows that

$$\bar{v} - \bar{V}\gamma = u \tag{2.11}$$

Hence

$$\frac{1}{\bar{T}}E(\bar{v}-\bar{V}\gamma)'(\bar{v}-\bar{V}\gamma) = \frac{1}{\bar{T}}E(u'u) = \sigma^2(\text{say}) \tag{2.12}$$

Now, γ can be determined uniquely from (2.10) iff

$$\rho(\Pi) = m \tag{2.13}$$

condition which implies

$$K_2 \geq m \qquad \text{(order condition)} \tag{2.14}$$

<u>Definition 1</u>: equation $y = Y\gamma + X_1\beta + u$ is said to be just identified or exactly identified if $\rho(\Pi) = m = K_2$.

<u>Definition 2</u>: equation $y = Y\gamma + X_1\beta + u$ is said to be over identified if $K_2 > m$.

Thus, in the case of just-identification, Π is a nonsingular square matrix; such that

$$\Pi\gamma = \pi \quad \text{or} \quad \gamma = \Pi^{-1}\pi \tag{2.15}$$

2.3 ORDINARY LEAST SQUARES AND LEAST SQUARES BIAS

The naive approach to estimating the parameters of a system of simultaneous equations is that of ordinary least squares (OLS). The equation to be estimated, the first of the complete structural equation system, can be written

$$y = Y\gamma + X_1\beta + u = (Y \vdots X_1)\begin{pmatrix}\gamma \\ \beta\end{pmatrix} + u = Z\delta + u \tag{2.16}$$

Here Z lumps together data on all $(m+K_1)$ included explanatory variables endogenous or exogenous.

$$Z = (Y \mid X_1) \tag{2.17}$$

and δ is a vector summarizing all coefficients to be estimated in the equation:

$$\delta = \begin{pmatrix} \gamma \\ \beta \end{pmatrix} \begin{matrix} m \\ K_1 \end{matrix} \tag{2.18}$$

Applying $\beta = (X'X)^{-1}X'y$ to (2.16) yields the estimator.

$$\hat{\delta} = (Z'Z)^{-1}Z'y \tag{2.19}$$

where the inverse exists if Z has rank $m+K_1$. In terms of the original notation, the OLS estimators can be written

$$\begin{pmatrix} \hat{\gamma} \\ \hat{\beta} \end{pmatrix} = \left(\begin{array}{c|c} Y'Y & Y'X_1 \\ \hline X_1'Y & X_1'X_1 \end{array}\right)^{-1} \begin{pmatrix} Y' \\ X_1' \end{pmatrix} y \tag{2.20}$$

In the case that (2.2) is the first equation of the system we have

$$\hat{\delta} = (Z'Z)^{-1}Z'(Z\delta + u) = \delta + (Z'Z)^{-1}Z'u \tag{2.21}$$

then, $E(\hat{\delta}) = \delta + E[(Z'Z)^{-1}Z'u]$. Since Z includes endogenous variables, Y, which are stochastic and not independent of the stochastic disturbance term, $E[(Z'Z)^{-1}Z'u]$ does not vanish. Thus, the OLS estimators are biased: $E(\hat{\delta}) \neq \delta$. The bias is given by

$$B(\hat{\delta}) = B\begin{pmatrix} \hat{\gamma} \\ \hat{\beta} \end{pmatrix}_{OLS} = E\left[\begin{pmatrix} \hat{\gamma} \\ \hat{\beta} \end{pmatrix} - \begin{pmatrix} \gamma \\ \beta \end{pmatrix}\right] = E\left[\left(\begin{array}{c|c} Y'Y & Y'X_1 \\ \hline X_1'Y & X_1'X_1 \end{array}\right) \begin{pmatrix} Y' \\ X_1' \end{pmatrix} u\right] \tag{2.22}$$

and this bias does not vanish, even asymptotically; that is the OLS estimators are also asymptotically biased.

$$\text{plim}(\delta) = \delta + \text{plim}(\frac{1}{n} Z'Z)^{-1}(\frac{1}{n}Z'u) \neq \delta \tag{2.23}$$

So, the OLS estimators are inconsistent. See [38] and [39].

2.4 INDIRECT LEAST SQUARES (ILS) ESTIMATOR OF PARAMETERS OF A JUST-IDENTIFIED STRUCTURAL EQUATION

Indirect least squares (ILS) is a limited-information technique that can be used to obtain consistent estimators of a just-identified equation. In the just-identified case, the structural parameters are uniquely determined from the reduced-form parameters, so the estimated reduced-form parameters can be used to infer estimated structural parameters indirectly, leading to the name "indirect least squares".

This approach involves two steps. The first is the estimation of the reduced-form parameters $\hat{\Pi}_W$ using least squares technique. Second step is possible iff the equation is just-identified and consists in the estimation of structural-form parameters $\hat{\Gamma}$ and $\hat{B}$ using the relationships between these parameters and the reduced-form parameters and the conditions of identifiability. (2.6) can be rewritten

$$y = X_1\pi^* + X_2\pi + \bar{v} \tag{2.24}$$

$$Y_1 = X_1\Pi^* + X_2\Pi + \bar{V} \tag{2.25}$$

$$Y_2 = X_1\Pi_2^* + X_2\Pi_2 + \bar{V}_2 \tag{2.26}$$

So that applying least squares to (2.24) and (2.25) we get

$$\begin{pmatrix}\hat{\pi}^* \\ \hat{\pi}\end{pmatrix} = (X'X)^{-1}X'y \quad \text{and} \quad \begin{pmatrix}\hat{\Pi}^* \\ \hat{\Pi}\end{pmatrix} = (X'X)^{-1}X'Y_1 \tag{2.27}$$

respectively. Here, $X = (X_1 \mid X_2)$ such that

$$\begin{pmatrix}\hat{\pi}^* \\ \hat{\pi}\end{pmatrix} = \left[(X_1 \mid X_2)'(X_1 \mid X_2)\right]^{-1}(X_1 \mid X_2)'y = \left[\begin{pmatrix}X_1' \\ X_2'\end{pmatrix}(X_1 \mid X_2)\right]^{-1}\begin{pmatrix}X_1' \\ X_2'\end{pmatrix}y =$$

$$\left(\begin{array}{c|c} X_1'X_1 & X_1'X_2 \\ \hline X_2'X_1 & X_2'X_2 \end{array}\right)^{-1}\left(\begin{array}{c} X_1' \\ \hline X_2' \end{array}\right)y$$

Now, the matrix $\left(\begin{array}{c|c} X_1'X_1 & X_1'X_2 \\ \hline X_2'X_1 & X_2'X_2 \end{array}\right)$ can be partitioned as follows:

$$\left(\begin{array}{c|c} X_1'X_1 & X_1'X_2 \\ \hline X_2'X_1 & X_2'X_2 \end{array}\right) = \left(\begin{array}{c|c} X_1'X_1 & 0 \\ \hline X_2'X_1 & X_2'X_2-(X_2'X_1)(X_1'X_1)^{-1}X_1'X_2 \end{array}\right)\left(\begin{array}{c|c} I & (X_1'X_1)^{-1}(X_1'X_2) \\ \hline 0 & I \end{array}\right)$$

such that

$$\left(\begin{array}{c|c} X_1'X_1 & X_1'X_2 \\ \hline X_2'X_1 & X_2'X_2 \end{array}\right)^{-1} = \left(\begin{array}{c|c} I & (X_1'X_1)^{-1}(X_1'X_2) \\ \hline 0 & I \end{array}\right)^{-1}$$

$$\times \left(\begin{array}{c|c} X_1'X_1 & 0 \\ \hline X_2'X_1 & X_2'X_2-(X_2'X_1)(X_1'X_1)^{-1}X_1'X_2 \end{array}\right)^{-1}$$

Therefore

$$\left(\begin{array}{c|c} X_1'X_1 & X_1'X_2 \\ \hline X_2'X_1 & X_2'X_2 \end{array}\right)^{-1} = \left(\begin{array}{c|c} I & -(X_1'X_1)^{-1}(X_1'X_2) \\ \hline 0 & I \end{array}\right) \times \left(\begin{array}{c|c} A & 0 \\ \hline & [X_2'X_2-(X_2'X_1)(X_1'X_1)^{-1}X_1'X_2]^{-1} \end{array}\right)$$

$$\times \left(\begin{array}{c|c} (X_1'X_1)^{-1}+(X_1'X_1)^{-1}(X_1'X_2)(X_2'M_1X_2)^{-1}X_2'X_1(X_1'X_1)^{-1} & \\ \hline -(X_2'M_1X_2)^{-1}X_2'X_1(X_1'X_1)^{-1} & B \end{array}\right\}$$

where

$$A \equiv \frac{(X_1'X_1)^{-1}}{-[X_2'X_2-X_2'X_1(X_1'X_1)^{-1}(X_1'X_2)]^{-1}X_2'X_1(X_1'X_1)^{-1}}$$

$$B \equiv \frac{-(X_1'X_1)^{-1}(X_1'X_2)(X_2'M_1X_2)^{-1}}{(X_2'M_1X_2)^{-1}}$$

and

$$M_1 = I - X_1'(X_1'X_1)^{-1}X_1'$$

Therefore

$$\left(\begin{array}{c|c} X_1'X_1 & X_1'X_2 \\ \hline X_2'X_1 & X_2'X_2 \end{array}\right)^{-1} \left(\begin{array}{c} X_1' \\ \hline X_2' \end{array}\right) y = \left\{ \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -(X_1'X_1)^{-1}(X_1'X_2) \\ I \end{bmatrix} \times \right.$$

$$\left. \left[X_2'(I - X_1'(X_1X_1)^{-1}X_1')X_2\right]^{-1} \times \begin{bmatrix} -(X_2'X_1)(X_1'X_1)^{-1} & I \end{bmatrix} \right\} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} y$$

$$= \left\{ \begin{pmatrix} (X_1'X_1)^{-1} + (X_1'X_1)^{-1}(X_1'X_2)(X_2'M_1X_2)^{-1}(X_2'X_1)(X_1'X_1)^{-1} & -(X_1'X_1)^{-1}(X_1'X_2)(X_2'M_1X_2)^{-1} \\ -(X_2'M_1X_2)^{-1}(X_2'X_1)(X_1'X_1)^{-1} & (X_2'M_1X_2)^{-1} \end{pmatrix} \right\} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} y$$

$$= \begin{bmatrix} (X_1'X_1)^{-1}X_1' + (X_1'X_1)^{-1}(X_1'X_2)(X_2'M_1X_2)^{-1}(X_2'X_1)(X_1'X_1)^{-1}X_1' & - X_1'X_1)^{-1}(X_1'X_2)(X_2'M_1X_2)^{-1}X_2' \\ -(X_2'M_1X_2)^{-1}(X_2'X_1)(X_1'X_1)^{-1}X_1' + & X_2'M_1X_2)^{-1}X_2' \end{bmatrix} y \tag{2.28}$$

The ILS estimators are generally biased estimators, as are the OLS estimators, but unlike the OLS estimators, ILS estimators are consistent, i.e:

$$\text{plim} \begin{pmatrix} \hat{\gamma} \\ \hat{\beta} \end{pmatrix}_{ILS} = \begin{pmatrix} \gamma \\ \beta \end{pmatrix} \tag{2.29}$$

The consistency of the ILS estimators follows from the fact that continuous functions of consistent estimators are also consistent estimators. The ILS estimators are obtained as continuous functions of the reduced-form estimators $\hat{\Pi}_W$ and the reduced-form estimators themselves are consistent from the least squares consistency theorem.

2.5 THE TWO - STAGE LEAST-SQUARES (2SLS) ESTIMATOR OF PARAMETERS IN A STRUCTURAL EQUATION

The technique of 2SLS, due to Theil [84], is a limited information technique that can be used to estimate either an overidentified or identified equation from a system of simultaneous equations. The difficulty in applying least squares directly to estimate (2.2) is, as mentioned, the presence of endogenous variables, Y, in the right hand side of (2.2), which are correlated with the structural disturbances,

u, even in the probability limit. The method of 2SLS overcomes this difficulty by using the estimated reduced form to replace Y by $\hat{Y}$. The least squares estimator of resulting equation is the two-stage least-squares (2SLS) estimator. The reduced form, for the sake of simplicity, is:

$$Y = X\hat{\bar{\Pi}} + \bar{V} \tag{2.30}$$

where $\bar{\Pi} = \begin{pmatrix} \Pi \\ \Pi \end{pmatrix}$ and X is as defined in (2.4). The least-squares estimators of the reduced-form coefficient $\hat{\bar{\Pi}}$ are given by $\hat{\bar{\Pi}} = (X'X)^{-1}X'Y$ as defined before. The estimates $\hat{\bar{\Pi}}$ of the endogenous variables $\hat{Y}$ are obtained from the estimated $\hat{\bar{\Pi}}$ and data on all exogenous variables of the model as

$$\hat{Y} = X\hat{\bar{\Pi}} = X(X'X)^{-1}X'Y \tag{2.31}$$

Replacing Y by $\hat{Y}$in (2.2) we get

$$y = \hat{Y}\gamma + X_1\beta + u \tag{2.32}$$

which can be written as

$$y = (\hat{Y} \vdots X_1)\begin{pmatrix} \gamma \\ \beta \end{pmatrix} + u \tag{2.33}$$

The 2SLS estimator is the least-squares estimator of (2.32) and given by

$$\begin{pmatrix} \hat{\gamma} \\ \hat{\beta} \end{pmatrix}_{2SLS} = \left[\begin{pmatrix} \hat{Y}' \\ X_1' \end{pmatrix}(\hat{Y} \quad X_1)\right]^{-1}\begin{pmatrix} \hat{Y}' \\ X_1' \end{pmatrix}y = \begin{pmatrix} \hat{Y}'\hat{Y} & \hat{Y}'X_1 \\ X_1'\hat{Y} & X_1'X_1 \end{pmatrix}^{-1}\begin{pmatrix} \hat{Y}'y \\ X_1'y \end{pmatrix} \tag{2.34}$$

using (2.31) we have

$$\hat{Y}'\hat{Y} = Y'X(X'X)^{-1}X'X(X'X)^{-1}X'Y = Y'X(X'X)^{-1}X'Y \tag{2.35}$$

and

$$X_1'\hat{Y} = X_1'X(X'X)^{-1}X'Y = X_1'Y \tag{2.36}$$

Therefore, we can express the 2SLS estimator as

$$\begin{pmatrix}\hat{\gamma}\\ \hat{\beta}\end{pmatrix}_{2SLS} = \begin{bmatrix} Y'X(X'X)^{-1}X'Y & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix}^{-1} \begin{pmatrix} Y'X(X'X)^{-1}X'y \\ X_1'y \end{pmatrix} \tag{2.37}$$

If we want to solve for $\hat{\gamma}_{2SLS}$ and $\hat{\beta}_{2SLS}$, let us write (2.37) as

$$\begin{bmatrix} Y'X(X'X)^{-1}X'Y & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix} \begin{pmatrix}\hat{\gamma}\\ \hat{\beta}\end{pmatrix}_{2SLS} = \begin{pmatrix} Y'X(X'X)^{-1}X'y \\ X_1'y \end{pmatrix} \tag{2.38}$$

or

$$\begin{cases} Y'X(X'X)^{-1}X'\hat{\gamma}_{2SLS} + Y'X_1\beta_{2SLS} = Y'X(X'X)^{-1}X'y & (2.39) \\ X_1'Y\,\gamma_{2SLS} + X_1'X_1\hat{\beta}_{2SLS} = X_1'y & (2.40) \end{cases}$$

From (2.40) we obtain

$$\hat{\beta}_{2SLS} = (X_1'X_1)^{-1}(X_1'y - X_1'Y\hat{\gamma}_{2SLS}) = (X_1'X_1)^{-1}X_1'(y - Y\hat{\gamma}_{2SLS}) \tag{2.41}$$

and substituting this in (2.39) we get

$$\hat{\gamma}_{2SLS} = (Y'NY)^{-1}Y'Ny \tag{2.42}$$

where

$$N = X(X'X)^{-1}X' - X_1(X_1'X_1)^{-1}X_1' \tag{2.43}$$

is an idempotent symmetric matrix ;

$$\rho(N) = \mathrm{tr}N = K - K_1 = K_2 \tag{2.44}$$

The OLS estimator of $\bar{V}$ in (2.30) is

$$V = Y - X\hat{\bar{\Pi}} = Y - X(X'X)^{-1}X'Y \tag{2.45}$$

and therefore,

$$Y'X(X'X)^{-1}X'Y = Y'Y-V'V \tag{2.46}$$

Hence, we may express the 2SLS estimator given in (2.37) as

$$\begin{pmatrix}\hat{\gamma}\\ \hat{\beta}_1\end{pmatrix}_{2SLS} = \begin{pmatrix}Y'Y - V'V & Y'X_1\\ X_1'Y_1 & X_1'X\end{pmatrix}\begin{pmatrix}Y' & - V'\\ & X'\end{pmatrix}y \tag{2.47}$$

There is a more elegant way, du to Basmann [11], [12], and [13] to derive the 2SLS estimator. Premultiplying the structural equation (2.2) by X′ leads to

$$X'y = X'Y\gamma + X'X_1\beta + X'u \tag{2.48}$$

$$X'y = (X'Y \quad X'X_1)\begin{pmatrix}\gamma\\ \beta\end{pmatrix} + X'u \tag{2.49}$$

Then applying GLS (General Least-Squares) to (2.49) we get

$$\begin{pmatrix}\hat{\gamma}\\ \hat{\beta}\end{pmatrix}_{GLS} = \left[\begin{pmatrix}Y'X\\ X_1'X\end{pmatrix}\left\{\sigma^2X'X\right\}^{-1}(X'Y \quad X'X_1)\right]^{-1}\begin{pmatrix}Y'X\\ X_1'X\end{pmatrix}\left\{\sigma^2X'X\right\}^{-1}X'y$$

$$= \left[\begin{array}{c|c} Y'X(X'X)^{-1}X'Y & Y'X(X'X)^{-1}X'X_1\\ \hline X_1'X(X'X)^{-1}X'Y & X_1'X(X'X)^{-1}X'X_1\end{array}\right]^{-1}\begin{pmatrix}Y'X(X'X)^{-1}X'y\\ X_1'X(X'X)^{-1}X\,y\end{pmatrix} \tag{2.50}$$

where

$$\sigma^2X'X = X'E(uu')X \tag{2.51}$$

Further, if we write

$$X_1 = (X_1 \;\vdots\; X_2)\left(\frac{I}{0}\right) = X\left(\frac{I}{0}\right) \tag{2.52}$$

where I is a $K_1 \times K_1$ unit matrix and 0 is $K_2 \times K_1$ zero matrix, then the GLS

estimator in (2.50) can be expressed as

$$\begin{pmatrix}\hat{\gamma}\\ \hat{\beta}\end{pmatrix}_{GLS} = \begin{pmatrix} Y'X(X'X)^{-1}X'Y & Y'X_1 \\ X_1'Y & X_1'X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y'X(X'X)^{-1}X'y \\ X_1'y \end{pmatrix} \tag{2.53}$$

which is identical with the 2SLS estimator defined previously. In the case that (2.2) is just identified and we have

$$K - K_1 = K_2 = m \tag{2.54}$$

the matrix $X'YX'X_1$ is a square matrix; and it follows that

$$\begin{pmatrix}\hat{\gamma}\\ \hat{\beta}\end{pmatrix}_{GLS} = \begin{pmatrix}\hat{\gamma}\\ \hat{\beta}\end{pmatrix}_{2SLS} = \left(X'Y \vdots X'X_1\right)^{-1}\left\{\sigma^2 X'X\right\}\begin{pmatrix}Y'X\\ X_1'X\end{pmatrix}^{-1}\begin{pmatrix}Y'X\\ X_1'X\end{pmatrix}\left\{\sigma^2 X'X\right\}^{-1} X'y$$

$$= \left(X'Y \vdots X'X_1\right)^{-1} X'y \tag{2.55}$$

writing $X = (X_1 \quad X_2)$ in (2.55) and premultiplying both sides of that equation by $(X'Y \quad X'X_1)$ we get

$$\left(\begin{array}{c|c} X_1'Y & X_1'X_1 \\ \hline X_2'Y & X_2'X_1 \end{array}\right)\begin{pmatrix}\hat{\gamma}\\ \hat{\beta}\end{pmatrix}_{2SLS} = \begin{pmatrix} X_1'y \\ X_2'y \end{pmatrix} \tag{2.56}$$

or

$$\begin{cases} X_1'Y\hat{\gamma} + X_1'X_1\hat{\beta} = X_1'y & (2.57) \\ X_2'Y\hat{\gamma} + X_2'X_1\hat{\beta} = X_2'y & (2.58) \end{cases}$$

From (2.57) one obtains

$$\hat{\beta}_{2SLS} = (X_1'X_1)^{-1}(X_1'y - X_1'Y\,\hat{\gamma}_{2SLS}) = (X_1'X_1)^{-1}X_1'(y - Y\,\hat{\gamma}_{2SLS}) \tag{2.59}$$

Substituting this in (2.58) we get

$$X_2'Y\hat{\gamma}_{2SLS} + X_2'X_1(X_1'X_1)^{-1}X_1'(y - Y\hat{\gamma}_{2SLS}) = X_2'y \tag{2.60}$$

Which gives

$$X_2'M_1Y\hat{\gamma}_{2SLS} = X_2'M_1y \tag{2.61}$$

where

$$M_1 = I - X_1(X_1'X_1)^{-1}X_1 \tag{2.63}$$

This shows that the 2SLS estimator reduces to the ILS estimator in the just-identified case.

It should be noted that the 2SLS estimator is asymptotically efficient within the class of all estimators that use the same a priori restrictions for a single equation, but it is not asymptotically efficient relative to the full-information technique of three-stage least squares.

2.6 THREE- STAGE LEAST SQUARES

The technique of three-stage least-squares (3SLS) is a full-information estimation technique which estimates all parameters of the structural equations simultaneously. The 3SLS estimator was developed by Zellner and Theil (1962) [88]. It is an extension of 2SLS and, consequently, an extension of the Basmann's GLS approach. The first stage of 3SLS consists in estimating all reduced-form coefficients using the least squares estimator. The second stage is the estimation of all structural coefficients by applying 2SLS to each of the structural equations. Finally the third stage generalizes least-squares estimation of all of the structural coefficients of the system, using a covariance matrix for the structural disturbances of

the equations that is estimated from the second-stage residuals. This use of covariance matrix makes 3SLS asymptotically more efficient than 2SLS. 3SLS technique is consistent.

The formulation of 3SLS requires the star notation. The j-th equation of the system can be written as

$$y_j = Y_j\gamma_j + X_j\beta_j + u_j \quad ; \qquad j = 1,\ldots,M \tag{2.63}$$

or

$$y_j = (Y_j \mid X_j) \begin{pmatrix} \gamma_j \\ \hline \beta_j \end{pmatrix} + u_j \tag{2.64}$$

Let δ_j summarizes all the coefficients to be estimated in the equation. Thus (2.64) becomes

$$y_j = Z_j\delta_j + u_j \quad ; j=1,\ldots,M \tag{2.65}$$

where

$$Z_j = (Y_j \mid X_j) \tag{2.66}$$

Assume all identities are eliminated and all equations are either just-identified or overidentified. Using star notation, we have

$$\underset{MT\times 1}{y^*} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} \qquad \underset{MT\times 1}{u^*} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{pmatrix} \tag{2.67}$$

Similarly, the M vectors of coefficients are stacked to give column vector of K^* coefficients:

$$\underset{K^*\times 1}{\delta^*} = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{pmatrix} \tag{2.68}$$

where $K^* = \sum_{j=1}^{M} (M_j - 1 + K_j)$ is the total number of coefficients to be estimated. Therefore, in star notation, all M equations of the system can be written

$$\underset{MTx1}{y^*} = \underset{MTxK^*}{Z^*} \underset{K^*x1}{\delta^*} + \underset{MTx1}{u^*} \tag{2.69}$$

where

$$Z = \left(\begin{array}{c|c|c|c} Z_1 & 0 & \cdots & 0 \\ \hline 0 & Z_2 & \cdots & 0 \\ \hline & & \ddots & \\ \hline 0 & 0 & 0 & Z_M \end{array}\right) = \left(\begin{array}{c|c|c|c|c} Y_1 & X_1 & 0 & \cdots & 0 \\ \hline 0 & & Y_2 & X_2 & \cdots & 0 \\ \hline & & & & \ddots & \\ \hline 0 & & 0 & & \cdots & Y_M \quad X_M \end{array}\right) \tag{2.70}$$

The data on explanatory variables in one equation is contained by a

matrix along the principal diagonal. It follows from iv) that

$$E(u^*) = 0 \tag{2.71}$$

$$E(u^* u^{*\prime}) = \begin{pmatrix} \sigma_{11} I & \sigma_{12} I \cdots\cdots \sigma_{1M} I \\ \sigma_{21} I & \sigma_{22} I \cdots\cdots \sigma_{2M} I \\ \sigma_{M1} I & \sigma_{M2} I \cdots\cdots \sigma_{MM} I \end{pmatrix} = \sum \otimes I \tag{2.72}$$

where $\sum \otimes I$ is the Kronecker product of these matrices.

The 3SLS estimator is a GLS estimator of (2.69) that considers the covariance matrix in (2.72). Let

$$X^{*\prime} = \begin{pmatrix} X' & 0 & \cdots & 0 \\ 0 & X' & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & X' \end{pmatrix} \tag{2.73}$$

Premultiplying (2.69) by $X^{*\prime}$ yields

$$X^{*\prime}y^{*} = X^{*\prime}Z^{*}\delta^{*} + X^{*\prime}u^{*} \tag{2.74}$$

$$\hat{\delta}^{*}_{3SLS} = \left\{Z^{*\prime}X^{*}[\mathrm{cov}(X^{*\prime}u^{*})]^{-1}X^{*\prime}Z^{*}\right\}^{-1} Z^{*\prime}X^{*}[\mathrm{cov}(X^{*}u^{*})]^{-1}X^{*\prime}y^{*} \tag{2.75}$$

It follows from (2.72) that

$$\mathrm{cov}(X^{*\prime}u^{*}) = X^{*\prime}\mathrm{cov}(u^{*})X^{*} = X^{*\prime}(\Sigma \otimes I)X^{*} \tag{2.76}$$

Thus the 3SLS estimator can be written

$$\hat{\delta}^{*}_{3SLS} = \left\{Z^{*\prime}X^{*}[X^{*\prime}(\Sigma^{*\prime}\otimes I)X^{*}]^{-1}X^{*\prime}Z^{*}\right\}^{-1} Z^{*\prime}X^{*}[X^{*\prime}(\Sigma^{*\prime}\otimes I)X^{*}]^{-1}X^{*\prime}y^{*} \tag{2.77}$$

From (2.77), the 3SLS estimator can be interpreted as taking all the 2SLS results and "correcting" them for the covariance matrix Σ.

2.7 LIMITED INFORMATION MAXIMUM LIKELIHOOD

The Limited-Information Maximum-Likelihood (LIML) estimator is obtained by maximizing the likelihood function for an individual equation subject only to the a priori restrictions imposed on the equation, without requiring information as to the specification of other equations of the system. The LIML technique can be used to estimate any just identified or over-identified equation, and, as in the case of 2SLS, it reduces to ILS in the just-identified case. The LIML is the only member of the k-class estimators which is invariant to the choice of which included endogenous variable is to be dependent variable.

The LIML estimator is asymptotically normally distributed and has the same limiting distribution as 2SLS in the case where the stochastic

terms are assumed to be normally distributed. Also the LIML has the asymptotically efficient property of minimum variance in the class of all estimators with the same a priori information.

The LIML technique is of historical importance. It has been used to estimate several major econometric models in the 1950's, but it has been forsaken in recent work because the findings in Monte Carlo studies show that the LIML as an estimator exhibits erratic and highly unstable behavior (see Theil [85]).

Consider the reduced form for all M jointly dependentvariables given by equation (1.9). Explicitly we have,

$$[y_1(t),\ldots,y_M(t)] = [x_1(t),\ldots,x_K(t)]\Pi_w + [u_1(t),\ldots,u_M(t)]\Gamma^{-1}$$

$$t = 1,\ldots,T \tag{2.78}$$

Assuming that the disturbance distribution is M-variate normal, means the conditional distribution of the row of $y_1(t),\ldots,y_M(t)$ given $x_1(t),\ldots,x_M(t)$ is also M-variate normal with mean equals to the row $[x_1(t)\ldots x_K(t)]\Pi_w$ and covariance matrix

$$\Gamma'^{-1}\sum \Gamma^{-1} = \Omega \qquad \text{(say)} \tag{2.79}$$

Let us assume that $\sum$, and consequently Ω, is non singular. The density function of the row $[y_1(t),\ldots,y_M(t)]$ is then equal to

$$(2\pi)^{-M/2}|\Omega|^{-\frac{1}{2}}\times\exp\left(-\frac{1}{2}\left\{\left([y_1(t),\ldots,y_M(t)]-[x_1(t)\ldots x_K(t)]\Pi_w\right)\Omega^{-1}\right.\right.$$

$$\times\left(\begin{bmatrix} y_1(t) \\ \cdot \\ y_M(t) \end{bmatrix} - \Pi'_w \begin{bmatrix} x_1(t) \\ \cdot \\ x_K(t) \end{bmatrix}\right)\Bigg)\Bigg)\Bigg) \tag{2.80}$$

$$= -\frac{1}{2}\mathrm{tr}\Omega^{-1}\left(\begin{bmatrix} y_1(t) \\ \cdot \\ y_M(t) \end{bmatrix} - \Pi'_w \begin{bmatrix} x_1(t) \\ \cdot \\ x_K(t) \end{bmatrix}\right)\left([y_1(t),\ldots,y_M(t)] - [x_1(t)..x_K(t)]\Pi_w\right) \tag{2.81}$$

Given that the rows of the matrix of structural disturbances are independent, the likelihood function for all observations is T times the density function given in (2.81). Hence the logarithmic likelihood is:

$$-\frac{1}{2}MT\log 2\pi - \frac{1}{2}T\log|\Omega| - \frac{1}{2}\mathrm{tr}\Omega^{-1}\sum_{t=1}^{T}\left(\begin{bmatrix} y_1(t) \\ \cdot \\ y_M(t) \end{bmatrix} - \Pi'_w \begin{bmatrix} x_1(t) \\ \cdot \\ x_K(t) \end{bmatrix}\right)$$

$$\times\left([y_1(t),\ldots,y_M(t)] - [x_1(t)..x_K(t)]\Pi_w\right)$$

$$= -\frac{1}{2}MT\log 2\pi - \frac{1}{2}T\log|\Omega| - \frac{1}{2}\mathrm{tr}\Omega^{-1}(Y_w - X\Pi_w)'(Y_w - X\Pi_w) \tag{2.82}$$

Then the maximum-likelihood estimator of Π is obtained by minimizing $\mathrm{tr}\Omega^{-1}(Y_w - X\Pi_w)'(Y_w - X\Pi_w)$. Consider an estimator

$$(X'X)^{-1}X'Y_w + A \tag{2.83}$$

where A is some KxM matrix. Then,

$$\mathrm{tr}\Omega^{-1}[Y - X(X'X)^{-1}X'Y - XA]'[Y - X(X'X)^{-1}X'Y - XA]$$

$$= \mathrm{tr}\Omega^{-1}Y'[I - X(X'X)^{-1}X'] + \mathrm{tr}\Omega^{-1}A'X'X\,A \tag{2.84}$$

Now $\mathrm{tr}\Omega^{-1}A'X'XA = \mathrm{tr}(XA)\Omega^{-1}(XA)'$ such that (2.93) is minimized for A = 0; that is the maximum-likelihood estimator of Π is $(X'X)^{-1}X'Y$.

Consider the jth structural equation for all observations:

$$y_j = Z_j\delta_j + u_j = [Y_j \vdots X_j]\begin{bmatrix}\gamma_j \\ \beta_j\end{bmatrix} + u_j \tag{2.85}$$

where $[y_j \quad Y_j]$ is a submatrix of Y_w and we may conclude that the distribution of each row of this submatrix is determined by the corresponding row of the right-hand side of

$$[y_j \quad Y_j] = [X_j \vdots X_j^*]\begin{bmatrix}\pi_j & \Pi_j \\ \pi_j^* & \Pi_j^*\end{bmatrix} + \bar{V}_j \tag{2.86}$$

where X_j^* is the $T\times(K-K_j)$ matrix of the values taken by the $K-K_j$ predetermined variables that are postulated to occur in the system but not in the jth equation. The column vectors π_j and π_j^* contain K_j and $K - K_j$ elements, respectively, Π_j is of order $K_j\times M_j$ and Π_j^* of order $(K-K_j)\times M_j$; they are all submatrices of the reduced-form parameters matrix $-B\Gamma^{-1}$.

Replacing in (2.82) M by $M_j + 1$, Ω by its principal matrix Ω_j of order $(M_j+ 1)\times(M_j+ 1)$ corresponding to $[y_j \quad Y_j]$ Y_w by $[y_j \quad Y_j]$, and Π_w by the partitioned matrix in (2.86) leads to

$$-\frac{1}{2}(M_j + 1)T\log 2\pi - \frac{1}{2}T\log|\Omega_j| - \frac{1}{2}\mathrm{tr}\Omega_j^{-1} D_j'D_j \tag{2.87}$$

where

$$D_j = [y_j \quad Y_j] - X\begin{bmatrix}\pi_j & \Pi_j \\ \pi_j^* & \Pi_j^*\end{bmatrix} \tag{2.88}$$

and follows that maximum-likelihood estimators of Π_j, Π_j^*, π_j, and π_j^* are submatrices of $(X'X)^{-1}X'Y$. Now, we know that

$$[\pi_j \ \Pi_j]\begin{bmatrix}1 \\ -\gamma_j\end{bmatrix} = \beta_j \text{ and } [\pi_j^* \quad \Pi_j^*]\begin{bmatrix}1 \\ -\gamma_j\end{bmatrix} = 0 \tag{2.89}$$

which gives

$$\pi_j^* = \Pi_j^*\gamma_j \tag{2.90}$$

Therefore, in the undetermined case no estimation is possible; and the just-identified case leads to the 2SLS estimator. When the equation is overidentified, (2.87) must be maximized subject to the constraint $\pi_j^* = \Pi_j^* \gamma_j$ using the Lagrange method. Lagrangian expression is then

$$\frac{1}{2}T\log|\Omega_j| + \frac{1}{2}tr\Omega_j^{-1}D_j'D_j - \lambda'(\pi_j^* - \Pi_j^*\gamma_j) \tag{2.91}$$

where λ is a $(K-K_j)$ element vector of lagrangian multipliers, and D_j the reduced-form disturbance matrix .

$$\frac{\partial\ \frac{1}{2}tr\Omega_j^{-1}D_j'D_j)}{\partial\ D_j} = D_j\Omega_j^{-1} \tag{2.92}$$

$$\frac{\partial\ (\frac{1}{2}tr\Omega_j^{-1}D_j'D_j)}{\partial \begin{bmatrix} \pi_j & \Pi_j \\ \pi_j^* & \Pi_j^* \end{bmatrix}} = -X'D_j\Omega_j^{-1} \tag{2.93}$$

$$\frac{\partial(\lambda'\pi_j^* - \lambda'\Pi_j^*\gamma_j)}{\partial \begin{bmatrix} \pi_j & \Pi_j \\ \pi_j^* & \Pi_j^* \end{bmatrix}} = \begin{bmatrix} 0 & 0 \\ \lambda & -\lambda\gamma_j'^2 \end{bmatrix} \tag{2.94}$$

$$\frac{\partial(2.91)}{\partial \begin{bmatrix} \pi_j & \Pi_j \\ \pi^* & \Pi^* \end{bmatrix}} = X'D_j\Omega_j^{-1} + \begin{bmatrix} 0 & 0 \\ \lambda & -\lambda\gamma_j'^2 \end{bmatrix} = 0 \tag{2.95}$$

$$\frac{\partial\ (2.91)}{\partial\lambda} = \pi_j^* - \Pi_j^*\gamma_j = 0 \tag{2.96}$$

$$\frac{\partial\ (2.91)}{\partial \gamma_j} = \Pi_j^{*\prime} \lambda = 0 \tag{2.97}$$

Let us suppose that the equation considered is the first of the system, that is $j = 1$. Introducing

$$Y_0 = [y_1 \quad Y_1] \begin{bmatrix} \Pi_0 \\ \Pi_0^* \end{bmatrix} = \begin{bmatrix} \pi_1 & \Pi_1 \\ \pi_1^* & \Pi_1^* \end{bmatrix} \text{ and } \gamma_0 = \begin{bmatrix} -1 \\ \gamma_1 \end{bmatrix} \tag{2.98}$$

the equation under consideration becomes

$$y_1 = Y_1 \gamma_1 + X_1 \beta_1 + u_1 \tag{2.99}$$

which takes the form

$$Y_0 \gamma_0 + X_1 \beta_1 + u_1 = 0 \tag{2.100}$$

Note that (2.100) is symetric in all $M_1 + 1$ jointly depend variables of the equation (2.99).

We need the following theorem due to Anderson and Rubin (1949-1950) [6].

<u>Theorem</u> (LIML). The estimator of the vector δ_j of equation (2.85) which maximizes the logarithmic likelihood function (2.87) subject to the constraints (2.89) is

$$\begin{bmatrix} Y_j' Y_j - \mu U_j' U_j & Y_j' X_j \\ X_j' Y_j & X_j' X_j \end{bmatrix}^{-1} \begin{bmatrix} Y_j' - \mu U_j' \\ X_j' \end{bmatrix} y_j \tag{2.101}$$

where μ is the smallest root of the determinantal equation

$$\left| \begin{bmatrix} y_j' \\ Y_j' \end{bmatrix} m_j \begin{bmatrix} y_j & Y_j \end{bmatrix} - \mu \begin{bmatrix} y_j' \\ Y_j' \end{bmatrix} m_j \begin{bmatrix} y_j & Y_j \end{bmatrix} \right| = 0 \tag{2.102}$$

where

$$m = I - X(X'X)^{-1}X' \quad \text{and} \quad m_j = I - X_j(X_j'X_j)^{-1}X_j' \tag{2.103}$$

which is real and larger than or equal to one. The vector (2.101) exists iff the equation (2.85) satisfies the rank condition for identification.

Returning to our problem, we realize that the determinantal equation is now reduced to

$$|Y_0'm_1Y_0 - \mu Y_0'mY_0| = 0 \tag{2.104}$$

D_j is now

$$D_1 = Y_0 - X_1\Pi_0 - X_1^*\Pi_0^* \tag{2.105}$$

where X_1^* is the observation matrix of the $K - K_1$ predetermined variables which are excluded form the first equation. Postmultiplying both sides of (2.95) (for $j = 1$) by Ω we get

$$X'D_1 + \begin{bmatrix} 0 & 0 \\ \lambda & -\lambda\gamma' \end{bmatrix}\Omega_1 = 0 \tag{2.106}$$

$$X'Y_0 - X'X_1\Pi_0 - X'X_1^*\Pi_0^* + \begin{bmatrix} 0 & 0 \\ \lambda & -\lambda\gamma' \end{bmatrix}\Omega_1 = 0 \tag{2.107}$$

$$X = [X_1 \quad X_1^*] \tag{2.108}$$

$$\begin{bmatrix} X_1'Y_0 \\ X_1^{*\prime}Y_0 \end{bmatrix} - \begin{bmatrix} X_1'X_1\Pi_0 \\ X_1^{*\prime}X_1\Pi_0 \end{bmatrix} - \begin{bmatrix} X_1'X_1^*\Pi_0^* \\ X_1^{*\prime}X_1^*\Pi_0^* \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \lambda\Omega_1 & -\lambda\gamma_1'\Omega_1 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2.109}$$

$$X_1'Y_0 - X_1'X_1\Pi_0 - X_1'X_1^*\Pi_0^* = 0 \tag{2.110}$$

$$X_1^{*\prime}Y_0 - X_1^{*\prime}X_1\Pi_0 - X_1^{*\prime}X_1^*\Pi_0^* - \lambda\gamma_0'\Omega_1 = 0 \tag{2.111}$$

Also

$$\lambda'(\pi_j^* - \Pi_j^* \gamma_j) \text{ is reduced to } -\lambda' \Pi_0^* \gamma_0 \tag{2.112}$$

Therefore (2.96) and (2.97) become

$$\Pi_0^* \gamma_0 = 0 \tag{2.113}$$

$$\Pi_0^{*\prime} \lambda = 0 \tag{2.114}$$

Replacing Π's by P's and γ_0 by C_0 we obtain

$$P_0 C_0 = 0 \tag{2.115}$$

and

$$P_0^{*\prime} \lambda = 0 \tag{2.116}$$

Considering equation (2.110) we can get P_0 as function of P_0

$$P_0 = (X_1' X_1)^{-1}(X_1' Y_0 - X_1' X_1^* P_0^*) = (X_1' X_1)^{-1} X_1' (Y_0 - X_1^* P_0^*) \tag{2.117}$$

Taking (2.117) into (2.111) leads to

$$X_1^{*\prime}(Y_0 - X_1^* P_0^*) - X_1^{*\prime} X_1^* P_0^* - \lambda \gamma_0' \Omega_1 = 0$$

$$X_1^{*\prime}(Y_0 - X_1^* P_0^*) - X_1^{*\prime} X_1 (X_1' X_1)^{-1} X_1' (Y_0 - X_1^* P_0^*) - \lambda C_0' \Omega_1 = 0$$

or

$$X_1^{*\prime}\left[I - X_1(X_1' X_1)^{-1} X_1'\right] Y_0 - X_1^*\left[I - X_1(X_1' X_1)^{-1} X_1'\right] X_1^* P_0^* - \lambda C_0' \Omega_1 = 0 \tag{2.118}$$

which is equivalent to

$$X_1^{*\prime} m_1 Y_0 - X_1^{*\prime} m_1 X_1^* P_0^* - \lambda C_0' \Omega_1 = 0 \tag{2.119}$$

and follows that

$$P_0^* = (X_1^{*\prime} m_1 X_1^*)^{-1}(X_1^{*\prime} m_1 Y_0 - \lambda C_0' \Omega_1) \tag{2.120}$$

Returning to (2.91), we have

$$\frac{1}{2}T\log|\Omega_1| + \frac{1}{2}\mathrm{tr}\Omega_1^{-1}D_1'D_1 - \lambda C_0'\Omega_1 = 0 \tag{2.121}$$

$$\frac{\partial(2.121)}{\partial\Omega_1^{-1}} = \frac{\partial(\frac{1}{2}\,T\log|\Omega_1|)}{\partial\Omega_1^{-1}} + \frac{\partial(\frac{1}{2}\,\mathrm{tr}\Omega_1^{-1}D_1'D_1)}{\partial\Omega_1^{-1}} = -\frac{1}{2}T\Omega_1 + \frac{1}{2}D_1'D_1 \tag{2.122}$$

therefore, the maximum likelihood estimator of Ω_1 is given by:

$$\hat{\Omega}_1 = \frac{1}{T}\hat{D}_1'\hat{D}_1 \tag{2.123}$$

where

$$\hat{D}_1 = Y_0 - X_1P_0 - X_1^*P_0^* = Y_0 - X_1(X_1'X_1)^{-1}X_1'(Y_0 - X_1^*P_0^*) - X_1^*P_0^*$$

$$= \left[I - X_1(X_1'X_1)^{-1}X_1'\right]Y_0 - \left[I - X_1(X_1'X_1)^{-1}X_1'\right]X_1^*P_0^* = m_1(Y_0 - X_1^*)P_0^* \tag{2.124}$$

Postmultiplying both sides of (2.120) by C_0 and using (2.115) and (2.116) we get :

$$\lambda = \frac{1}{C_0'\hat{\Omega}_1C_0}X_1^{*\prime} m_1 Y_0 C_0 \tag{2.125}$$

Taking this back into (2.120) leads to

$$P_0^* = (X_1^{*\prime} m_1 X_1^*)^{-1}(X_1^{*\prime} m_1 Y_0)(I - \frac{1}{C_0'\hat{\Omega}_1C_0}\,C_0C_0'\Omega_1) \tag{2.126}$$

then (2.124) becomes:

$$\hat{D}_1 = \left[m_1 - m_1 X_1^*(X_1^{*\prime} m_1 X_1^*)^{-1} X_1^{*\prime} m_1\right] Y_0 + \frac{1}{C_0' \hat{\Omega}_1 C_0} m_1 X_1^*(X_1^{*\prime} m_1 X_1^*)^{-1} X_1^{*\prime} m_1 Y_0 C_0 C_0' \hat{\Omega}_1 \quad (2.127)$$

Now the matices m, m_1 and $m_1 - m$ have interesting properties; especially $m^2 = m$, $m_1^2 = m$, $(m_1 - m)^2 = m_1 - m$ and $m(m_1 - m) = m_1(m_1 - m) - (m_1 - m)^2 = m_1 - m - (m_1 - m) = 0$. Hence, the expression for $\hat{\Omega}$ reduces to

$$\hat{\Omega}_1 = \frac{1}{T} Y_0' m Y_0 + \frac{C_0' Y_0' (m_1 - m) Y_0 C_0}{T(C_0' \hat{\Omega}_1 C_0)^2} \hat{\Omega}_1 C_0 C_0' \hat{\Omega}_1 \quad (2.128)$$

which gives

$$\hat{\Omega}_1 C_0 = \frac{1}{T} Y_0' m Y_0 C_0 + \frac{C_0' Y_0' (m_1 - m) Y_0 C_0}{T(C_0' \hat{\Omega}_1 C_0)^2} \hat{\Omega}_1 C_0 C_0' \hat{\Omega}_1 C_0$$

$$= \frac{1}{T} Y_0' m Y_0 C_0 + \frac{C_0' Y_0' (m_1 - m) Y_0 C_0}{T(C_0' \hat{\Omega}_1 C_0)} \hat{\Omega}_1 C_0 \quad (2.129)$$

and we see that $\Omega_1 C_0$ is of the form :

$$\hat{\Omega}_1 C_0 = \frac{1}{T(1-\alpha)} Y_0' m Y_0 C_0 \quad (2.130)$$

with

$$\alpha = \frac{C_0' Y_0' (m_1 - m) Y_0 C_0}{T(C_0' \hat{\Omega}_1 C_0)} \geq 0 \quad (2.131)$$

Now if we pre and post multiply (2.128) by C_0 and C respectively, we would obtain

$$C_0' \hat{\Omega}_1 C_0 = \frac{1}{T} C_0' Y_0' m Y_0 C_0 + \frac{C_0' C_0' Y_0' (m_1 - m) Y_0 C_0}{T\, C_0' \hat{\Omega}_1 C_0} \hat{\Omega}_1 C_0 = \frac{1}{T(1-\alpha)} C_0' Y_0' m Y_0 C_0 \quad (2.132)$$

Therefore,

$$\alpha = 1 - \frac{C_0'Y_0'm\,Y_0C_0}{C_0'Y_0'm_1Y_0C_0} \tag{2.133}$$

Then (2.133) and (2.131) ensure that:

$$0 \le \alpha \le 1 \tag{2.134}$$

Combining (2.128) and (2.130), one obtains:

$$\hat{\Omega}_1 = \frac{1}{T}Y_0'mY_0 + \frac{C_0'Y_0'(m_1-m)Y_0C_0}{T(C_0'\hat{\Omega}_1C_0)^2}\hat{\Omega}_1C_0C_0'\hat{\Omega}_1 = \frac{1}{T}Y_0'mY_0 + \frac{\alpha}{C_0'\hat{\Omega}_1C_0}\hat{\Omega}_1C_0C_0'\hat{\Omega}_1$$

$$= \frac{1}{T}Y_0'mY_0 + \frac{T\alpha(\alpha-1)}{C_0'Y_0'mY_0C_0}\hat{\Omega}_1C_0C_0'\hat{\Omega}_1 = \frac{1}{T}Y_0'mY_0 + \alpha\frac{Y_0'mY_0C_0(Y_0'mY_0'C_0)}{T(1-\alpha)C_0'Y_0'mY_0C_0} \tag{2.135}$$

Now it should be remembered that $P_0^{*\prime}\lambda = 0$. Applying (2.125) and (2.126) we get

$$\left(I - \frac{1}{C_0'\hat{\Omega}_1C_0}\hat{\Omega}_1C_0C_0'\right)Y_0'm_1X_1^*(X_1^*m_1X_1^*)^{-1}X_1^{*\prime}m_1Y_0C_0 = 0 \tag{2.136}$$

but we have

$$m_1 - m = X(X'X)^{-1}X' - X_1(X_1'X_1)^{-1}X_1' = m_1X_1^*(X_1^{*\prime}m_1X_1^*)^{-1}X_1^{*\prime}m_1 \tag{2.137}$$

Applying (2.137) to (2.136) leads to

$$\left[I - \frac{1}{C_0'\hat{\Omega}_1C_0}\hat{\Omega}_1C_0C_0'\right]Y_0'(m_1 - m)Y_0C_0 = 0 \tag{2.138}$$

using (2.130) and (2.132) we obtain

$$\frac{1}{C_0'\hat{\Omega}_1C_0}\hat{\Omega}_1C_0 = \frac{1}{C_0'Y_0'mY_0C_0}Y_0'mY_0C_0 \tag{2.139}$$

which is substituted into (2.138) to produce

$$Y_0'(m_1 - m)Y_0C_0 - \frac{C_0'Y_0'(m_1 - m)Y_0C_0}{C_0'Y_0'mY_0C_0} Y_0'mY_0C_0 = 0 \tag{2.140}$$

which can be written

$$Y_0'(m_1 - m)Y_0C_0 - \frac{C_0'Y_0' m_1Y_0C_0}{C_0'Y_0'mY_0C_0} Y_0'mY_0C_0 + \frac{C_0'Y_0'm Y_0C_0}{C_0'Y_0'mY_0C_0} Y_0'mY_0C_0 = 0$$

or

$$Y_0'm_1Y_0C_0 - \frac{C_0'Y_0' m_1Y_0C_0}{C_0'Y_0'mY_0C_0} Y_0'mY_0C_0 = 0$$

$$Y_0'm_1Y_0C_0 - \frac{1}{(1-\alpha)}Y_0'mY_0C_0 = 0$$

or

$$\left(Y_0'm_1Y_0 - \frac{1}{(1-\alpha)}Y_0'mY_0\right)C_0 = 0 \tag{2.141}$$

thus the matrix in parentheses must be singular and $\frac{1}{(1-\alpha)}$ corresponds to μ of the determinantal equation (2.104). Hence,

$$\mu = \frac{C_0'Y_0' m_1Y_0C_0}{C_0'Y_0'mY_0C_0} \tag{2.142}$$

Then the determinantal equation

$$|Y_0m_1Y_0 - (1-\alpha)^{-1}Y_0m Y_0| = 0 \tag{2.143}$$

has M_1+1 roots and the function to be minimized is: $\frac{1}{2}T\log|\Omega_1| + \frac{1}{2}$ $tr\Omega_1^{-1}D_1'D_1$. It follows from (2.123) that, for each of $M_1 + 1$ roots, $\frac{1}{2}tr\Omega_1^{-1}D_1'D_1 = \frac{T}{2}(M_1 + 1)$. That is, in order to minimize $\frac{1}{2}\log|\Omega_1| + \frac{1}{2}$ $tr\Omega_1^{-1}D_1'D_1$, we must minimize the determinant of (2.135) in which C_0 is determined only up to a multiplicative scalar. This provides a freedom to normalize C_0 so that $C_0'Y_0'mY_0C_0 = 1$. Let $\mu_2,\ldots,\mu_{M_1+1}$ be other roots of the determinantal equation and assume they are all distinct. Define $g_1 = C_0$, $g_2,\ldots,$ g_{M_1+1} for corresponding characteristic vectors, all normalized so that $g_i'Y_0'mY_0g_i = 1$. We have $g_i'Y_0'mY_0g_k = 0$ for $i\neq k$. Follows that

$$G'Y_0'mY_0G = I \tag{2.144}$$

where G is a square matrix of order M_1+1 whose ith column is g_i. Considering the normalisation and the fact that $\mu = (1-\alpha)^{-1}$ equation(2.135) for the kth root (μ_k) is :

$$T\hat{\Omega}_1 = Y_0'mY_0 + (\mu_k - 1)Y_0'mY_0g_k(Y_0'mY_0g_k)'$$

$$\Leftrightarrow TG'\hat{\Omega}_1G = I + (\mu_k-1)i_ki_k' \tag{2.145}$$

where i_k is the unit matrix of order $M_1 + 1$. The right-hand of matrix (2.145) is equal to a unit matrix except that the kth diagonal element is μ_k rather than 1. That is $|I + (\mu_k-1)i_ki_k'| = \mu_k$; follows that $|T\hat{\Omega}_1| = \mu_k|G|^{-2}$. Now, $G'Y'mYG = I$ such that

$$|G|^2|Y_0'mY_0| = 1 \tag{2.146}$$

therefore, in order to minimize $\frac{1}{2}T\mathrm{Log}|\Omega_1| + \frac{1}{2}\mathrm{tr}\Omega_1 D_1 D_1$, we should consider the smallest root of the determinantal equation.

2.8 FULL-INFORMATION MAXIMUM LIKELIHOOD

This approach maximizes the likelihood function for the entire system by choice of all structural parameters, subject to all a priori identifiability conditions ([81], [85]). Obtained estimators are consistent and asymptotically efficient. See [78]. The technique of full-information maximum likelihood (FIML) allows the use of a wide range of a priori information, pertaining not only to each individual equation but also to several equations simultaneously. However, the main disavantage of this technique resides in the difficulty and the cost of computation.

It is convenient to use the star notation in developing the FIML estimator. We have

$$\underset{MT\times 1}{y^*} = \underset{MT\times K^*}{Z^*}\ \underset{K^*\times 1}{\delta^*} + \underset{MT\times 1}{u^*} \tag{2.147}$$

Then we assume that

$$E(u^*) = 0 \tag{2.148}$$

$$\mathrm{cov}(u^*) = \Sigma\otimes I \tag{2.149}$$

$$u^* \sim N(0\ ,\Sigma\otimes I) \tag{2.150}$$

The logarithm of the likelihood function is given by

$$\ln L(u^*) = -\frac{TM}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma\otimes I| - \frac{1}{2}(y^* - Z^*\delta^*)'(\Sigma^{-1}\otimes I)(y^* - Z^*\delta^*) \qquad (2.151)$$

Letting L(y) denote the likelihood function for y , we have

$$\ln L(y^*) = \ln L(u^*) + \ln\left|\frac{\partial u^*}{\partial y^*}\right| \qquad (2.152)$$

where $\left|\frac{\partial}{\partial y^*}\right|$ is the determinant of the jacobian matrix of all first-order partial derivatives of elements of u^* with respect to elements of y^*. According to star notation,

$$\left|\frac{\partial u^*}{\partial y^*}\right| = \left|\begin{pmatrix} \Gamma & 0 & \cdots & \cdots & 0 \\ & \Gamma & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & \Gamma \end{pmatrix}\right| = |\Gamma|^T \qquad (2.153)$$

We can replace $-\frac{1}{2}\ln|\Sigma\otimes I|$ by $-\frac{T}{2}\ln|\Sigma|$ since

$$-\frac{1}{2}\ln|\Sigma\otimes I| = -\frac{1}{2}\ln|\Sigma|^T = -\frac{T}{2}\ln|\Sigma| \qquad (2.154)$$

Turns out that

$$\ln L(y^*) = -\frac{TM}{2}\ln 2\pi - \frac{T}{2}\ln|\hat{\Sigma}| + T\ln|\Gamma| - \frac{1}{2}(y^* - Z^*\delta^*)'(\Sigma^{-1}\otimes I)(y^* - Z^*\delta^*) \qquad (2.155)$$

And we note that

$$-\frac{1}{2}(y^* - Z^*\delta^*)'(\Sigma^{-1}\otimes I)(y^* - Z^*\delta^*) = -\frac{1}{2}\sum_{i=1}^{M}\sum_{i'=1}^{M}\sigma^{ii'}(y_i - Z_i\delta_i)'(y_{i'} - Z_{i'}\delta_{i'}) \qquad (2.156)$$

where

$$\Sigma = (\sigma_{ii'}) \quad \text{and} \quad \Sigma^{-1} = (\sigma^{ii'}) \qquad (2.157)$$

Now,

$$\frac{\partial \ln L(y^*)}{\partial \sigma^{ii'}} = \frac{T}{2}\sigma_{ii'} - \frac{1}{2}(y_i - Z_i\delta_i)'(y_{i'} - Z_{i'}\delta_{i'}) = 0 \qquad (2.158)$$

which leads to the FIML estimator of the elements of the covariance

matrix :

$$\hat{\Sigma}=(\hat{\sigma}_{ii'})$$

where

$$\hat{\sigma}_{ii'} = \frac{1}{T}(y_i - Z_i\hat{\delta}_i)'(y_{i'} - Z_{i'}\hat{\delta}_{i'}) \tag{2.159}$$

$$i, i' = 1,2,\ldots M$$

Therefore

$$\hat{\Sigma}\otimes I = \frac{1}{T}(y^* - Z^*\hat{\delta}^*)(y^* - Z^*\hat{\delta}^*)' \tag{2.160}$$

Taking (2.160) in (2.155) we get

$$\ln L(y^*) = -\frac{TM}{2}\ln 2\pi - \frac{T}{2}\ln|\hat{\Sigma}| + T\ln|\Gamma| - \frac{T}{2} \tag{2.161}$$

and in the unrestricted case,

$$\frac{\partial \ln L(y^*)}{\partial \delta^*} = -\frac{T\partial \ln|\hat{\Sigma}|}{2\,\partial \delta^*} + T\frac{\partial \ln|\Gamma|}{\partial \delta^*} = 0 \tag{2.162}$$

Since $|\Gamma|$ is a function of the coefficient of endogenous variables in all equations, the system of nonlinear equations in unknown coeffients is particularly awkward to solve. However, the difficulties of solving for FIML estimators vanish in the recursive case where Γ is triangular and Σ is diagonal. Thus in this case δ^* is estimated from

$$\frac{\partial \ln|\hat{\Sigma}|}{\partial \delta^*} = 0 \tag{2.163}$$

It follows from the fact that Σ is diagonal that

$$\ln|\hat{\Sigma}| = \sum_{i=1}^{M} \ln\hat{\sigma}_{ii'} = \sum_{i=1}^{M} \ln\frac{1}{T}(y_i - Z_i\hat{\delta}_i)'(y_i - Z_i\hat{\delta}_i)$$

$$= \sum_{i=1}^{M} \ln\frac{1}{T}(y_i'y_i - 2\hat{\delta}_i Z_i' y_i + \hat{\delta}_i Z_i' Z_i \hat{\delta}_i) \tag{2.164}$$

and

$$\hat{\delta}_i = (Z_i' Z_i)^{-1} Z_i' y_i \qquad i = 1,2,\ldots M$$

It should be realized that in the recursive case, the OLS estimators are also the FIML estimators and hence they are consistent and asymptotically efficient.

2.9 INSTRUMENTAL VARIABLE (IV) ESTIMATOR

The technique of instrumental variables (IV) is a general approach to estimating a single equation in a system of M linear structural equations. The IV estimator is extremely useful because it represents a whole class of estimators, each defined by F, the matrix of data on the instumental variables. See Sargan (1958) [79], Brundy and Jorgenson (1971, 1973, 1974) [18], and Madansky (1976) [49].

Consider the equation

$$y = Y\gamma + X_1\beta + u \tag{2.165}$$

can be expressed

$$y = Z\delta + u \tag{2.166}$$

where

$$\left.\begin{aligned} Z &= (\, Y \;\vdots\; X_1) \\ \delta &= \begin{pmatrix} \gamma \\ \beta \end{pmatrix} \end{aligned}\right\}$$

Premultiplying equation (2.166) by the transpose of a $Tx(m+K_1)$ transformation matrix F, yields

$$F'y = F'Z\,\delta + F'u \tag{2.167}$$

Suppose that

$$\text{i) } \rho(F) = m + K_1 < T \tag{2.168}$$

$$\text{ii) } \underset{T \to \infty}{\rho\lim} \frac{1}{T} F' u = 0 \tag{2.169}$$

$$\text{iii) } \underset{T \to \infty}{\rho\lim} \frac{1}{T} F' F = \sum_{FF} \text{ exists} \tag{2.170}$$

then

$$\hat{\delta}_{IV} = \begin{pmatrix} \hat{\gamma} \\ \hat{\beta} \end{pmatrix}_{IV} = ((F'Z)'F'Z)^{-1}(F'Z)'F'y = (F'Z)^{-1}F'y \tag{2.171}$$

is the IV of $\delta = \begin{pmatrix} \gamma \\ \beta \end{pmatrix}$. Let us choose

$$F = (Y-V \mid X_1) \tag{2.172}$$

where

$$V = Y - X(X'X)^{-1}X'Y \tag{2.173}$$

Equation (2.167) becomes

$$\begin{pmatrix} Y' - V' \\ X_1' \end{pmatrix} y = \begin{pmatrix} Y' - V' \\ X_1' \end{pmatrix} Z\delta + \begin{pmatrix} Y' - V' \\ X_1' \end{pmatrix} u \tag{2.174}$$

Then

$$\hat{\delta} = \left[\begin{pmatrix} Y' - V' \\ X_1' \end{pmatrix} Z \right]^{-1} \begin{pmatrix} Y' - V' \\ X_1' \end{pmatrix} y \tag{2.175}$$

It is instructive to note that

$$Y'V = Y'[Y - X(X'X)^{-1}X'Y] = Y'Y - Y'X(X'X)^{-1}X'Y \tag{2.176}$$

and

$$V'X = Y'X - Y'X(X'X)^{-1}X'X = Y'X - Y'X = 0 \tag{2.177}$$

It follows from (2.177) that

$$V'X_1 = 0 \tag{2.178}$$

Hence,

$$\hat{\delta} = \begin{pmatrix} Y'Y-V'V & Y'X_1 \\ X_1'Y & X_1'X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y' - V' \\ X_1' \end{pmatrix} y = \begin{pmatrix} Y'X(X'X)^{-1}X'Y & Y'X_1 \\ X_1'Y & X_1'X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y'X(X'X)^{-1}X'y \\ X_1'y \end{pmatrix} \tag{2.179}$$

is the same as the 2SLS estimator o[illegible] β) as given by (2.49). Thus, for this particular choice of F it [illegible]d that the 2SLS estimator of (γ β)′ can be interpreted as an IV esti[illegible]tor. It should be noted that IV estimators depend on the choice of instruments and the data on these instruments.

2.10 FAMILIES OF GENERAL k-CLASS, h-CLASS AND DOUBLE k-CLASS ESTIMATORS

The method is due to Henri Theil. Consider the OLS estimator in (2.20) and compare this with the 2SLS estimator in (2.47). Let define

$$V = Y - X(X'X)^{-1}X'Y \tag{2.180}$$

Then,

$$\begin{pmatrix} \hat{\gamma} \\ \hat{\beta} \end{pmatrix}_{2SLS} = \begin{bmatrix} Y'Y-V'VY'X_1 \\ X_1'Y \quad X_1'X_1 \end{bmatrix}^{-1} \begin{pmatrix} Y' - V' \\ X_1' \end{pmatrix} y \tag{2.181}$$

Turns out that the two estimators differ only in the leading matrices. And equation (2.178) is the OLS estimator of $\bar{V}$ in (2.30). So, alternatively, one can subtract only a part of V′V and a part of V′ from Y′Y and Y respectively. Therefore for arbitrary scalar k

(deterministe or stochastic) we have

$$\begin{pmatrix}\hat{\gamma}\\ \hat{\beta}\end{pmatrix}_k = \begin{bmatrix} Y'Y-kV'V & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix}^{-1} \begin{pmatrix} Y' - kV' \\ X_1' \end{pmatrix} y \tag{2.182}$$

as an estimator of $\begin{pmatrix}\gamma\\ \beta\end{pmatrix}$. By specifying different values for k one obtains different estimators. In particular for k = 0, (2.182) provides the OLS and for k = 1 we have the 2SLS estimator. This is expressed by saying that (2.182) provides the family of general k-class estimators of the structural parameters.

Observing that the 2SLS estimator was obtained by applying OLS to the structural equation after substituting $\hat{Y}$ for Y where

$$\hat{Y} = Y - V = X(X'X)^{-1}X'Y \tag{2.183}$$

instead replacing Y by Y - V we can consider Y - hV, where h is an arbitrary either determinist or stochastic scalar. Then

$$y = (Y - hV)\gamma + X_1\beta + u \tag{2.184}$$

is the new form of the structural equation. Applying OLS to (2.184) we get

$$\begin{pmatrix}\hat{\gamma}\\ \hat{\beta}\end{pmatrix}_h = \left[\begin{pmatrix} Y' - hV' \\ X_1' \end{pmatrix}(Y - hV \quad X_1)\right]^{-1} \begin{pmatrix} Y' - hV' \\ X_1' \end{pmatrix} y \tag{2.185}$$

and this is known as the family of h-class estimators and was also proposed by Henri Theil. For h = 0 we have the OLS and for h = 1 we obtain the 2SLS. Thus the OLS and the 2SLS estimators are said to belong to the family of h-class. The two families of general k-class and h-class can be merged into one family by defining the double

k-class estimator as follows:

$$\begin{pmatrix}\hat{\gamma}\\ \hat{\beta}\end{pmatrix}_{k_1k_2} = \begin{bmatrix} Y'Y - k_1V'V & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix}^{-1} \begin{pmatrix} Y' & - k_2V' \\ & X_1 \end{pmatrix} y \tag{2.186}$$

where k_1 and k_2 are either stochastic or non-stochastic arbitrary scalars. This was originally proposed by A.L. Nagar [60]. Again, for $k_1 = k_2 = 0$ we have OLS estimators while (2.186) provides 2SLS estimators for $k_1 = k_2 = 1$. Thus the OLS and 2SLS estimators are members of the family of double k-class estimators. If $k_1 = k_2 = k$ we have the family of general k-class estimators and $k_1 = 2h - h^2$ and $k_2 = h$ provides the family of h-class estimators.

Oi (1969) [62] had proved a mathematical identity between the k-class and 2SLS estimators. To this end he derived the k-class estimator by applying OLS to a structural equation after its jointly dependent variables have been transformed in a suitable manner. Also, Oi showed that the k-class estimator is a weighted average of the OLS and the 2SLS estimator. Following the method of Oi, Dhrymes (1969) [22] derived a similar relationship between the double k-class and the 2SLS estimator. However, what Dhrymes called the double k-class is only a subset of the entire family of double k-class. And its result is comparable with one obtained by Srivastava and Tiwari. (See [28] for reference).

Finally, let us mention that there is an other estimation process called Least Variance Ratio (LVR) method which provides estimators identical to the Limited Information Maximum Likelihood estimators.

Chapter 3

SURVEY OF SOME RECENT DEVELOPMENTS (1973-1988) ON THE DISTRIBUTION OF ECONOMETRIC ESTIMATORS

3.1 INTRODUCTION

We have seen earlier that the estimators satisfy only the consistency property. In order to make any meaningful inference about the parameters we need to know the sampling distribution of the estimators. The presence of endogenous variables makes it quite complicated. Several authors, in the past decade, have tried to find the distribution of the estimators. In this chapter we present a survey of the results available to date in this area for ready reference for some further work. It must also be mentioned that a lot of simulation work in this area is equally available; however, we have only concentrated on theoritical results. The exact distributions are also complicated; some times it is very hard to find the percentage points. In any case, the presentation in this chapter throws a light on the distributional aspects of the estimators.

Given that the expression derived for the distribution is quite complicated it is suggested that some approximations may be evolved which may prove to be useful for the inferential purposes. One such approximation is obtained by Mariano [51] for the exact distribution obtained by Sawa [83], and Richardson [75]. We will first describe the

exact distribution of estimators as available in the literature and then talk about their approximations so that they can be used in making inference on the coefficients of structural equations.

3.2 DEVELOPMENTS ON THE EXACT DISTRIBUTION THEORY

Richardson [75] found the exact distribution when only one endogenous variable was present on the right hand side of the equation. He considered the following equation:

$$y_1 = y_2\beta + Z_1\gamma_1 + Z_2\gamma_2 + e_1 \tag{3.1}$$

where y_1 and y_2 are Nx1 vectors of endogenous variables and Z_1 and Z_2 are, respectively, NxK_1 and NxK_2 matrices of exogenous variables where $K_2 \geq 2$. Also e_1 is a Nx1 vector of disturbances independently identically distributed (iid). The reduced form equations for y_1 and y_2 are

$$y_1 = Z_1\Pi_{11} + Z_2\Pi_{12} + \eta_1 \tag{3.2}$$

$$y_2 = Z_1\Pi_{21} + Z_2\Pi_{22} + \eta_2 \tag{3.3}$$

where Π_{11}, Π_{21} and Π_{21}, Π_{22} are column vectors of dimension K_1 and K_2 respectively. η_1 and η_2 are Nx1 column vectors of random disturbance terms whose ith components are independent of other components and distributed as bivariate normal with mean zero and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{21} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \tag{3.4}$$

Using 2SLS method; an estimator for β is derived, say v_1, then the probability density function (pdf) of v_1, $f(v_1)$, is

$$f(v_1) = \frac{1}{\beta(\frac{1}{2},\frac{n}{2})} \frac{e^{-(\mu^2/2)(1+\beta_1^2)}}{(1+v_1^2)^{(n+1)/2}} \sum_{j=0}^{\infty} \frac{(\frac{n+1}{2})_j \, (\frac{X^2}{2})^j}{(\frac{n}{2})_j \; j!} \, {}_1F_1(\tfrac{n-1}{2}, j+\tfrac{n}{2}; 2Z) \qquad (3.5)$$

where:

$$X^2 = \frac{\mu^2(1+\beta_1 v_1)^2}{1+v_1^2} \qquad (3.6)$$

$$2Z = \frac{\mu^2\beta_1^2}{2} \qquad (3.7)$$

$$\mu^2 = \Pi'_{22} - \frac{S}{\sigma_{22}}\Pi_{22} \qquad (3.8)$$

$$S = Z'_2Z_2 - Z'_2Z_1(Z'_1Z_1)^{-1}Z'_1Z_2 \qquad (3.9)$$

The parameter μ^2 is called the concentration parameter, $\beta(a,b)$ is the Beta function, and ${}_1F_1(a;\ b;\ x)$ is the confluent hypergeometric function.

Sawa [83] considered the case with two endogenous variables in a complete system of stochastic equation and worked out the exact sampling distribution of ordinary and two stages least squares estimators of a structural parameter from the equation:

$$y_{2t} = \alpha + \beta y_{1t} + u_t \quad , \; t = 1,\ldots,N \qquad (3.10)$$

where y_{1t} and y_{2t} are both endogenous variables. The number of exogenous variables in this system is finite and they are Z_{1t}, $Z_{2t},\ldots,Z_{Kt}$ $(t = 1,2,\ldots N)$. The system is assumed to include no lagged variables. The reduced form equations of y_{1t} and y_{2t} are given by:

$$y_{1t} = \pi_{10} + \sum_{j=0}^{K} \pi_{1j}Z_{jt} + v_{1t} \qquad (3.11)$$

$$y_{2t} = \pi_{20} + \sum_{j=0}^{K} \pi_{2j} Z_{jt} + v_{2t} \tag{3.12}$$

where π_{ij}'s are unknown constants and the disturbance terms (v_{1t}, v_{2t}) are supposed to be mutually iid according to the bivariate normal distribution with mean zero and non singular variance-covariance matrix

$$\Sigma = \|\sigma_{ij}\| = \|E(v_{it} v_{jt})\| \tag{3.13}$$

It is assumed that the sample size N (>2) is greater than the number of exogenous variables. Also, this system is more general than one by Basmann (see [83] for reference). Assuming the orthonormality of exogenous variables the sampling distribution of OLS estimator of the structural parameter β is as follows:

when $\beta \neq \rho$

$$dF = \frac{\sigma \xi^{N-1} e^{-\tau^2/2\sigma^2}}{\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{\Gamma(\frac{N}{2} + j)}{j!\,\Gamma((N-1)/2)+j)} \left\{ \sigma(\hat{\beta}-\rho) + \frac{\xi^2}{\sigma^2(\beta-\rho)} \right\}^{2j}$$

$$\times \left\{ \frac{1}{\sigma^2(\hat{\beta}-\rho)^2+\xi^2} \right\}^{(N/2)+1} \theta^j {}_1F_1(j+\tfrac{1}{2}, \tfrac{N-1}{2}+j; -\theta) d\hat{\beta} \qquad -\infty \le \hat{\beta} \le \infty \tag{3.14}$$

when $\beta = \rho$

$$dF = \frac{\sigma e^{-\tau^2/2\sigma^2}}{\sqrt{\pi}\,\xi} \sum_{j=0}^{\infty} \frac{\Gamma(\frac{N}{2} + j)}{j!\,\Gamma((N-1)/2)+j)} \left\{ \frac{\tau^2}{2\sigma^2} \right\}^{j} \left\{ \frac{\xi^2}{\sigma^2(\hat{\beta}-\beta)^2 + \xi^2} \right\}^{(N/2)+j} d\hat{\beta}$$

$$-\infty \le \hat{\beta} \le \infty \tag{3.15}$$

where

$$\xi^2 = \sigma \quad - \frac{\sigma_{12}^2}{\sigma_{11}} \tag{3.16}$$

$$\sigma^2 = \sigma_{11} \tag{3.17}$$

$$\rho = \frac{\sigma_{12}}{\sigma_{11}} \tag{3.18}$$

$$\theta = \frac{\tau^2(\beta-\rho)^2}{2\xi^2} \tag{3.19}$$

and ${}_1F_1(\ldots)$ is the confluent hyper-geometric function. The 2SLS estimator $\tilde{\beta}$ of the coefficient β may be obtained by applying the OLS method directly to the equation

$$y^*_{2j} = \beta y^*_{1j} + u^o_j \qquad , \; j = 1,\ldots,K \tag{3.20}$$

where

$$y^*_{ij} = \frac{1}{\sqrt{N}}\sum_{t=1}^{N}(Z_{jt}-Z_j)y_{it} = \sqrt{N}\,\pi_{ij} + \frac{1}{\sqrt{N}}\sum_{t=1}^{N}(Z_{jt}-Z_j)v_{it} \tag{3.21}$$

$$i = 1,2 \qquad \text{and} \quad j = 1,\ldots,K$$

Therefore the distribution of the 2SLS estimator given by

$$\tilde{\beta} = \frac{\sum_{j=1}^{K} y^*_{1j} y^*_{2j}}{\sum_{j=1}^{K} y^*_{1j}}$$

is essentially the same as that of the OLS where N is replaced by K + 1.

For the matrix representation (3.23a-b) of a class of structural stochastic economic hypotheses:

$$B'y'_{t.} + \Gamma'Z'_{t.} + e'_{t.} = 0 \tag{3.23a}$$

$$f(e_{t.}) = (2\pi)^{-G/2}|\Omega|^{-1/2}\exp\{-\tfrac{1}{2}e_{t.}\Omega^{-1}e'_{t.}\} \tag{3.23b}$$

where all matrices and vectors are real. B is square and nonsingular

with G rows; Γ has dimension KxG, Ω has G rows and is symmetric positive definite. The row vectors $y_{t.}$, $Z_{t.}$ and $e_{t.}$ have G, K, and G components respectively. The vectors $e_{t.}$ are iid for all t = 0, ±1, ±2,

Basmann et al. [16] supposed that the first equation of (3.23a) is:

$$-y_{t1} + \beta_1 y_{t2} + \beta_2 y_{t3} + \gamma_3 Z_{t3} + \ldots + \gamma_K Z_{tK} + e_{t1} = 0 \qquad (3.24)$$

and presented the exact finite joint sample density function of GCL estimators and its derivation. Under the standardizing transformation, (3.23a) becomes:

$$\bar{B}' y_{t.} + \bar{\Gamma}' Z_{t.} + \bar{e}'_{t.} = 0 \qquad (3.25)$$

The corresponding reduced form is:

$$\bar{\Pi} = (-\bar{\Gamma}\,\bar{B}^{-1})' \qquad (3.26)$$

$$\bar{\Sigma} = I_3 \qquad (3.27)$$

$$\sum_{t=1}^{N} \bar{Z}_{tj}\bar{Z}_{tk} = \delta_{jk} \qquad j,k = 1,2 \qquad (3.28)$$

where δ_{jk} is Kronecker's delta. Furthermore, the first equation of (3.25) is

$$-\bar{y}_1 + \bar{\beta}_1 \bar{y}_{t2} + \bar{\beta}_2 \bar{y}_{t3} + \bar{\gamma}_3 \bar{Z}_{t3} + \ldots + \bar{\gamma}_K \bar{Z}_{tK} + \bar{e}_{t1} = 0 \qquad (3.29)$$

It follows from (3.26) and (3.29) that

$$\bar{\pi}_{11} = \bar{\beta}_1 \bar{\pi}_{21} + \bar{\beta}_2 \bar{\pi}_{31} \qquad (3.30)$$

$$\bar{\pi}_{12} = \bar{\beta}_1 \bar{\pi}_{22} + \bar{\beta}_2 \bar{\pi}_{32} \qquad (3.31)$$

and

$$\bar{\pi}_{21}\bar{\pi}_{32} = \bar{\pi}_{31}\bar{\pi}_{22} \qquad (3.32)$$

By letting v_1 and v_2 designate the 2SLS estimators of $\bar{\beta}_1$, $\bar{\beta}_2$ respectively in (3.29). The joint probability density function of v_1 and v_2 is given by

$$f(v_1, v_2) = \frac{e^{\lambda^2/2}}{2\pi(1+v_1^2)^{3/2}(1+v_2^2)^{3/2}}$$

$$\sum_{n=0,k=0,m=0,j=0,p=0}^{\infty} \sum^{2j} \frac{(1/2)_j (1)_n (3/2)_{n+j+k} (3/2)_{n+j+m} 2^{2n+2j}}{(1)_{2n} (1)_{2n+2j-p+m+k} (3/2)_{n+j} k!\, m!}$$

$$\times \frac{(\bar{\mu}_2^2/2)^k (\bar{\mu}_3^2/2)^m (\bar{\mu}_2\bar{\mu}_3/2)^{2n+2j-p} (-)^p (v_1 v_2)^p (\cos\bar{\alpha})^{2j-p} (\sin\bar{\alpha})^{2n}}{(2j-p)!\, p!\, (1+v_1^2)^{n+j+k} (1+v_2^2)^{n+j+m}}$$

$$-\infty < v_i < \infty \qquad i = 1,2 \tag{3.33}$$

where

$$\lambda^2 = (1+\bar{\beta}_1^2)\mu_2^2 + (1+\bar{\beta}_2^2)\mu_3^2 + 2\bar{\beta}_1\bar{\beta}_2\mu_2\mu_3\cos\alpha \tag{3.34}$$

$$\bar{\mu}_2^2 = [(1+\bar{\beta}_1 v_1)\bar{\pi}_{21} + \bar{\beta}_2 v_1 \bar{\pi}_{31}]^2 + [(1+\bar{\beta}_1^2 v_1)\bar{\pi}_{22} + \bar{\beta}_2 v_1 \bar{\pi}_{32}]^2 \tag{3.35}$$

$$\bar{\mu}_3^2 = [(1+\bar{\beta}_2 v_2)\bar{\pi}_{31} + \bar{\beta}_1 v_2 \bar{\pi}_{21}]^2 + [(1+\bar{\beta}_2^2 v_2)\bar{\pi}_{32} + \bar{\beta}_1 v_2 \bar{\pi}_{22}]^2 \tag{3.36}$$

$$\bar{\mu}_2\bar{\mu}_3\cos\bar{\alpha} = [(1+\bar{\beta}_1 v_1)\bar{\pi}_{21} + \bar{\beta}_2 v_1 \bar{\pi}_{31}] x [(1+\bar{\beta}_2 v_2)\bar{\pi}_{31} + \bar{\beta}_1 v_2 \bar{\pi}_{21}] +$$

$$[(1+\bar{\beta}_1 v_1)\bar{\pi}_{22} + \bar{\beta}_2 v_1 \bar{\pi}_{32}] x [(1+\bar{\beta}_2 v_2)\bar{\pi}_{32} + \bar{\beta}_1 v_2 \bar{\pi}_{22}] \tag{3.37}$$

Then the joint and marginal density functions of $\hat{\beta}_1$ and $\hat{\beta}_2$ the 2SGCL estimators of β_1 and β_2 in (3.24) can be recovered with help of the inverse standardizing transformation.

Under the assumptions: that the equation being estimated contains two endogenous variables and an arbitrary number of exogenous variables; that the system contains no lagged endogenous variables, and

that the number of exogenous variables excluded from the equation as well as the number of equations in the system are arbitrary, Mariano and Sawa (1972) [54] derived the exact probability density function of the LIML estimator and then drew the important conclusion that for arbitrary values of the parameters in the model, the LIML estimator does not possess moments of order greater than or equal to one. They considered the structural equation

$$y_1 = y_2\beta + Z_1\gamma_1 + Z_2\gamma_2 + u \tag{3.38}$$

where y_1 and y_2 are Nx1 vectors of independent observations on the two included endogenous variables; Z_1 is an NxK_1 matrix of observations on K_1 exogenous variables; Z_2 is an NxK_2 matrix of N observations on K_2 (= $K - K_1$) exogenous variables; β, γ_1, and γ_2 are unknown structural parameters with β a scalar, γ_1 a K_1x1 vector and γ_2 a K_2x1 vector; and u is an Nx1 vector of disturbance terms. The associated reduced form is:

$$Y = Z_1\Pi_1 + Z_2\Pi_2 + V \tag{3.39}$$

where $Y = (y_1, y_2)$; $\Pi_1 = (\pi_{11}, \pi_{12})$ and $\Pi_2 = (\pi_{21}, \pi_{22})$ are, respectively, K_1x2 and K_2x2 matrices of unknown coefficients; V is an Nx2 matrix of random disturbance terms whose rows are assumed to be iid normal with mean vector 0 and 2x2 positive definite covariance matrix

$$\Omega = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \tag{3.40}$$

The exact distribution of the LIML estimator is given by the following theorem whose proof can be found in [54].

Theorem 1: For the case where the structural equation being estimated

contains two endogenous variables and K_1 exogenous variables such that $N-K_1$ is an even number greater than one, the probability density function of the LIML estimator $\hat{\beta}$ is

$$e^{-\frac{1}{2}\xi^2} \times \sum_{h=0}^{\frac{1}{2}(N-K_1-2)} \sum_{i,j=0}^{\infty} \delta(i,j,h)\xi^{2(i+j)} \times \frac{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)^{\frac{1}{2}(N-K_1-1)-h}}{(\sigma_{11} - 2\beta\sigma_{12} + \beta^2\sigma_{22})^{\frac{1}{2}(N-K_1)+i-1}}$$

$$\times \frac{[(\hat{\beta}-\beta)^2]^{\frac{1}{2}(N-K_1)-h-1} [\sigma_{11} - \sigma_{12}(\beta+\hat{\beta}) + \sigma_{22}\beta\hat{\beta}]^{2(h+i)}}{(\sigma_{11} - 2\sigma_{12}\hat{\beta} + \sigma_{22}\hat{\beta}^2)^{\frac{1}{2}(N-K_1)+i}} \tag{3.41}$$

$$\text{for} \quad -\infty < \hat{\beta} < \infty \ ;$$

where

$$\xi^2 = \frac{\mu^2(\sigma_{11} - 2\beta\sigma_{12} + \beta^2\sigma_{22})}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \tag{3.42}$$

$$\mu^2 = \pi_{22}' Z_2' (I_N - Z_1(Z_1'Z_1)^{-1}Z_1')Z_2\pi_{22} \tag{3.43}$$

$$\delta(i,j,h) = 2^{N-K_1-i-j} C_1 \binom{N-K_1-2}{2h} w(i,j) \times \frac{\Gamma(h+1/2)\Gamma[((N-K_1)/2)+i]}{\Gamma(i+1)\Gamma(j+i)}$$

$$\times \frac{\Gamma[((N-K_1-1)/2)+j-h]}{\Gamma((K_2/2)+i+j)} \tag{3.44}$$

with

$$w(i,j) = \frac{B(2p+3,\ j+q+i)}{(p+i)(p+2)} \times {}_3F_2(p+i,\ 2p+3,\ -i-q;\ 2p+j+q+4, p+3; 1) \tag{3.45}$$

and

$$p = \frac{1}{2}(N - K - 3) \text{ and } q = \frac{1}{2}(K_2 - 3) \tag{3.46}$$

Considering the model:

$$YB + Z_1\Gamma_1 + Z_2\Gamma_2 + E = 0 \tag{3.47}$$

where $\Gamma = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix}$ and $Z = [Z_1 \; Z_2]$ are defined in (3.23a-b), Basmann and Richardson (1973) [15], briefly, examined the exact finite sample properties of an asymptotically unbiased but nonconsistent estimator of a structural variance, $\bar{w}_{11}$, defined by

$$\bar{w}_{11} = \frac{G_1(\hat{\beta}_{.1}) - G_2(\hat{\beta}_{.1})}{\nu} \tag{3.48}$$

where the quadratic forms $G_1(\beta_{.1})$ and $G_2(\beta_{.1})$ are given by

$$G_1(\beta_{.1}) = \beta'_{.1}Y'[I_N - Z_1(Z'_1Z_1)^{-1}Z'_1]Y\beta_{.1} \tag{3.49}$$

$$G_2(\beta_{.1}) = \beta'_{.1}Y'[I_N - Z(Z'Z)^{-1}Z']Y\beta_{.1} \tag{3.50}$$

and

$$\nu = K_2 - G^{(1)} + 1 \quad \geq 0 \tag{3.51}$$

the GCL estimator $\hat{\beta}_{.1}$ in (3.48) is defined by

$$\frac{\partial}{\partial\beta_{.1}}[G_1(\beta_{.1}) - G_2(\beta_{.1})] = 0 \tag{3.52}$$

$$\beta \quad = -1 \tag{3.53}$$

$$\beta_{i1} = 0 \qquad (i = G^{(1)} + 1,\ G^{(1)} + 2, \ldots, G) \tag{3.54}$$

Since $\hat{\beta}_{.1}$ minimizes $Q(\beta_{.1})$, for every $\beta_{.1}$ that satisfies (3.47) and (3.54) it is shown that

$$\frac{\nu\bar{w}_{11}}{w_{11}} = \frac{G_1(\hat{\beta}_{.1}) - G_2(\hat{\beta}_{.1})}{w_{11}} \leq \frac{G_1(\beta_{.1}) - G_2(\beta_{.1})}{w_{11}} \tag{3.55}$$

Let

$$\Pi_2 = -\Gamma_2 B^{-1} \tag{3.56}$$

$$\Sigma = (B')^{-1}\Omega B^{-1} \tag{3.57}$$

The $K_2 \times 2$ matrix Π_{2*} is formed by the first two columns of Π_2. Furthermore Π_{22} is the second column of Π_{2*}. $\Pi_{22} \neq 0$. The concentration parameter associated with the first equation in the system (3.47) is

$$\bar{\mu}^2 = \frac{\Pi_{22}' Z_2' [I - Z_1 (Z_1' Z_1)^{-1} Z_1'] Z_2 \Pi_{22}}{\sigma_{22}} \tag{3.58}$$

β_1, V_1, V_2, V_3 are defined by

$$\beta_{22} = \frac{\sigma_{12}}{\sigma_{22}} + \frac{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)^{1/2}}{\sigma_{22}} \bar{\beta}_1 \tag{3.59}$$

$$\hat{\beta}_{21} = \frac{\sigma_{12}}{\sigma_{22}} + \frac{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)^{1/2}}{\sigma_{22}} V_1 \tag{3.60}$$

$$G_1(\hat{\beta}_{.1}) - G_2(\hat{\beta}_{.2}) = \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22}} V_2 \tag{3.61}$$

and

$$G_2(\hat{\beta}_{.1}) = \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22}} V_3 \tag{3.62}$$

As the transformation of variables (3.60)-(3.62) is one-to-one, expressions for the joint and marginal distribution functions of $\hat{\beta}_{12}$, $G_1(\hat{\beta}_{.1}) - G_2(\hat{\beta}_{.2})$, and $G_2(\hat{\beta}_{.1})$ may be recovered by mean of (3.59)-(3.62) as needed. The joint density of V_1 and V_2 is

$$g(V_1, V_2) = \frac{e^{-(\bar{\mu}^2/2) - 2z}}{\sqrt{\pi}(1+V_1^2)^{(\nu+2)/2}} \times \frac{\Gamma[(\nu+2)/2] e^{-(V_2^2/2)}}{\Gamma[(\nu+1)/2] 2\Gamma(\nu/2)} \times \left(\frac{V_2}{2}\right)^{(\nu/2)-1}$$

$$\times \sum_{j=0}^{\infty} \frac{[(\nu+2)/2]_j (\bar{\mu}^2/2)^j (1+\bar{\beta}_1 V_1)^{2j}}{[(\nu+1)/2]_j \, j! \, (1+V^2)^{2j}} \times {}_0F_1(-; \frac{\nu+1}{2} + j; zV_2),$$

$$-\infty < V_1 < \infty, \quad V_2 > 0 \quad (\nu = 1, 2, \ldots) \tag{3.63}$$

The marginal density function of V_2 is defined by

$$f_2(V_2) = e^{-(\bar{\mu}^2/2)-2Z} \frac{e^{-(V_2/2)}(V_2/2)^{(\nu/2)-1}}{2\Gamma(\nu/2)} \sum_{j=0}^{\infty} \frac{(\bar{\mu}^2/2)^j}{j!}$$

$$\times \sum_{k=0}^{\infty} \frac{(\frac{1}{2}+j)_k (2Z)^k}{[((\nu+1)/2)+j]_k k!} \times {}_0F_1(-;\frac{\nu+1}{2}+j+k;2ZV_2/2) \quad \text{for } V_2>0 \tag{3.64a}$$

$$= 0 \quad \text{otherwise} \tag{3.64b}$$

For $\bar{\beta}_1 = 0$ the density function $f_2(V_2)$ is that of a χ^2 variable with ν degrees of freedom. In concluding their paper, Basmann and Richardson showed that the distribution function $F_2(U)$ of the statistic

$$U = \frac{V_2}{1+\bar{\beta}_1^2} \tag{3.65}$$

$$= \frac{\nu\bar{W}_{11}}{W_{11}} \tag{3.66}$$

converges to the distribution of χ^2 with ν degrees of freedom as $\bar{\mu}^2 \rightarrow \infty$, sample size N being fixed for this convergence.

In a thick survey and appraisal, Basmmann (1974) [20] presented in a chronological order the developments of exact distributions of estimators since Haavelmo's 1947 article. Haavelmo in his paper did not present exact distribution function of any estimators; however the exact marginal distribution functions of maximum likelihood estimators for two of the structural constants in his first model can be determined by inspection and derivation of the marginal distribution function of the maximum likelihood estimator of the marginal propensity

to consume is quite straightforward. Basmann divided his review of 25 years in three periods corresponding to a classification of developments.

Carter (1976) [21] derived the exact distribution of the IV estimator when the instruments are non-stochastic. It is assumed that the equation of interest contains only two endogenous variables and is of the form:

$$y_1 = y_2\beta + X_1\gamma_1 + u_1 \tag{3.67}$$

where y_1 and y_2 are Tx1 vectors of endogenous variables and X_1 TxK_1 matrix of exogenous variables, β is an unknown scalar parameter, γ_1 is K_1x1 vector of unknown parameters and u_1 is a Tx1 vector of disturbances. Assume that two of the reduced form equations are

$$y_1 = X\pi_1 + v_1 \tag{3.68}$$

$$y_2 = X\pi_2 + v_2 \tag{3.69}$$

where $X = [X_1\ X_2]$ with X_2 a TxK_2 matrix of observations on K_2 exogenous variables which have been excluded, a priori, from (3.67). π_1 and π_2 are Kx1 vectors of parameters, v_1 and v_2 are Tx1 vectors of disturbances, and (3.67) is identified i.e., $K_2 \geq 1$. Furthermore the observations on v_1 and v_2 ($v_{1(t)}$ and $v_{2(t)}$) are independent and distributed according to normal distribution $N(0,\Omega)$ where contemporaneous covariance matrix $\Omega = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}$ is positive definite and constant over all t. Consider the set of non-stochastic instruments w and X_1 where w is a Tx1 vector which is a non-stochastic linear combination of the columns of X; w = Xp with p a non-stochastic Kx1

vector. The IV estimator of β is then given by

$$\hat{\beta} = \frac{w'M_1y_1}{w'M_1y_2} = \frac{z_1}{z_2} \tag{3.70}$$

where $M_1 = I_T - X_1(X_1'X_1)^{-1}X_1$. Follows that

$$f(\hat{\beta}) = \frac{1}{\pi a}\left\{\sigma_1\sigma_2\sqrt{1-\rho^2}\exp\left[\frac{-1}{2(1-\rho^2)}\left(\frac{\mu_2^2}{\sigma_2^2} - \frac{2\rho\mu_1\mu_2}{\sigma_1\sigma_2} + \frac{\mu_1^2}{\sigma_1^2}\right)\right] + \frac{b}{\sqrt{a}}\exp\left[\frac{-k^2}{2}\right]\right.$$

$$\left.\times\int_0^{\frac{b}{\sigma_1\sigma_2\sqrt{a(1-\rho^2)}}}\exp\left[-\frac{1}{2}t^2\right]dt\right\} \tag{3.71}$$

where

$$\sigma_1^2 = w_{11}w'M_1w \tag{3.72}$$

$$\sigma_2^2 = w_{22}w'M_1w \tag{3.73}$$

$$\sigma_{12} = w_{12}w'M_1w \tag{3.74}$$

$$\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2} \tag{3.75}$$

$$a = \sigma_1^2 - 2\hat{\beta}\sigma_{12} + \hat{\beta}^2\sigma_2^2 \tag{3.76}$$

$$b = (\mu_1\sigma_{12} - \mu_2\sigma_1^2) + (\mu_2\sigma_{12} - \mu_1\sigma_2^2)\hat{\beta} \tag{3.77}$$

$$k = \frac{\mu_1 - \mu_2\hat{\beta}}{\sqrt{a}} \tag{3.78}$$

$$t = \frac{az_2 - \rho\sigma_1\mu_1 + (\rho\sigma_1\mu_2 - \sigma_2\mu_1)\hat{\beta} - \mu_2\sigma_2\hat{\beta}^2}{\sigma\sqrt{a(1 - \rho^2)}} \tag{3.79}$$

Given that (3.71) has no moments of any order, the author applied, in the case where $\frac{\mu_2}{\sigma_2} > 3$, Geary's (see [21] for reference) approximation

$$f(\hat{\beta}) = \frac{-b}{\sqrt{2\pi a^3}} \exp\left[\frac{-k^2}{2}\right] \tag{3.80}$$

Turns out that one can approximate the distribution of -k by

$$f(\hat{\beta}) = \frac{-1}{\sqrt{2\pi}} \exp\left[\frac{-(-k^2)}{2}\right] \tag{3.81}$$

considering the fact that $\frac{d(-k)}{d\hat{\beta}} = \frac{-b}{a^{3/2}}$.

Phillips (1980a) [64] generalized the presently known results for single equation instrumental variable estimators in the simultaneous equations settings. His main result is the derivation of the exact pdf of instrumental variable estimators of the coefficient vector of the endogenous variables in a structural equation containing n + 1 endogenous variables and N degrees of overidentification. Consider a structural equation

$$y_1 = Y_2\beta + Z_1\gamma + u \tag{3.82}$$

where y_1(Tx1) and Y_2(Txn) are an observation vector and observation matrix, respectively, of n+1 included endogenous variables, Z_1 is a TxK_1 matrix of included exogenous variables, and u is a normally distributed disturbance with zero mean and covariance $\sigma^2 I$. Corresponding reduced form is given by

$$[y_1 \vdots Y_2] = [Z_1 \vdots Z_2]\begin{bmatrix}\pi_{11} & \Pi_{12}\\ \pi_{21} & \Pi_{22}\end{bmatrix} + [v_1 \vdots V_2] = Z\Pi + V \tag{3.83}$$

The rows of the reduced form disturbance matrix V are assumed to be iid, normal random vectors. It is also assumed that $K_2 \geq n$ and the matrix Π_{22} $(K_2 x n)$ in (3.83) has full rank so that (3.82) is identified. The parameter $N = K_2 - n$ measures the degree of overidentification. Let $H = [Z_1 \vdots Z_3]$, where $Z_3(TxK_3)$ is a submatrix of Z_2 and $K_3 \geq n$, be a matrix of instrumental variables to be used in the estimation of (3.82). Define

$$A = \begin{bmatrix}a_{11} & a'_{12}\\ a_{21} & A_{22}\end{bmatrix} = T^{-1}\begin{bmatrix}y'_1Z_3Z'_3y_1 & y'_1Z_3Z'_3Y_2\\ Y'_2Z_3Z'_3y_1 & Y'_2Z_3Z'_3Y_2\end{bmatrix} \tag{3.84}$$

and then $\beta_{IV} = A_{22}^{-1}a_{21}$ (β_{IV} denotes the instrumental variable estimator of the parameter vector β). After use of the inverse transform representation of ${}_0F_1$ function and completion of required integration, the obtained joint pdf of β_{IV} is given by

$$\text{pdf}(r) = \frac{\text{etr}\{-\frac{T}{2}(I+\beta\beta')\bar{\Pi}'_{22}\bar{\Pi}_{22}\}\Gamma_n(\frac{L+n+1}{2})}{\pi^{\frac{n}{2}}[\det(I+rr')]^{\frac{L+n+1}{2}}} \times\sum_{j=0}^{\infty}\frac{(\frac{1}{2})_j}{j!\,\Gamma_n((L+n)/2)+j)}$$

$$\times\Big[(\frac{T}{2}\beta'\bar{\Pi}'_{22}(\text{adj}\frac{\partial}{\partial W})\bar{\Pi}_{22}\beta)^j\times(\det(I+W))^{(L-1)/(2+j)}$$

$$\times{}_1F_1(\frac{L+n+1}{2},\frac{L+n}{2}+j;\frac{T}{2}(I+W)\bar{\Pi}_{22}(I+\beta r')(I+rr')^{-1}(I+r\beta')\bar{\Pi}'_{22})\Big]_{W=0} \tag{3.85}$$

where $r = A_{22}a_{21}$, $L = K_3 - n$, Π_{22} is an nxn nonsingular matrix and $w = a_{11} - r'A_{22}r$. If we set $L = N$ so that $Z_3 = Z_2$, β_{IV} is the 2SLS estimator and the matrix W is a scalar; therefore the pdf is given by (3.86). When $L = 0$ in (3.85), we have

$$\mathrm{pdf}(r) = \frac{\mathrm{etr}\{-\frac{T}{2}(I+\beta\beta')\bar{\Pi}'_{22}\bar{\Pi}_{22}\}\Gamma_n(\frac{n+1}{2})}{\pi^{\frac{n}{2}}[\det(I+rr')]^{\frac{n+1}{2}}\Gamma_n(\frac{n}{2})}$$

$$\times {}_1F_1(\frac{n+1}{2}, \frac{n}{2}+j; \frac{T}{2}\bar{\Pi}_{22}(I+\beta r')(I+rr')^{-1}(I+r\beta')\bar{\Pi}'_{22}) \qquad (3.86)$$

That is, a single term involving argument hypergeometric function as obtained by Sargan (see [64] for reference). This, also, generalizes the formula derived by Richardson [75].

As pointed out by Basmann (1974), detailed information on the small sample distribution of estimators can be used to help determine how accurate our data needs to be before we can confidently express a preference for one estimator or another in a given situation. This argument underlines our need for useful knowledge about the finite sample effets of misspecification; in this context, measurement error misspecification. Phillips (1980b) [65] discussed some of the results that have so far been obtained and the work that is under way in the area of mathematical study of the effets of specification error on sampling distributions in econometrics. He proposed to extend the range of realistic models to which comparative studies of econometric estimators refer and to develop techniques and associated computer software to enable an empirical researcher to extract information about the small sample behavior of various estimators and statistics he may be considering for use and to do so explicitly in the context of the sample size, the particular model specification and the exogenous series with which he may be working. In a stimulating quasi general survey he provided some background on the nature of the Edgeworth

approximation and associated expansions and an application of the theory to the distribution of OLS and 2SLS estimators of the coefficients in a simple consumption function that involves lagged consumption in the set of regressors. The paper is concentrated on the small sample distributions of the OLS and 2SLS of the coefficients of the kernel of many macroeconometric models:

$$C_t = \alpha Y_t + \beta C_{t-1} + u_t \qquad (t = \ldots, -1, 0, 1, 2, \ldots) \tag{3.87}$$

$$Y_t = C_t + I_t \qquad (t = \ldots, -1, 0, 1, 2, \ldots) \tag{3.88}$$

where the variables C_t, Y_t, and I_t represent consumption, income and investment, respectively. It is assumed that the distribances u_t are serially independent and identically distributed as $N(0, \sigma_u^2)$, and I_t is taken to be a non-random exogenous variable whose sample second moment converges tc a finite positive constant as the sample size tends to infinity. Writing $\delta = \beta/(1 - \alpha)$ and under the stability condition $|\delta| < 1$ Phillips (1980b) derived from (3.87) and (3.88) the final form equations

$$C_t = \left(\frac{\alpha}{1-\alpha}\right) \sum_{s=0}^{\infty} \delta^s I_{t-s} + \left(\frac{1}{1-\alpha}\right) \sum_{s=0}^{\infty} \delta^s u_{t-s} = m_t + w_t, \quad \text{say,} \tag{3.89}$$

and

$$Y_t = I_t + m_t + w_t .$$

Define Tx1 observation vectors $c' = (C_1, C_2, \ldots, C_T)$, $c'_{-1} = (C_0, C_1, C_2 \ldots, C_{T-1})$, $y' = (Y_1, Y_2, \ldots Y_T)$ and $d' = (I_1, I_2, \ldots, I_T)$ so that the OLS estimators α^* and β^* of α and β in (3.87) are defined by

$$\alpha^* = \frac{(c'_{-1}c_{-1})(c'c + d'c) - (c'c_{-1} + d'c_{-1})(c'_{-1}c)}{(c'c + 2d'c + d'd)(c'_{-1}c_{-1}) - (c'c_{-1} + d'c_{-1})^2} \tag{3.90}$$

and

$$\beta^* = \frac{(c'c_{-1})(c'c+2d'c + d'd)-(c'c_{-1} + d'c_{-1})(c'c +d'c)}{(c'c + 2d'c+d'd)(c'_{-1}c_{-1})-(c'c_{-1} + d'c_{-1})^2} \tag{3.91}$$

Next, α^* and β^* are written in terms of standardized sample moments. Let $X' = (C_0, C_1, C_2 \ldots, C_T)$ and define $(T+1)x(T+1)$ matrices

$$A_1 = \begin{bmatrix} 1 & 0 & .. & 0 & 0 \\ 0 & 1 & .. & 0 & 0 \\ . & . & .. & . & . \\ 0 & 0 & & 1 & 0 \\ 0 & 0 & & 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & \frac{1}{2} & .. & 0 & 0 \\ \frac{1}{2} & 0 & .. & 0 & 0 \\ . & . & .. & . & . \\ 0 & 0 & & 0 & \frac{1}{2} \\ 0 & 0 & & \frac{1}{2} & 0 \end{bmatrix}, \text{ and } A_3 = \begin{bmatrix} 0 & 0 & .. & 0 & 0 \\ 0 & 1 & .. & 0 & 0 \\ . & . & .. & . & . \\ 0 & 0 & & 1 & 0 \\ 0 & 0 & & 0 & 1 \end{bmatrix}$$

and introduce the variables $q_i = \left\{x'A_ix - E(x'A_ix)\right\}/T$, $(i=1,2,3)$ and $q_i = \left\{b_i'x - E(b_i'x)\right\}/T$, $(i=4,5)$ where $b_4' = (d' \ 0)$ and $b_5' = (0 \ d')$. Note that $c'_{-1}c_{-1} = x'A_1x$, $c'c_{-1} = x'A_2x$, $c'c = x'A_3x$, $d'c_{-1} = b_4'x$ and $d'c = b_5'x$. Then α^* and β^* become:

$$\begin{aligned} \alpha^* = {} & \{(q_1 + \mu_1)(q_3+ q_5+ \mu_3+ \mu_5) - (q_2+ q_4+ \mu_2+ \mu_4)(q_2+ \mu_2)\} \\ & +\{(q_3+ 2q_5+ \mu+ \mu_3+ 2\mu_5)(q_1+ \mu_1) - (q_2+ q_4+ \mu_2+ \mu_4)^2\} \end{aligned} \tag{3.92}$$

$$\begin{aligned} \beta^* = {} & \{(q_2+\mu_2)(q_3+2q_5+\mu+\mu_3+2\mu_5) - (q_2+q_4+\mu_2+\mu_4)(q_3+q_5+\mu_3+\mu_5)\} \\ & +\{(q_3+ 2q_5+ \mu+ \mu_3+ 2\mu_5)(q_1+ \mu_1) - (q_2+q_4+\mu_2+\mu_4)^2\} \end{aligned} \tag{3.93}$$

where $\mu = T^{-1}d'd$, $\mu_i = T^{-1}E(x'A_ix)$, $(i=1,2,3)$ and $\mu_p = T^{-1}E(b_i'x)$, $(p = 4,5)$. Similarly, the representations of the 2SLS estimators $\hat{\alpha}$ and $\hat{\beta}$ of α and β are given by

$$\hat{\alpha} = \frac{(q_1 + \mu_1)(q_5 + \mu_5) - (q_2 + \mu_2)(q_4+\mu_4)}{(q_1+\mu_1)(q_5+\mu_5+\mu)-(q_4+\mu_4)(q_2+q_4+\mu_2+\mu_4)} \tag{3.94}$$

and

$$\hat{\beta} = \frac{(q_2+\mu_2)(q_5+\mu_5+\mu)-(q_5+\mu_5)(q_2+q_4+\mu_2+\mu_4)}{(q_1+\mu_1)(q_5+\mu_5+\mu)-(q_4+\mu_4)(q_2+q_4+\mu_2+\mu_4)} \tag{3.95}$$

follows that estimators α^*, $\hat{\alpha}$, β^* and $\hat{\beta}$ are suitable for the application of the algorithm proposed by Sargan to derive the Edgeworth approximation. Phillips pointed out that to calculate the Edgeworth approximation to the OLS and 2SLS estimators of α and β in (3.87) one needs first to specify a series for I_t. Once the process generating the series I_t is specified the limits in probability of the OLS estimators of α and β can be readily computed. Numerical computations of the Edgeworth approximation to the small-sample distributions of the OLS and 2SLS estimators of α and β in (3.87) are possible once we have specified values of the underlying parameters α, β, and σ_u^2 in the model (3.87)-(3.88) as well as a series for the exogenous variables, I_t over the relevant sample periode, and enough of the past history of I_t to accurately compute the components m_t as in (3.89). Consider the single structural equation (3.82). The reduce: form equations are:

$$[y_1 \vdots y_2] = [Z_1 \vdots Z_2]\begin{bmatrix}\pi_{11} & \pi_{12}\\ \pi_{21} & \pi_{22}\end{bmatrix} + [v_1 \vdots v_2] \tag{3.96}$$

where Z_2 is a TxK_2 matrix of exogenous variables excluded from (3.82). The rows of the reduced form disturbance matrix V are assumed to be iid, normal random vectors. Assume standardizing transformations have been carried out so that the covariance matrix of rows of V is the

identity matrix and $T^{-1}Z'Z = I_K$ where $K = K_1 + K_2$ and $Z = [Z_1 \vdots Z_2]$. Suppose $K_2 \geq 1$. The 2SLS estimator of β in (3.82) is given by

$$\hat{\beta} = y_2'Ry_1/y_2'Ry_2 \tag{3.97}$$

where $R = Z_2(Z_2'Z_2)^{-1}Z_2' = Z_2Z_2'$. The density of $\hat{\beta}$ is

$$\mu df(r) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} B(w)\exp\{\frac{\mu^2}{2}\psi(w)\}dw \tag{3.98}$$

where $B(w) = \left[K_2 + \frac{\mu^2(1 + \beta w)^2}{1 + 2rw - w^2}\right](1 + 2rw - w^2)^{-(K_2+2)/2}$ and $\psi(w) = \frac{w^2(1 + \beta^2) + 2w(\beta - r)}{1 + 2rw - w^2}$ with $\mu^2 = T\pi_{22}'\pi_{22}$. If K_2 is even, the density of $\hat{\beta}$ is reduced to

$$pdf(r) = \text{residue}_{w=a_1}[B(w)\exp\{(\mu^2/2)\psi(w)\}] \tag{3.99}$$

while for K_2, an odd number we have

$$pdf(r) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} f(w)dw = \frac{1}{2\pi i}\int_{\gamma^*} f(w)dw \tag{3.100}$$

where $f(w) = B(w)\exp\{(\mu^2/2)\psi(w)\}$ and γ^* is the particular associated contour. Dealing with an equation with n + 1 endogenous variables , we have (3.82) and the reduced form of the endogenous variables of this equation is

$$[y_1 \vdots Y_2] = [Z_1 \vdots Z_2]\begin{bmatrix}\pi_{11} & \Pi_{12} \\ \pi_{21} & \Pi_{22}\end{bmatrix} + [v_1 \vdots V_2] \tag{3.101}$$

where Π_{22} is a K_2xn matrix of rank n, ($K_2 \geq n$). The rows of $[v_1 \vdots V_2]$ are independent normal with mean vector zero and covariance matrix I_n. Let $\Pi_{22}'\Pi_{22} = \bar{\Pi}_{22}\bar{\Pi}_{22}$ where $\bar{\Pi}_{22}$ is an nxn nonsingular matrix. The joint

probability density function of the 2SLS is given by

$$\mathrm{pdf}(r) = \frac{\mathrm{etr}\{-\frac{T}{2}(I+\beta\beta')\bar{\Pi}'_{22}\bar{\Pi}_{22}\}\Gamma_n(\frac{N+n+1}{2})}{\pi^{\frac{n}{2}}[\det(I+rr')]^{\frac{N+n+1}{2}}} \times\sum_{j=0}^{\infty}\frac{(\frac{1}{2})_j}{j!\,\Gamma_n(((N+n)/2)+j)}$$

$$\times\Big[(\tfrac{T}{2}\beta'\bar{\Pi}'_{22}(\mathrm{adj}\tfrac{\partial}{\partial W})\bar{\Pi}_{22}\beta)^j\times(\det(I+W))^{(N+n)/2+j-(n+1)}$$

$$\times{}_1F_1(\tfrac{N+n+1}{2},\tfrac{N+n}{2}+j;\tfrac{T}{2}(I+W)\bar{\Pi}_{22}(I+\beta r')(I+rr')^{-1}(I+r\beta')\bar{\Pi}'_{22})\Big]_{W=0} \qquad (3.102)$$

where $\mathrm{etr}(.) = \exp\{\mathrm{tr}(.)\}$. When $n = 1$, the multivariate density reduces to the univariate density function for the 2SLS estimator in the two endogenous variables case, i.e.

$$\mathrm{pdf}(r) = \frac{\exp\left\{\frac{\mu^2}{2}(1+\beta^2)\right\}}{B\left(\frac{1}{2},\frac{N+1}{2}\right)(1+r^2)^{(N+2)/2}} \times\sum_{j=0}^{\infty}\frac{\left(\frac{1}{2}\right)_j}{j!\left(\frac{N+1}{1}\right)_j}\left(\frac{\mu^2}{2}\beta^2\right)^j$$

$$\times{}_1F_1\left(\frac{N+2}{2},\frac{N+1}{2}+j,\frac{\mu^2}{2}\frac{(1+\beta r)^2}{1+r^2}\right) \qquad (3.103)$$

where $\mu^2 = T\bar{\pi}^2_{22} = T\pi'_{22}\pi_{22}$ is the concentration parameter and $N = K_2 - 1$ in this case. The term involves the factor $[\det(I + rr')]^{-(N+n+1)/2} = (1 + r'r)^{-(N+n+1)/2}$ which is similar in form to the principal factor of a multivariate t-density when $N > 0$ and a multivariate Cauchy-density when $N = 0$. Phillips, however, pointed out that virtually all the results established so far in the probability literature apply only in the case of standardized sums of independent random variables; this is an immediate limitation to the application in econometrics of the large deviation limit theory and its associated expansions. Using Sargan's

theorem, he came to the conclusion that suitably adjusted, the theorem applies to the important cases of econometric estimators whose finite sample moments may exist only up to a certain order.

In an elegant survey paper, Mariano (1982) [53] dealt exclusively with analytical finite-sample results in simultaneous-equations models with fixed parameters. Most of the results for the simultaneous-equations case are confined to limited-information instrumental variable estimators such as the k-class and the modified two-stage least squares.

For the set up (3.82), Phillips (1983a) [66] obtained an approximation to the marginal density of instrumental variable estimators in the general single equation by using the multivariate version of the Laplace formula. Let $H = [Z_1 \vdots Z_3]$, where $Z_3(TxK_3)$ is a submatrix of Z_2 and $K_2 \geq n$, be a matrix of IVs to be used in the estimation of (3.82). The IV estimator of the parameter vector β in (3.82) is then

$$\beta_{IV} = (Y_2'M_H Y_2)^{-1}(Y_2 M_H y_1) \tag{3.104}$$

where

$$M_H = H(H'H)^{-1}H' - Z_1(Z_1'Z_1)^{-1}Z_1' \tag{3.105}$$

The exact joint probability of β_{IV} is given by (3.85). Using the asymptotic representation of the ${}_1F_1$ function, namely, as $T\to\infty$ and for a nonsingular matrix R

$${}_1F_1(a,b;TR) = [\Gamma_n(b)/\Gamma_n(a)]\mathrm{etr}(TR)(\det TR)^{a-b}[1 + O(T^{-1})] \tag{3.106}$$

Phillips obtains the following asymptotic approximation of the joint

density of β

$$\sim \frac{T^{n/2}\operatorname{etr}-\frac{T}{2}\left\{\frac{M(r-\beta)(r-\beta)'}{1+r'r}\right\}(\det M)^{1/2}}{2^{n/2}\pi^{n/2}(1+r'r)^{(L+n+2)/2}(1+2\beta'r-\beta'\beta)^{L/2}}\frac{(1+\beta'r)^{L+1}}{} \tag{3.107}$$

where $M = \bar{\Pi}'_{22}\bar{\Pi}_{22}$. And then a direct application of the Laplace's method permits to extract densities from (3.107). Let us consider the following partitions of the matrix M and vectors β and r.

$M = \begin{pmatrix} m_{11} & m'_{12} \\ m_{21} & M_{22} \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ and $r = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$ where m_{11}, β_1 and r_1 are scalars.

Define

$$\bar{m}_{11} = m_{11} - m'_{21}M_{22}^{-1}m_{21}$$

$$\bar{r}_2 = r_2(r_1) = \beta_2 - (r_1 - \beta_1)M_{22}^{-1}m_{21}$$

$$\phi(r_1, r_2) = \frac{(r_1-\beta_1)^2\bar{m}_{11} + (r_2-\bar{r}_2)'M_{22}(r_2-\bar{r}_2)}{1+r_1^2+r'_2r_2}$$

$$g(r_1, r_2) = \frac{(1+\beta_1r_1+\beta'_2r_2)^{L+1}}{(1+r_1^2+r'_2r_2)^{(L+n+2)/2}(1+2\beta_1r_1+2\beta'_2r_2-\beta'\beta)^{L/2}}$$

The author obtained the following asymptotic approximation to the marginal density

$$\operatorname{pdf}(r_1) \sim (T/2\pi)^{1/2}(\det M)^{1/2}[\det H(r_1, r_2^*(r_1)]^{-1/2}$$

$$\times g[r_1, r_2^*(r_1)]\exp\left\{-\frac{T}{2}\phi[r_1, r_2^*(r_1)]\right\} \tag{3.108}$$

where $H(r_1, r_2) = \partial^2[(1/2)\phi(r_1, r_2)]/\partial r_2\partial r'_2 = [\alpha_1(r_1, r_2)I +$

$$\alpha_2(r_1,r_2)r_2(r_2 - \bar{r}_2)']M_{22} + \alpha_3(r_1,r_2)I + \alpha_4(r_1,r_2)r_2r_2'$$

and the α are scalar functions given by

$$\alpha_1(r_1,\ r_2) = (1 + r_1^2 + r_2'r_2)^{-1}$$

$$\alpha_2(r_1,\ r_2) = -4(1 + r_1^2 + r_2'r_2)^{-2}$$

$$\alpha_3(r_1,\ r_2) = -(1+ r_1^2+r_2'r_2)^{-2}[(r_1-\beta_1)^2\bar{m}_{11}+(r_2-\bar{r}_2)'M_{22}(r_2-\bar{r}_2)$$

$$\alpha_4(r_1,\ r_2) = 4(1+r_1^2+r_2'r_2)^{-3}[(r_1-\beta_1)^2\bar{m}_{11}+(r_2-\bar{r}_2)'M_{22}(r_2-\bar{r}_2)$$

And exact densities in some leading cases are given in [66].

Anderson (1982) [5] gave an excellent summary of some of works on the two-stage least squares (2SLS) estimator and the limited information maximum likelihood (LIML) estimator which have involved its associates Takamitsu Sawa, Naoto Kunitomo, and Kimio Morimune. The emphasis is on comparison of the 2SLS and LIML estimators based on finite-sample distributions. However, he commented on the higher-order efficiency of the LIML estimator and some improvements.

Phillips (1983b) [67], in a chapter devoted to acquaint the reader with the main strands of thought in the literature leading up to advancements, attempted to foster an awareness of the methods that have been used or that are currently being developed to solve problems in distribution theory and considered their suitability and scope in transmitting results to empirical researchers. After having provide a general framework for the distribution problem and details formulae that are frequently useful in the derivation of sampling distributions and moments, he dealt with the exact theory of single equation

estimators, commencing with a general discussion of the standardizing transformations, which provide research economy in the derivation of exact distribution theory in this context and which simplify the presentation of final results without loss of generality. Then he outlined the essential features of a new approach to small sample theory that seems promising for future research.

Richardson and Rohr (1983) [77] presented the results of an experimental study of the exact finite-sample distribution functions of two-stage least squares estimators in equations containing three endogenous variables. The system of simultaneous equations under consideration is

$$YB + X\Gamma = E \tag{3.109}$$

where Y is an NxG matrix of observations on G endogenous variables; X is an NxK matrix of nonstochastic exogenous variables; and E is an NxG matrix of unobservable disturbance terms. The rows of E are assumed to be iid according to the multivariate normal distribution with mean zero and covariance matrix Ω. Elements of B and Γ are real constants, B is non singular and the associated reduced form is

$$Y = X\Pi + V \tag{3.110}$$

where $\Pi = -\Gamma B^{-1}$, $V = EB^{-1}$, and the covariance matrix of the rows of V is $\Sigma = (B')^{-1}\Omega B^{-1}$. Let

$$y_i = Y_i\beta_i + X_i\gamma_i + \varepsilon_i \tag{3.111}$$

be the ith structural equation of (3.109) where y_i(Tx1) is a vector of observations on the normalized endogenous variables; Y_i(TxG_i-1) is a

matrix of observations on the G_i-1 nonnormalized endogenous variables included in the equation; and $X_i(TxK_i)$ is a matrix of included exogenous variables. Equation (3.110) can be partitioned as

$$y_i = X_i\pi_1 + X_i^*\pi_2^* + v_{1i} \tag{3.112}$$

$$Y_i = X_i\Pi_2 + X_i^*\Pi_2^* + V_{2i} \tag{3.113}$$

where $X = (X_i \vdots X_i^*)$. Richardson and Rohr have, particularly, investigated how the distribution functions of the estimators and test statistics in the $G_i = 3$ case are affected by values of the concentration matrix

$$\bar{M}_i = \Sigma_{22}^{-1}\Pi_2^{*\prime}S_i\Pi_2^* \tag{3.114}$$

where

$$S_i = X_i^*[I - X_i(X_i'X_i)^{-1}X_i']X_i^* \tag{3.115}$$

and Σ_{22} is the covariance matrix of nonnormalized endogenous variables included in the structural equation. For the standardized coefficient estimators

$$z = (\hat{\beta}_{ji} - \beta_{ji})/(w_{ii}\bar{M}_i^{jj})^{1/2} \tag{3.116}$$

where $\bar{M}_i^{jj}$ is the jjth element of $\bar{M}_i^{-1}$; it is found that it can be approximated by the standard normal distribution for large values of $\bar{M}_i$.

Phillips (1984a) [68] proved that the exact finite sample distribution of the Limited Information Maximum Likelihood (LIML) estimator in a general and leading single equation case is multivariate Cauchy. A sequel paper, [70], gives results which characterize, in the

Basmann's notation, the complete class of distributions

$$M = \bigcup_{n=1}^{\infty} \bigcup_{L=1}^{\infty} M_{n,L} \tag{3.117}$$

corresponding to a structural equation containing any number of endogenous variables, even or odd degrees of freedom and an arbitrary degree of overidentification. Phillips worked with the set up (3.82). He considered the leading subcase of (3.82) and (3.83) in which $\Pi_{22} = 0$. Then the IV estimator of β is defined to be

$$\beta_{IV} = [Y_2'Z_3Z_3'Y_2]^{-1}[Y_2'Z_3Z_3'y_1] \tag{3.118}$$

where $Z_3(TxK_3)$ is a submatrix of Z_3 forming instruments additional to Z_1 and where it is assumed that $K_3 \geq n$. And the pdf of β_{IV} is given by

$$\text{pdf}(r) = \frac{\Gamma\left(\frac{K_3 + 1}{2}\right)}{\pi^{n/2}\Gamma\left(\frac{K_3 - n + 1}{2}\right)(1 + r'r)^{(K_3 + 1)/2}}$$

$$= \frac{\Gamma\left(\frac{L + n + 1}{2}\right)}{\pi^{n/2}\Gamma\left(\frac{L + 1}{2}\right)(1 + r'r)^{(L + n + 1)/2}} \tag{3.119}$$

where $L = K_3 - n$ is the number of surplus instruments used in the estimation of β and $\beta_{IV} = r$. In the overidentified case ($K_2 \geq n + 1$), the LIML estimator, β_{LIML}, of β minimizes the ratio $\beta_\Delta'W\beta_\Delta/\beta_\Delta'S\beta_\Delta$ where $\beta_\Delta' = (1, -\beta')$, $W = X'(P_Z - P_{Z_1})X$, $S = X'(I - P_Z)X$ and $X = [y_1 \vdots Y_2]$ and $P_A = A(A'A)^{-1}A'$. When the covariance matrix of the rows of X is known, the corresponding estimator is called LIMLK and will be denoted by β_{LIMLK}. Setting $\beta_{LIMLK} = r$, Phillips (1984a) found that the pdf of

β_{LIMLK} takes the form

$$\mathrm{pdf}(r) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{(n+1)/2}(1 + r'r)^{(n+1)/2}} \tag{3.120}$$

that is, a multivariate Cauchy distribution. The paper [68] ended with the distribution of LIML which is given by

$$\mathrm{pdf}(r) = \frac{2^{(m+g)/2-1}\pi^{(m-1)/2}}{2^{(T-K_1)/2-1}\pi^{n+1/2}} \times \frac{\Gamma\left(\frac{g+1}{2}\right)\Gamma\left(\frac{n+1}{2}\right)(1 + r'r)^{g/2}}{\Gamma\left(\frac{T-1}{2}\right)(1 + r'r)^{(T-K_1)/2}} \tag{3.121}$$

Replacing g by its expression $g = T - K_1 - m = T - K_1 - n - 1$ reduces (3.121) to (3.120) and the distribution of β_{LIML} is therefore multivariate Cauchy. In the sequel paper [70] Phillips had used the series representation of ${}_0F_1$ in zonal polynomials and the multinomial expansion of a sum of matrices in terms of invariant polynomials of several matrix arguments to obtain that the pdf of LIML is

$$\mathrm{pdf}(r) = \frac{\pi^{nm/2}\mathrm{etr}\left\{-\frac{T}{2}(I + \beta\beta')\bar{\Pi}_{22}\bar{\Pi}_{22}\right\}\Gamma\left(\frac{T-K_1-n}{2}\right)}{\Gamma_n(n/2)\Gamma_m(K_2/2)\Gamma_m(\frac{T-K}{2})(1 + r'r)^{(n+1)/2}}$$

$$\times \sum_{f=0}^{\infty}\sum_{\varphi}\frac{1}{(K_2/2)_\varphi}\sum_{\substack{j_1,j_2\\(\varphi \in j_1 \cdot j_2)}}\frac{\Theta_\varphi^{[j_1],j_2}}{j_1!j_2!}W_n(j_1,j_2)\Gamma_n\left(\frac{T-K_1}{2}, j_2\right)\left[C_\varphi^{[j_1],j_2}\right.$$

$$\times\left(\frac{T}{4}\bar{\Pi}_{22}BD(r)\partial x\partial x'D(r)'B'\bar{\Pi}'_{22}, \frac{T}{2}\bar{\Pi}_{22}(I + \beta r')(I + rr')^{-1}\right.$$

$$\left.\left.\times(I + r\beta')\bar{\Pi}'_{22}\right)e^{x_2'x_2/2}{}_1F_1\left(\frac{T-K_1-n}{2}, \frac{1}{2}; \frac{x_1^2}{2}\right)\right]_{x=0} \tag{3.122}$$

where $e_{\varphi}^{J[2]} = C_{\varphi}^{J[2]}(I,I)/C_{\varphi}(I)$, φ represents the order of partition of $f = j_1 + j_2$ into at most n parts, $J[2] = (J_1, J_2)$ where J_i represents an ordered partition of nonnegative integer j_i $(i = 1,2)$ into $\leq n$ parts. $\bar{\Pi}_{22}$ is an nxn matrix. $\beta_{LIML} = r$. $D = D_{22}'D_{22}$ according to the unique decomposition $D_{22} = HD^{1/2}$ where H is orthogonal. $B = [\beta \vdots I_n]$ and $D = D(r)$ means D depends on r. When $\bar{\Pi}_{22} = 0$ (3.122) reduces to

$$\text{pdf}(r) = \frac{\pi^{nm/2}\Gamma_n\left(\frac{T-K_1}{2}\right)\Gamma_n\left(\frac{T-K_1}{2}-\frac{n}{2}\right)W_n(0,0)}{\Gamma_n\left(\frac{n}{2}\right)\Gamma_m\left(\frac{K_2}{2}\right)\Gamma_m\left(\frac{T-K}{2}\right)(1+r'r)^{(n+1)/2}} \propto \frac{1}{(1+r'r)^{(n+1)/2}} \tag{3.123}$$

let n =1. Then (3.123) becomes

$$\text{pdf}(r) = \frac{2\Gamma(T-K)\Gamma(\frac{T-K}{2}+1)}{\pi\Gamma(T-K+1)\Gamma(\frac{T-K}{2})(1+r^2)} = \frac{2}{\pi}\cdot\frac{1}{2}\cdot\frac{\Gamma(\frac{T-K}{2})}{\Gamma(\frac{T-K}{2}+1)}\;\frac{\Gamma(\frac{T-K}{2}+1)}{\Gamma(\frac{T-K}{2})}\;\frac{1}{1+r^2}$$

$$= \frac{1}{\pi(1+r^2)} \tag{3.124}$$

Hillier (1985) [37] provided an alternative derivation of Phillips' results on the joint density of the IV estimator for the endogenous coefficients, and derived an expression for the marginal density of a linear combination of these coefficients. He extended Phillips' approximation to the joint density to $o(T^{-2})$, and showed how his result can be used to improve the approximation to the marginal density. Consider a structural equation

$$y = Y\beta^* + Z_1\gamma + u \tag{3.125}$$

where $(y, Y)(T\times(n+1))$ contains observations on the included endogenous

variables and $Z_1(T\times K_1)$ contains observations on the included exogenous variables. The reduced form for (y,Y) is denoted by

$$(y,\ Y) = (Z_1,\ Z_2)\begin{pmatrix}\pi_1 & \Pi_1\\ \pi_2 & \Pi_2\end{pmatrix} + (v,\ V) \tag{3.126}$$

where $Z_2(T\times K_2)$ contains observations on the exogenous variables excluded from (3.125) but appearing elsewhere in the system. Assume that $Z = (Z_1,\ Z_2)(T\times K;\ K = K_1 + K_2)$ is fixed, and that the rows of (v,V) are independent normal vectors with mean zero and common covariance matrix

$\Omega = \begin{pmatrix}w_{11} & w_{12}\\ w_{21} & w_{22}\end{pmatrix}$. The OLS, 2SLS, and IV estimators for β^* are of the form

$$b = (Y'PY)^{-1}Y'Py \tag{3.127}$$

with, in each case, P an idempotent matrix of rank ν (say) such that $PZ_1 = 0$. Let

$$r = (\Omega_{22}^{1/2}b - \Omega_{22}^{-1/2}w_{12})/w \tag{3.128}$$

where $w^2 = w_{11} - w_{12}'\Omega_{22}^{-1}w_{12}$. The conditional density of r given Y may be reduced to the form

$$r/x \sim N((X'X)^{-1}X'M\beta,\ (X'X)^{-1}) \tag{3.129}$$

where

$$\beta = (\Omega_{22}^{1/2}\beta^* - \Omega_{22}^{-1/2}w_{12})/w \tag{3.130}$$

$$M = E(X) = C'Z_2\Pi_2\Omega_{22}^{-1/2} \tag{3.131}$$

$X = C'Y\Omega_{22}^{-1/2}$, and $C(T\times\nu)$ is such that $P = CC'$ and $C'C = I_\nu$. The rows of X are independent normal vectors with covariance matrix I_n. The joint density of the elements of r may be obtained by averaging the

conditional density (3.129) with respect to the density of X:

$$f(r) = \int_X f(r|X)g(X;M)dX \tag{3.132}$$

with

$$g(X;M) = (2\pi)^{-(n\nu)/2} \operatorname{etr}\left\{-\frac{1}{2}(X - M)'(X - M)\right\} \tag{3.133}$$

After evaluation of required integrals (and expansion of the involved confluent hypergeometric function and the exponential function in terms of zonal polynomials) Hillier obtains

$$f(r) = \frac{\Gamma_n\left(\frac{1}{2}(\nu+1)\right)\operatorname{etr}\left\{-T\Lambda(I+\beta\beta')\right\}}{\pi^{n/2}(1 + r'r)^{(\nu+1)/2}\Gamma_n(\nu/2)} \times \sum_{\alpha,[k];\phi} \frac{\left(\frac{1}{2}(\nu+1)_\alpha\right)\left(\frac{1}{2}(\nu-n)\right)_k}{j!\,k!\,(\nu/2)_\phi}$$

$$\times\Theta_\phi^{\alpha,[k]}C_\phi^{\alpha,[k]}(T\Lambda(r), T\delta\delta') \tag{3.134}$$

where [k] denotes the partition (k, 0, ... ,0), $C_\phi^{\alpha,\lambda}(.,.)$ is the invariant zonal polynomial (see [37] for reference), and $\Theta_\phi^{\alpha,\lambda} = C_\phi^{\alpha,\lambda}(I,I)/C_\phi(I)$, and $\Lambda = \frac{1}{2}T^{-1}M'M = \frac{1}{2}T^{-1}\Omega_{22}^{-1/2}\Pi_2'Z_2'PZ_2\Pi_2\Omega_{22}^{-1/2}$ It should be noticed that equation (3.134) is Phillips' [63] equation B3. The density of r in the leading case corresponding to $\beta = 0$ is given by

$$f(r)= \frac{\Gamma(\nu/2)\operatorname{etr}\{-T\Lambda(I+\beta\beta')\}}{\pi(1+r'r)^{(\nu+1)/2}} \sum_{j,k,u,v=0}^{\infty} \frac{((\nu-2)/2)_{j+v}(j+v-1/2)_k(-1)^k}{j!\,k!\,u!\,v!}$$

$$\frac{\Gamma(k+u+v+\frac{1}{2}(\nu+1))(T\delta'\delta)^j}{\Gamma(k+v+\frac{1}{2}(\nu-1))\Gamma(j+u+2v+2k+\frac{1}{2}\nu)}(\delta'\Lambda(r)^{-1}\delta)^v(\operatorname{tr}T\Lambda(r))^u|T\Lambda(r)|^{k+v} \tag{3.135}$$

where

$$(\delta'\Lambda(r)\delta) = \beta'(I + rr')\beta/(1 + r'\beta)$$

$$|T\Lambda(r)| = |T\Lambda|(1 + r'\beta)^2/(1 + r'r)$$

$$\mathrm{tr}T\Lambda(r) = \mathrm{tr}\{T\Lambda(I+\beta\beta')\} - T(r - \beta)'\Lambda(r - \beta)/(1+r'r) \tag{3.136}$$

Approximations based on inverse Laplace transform expression for the joint density are given by

$$f(r) = \frac{\exp\{-T(r-\beta')\Lambda(r-\beta)/(1+r'r)\}|T\Lambda(r)|^{1/2}}{\pi^{n/2}(1+r'r)^{(\nu+1)/2}(1-\delta'\Lambda(r)^{-1}\delta)^{(\nu-n)/2}}[1+o(T^{-1})] \tag{3.137}$$

where

$$1 + 2r'\beta - \beta'\beta > 0 \tag{3.138}$$

Next, consider the following

$$(1 - \delta'\Lambda(r)^{-1}\delta)^{-(\nu-n)/2}\Gamma_n\left(\frac{1}{2}(\nu+1)\right)a_n$$

$$\times\int_{Re(W)=W_1>0}\mathrm{etr}(W)|W|^{-(\nu+1)/2}[1- \frac{1}{2}T^{-1}(\nu - n)$$

$$\times\mathrm{tr}\{(\Lambda(r) - \delta\delta')^{-1}W\} + \frac{1}{2}T^{-1}(\nu - n + 1)\mathrm{tr}\{\Lambda(r)^{-1}W\}]dW \tag{3.139}$$

Now, the integral $\Gamma_n(t)a_n\int_{Re(W)=W_1>0}\mathrm{etr}(W)|W|^{-t}(\mathrm{tr}AW)dW$ is the coefficient of q in the expansion of $\Gamma_n(t)a_n\int_{Re(W)=W_1>0}\mathrm{etr}\{(I + qA)W\} \times |W|^{-t}dW = |I + qA|^{t-(n+1)/2}$ and is therefore equal to $(t - \frac{1}{2}(n + 1)\mathrm{tr}(A)$. Hence we have $(1 - \delta'\Lambda(r)^{-1}\delta)^{-(\nu-n)/2}[1 - \frac{1}{4}T^{-1}(\nu - n) \times\{(\nu - n)\mathrm{tr}(\Lambda(r) - \delta\delta')^{-1} - (\nu - n + 1)\mathrm{tr}\Lambda(r)^{-1}\}]$. Therefore, $f(r) = f_1(r)[1 - \frac{1}{4}T^{-1}(\nu - n)\times\{(\nu - n)\mathrm{tr}(\Lambda(r) - \delta\delta')^{-1} - (\nu - n + 1)\mathrm{tr}\Lambda(r)^{-1}\} + o(T^{-2})]$ (3.140)

where $f_1(r)$ denotes the $o(T^{-1})$ approximation to the right in (3.137). The density of a fixed linear combination $\tilde{\alpha} = a'r$ of the elements of r is obtained by inverting the characteristic function $\phi(t) = E_X[\exp\{ita'(X'X)^{-1}X'M\beta - (t^2a'(X'X)^{-1}a)/2\}]$. Inverting the characteristic function leads to

$$f(\tilde{\alpha}) = \frac{\mathrm{etr}(-T\Lambda)2^{-n\nu/2}}{(2\pi)^{1/2}} a_n \int_{R>0}\int_{\mathrm{Re}(W)=W_0>0} \mathrm{etr}(W)|W|^{-\nu/2} \times \mathrm{etr}\{-\tfrac{1}{2}(I - T\Lambda^{1/2}W^{-1}\Lambda^{1/2})R\}|R|^{(\nu-n-1)/2} \times [a'R^{-1}a(1 + T\delta'W^{-1}\delta]^{-1/2} T\delta'W^{-1}\delta]^{-1/2} \times$$

$$\exp\{-\tfrac{1}{2}(\tilde{\alpha} - Tc'W^{-1}\delta)^2/a'R^{-1}a(1 + T\delta'W^{-1}\delta)\}dWdR \qquad (3.141)$$

Expanding the exponential term in the integrand of (3.141) and integrating term by term over R > 0 using some results from complex symmetric matrix yields the density in the form

$$f(\tilde{\alpha}) = \frac{\mathrm{etr}(-T\Lambda)\Gamma(\frac{1}{2}(\nu-n+2))}{\pi^{1/2}\Gamma(\frac{1}{2}(\nu-n+1))}\Gamma_n(\nu/2)a_n\int_{\mathrm{Re}(W)=W_0}\mathrm{etr}(W)|W|^{-\nu/2}\times|I-T\Lambda W^{-1}|^{-\nu/2}$$

$$(1 + T\delta'W^{-1}\delta)^{-1/2}(1 - Tc'W^{-1}c)^{-1/2}$$

$$\times\left[1+\frac{(\tilde{\alpha}-Tc'W^{-1}\delta)^2}{(1+T\delta'W^{-1}\delta)(1-Tc'W^{-1}c)}\right]^{-(\nu-n+2)/2} dW \qquad (3.142)$$

which can be written in the form

$$f(\tilde{\alpha}) = \int_{\mathrm{Re}(W)>0} g(W)f(\tilde{\alpha}|W)dW \qquad (3.143)$$

with

$$g(W) = \Gamma_n(\nu/2)a_n\mathrm{etr}(W - T\Lambda)|W - T\Lambda|^{-\nu/2} \qquad (3.144)$$

and

$$f(\tilde{\alpha}|W) = \frac{\Gamma\left(\frac{1}{2}(\nu+1)\right)[(1+T\delta'W^{-1}\delta)(1-Tc'W^{-1}c]^{1/2}}{\pi^{1/2}\Gamma(\frac{1}{2}(\nu-n+1))} \times$$

$$\left[1+\frac{(\tilde{\alpha}-Tc'W^{-1}\delta)^2}{(1+T\delta'W^{-1}\delta)(1-Tc'W^{-1}c)}\right]^{-(\nu-n+2)/2} \tag{3.145}$$

which is a (complex) t-distribution with mean $\tilde{\mu} = Tc'W^{-1}\delta$ and variance $\tilde{\sigma}^2 = (1 + T\delta'W^{-1}\delta)(1 - Tc'W^{-1}c)/(\nu - n -1)$. If instead one considers the transformation $W\ W - T\delta\delta'$ it follows that

$$f(r|W) = \frac{\Gamma\left(\frac{1}{2}(\nu+1)\right)|I - T\Lambda^{1/2}W^{-1}\Lambda^{1/2}|^{-1/2}}{\pi^{n/2}\,\Gamma\,(\frac{1}{2}(\nu-n+1))(1+T\delta'W^{-1}\delta)^{n/2}}$$

$$\times \left[1 + \frac{(r-T\Lambda^{1/2}W^{-1}\delta)'(I-T\Lambda^{1/2}W^{-1}\Lambda^{1/2})^{-1}(r-T\Lambda^{1/2}W^{-1}\delta)}{1 + T\delta'W^{-1}\delta}\right] \tag{3.146}$$

which is a multivariate t-distribution with mean $\tilde{\mu} = T\Lambda^{1/2}W^{-1}\delta$ and covariance matrix $\tilde{\Sigma} = (1 + T\delta'W^{-1}\delta)[I - T\Lambda^{1/2}W^{-1}\Lambda^{1/2}]/(\nu - n - 1)$. A variety of different expressions for the exact density may be obtained from (3.142), depending on how the last term in the integrand is treated.

For the model

$$B'y_t' + \Gamma'z_t' + e_t' = 0 \tag{3.147}$$

whose first equation is

$$y_{t1} = \beta_2 y_{t2} + \beta_3 y_{t3} + \sum_{k=0}^{K_1} \gamma_k z_{tk} + e_{t1} \tag{3.148}$$

Richardson (1986) [76] gave a full statement of the standardizing transformations and showed explicitly how the standardized parameters of the exact density function are related to the parameters of the original model. Hence he derived the exact joint density function of the estimates of β_2 and β_3. He also considered computations of the joint and marginal densities for the special case in which the estimators are exactly unbiased. The reduced form associated with (3.147) is

$$y_t' = \pi' z_t' + \eta_t' \tag{3.149}$$

where

$$\pi = -\Gamma B^{-1}$$

$$\eta = -e_t B^{-1}$$

$$t = 0, \pm 1, \pm 2, \ldots$$

The vectors η_t are iid with mean zero and covariance matrix $\Sigma = (B')^{-1}\Omega B^{-1}$. The maximum likelihood estimate of π is

$$\hat{\pi} = (Z'Z)^{-1}Z'Y \tag{3.150}$$

where the sample matrices Y and Z are defined by

$$Y = [y_{ti}] \ , \ t = 1, \ldots, N \quad i = 1, \ldots, G \tag{3.151}$$

$$Z = [z_{tk}] \ , \ t = 1, \ldots, N \quad k = 1, \ldots, K \tag{3.152}$$

The GCL (2SLS) estimate of β is

$$\hat{\beta} = (P_{22}'SP_{22})^{-1}P_{22}'SP_{21} \tag{3.153}$$

where P denotes the submatrix of (2.9) which corresponds to

$$\pi_* = \begin{bmatrix} \pi_{*1} \\ \pi_{*2} \end{bmatrix} \tag{3.154}$$

and partitioned conformably with the partition of π_* given by (3.154).

$$S = Z_2'[I - Z_1(Z_1'Z_1)^{-1}Z_1']Z_2 \tag{3.155}$$

Let y_t denote the subvector of y_t consisting of the first three elements of y_t, and let Σ_* denote the covariance matrix of y_t^*. The vector y_t^* is transformed according to

$$\bar{y}_t^* = y_t^* R \tag{3.156}$$

where R is non singular matrix such that $R'\Sigma_* R = I$. Let R take the form

$$R = \begin{bmatrix} r_{11} & 0 \\ r_1 & R_{22} \end{bmatrix} \tag{3.157}$$

where R_{22} is lower triangular. We have $\bar{\Sigma}_* = I_3$. And the vector β is transformed according to

$$\bar{\beta} = R_{22}^{-1}(\beta r_{11} + r_1) \tag{3.158}$$

Therefore (3.133) becomes

$$\bar{y}_{t1}^* = \bar{\beta}_2 \bar{y}_{t2}^* + \bar{\beta}_3 \bar{y}_{t3}^* + \sum_{j=1}^{K_1} \gamma_j Z_{tj} + \bar{e}_{t1} \tag{3.159}$$

where

$$\bar{\beta}_2 = \frac{(\sigma_{22}\sigma_{33} - \sigma_{23}^2)\beta_2 - (\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23})}{\sigma_{33}|\Sigma_*|^{1/2}} \tag{3.160}$$

$$\bar{\beta}_3 = \frac{(\sigma_{22}\sigma_{33} - \sigma_{23}^2)^{1/2}}{\sigma_{33}|\Sigma_*|^{1/2}} (\sigma_{33}\beta_3 + \sigma_{23}\beta_2 - \sigma_{13}) \times \frac{(\sigma_{22}\sigma_{33} - \sigma_{23}^2)^{1/2}}{|\Sigma_*|^{1/2}} e_t \tag{3.161}$$

The reduced form coefficient matrices π_{*1} and π_{*2} are transformed to

$$\bar{\pi}_{*1} = \Lambda_1 O_1' [\pi_{*1} + (Z_1'Z_1)^{-1} Z_1'Z_1 \pi_{*2}] R \tag{3.162}$$

$$\bar{\pi}_{*2} = \Lambda_2 O_2' \pi_{*2} R \tag{3.163}$$

where O_1 and O_2 are orthogonal matrices such that

$$O_1'Z_1'Z_1O_1 = \Lambda_1'\Lambda_1 \tag{3.164}$$

$$O_2'V_2'V_2O_2 = \Lambda_2'\Lambda_2 \tag{3.165}$$

Λ_1 and Λ_2 being diagonal matrices whose diagonal elements are the square roots of the characteristic roots of $Z_1'Z_1$ and $V_2 = [I - Z_1 (Z_1'Z_1)^{-1} Z_1']Z_2$. Let $\bar{P}$ and $\bar{S}$ denote the maximum likelihood estimators of $\bar{\pi}_*$ and $\bar{\Sigma}_*$, respectively, the 2SLS of $\bar{\beta}$ is

$$v = \begin{bmatrix} v_2 \\ v_3 \end{bmatrix} = (\bar{P}_{22}'\bar{P}_{22})^{-1} \bar{P}_{22}'\bar{P}_{21} \tag{3.166}$$

The standardized estimators v_2 and v_3 are related to $\hat{\beta}_2$ and $\hat{\beta}_3$ of (3.153) by

$$v_2 = \frac{(\sigma_{22}\sigma_{33} - \sigma_{23}^2)\hat{\beta}_2 - (\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23})}{\sigma_{33}|\Sigma_*|^{1/2}} \tag{3.167}$$

$$v_3 = \frac{(\sigma_{22}\sigma_{33} - \sigma_{23}^2)^{1/2}}{\sigma_{33}|\Sigma_*|^{1/2}} (\sigma_{33}\hat{\beta}_3 + \sigma_{23}\hat{\beta}_2 - \sigma_{13}) \tag{3.168}$$

The joint density function of $v' = (v_2, v_3)$ is

$$h(v) = \frac{\operatorname{etr}\left\{-\frac{1}{2}(I + \bar{\beta}\bar{\beta}')T_{22}\right\}}{\pi(1 + v'v)^{(\nu+3)/2}B\left(\frac{\nu+1}{2}, 1\right)} \sum_{j=0}^{\infty} \frac{(\nu/2)_j(\bar{\beta}'T_{22}\bar{\beta})^j}{\left(\frac{\nu+2}{2}\right)_j j!}$$

$$\times \sum_{k=0}^{\infty}\sum_{q=0}^{\infty} \left(\frac{\nu}{2}+j\right)_{k+q} \times \left(\frac{\nu+3}{2}\right)_{k+q} \left(\frac{1}{4}|T_{22}|\right)^{k+q} \left[\bar{\beta}'(I + vv')\bar{\beta}\right]^k$$

$$\times \frac{\left[\frac{1}{2}\bar{\beta}'(I + v\beta)\bar{T}(I + \bar{\beta}v')\bar{\beta}\right]^q}{\left(\frac{\nu+2}{2} + j\right)_{2k+3q}\left(\frac{\nu+1}{2}\right)_{k+q}} \times \frac{1}{k!q!(1 + v'v)^{k+q}}$$

$$\times \sum_{p=0}^{\infty} \frac{\left(\frac{\nu+3}{2} + k + q\right)_p \left(k + q - \frac{1}{2}\right)_p (-1)^p \left[\frac{1}{2}(1 + \beta'v)^2|T_{22}|\right]^p}{\left(\frac{\nu+2}{2} + j + 2k + 3q\right)_{2p}\left(\frac{\nu+1}{2} + k + q\right)_p p!(1 + v'v)^p}$$

$$\times {}_1F_1\left[\begin{matrix}\frac{\nu+3}{2} + k + q + p; \\ \frac{\nu+2}{2} + j + 2k + 3q + 2p;\end{matrix} \quad \frac{1}{2}\operatorname{tr}(I + v\bar{\beta}')(I + vv')(I + \bar{\beta}v')T_{22}\right]$$

$$-\infty < v < \infty \tag{3.169}$$

where $T_{22} = R_{22}'\pi_{22}'V_2'V_2\pi_{22}R_{22}$. For $\beta = 0$, we have

$$n(v) = \frac{\operatorname{etr}\left(-\frac{1}{2}T_{22}\right)}{\pi B\left(\frac{\nu+1}{2}, 1\right)(1 + v'v)^{(\nu+3)/2}}$$

$$\times \sum_{p=0}^{\infty} \frac{\left(\frac{\nu+3}{2}\right)_p \left(-\frac{1}{2}\right)_p (-1)^p \left(\frac{1}{4}|T_{22}|\right)^p}{\left(\frac{\nu+1}{2}\right)_p \left(\frac{\nu+2}{2}\right)_{2p} p!\,(1 + v'v)^p}$$

$$\times {}_1F_1\left[\frac{\nu+3}{2}+p;\ \frac{\nu+2}{2}+2p;\ \frac{1}{2}\operatorname{tr}(I+vv')^{-1}T\right], \quad -\infty < v < \infty \tag{3.170}$$

Using the results obtained by Phillips (1980, 1983), Richardson obtained the following approximation to the exact density (3.170)

$$\bar{h}(v) = \frac{\exp\left(-\frac{1}{2}\frac{v'T_{22}v}{1 + v'v}\right)|T_{22}|}{2\pi(1 + v'v)^{(\nu+4)/2}}, \quad -\infty < v < \infty \tag{3.171}$$

The marginal density function of the standardized estimator v_2 for the special case $\bar{\beta} = 0$ is

$$h_2(v_2) = \frac{\operatorname{etr}\left(-\frac{1}{2}T_{22}\right)}{B\left(\frac{\nu+1}{2}, \frac{1}{2}\right)(1 + v_2^2)^{(\nu+3)/2}}$$

$$\times \sum_{p=0}^{\infty} \frac{\left(\frac{\nu+2}{2}\right)_p \left(-\frac{1}{2}\right)_p (-1)^p \left(\frac{1}{4}|T_{22}|\right)^p}{\left(\frac{\nu+1}{2}\right)_p \left(\frac{\nu+2}{2}\right)_{2p} p!\,(1 + v_2^2)^p}$$

$$\times \sum_{n=0}^{\infty}\sum_{j=0}^{\infty} \frac{\left(\frac{\nu+3}{2}+p\right)_{n+j} \left(\frac{\nu+2}{2}+p\right)_n \left(\frac{1}{2}t_{22}\right)^j \left(\frac{1}{2}t_{23}\right)^n}{\left(\frac{\nu+2}{2}+2p\right)_{n+j} \left(\frac{\nu+3}{2}+p\right)_n (1 + v_2^2)^j\, j!\, n!}$$

$$\times\sum_{k=0}^{\infty}\frac{\left(\frac{\nu+3}{2}+p+n+j\right)_{2k}\left(\frac{\nu+2}{2}+p+n\right)_{k}\left(\frac{1}{2}\right)_{k}(v_2t_{23})^{2k}}{\left(\frac{\nu+2}{2}+j+2p+n\right)_{2k}\left(\frac{\nu+3}{2}+p+n\right)_{2k}(1+v_2^2)k!}$$

$$\times{}_2F_1\left[\begin{matrix} k+\frac{1}{2},\ \frac{\nu+3}{2}+p+n+j+2k; \\ \frac{\nu+3}{2}+p+n+2k; \end{matrix}\ \frac{v_2^2}{1+v_2^2}\right],\quad -\infty < v_2 < \infty \tag{3.172}$$

where ν is the degree of freedom of $\chi_\nu(v_1)$ with $v_1 = P'_{21}P_{21} - v'(\bar{P}'_{22}\bar{P}_{22})^{-1}v$ is the numerator of the identifiability test statistic in standardized form. The marginal density of v_3 is of the same form as (3.172) with t_{22} and t_{33} and the series in j and n interchanged. It should be noticed that the exact marginal density function (3.172) is symmetric about the origine. Follows that the exact marginal density of $\hat{\beta}_2(\hat{\beta}_3)$ is symmetric about β_2 (β_3) when $\bar{\beta} = 0$. The marginal marginal density function is too complicated for numerical computation.

3.3 SURVEY OF DEVELOPMENTS ON THE APPROXIMATION TO THE DISTRIBUTION OF ECONOMETRIC ESTIMATORS IN SIMULTANEOUS EQUATIONS MODELS

Consider equation (3.38). Working directly with distribution functions, Mariano (1973) [51], presented an approximation to the distribution function of the 2SLS estimators up to terms whose order of magnitude is $1/\sqrt{T}$, where the sample size T is held fixed as the noncentrality parameter increases. It is assumed that all of the K predetermined variables are exogenous, the equation to be estimated is identified by zero-restrictions on the structural coefficients in the

model, the sample size $T \geq G + K$ (where G is the number of endogenous variables), X is a full rank matrix of constants, $\frac{X'X}{T}$ tends to a finite positive definite matrix as $T \longrightarrow \infty$, and the rows of V are mutually independent and identically distributed as bivariate normal random vectors with 0 mean and covariance matrix Σ.

$$\hat{\beta} = \frac{Y'Py}{Y'PY} \tag{3.173}$$

is then reduced to canonical form $\hat{\beta} = \beta + \frac{w}{\sqrt{\sigma_{22}}}\hat{\beta}^*$ with

$$\hat{\beta}^* = \frac{\sum_{i=1}^{m} x_i^* y_i^*}{\sum_{i=1}^{m} y_i^{*2}} \tag{3.174}$$

where

$$P = X(X'X)^{-1}X' - X_1(X_1'X_1)^{-1}X_1' \tag{3.175}$$

$$\Sigma = (\sigma_{ij}) \; ; \; w^2 = \sigma_{11} - 2\beta\sigma_{21} + \beta^2\sigma_{22} \tag{3.176}$$

$$x_i^* = \sqrt{1-\rho^2}x_i + \rho y_i \qquad i = 1,2,\ldots,K_2$$

$$y_i^* = \begin{cases} y_i & i = 1,2,\ldots,K_2-1 \\ y_i + \mu & i = K_2 \end{cases}$$

and (x_i^*, y_i^*) mutually independent normal vectors having common covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. For an arbitrary real number b, Mariano stated and proved that in the case of two included endogenous variables in the equation being estimated, the 2SLS distribution can be approximated by:

$$\Pr\left\{(\hat{\beta} - \beta) \le \frac{bw}{\mu\sqrt{\sigma_{22}}}\right\} = \Phi(b) + \frac{\rho}{\mu}\phi(b)(b^2 - K_2 + 1) + O(\mu^{-2}) \text{ as } \mu \to \infty \text{ here } \rho =$$

$\frac{1}{w\sqrt{\sigma_{22}}}(\sigma_{21} - \beta\sigma_{22})$, $w^2 = \sigma_{11} - 2\beta\sigma_{21} + \beta^2\sigma_{22}$ and

$$\mu^2 = \frac{1}{\sigma_{22}}\{\Pi_{22}Z_2'[I - Z_1(Z_1'Z_1)^{-1}Z_1']Z_2\Pi_{22}'\} = o(T) \tag{3.177}$$

Then follows the main result which stipulates that if the equation being estimated contains two endogenous variables and is just-identified we have

$$\Pr\{(\hat{\beta} - \beta) \le \frac{bw}{\mu\sqrt{\sigma_{22}}}\} = \Phi(b\xi) + R. \tag{3.178}$$

where $|R| \le \frac{1}{\sqrt{2\pi}\mu} e^{-\frac{1}{2}\mu^2}$, and $\xi^{-2} = 1 - \frac{2b\rho}{\mu} + \frac{b^2}{\mu^2}$. For fixed T Mariano found that in the case of two endogenous variables present in the equation to be estimated if $\tilde{\beta}$ is the OLS estimator of β then

$$\Pr\left\{(\tilde{\beta} - \beta) \le \frac{bw}{\mu\sqrt{\sigma_{22}}}\right\} = \Phi(b) + \frac{\rho}{\mu}\phi(b)(b^2 - N + K_1 + 1) + O(\mu^{-2}) \text{ as } \mu \to \infty \tag{3.179}$$

Keeping the sample size T fixed and considering for the $G_1 + 1$ endogenous variables included in (3.38) the reduced form equations

$$Y = Z\Pi' + V \tag{3.180}$$

where $Y = (y\ Y_1)$ is the $N\times(G_1+1)$ matrix of included endogenous variables, $Z = (Z_1\ Z_2)$ is the NxK matrix of predetermined variables, V

is the $N\times(G_1+1)$ matrix of reduced form disturbance terms and Π is the $(G_1+1)\times K$ matrix of reduced form coefficients. Mariano [52] presented a large μ asymptotic approximation to the distribution function of the k-class estimator summarized in the following theorem whose proof can be found in [52].

Theorem 2: In the case of two included endogenous variables in the equation being estimated, when the sample size T is fixed an approximation to the distribution function of the k-class estimator is given by:

$$\Pr\left\{(\hat{\beta}_{(k)}-\beta)\leq\frac{bw}{\sqrt{\sigma_{22}}}\right\}=\Phi(b)+\frac{\rho}{\mu}\phi(b)[b^2-K_2+1+(1-k)(N-K)]+O(\mu^{-2}) \text{ as } \mu \quad \infty \tag{3.181}$$

Using the results obtained by Mariano (1973), Anderson and Sawa [7] considered the following partition:

$$(y \vdots Y) = (X_1 \vdots X_2)\begin{pmatrix}\pi_{11} & \pi_{12}\\ \pi_{21} & \pi_{22}\end{pmatrix} + (v_1 \vdots v_2) \tag{3.182}$$

Assuming independent normal distribution of the rows of $V = (v_1 \vdots v_2)$, full-column rank of the matrix of predetermined variables, and rank 1 for the matrix $(\pi_{21} \quad \pi_{22})$; they obtained two terms of the asymptotic expansion of the distribution of the 2SLS estimate for increasing sample size and noncentrality parameter and approximated the cumulative distribution function of 2SLS estimate, $\hat{\beta}_{2SLS}$, by:

$$P\left\{\frac{\sqrt{\pi'_{22}A_{22.1}\pi_{22}}}{\sigma}(\hat{\beta}_{2SLS}-\beta)\leq w^*\right\}=\Phi(w^*)-\frac{\beta w_{22}-w_{12}}{\sqrt{\pi'_{22}A_{22.1}\pi_{22}}}$$

$$\times \left[w^{*2} - (K_2 - 1)\right]\phi(w^*) + O\left(\frac{1}{\pi'_{22}A_{22.1}\pi_{22}}\right) \tag{3.183}$$

where $\Omega = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$ is the covariance matrix of each row of V and $\frac{\pi'_{22}A_{22.1}\pi_{22}}{W_{22}}$ is the noncentrality parameter and

$$A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12} \tag{3.184}$$

Once the problem is reduced to canonical form, attention is focused on canonical representation of quadratic forms.

$$P\left\{\hat{\alpha} \le r\right\} = P\left\{w'_2x_2 + hw'_3x_3 \le r(x'_2x_2 + hx'_3x_3)\right\} = P\left\{Q(r) \le 0\right\}$$

where

$$Q(r) = 2[w'_2x_2 + hw'_3x_3 - rx'_2x_2 - rhx'_3x_3] \tag{3.185}$$

and H is a TxT orthogonal matrix such that

$$Hy^* = w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \tag{3.186}$$

$$Hy^* = x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{3.187}$$

The authors assumed that there exists a finite positive number ψ such that

$$\frac{\eta'\eta}{T} = \psi + O(T^{-1}) \tag{3.188}$$

(where $\eta'\eta = \frac{\pi'_{22}A_{22.1}\pi_{22}}{W_{22}}$ and ψ called the reduced noncentrality parameter) which ensures that

$$\operatorname{plim}_{T\to\infty}\hat{\alpha} = \frac{\alpha\psi}{\psi} = \alpha \tag{3.189}$$

Then,

$$P\left\{\sqrt{T}\,(\hat{\alpha} - \operatorname{plim}\alpha) \le x\right\} = P\left\{\frac{Q(r)}{\sqrt{T}} \le 0\right\}$$

(with $r = \alpha + \frac{x}{\sqrt{T}}$). Let $\theta(t)$ denotes the characteristic function of $\frac{Q(r)}{\sqrt{T}}$. The Taylor series of the logarithm of $\theta(t)$ is

$$\operatorname{Log}\theta(t) = -2it\psi x \;-2t(1+\alpha)\psi - \frac{}{\sqrt{T(h+\psi)}}\{2it\alpha\psi K_2 + 8t\alpha\psi x - 8it\alpha\psi$$

$$[(1+\alpha)\psi]\} + O(T^{-1}) \tag{3.190}$$

Therefore, $\frac{Q(r)}{\sqrt{T}}$ has a limiting normal distribution with mean $\mu = -2\psi x$ and variance $4\xi^2$ where

$$\xi^2 = (1 + \alpha^2)\psi \tag{3.191}$$

From here a variable W is defined to be

$$W = \frac{\frac{Q}{\sqrt{T}} - \mu}{2\xi} = \frac{Q + 2(0 + \psi)x\sqrt{T}}{2\sqrt{T}\xi} \tag{3.192}$$

that has limiting distribution N(0, 1). Let $\nu(t)$ denotes the characteristic function of W.

$$\nu(t)=\exp\left[\operatorname{Log}\theta\left(\frac{t}{2\xi}\right)-\frac{it\mu}{2\xi}\right]=e^{-\frac{t^2}{2}}\left\{1+\frac{1}{\sqrt{T}}\left[\kappa_1 it+\kappa_2\left(\frac{x}{\nu}\right)(it)^2 + \kappa_3(it)^3\right] + O(T^{-1})\right. \tag{3.193}$$

where

$$\nu = \frac{\xi}{\psi} \tag{3.194}$$

$$\kappa_1 = \frac{\alpha\psi(-K_2)}{\xi\psi} = \frac{-\alpha K_2}{\xi} \tag{3.195}$$

$$\kappa_2 = \frac{2\alpha\psi^2}{\xi\psi^2} = \frac{2\alpha}{\xi} \tag{3.196}$$

$$\kappa_3 = \frac{\alpha\psi}{\xi^3(0 + \psi)}[- (1 + \alpha^2)\psi] = \frac{-\alpha(1 + \alpha^2)\psi}{\xi^3} \tag{3.197}$$

Then, combined used of the inversion theorem with the well-known result:

$$H_m(w)\phi(w) = \frac{1}{2\pi}\int_{-\infty}^{\infty}(it)^m e^{-t^2/2}e^{-itw}dt \tag{3.198}$$

leads to an expansion of the cumulative distribution function of

$$P\left\{W \le w^*\right\} = \int_{-\infty}^{\infty}\phi(w)dw + \frac{1}{\sqrt{T}}\left[\kappa_1\int_{-\infty}^{\infty}H_1(w)\phi(w)dw + \frac{x}{\nu}\ \kappa_2\int_{-\infty}^{\infty}H_2(w)\phi(w)dw\right.$$

$$\left. + \kappa_3\int_{-\infty}^{\infty}H_3(w)\phi(w)dw\right] + O(T^{-1}) \tag{3.199}$$

where $\phi(w)$ is the density function of $N(0, 1)$; $H_m(w)$'s are the Tchebycheff-Hermite polynomials,

$$H_m(w)\phi(w) = (- \frac{d}{dw})^m\phi(w) \tag{3.200}$$

The approximate density of the k-class estimate is

$$f_T(w^*) = \left\{1 + \frac{1}{\sqrt{T}}[(\kappa_3 + \kappa_2)]w^{*3} + (\kappa_1 - 2\kappa_2 - 3\kappa_3)w^*\right\}\phi(w^*) \tag{3.201}$$

while the corresponding approximate cumulative distribution is

$$F_T(w^*) = \phi(w^*) - \frac{1}{\sqrt{T}}[(\kappa_3 + \kappa_2)w^{*2} + (\kappa_1 - \kappa_3)]\phi(w^*) \tag{3.202}$$

Follows that for the 2SLS (i.e., h = 0) estimate the approximate density and cumulative distribution functions are respectively

$$f_T(w^*) = \left\{1 + \frac{\alpha}{\sqrt{T\psi(1+\alpha^2)}}\left[w^{*3} - (\kappa_2 + 1)w^*\right]\right\}\phi(w^*) \tag{3.203}$$

and

$$F_T(w^*) = \phi(w^*) - \frac{\alpha}{\sqrt{T\psi(1+\alpha^2)}}\left[w^{*2} - (\kappa_2 - 1)\right]\phi(w^*) \tag{3.204}$$

Letting the noncentrality parameter, $\eta'\eta$, increases permits to the authors to obtain the asymptotic expansion of the distribution of 2SLS estimate. Their findings are also that the expression of the distribution of the 2SLS and OLS estimates can be obtained in terms of a doubly-noncentral F distribution. They also provided several forms for the expressions of the exact distribution.

Anderson (1974) [2] extended to the Limited Information Maximum Likelihood (LIML) case, the framework used by Anderson and Sawa in 1973 (to derive the distribution of estimates of coefficients of a single equation in a simultaneous system and their asymptotic expansions). He developed an expansion of the distribution of the LIML estimate of the coefficient of one endogenous variable in an equation with two endogenous variables when the coefficient of the other endogenous variable is prescribed to be unity . The expansion is carried out to terms of the order of the three-halves power of the reciprocal of the

noncentrality parameter when this parameter increases with the sample size. The problem of finding the distribution of the estimate is reduced to its canonical form. Considering

$$Q = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = \frac{1}{\sigma}\begin{pmatrix} 1 & -\beta \\ \dfrac{\beta w_{22} - w_{12}}{\sqrt{|\Omega|}} & \dfrac{w_{11} - \beta w_{12}}{\sqrt{|\Omega|}} \end{pmatrix} \tag{3.205}$$

such that

$$Q\Omega Q' = I \tag{3.206}$$

$$Q\Theta_T Q' = Q\begin{pmatrix} \pi'_{21} \\ \pi'_{22} \end{pmatrix} A_{22.1}(\pi_{21} \quad \pi_{22})Q' = \begin{pmatrix} 0 & 0 \\ 0 & T\nu_T^2 \end{pmatrix} \tag{3.207}$$

where

$$\nu_T^2 = \psi_T w_{22}(\beta \ 1)\Omega^{-1}\begin{pmatrix} \beta \\ 1 \end{pmatrix} = \frac{\psi_T w_{22}\sigma^2}{|\Omega|} \tag{3.208}$$

with

$$\psi_T = \pi'_{22} A_{22.1}\pi_{22}/Tw_{22} \tag{3.209}$$

and letting $b = (b_1 \ b_2)'$ be any (nontrivial) solution to

$$[(1/T)G - \lambda_1\hat{\Omega}]b = 0 \tag{3.210}$$

where λ_1 is the smallest root of

$$|(1/T)G - \lambda_1\hat{\Omega}| = 0$$

$$G = \begin{pmatrix} p'_{21} \\ p'_{22} \end{pmatrix} A_{22.1}(p_{21} \quad p_{22})$$

and

$$P = \left[\begin{pmatrix} Z'_1 \\ Z'_2 \end{pmatrix}(Z_1 \quad Z_2)\right]^{-1}\begin{pmatrix} Z'_1 \\ Z'_2 \end{pmatrix}Y = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}.$$

Anderson obtained that the LIML estimate of β is

$$\hat{\beta}_{LIML} = -\frac{q_{12}b_1^* + q_{22}b_2^*}{q_{11}b_1^* + q_{21}b_2^*} = -\frac{q_{12} + q_{22}(b_2^*/b_1^*)}{q_{11} + q_{21}(b_2^*/b_1^*)} \tag{3.211}$$

where $b = Q'b$. He then proposed

$$\Pr\left\{\frac{\sqrt{T\psi w_{22}}}{\sigma}(\hat{\beta}_{LIML} - \beta) \le w\right\} = \Phi(w) + \left\{-\frac{\alpha}{\mu}w^2 + \frac{1}{2\mu^2}[-(K_2 - 1)w + (2\alpha - 1)w - \alpha w] + \frac{1}{6\mu^3}[3(K_2 - 1)w - (6\alpha + 3K_2 - 12)w + (7\alpha^2 - 3)w^6 - \alpha^2 w^8\right\} + O(\mu^{-4}) \tag{3.212}$$

as cumulative distribution whose approximate density is

$$\left\{1 + \frac{\alpha}{\mu}(w^3 - 2w) + \frac{1}{2\mu^2}[-(K_2 - 1) + (K_2 - 4 + 6\alpha^2)w^2 + (1 - 7\alpha^2)w^4 + \alpha^2 w^6] + \frac{\alpha}{6\mu^3}[6(K_2 - 1)w - (24\alpha^2 + 15K_2 - 51)w^3 + (48\alpha^2 + 3K_2 - 30)w^5 - (15\alpha^2 - 3)w^7 + \alpha^2 w^9]\right\}\phi(w) \tag{3.213}$$

Sargan (1975) [82] published a seminal paper in which a very general theorem on the validity of Edgeworth expansions for sample distributions is established. The approach consists to write the bias of an estimator as a function of a more basic statistics comprising the errors in the sample moments of the data.

Anderson (1976) [3] established an equivalence between the estimation of the slope of a linear functional relationship and the estimation of a coefficient in one equation of a simultaneous system of

stochastic equations, as developed in econometrics. Therefore, one will draw on theorems in the econometric literature to obtain results for the functional relation. Mariano (1973), Anderson and Sawa (1973), and Anderson (1974) have obtained, in the econometric context, asymptotic expansions of the distribution of the 2SLS and the LIML estimates which can be easily translated into terms of the linear functional relation.

Anderson (1977) [4] pursuing in the same order of the idea underlying his paper with Sawa (1973), gave the expansions in the case of the covariance matrix of the disturbances known and alternatively the case of the sample size held fixed. (p_{21}, p_{22}) has multivariate normal distribution with mean (π_{21}, π_{22}) and and covariances

$$\mathcal{E}(p_{21} - \pi_{21})(p_{2j} - \pi_{2j})' = w_{ij}A_{22.1}^{-1}, \quad i,j = 1,2 \tag{3.214}$$

The matrix $G = \begin{pmatrix} p'_{21} \\ p'_{22} \end{pmatrix} A_{22.1}(p_{21} \quad p_{22})$ has a noncentral Wishart distribution with K_2 degrees of freedom, covariance matrix Ω, and noncentrality matrix

$$\begin{pmatrix} \pi'_{21} \\ \pi'_{22} \end{pmatrix} A_{22.1}(\pi_{21} \quad \pi_{22}) = \pi'_{22}A_{22.1}\pi_{22}\binom{\beta}{1}(\beta \quad 1) \tag{3.215}$$

Define

$$C = Y'Y - P'AP = Y'Y - Y'Z(Z'Z)^{-1}Z'Y = Y'[I - Z(Z'Z)^{-1}Z']Y$$

Clearly, C has a Wishart distribution with T - K degrees of freedom and covariance matrix Ω. Now,

$$\hat{\beta}_{2SLS} = \frac{p'_{21}A_{22.1}p_{22}}{p'_{22}A_{22.1}p_{22}} \tag{3.216}$$

and the asymptotic expansion of its distribution is given by

$$\Pr\left\{\frac{\sqrt{\pi'_{22}A_{22.1}\pi_{22}}}{\sigma}(\hat{\beta}_{2SLS} - \beta) \le w\right\} = \Phi(w) - \left\{\frac{\alpha}{\mu}[w^2 - (K_2 - 1)] + \frac{1}{2\mu^2}\right.$$

$$[((K_2 - 1)^2\alpha^2 - (K_2 - 1))w + (1 - 2K_2\alpha^2)w^3 + \alpha^2 w^5] + \frac{1}{6\mu^3}[(K_2^2 - 1)(3K_2 +$$

$$19)\alpha^2 + 3(K_2^2 - 4K_2 + 3) + (11(K_2^2 - 1)\alpha^2 + 3(K_2 - 1)^2)w^2 + 3((K_2 + 1)K_2\alpha^2$$

$$\left. - (2K_2 + 1))w^4 + (3 - 3(3K_2 + 4)\alpha^2)w^6 + \alpha^2 w^8]\right\}\phi(w) + O(\mu^{-4}) \qquad (3.217)$$

where

$$\mu^2 = \pi'_{22}A_{22.1}\pi_{22}\frac{\sigma^2}{|\Omega|} \qquad (3.218)$$

$$\alpha = \frac{\beta w_{22} - w_{12}}{\sqrt{|\Omega|}} \qquad (3.219)$$

Assuming that $\pi'_{22}A_{22.1}\pi_{22}/T$ is bounded, Anderson obtained

$$\Pr\left\{\frac{\sqrt{\pi'_{22}A_{22.1}\pi_{22}}}{\sigma}(\hat{\beta}_{LIML} - \beta) \le w\right\} = \phi(w) - \left\{\frac{\alpha}{\mu}w^2 + \frac{1}{2\mu^2}[(K_2 - 1)w + (1 -\right.$$

$$2\alpha^2)w^3 + \alpha^2 w^5] + \frac{\alpha}{6\mu^3}[-3(K_2 - 1)w^2 + (6\alpha^2 + 3k_2 - 12)w^4 + (3 - 7\alpha^2)w^6 +$$

$$\left.\alpha^2 w^8]\right\}\phi(w) + O(\mu^{-4}) \qquad (3.220)$$

When the noncentrality parameter increases and the covariance matrix is known, $(1/T)C$ is replaced by $C^* = TQ\hat{\Omega}Q'$ In the case of an increasing noncentrality parameter and constant sample size, the elements of C are modified to be as having the Wishart distribution. Follows that

$$\Pr\left\{\frac{\sqrt{\pi'_{22}A_{22.1}\pi_{22}}}{\sigma}(\hat{\beta}_{LIML} - \beta) \le w\right\} = \phi(w) - \left\{\frac{\alpha}{\mu}w^2 + \frac{1}{2\mu^2}[(K_2 - 1 +\right.$$

$$\frac{K_2^2 - 1}{T - K - 2})w + (1 - 2\alpha^2)w^3 + \alpha^2 w^5]\Big\}\phi(w) + O(\mu^{-3}) \tag{3.221}$$

where $\frac{1}{2\mu^{2T - K - 2}}\frac{K_2^2 - 1}{}w\phi(w)$ represents the effect of holding T fixed. The approximate density is

$$\phi(w)\Big\{1 - \frac{\alpha}{\mu}[2w - w^3] + \frac{1}{2\mu^2}[-(K_2 - 1 + \frac{K_2^2 - 1}{T - K - 2}) + (K_2 - 4 + \frac{K_2^2 - 1}{T - K - 2} +$$

$$6\alpha^2)w^2 - (7\alpha^2 - 1)w^4 + \alpha^2 w^6]\Big\}\phi(w) \tag{3.222}$$

from which the MSE is $1 + \frac{1}{\mu^2}[K_2 + 2 + \frac{K_2^2 - 1}{T - K - 2} + 9\alpha^2]$ Turns out that neither the distribution of $\hat{\beta}_{2SLS}$ nor its asymptotic expansion are affected by substitution of the estimate of Ω by Ω (case of the covariance matrix of disturbances known). Also, when the noncentrality parameter increases and the sample size is held fixed (case of sample size fixed), the distribution of the 2SLS estimate is unaffected by the conditions on C, of which estimate of Ω is function. Anderson related these cases to the approach of letting the disturbances decrease. The method used is due to Kadane. Given a positive definite matrix Ψ, he defined

$$\Omega = \tau^2\Psi \tag{3.223}$$

Then G/τ^2 has noncentral Wishart distribution with K_2 degrees of freedom, covariance matrix Ψ, and noncentrality matrix

$$\frac{\pi'_{22}A_{22.1}\pi_{22}}{\tau^2}\binom{\beta}{1}(\beta\ 1) \tag{3.224}$$

Follows that the behavior of $\hat{\beta}_{2SLS}$ as $\tau^2 \longrightarrow 0$ is identical to its behavior as $\pi'_{22}A_{22.1}\pi_{22} \longrightarrow \infty$. Let $b = (1\ -\hat{\beta})'$ minimizes

$$\frac{b'Gb}{b'\frac{1}{T}Cb} = \frac{b'\frac{1}{\tau^2}Gb}{b'\frac{1}{\tau^2}\ \frac{1}{T}Cb} \tag{3.225}$$

Then, the behavior of $\hat{\beta}_{LIML}$ as $\tau^2 \rightarrow 0$ is identical to its behavior as $\pi'_{22}A_{22.1}\pi_{22} \rightarrow \infty$ with T fixed.

Holly and Phillips (1979) [63] proposed a new approximation based on the saddlepoint method of approximating integrals. The method was applied to derive the probability density of the k-class estimator in the case of the equation with two endogenous variables. The two tails of the density are approximated by different functions, each of which bears a close relationship with the exact density in the same region of the distribution. Given the single structural equation:

$$y_1 = \beta y_2 + Z_1\gamma_1 + u \tag{3.226}$$

where y_1 and y_2 are vectors of T observations on two endogenous variables, Z_1 is a $T \times K_1$ matrix of observations on K_1 exogenous variables, and u is a vector of random disturbances. The structural coefficients are the scalar parameter β and the the parameter vector γ_1. The reduced form for the two endogenous variables in (3.226) is given by

$$Y = Z\Pi + V \tag{3.227}$$

where $Y = [y_1 \vdots y_2]$, $Z = [Z_1 \vdots Z_2]$ is a $T \times K$ $(K = K_1 + K_2)$ matrix of exogenous variables, and $V = [v_1 \vdots v_2]$ is a matrix of reduced form disturbances. (3.227) can be written in the form

$$[y_1 \vdots y_2] = [Z_1 \vdots Z_2]\begin{bmatrix}\pi_{11} & \Pi_{12} \\ \pi_{21} & \Pi_{22}\end{bmatrix} + [v_1 \vdots v_2]$$

$$= Z_1[\pi_{11} \vdots \Pi_{12}] + Z_2[\pi_{21} \vdots \Pi_{22}] + [v_1 \vdots v_2] \tag{3.228}$$

It is assumed: that each row of $[v_1 \vdots v_2]$ is iid as a normal vector with zero mean and nonsingular covariance matrix

$$\Omega = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$$

that equaton (3.226) is identified by zero restrictions in the structural coefficients; that the observation matrix is nonstochastic, of rank K, and T > K; and that, for simplicity, $Z_1'Z_2 = 0$. The k-class estimator $\hat{\beta}_k$ of β in equation (3.226) has the form

$$\hat{\beta}_k = \frac{y_2'A_k y_1}{y_2'A_k y_2} \tag{3.229}$$

where

$$A_k = (1 - k)\{I - Z_1(Z_1'Z_1)^{-1}Z_1' - Z_2(Z_2'Z_2)^{-1}Z_2'\} + Z_2(Z_2'Z_2)^{-1}Z_2' \tag{3.230}$$

and k is nonstochastic. Define

$$\beta^* = \sqrt{\frac{W_{22}}{W_{11} - W_{12}^2/W_{22}}}\,(\beta - \frac{W_{12}}{W_{22}}) \tag{3.231}$$

The k-class estimator of β^* in the transformed system is given by

$$\hat{\beta}_k^* = \sqrt{\frac{W_{22}}{W_{11} - W_{12}^2/W_{22}}}\,(\hat{\beta}_k - \frac{W_{12}}{W_{22}}) \tag{3.232}$$

(for convenience, we will drop the asterisk on $\hat{\beta}_k^*$). Let $L(w_1, w_2)$ denote the Laplace transform of the joint density of $y_2'A_k y_1$ and $y_2'A_k y_2$. We have

$$L(w_1, w_2) = (1 - 2w_1 - w_2^2)^{-K_2/2}(1 - 2hw_1 - h^2w_2^2)^{-(T-K)/2}$$

$$\times\exp\left[\frac{\mu^2}{2}\left\{\frac{(1 + \beta w_2)^2}{1-2w_1-w_2^2}\right\}\right]$$

where

$$\mu^2 = \frac{\pi_{22}'(Z_2'Z_2)\pi_{22}}{w_{22}} \quad \text{and } h = 1 - k \tag{3.233}$$

Droping the subscript on w_2, we find that

$$\left.\frac{\partial L(u-rw,\ w)}{\partial u}\right|_{u=0} = \left\{K_2(1 + 2hrw - h^2w^2) + h(T - K)(1 + 2rw - w^2) + \mu^2 \frac{(1 + 2hrw - h^2w^2)(\beta w + 1)^2}{(1 + 2rw - w^2)}\right\} \times (1 + 2rw - w^2)^{-(K_2+2)/2} \times\exp\left\{\frac{\mu^2}{2}\frac{w^2(1 + \beta^2) + 2w(\beta - r)}{1 + 2rw - w^2}\right\} \tag{3.234}$$

which can be written in the form

$$B(w)\exp\left\{\frac{\mu^2}{2}\psi(w)\right\} \tag{3.235}$$

and then

$$f(r) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} B(w)\exp\left\{\frac{\mu^2}{2}\psi(w)\right\}dw \tag{3.236}$$

The essence of the saddlepoint method is to select the path of integration in (3.237) in such a way that the major contribution to the value of the integral comes from the value of the integrand in a region

of a saddlepoint on the real axis. Changing the variable of integration in (3.236) from w to y in $w = w^0 + iy$ -w^0 being a suitable saddlepoint- we have

$$h(r) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} B(w^0 + iy)\exp\left\{\frac{\mu^2}{2}\psi(w^0 + iy)\right\}dy \tag{3.237}$$

Since B(.) and ψ(.) are analytic on the path of integration, their Taylor expansions exist and can be utilized to reduce (3.237) to

$$\frac{B(w^0)\exp\left\{\frac{\mu^2}{2}\psi(w^0)\right\}}{\pi^{1/2}\mu(\psi''(w^0))^{1/2}}\left[1 + \frac{1}{\mu^2}\left\{-\frac{B''(w^0)}{B(w^0)\psi''(w^0)} + \frac{1}{4}\frac{\psi^{(4)}(w^0)}{\psi''(w^0)} + \frac{\psi^{(3)}(w^0)\ B'(w^0)}{(\psi''(w^0))^2 B(w^0)} - \frac{5}{12}\frac{(\psi^{(3)}(w^0))^2}{(\psi''(w^0))^3}\right\} + o(\mu^{-4})\right] \tag{3.238}$$

where

$$\psi'(w) = \frac{2\beta(\beta r + 1)(w - w_1^0)(w - w_2^0)}{(1 + 2rw - w^2)^2} \tag{3.239}$$

(with $w_1^0 = (r - \beta)/(1 + \beta r)$ and $w_2^0 = -1/\beta$)

$$\psi''(w_1^0) = \frac{2(\beta r + 1)^4}{(1 + 2\beta r - \beta^2)(r^2 + 1)^2} \tag{3.240}$$

and

$$\psi''(w_2^0) = \frac{2\beta^2}{\beta^2 - 2\beta r - 1} \tag{3.241}$$

The first equation in (3.238) is the saddlepoint approximation and the series is sometimes referred to as the saddlepoint expansion. To specify the first factor of equation (3.238), the evaluation of $B(w^0)$

and $\psi(w^0)$ is needed. In the case of $\beta > 0$ the saddlepoint approximation, when $r < \frac{\beta^2 - 1}{2\beta}$, is given by

$$h(r) = \frac{1}{\sqrt{2\pi}\mu}\{K_2(\beta^2 - 2h\beta r - h^2) + h(T - K)(\beta^2 - 2\beta r - 1)\}\times\beta^{T-K_1}(\beta^2 - 2\beta r - 1)^{-(K_2+1)/2}(\beta^2 - 2h\beta r - h^2)^{-(T-K+2)/2}e^{-\mu^2/2}$$

and for $r > \frac{\beta^2 - 1}{2\beta}$, the saddlepoint approximation is:

$$h(r) = \frac{1}{\sqrt{2\pi}\mu}[K_2\{(\beta r + 1)^2 + 2rh(r - \beta)(\beta r + 1) - h^2(r - \beta)^2\} + h\{T + K)(1 + 2\beta r - \beta^2)(r^2 + 1) + \mu^2\{(\beta r + 1)^2 + 2rh(r - \beta)(\beta r + 1) - h^2(r - \beta)^2\}(1 + 2\beta r - \beta^2) \times (r^2 + 1)^{-1}](\beta r + 1)^{T-K_1}(1 + 2\beta r - \beta^2)^{-(K_2+1)/2}(r^2 + 1)^{-K_2/2} \times \{(\beta r + 1)^2 + 2rh(r - \beta)(\beta r + 1) - h^2(r - \beta)^2\}^{-(T-K+2)/2}\times \exp\{-\mu^2(r - \beta)^2/2(r^2 + 1)\}$$

Morimune and Kunitomo (1980) [58'] gave a sort of a fusion of the papers by Morimune (1978) and Anderson (1976). They consider a structural equation

$$y_{1t} - \beta y_{2t} = \gamma_{11} + \sum\gamma_{1k}z_{kt} + u_{1t} \quad t = 1,\ldots,T \tag{3.242}$$

where u_{1t} is arandom variable with mean 0 and variance ρ. Corresponding reduced-form equations are:

$$y_{1t} = \pi_{11} + \sum\pi_{1k}z_{kt} + v_{1t} \tag{3.243}$$

$$y_{2t} = \pi_{21} + \sum \pi_{2k} z_{kt} + v_{2t} \quad , \; k = 2, \ldots, K \tag{3.244}$$

such that v_{1t} and v_{2t} are normally distributed with mean 0 and covariance matrix

$$\Omega = \sigma^2 \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \tag{3.245}$$

$$\text{let } L = K - K_1 - 1 \tag{3.246}$$

be the degree of overidentification. Morimune and Kunitomo considered the less biased estimate than $\hat{\beta}_{LIML}$ and $\hat{\beta}_{2SLS}$,

$$\hat{\beta} = ((L-1)/L)\hat{\beta}_{LIML} + (1/L)\hat{\beta}_{2SLS} \tag{3.247}$$

and proved that $\hat{\beta}$ improves uniformly $\hat{\beta}_{LIML}$ in terms of $AM_{\lambda}2$ given L large and Ω (unknown).

Kunitomo (1980) [43]; pursuing from the above result, derived the asymptotic expansions of the distributions of the estimators in a linear functional relationship and discussed some implications for econometrics. Especially, he had pointed out that if the sample size T is large enough, the asymptotic expansions obtained under assumption that Ω is known to within a proportionality constant may still be valid since one may estimate the covariance matrix by using the residual matrix of the regression estimators of the reduced form parameters π_{ij}.

Morimune (1981) [58] derived the asymptotic expansions of the distribution of the improved limited information maximum likelihood estimator proposed by the author in 1978. Consider (3.226) and assume the rows of V are independently normally distributed with mean 0 and

covariance matrix $\Omega = \|w_{ij}\|$ $i,j = 1,2$. The improved estimator is defined to be a combination of LIML and k-class estimators. It is given by

$$\hat{\beta}_{com} = \frac{L - 1 + (T - K)h}{L + (T - K)h}\hat{\beta}_{LIML} + \frac{1}{L + (T - K)}\hat{\beta}_k \tag{3.248}$$

where β_k is the fixed k-class estimator of β and h is (1-k). The large-μ asymptotic expansion of the distribution of $\hat{e}_{com}$ for $T - K > 2$ is given by

$$P\{\hat{e}_{com} \le \xi\} = \Phi(\xi) - \frac{\alpha}{\mu}(\xi^2 - 1)\phi(\xi) + \frac{\xi}{2\mu^2}\{(2 - \xi^2) + \alpha^2(1 + 4\xi^2 - \xi^4)$$

$$- \frac{1}{m}[\frac{L(L + T - K)(m - 1)^2}{(T - K - 2)} + (L + h^2(T - K))(1 + 2\alpha^2) + 2L(m - 1)(1 -$$

$$h)]\}\phi(\xi) + O(\mu^{-3}) \tag{3.249}$$

as μ increases while $T - K$ stay fixed. m is $L + (T - K)h$

$$\alpha = (\beta w_{22} - w_{12})/(|\Omega|^{1/2}) \tag{3.250}$$

$$\mu^2 = (1 + \alpha)\pi'_{22}X'_2(I - X_1(X'_1X_1)^{-1}X'_1)X_2\pi_{22}/w_{22} \tag{3.251}$$

and $\hat{e} = \mu\frac{\sqrt{|\Omega|}}{\sigma^2}(\hat{\beta} - \beta)$ is the standardized estimator. Let $AMSE_{\mu^2}$ be the asymptotic MSE based on the large-μ expansion. Morimune obtained

$$AMSE_{\mu^2}(\hat{e}_{com}) = 1 + \{(1 + 2\alpha^2)(m^2 + L + (T - K)h^2) + L(L + T- K)(m$$

$$-1)^2/(T - K - 2) + 2L(m - 1)(1 - h)\}/(\mu^2m^2) \le AMSE_{\mu^2}(\hat{e}_{LIML}) = 1 + \{3 +$$

$$9\alpha^2 + L(L + T - K)/(T - K - 2)\}/\mu^2 \tag{3.252}$$

He, then, concluded that (3.219) is a convex combination of the two

estimators.

Fujikoshi, Morimune, Kunitomo, and Taniguchi (1982) [36] derived asymptotic expansions for the density functions of the 2SLS and LIML estimates of coefficients in a simultaneous equation model when the sample size as well as the effect of the exogenous variables increase. Their results are somewhat a direct generalization of the results obtained Anderson and Sawa (1973). Unlike the latters, the authors used the perturbation method and applied the Fourrier inverse transformation. Model (3.74) is considered and it is assumed that the rows of $(v_1 \quad v_2)$ are independently and normally distributed, each row having mean 0 and (non-singular) matrix

$$\Omega = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & \Omega_{22} \end{pmatrix} \tag{3.253}$$

Furthermore, the authors assume the matrix $(\pi_{21} \quad \pi_{22})$ is of rank G_1, π_{22} also is of rank G_1, and $A = O(1)$. Define

$$L = K_2 - G_1 \equiv \text{degree of overidentification} \tag{3.254}$$

$$q' = \frac{1}{\sigma^2}(w_{12} - \beta'\Omega_{22}0') \; : \; 1\times(G_1 + K_1) \tag{3.255}$$

$$C_1 = qq' \; : (G_1 + K_1)\times(G_1 + K_1) \tag{3.256}$$

$$C_2 = \begin{pmatrix} \frac{1}{\sigma^2}\Omega_{22} & 0 \\ 0 & 0 \end{pmatrix} - C_1 \; : (G_1 + K_1)\times(G_1 + K_1) \tag{3.257}$$

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \frac{1}{T}(X_1X_2)'(X_1X_2) \; : \; (K_1 + K_2)\times(K_1 + K_2)$$

$$P_F = F(F'F)^{-1}F' \text{ and } \bar{P}_F = I - P_F \tag{3.258}$$

and

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} : (G_1 + K_1)\times(G_1 + K_1) \tag{3.259}$$

where

$$Q_{11} = (\Pi'_{22}A_{22.1}\Pi_{22})^{-1}, \quad Q_{21} = -A_{11}^{-1}(A_{11} \;\; A_{12})\Pi_2 Q_{11}$$

$$Q_{22} = A_{11}^{-1} + Q_{21}Q_{11}^{-1}Q_{12} \text{ and } A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12} \tag{3.260}$$

Then the 2SLS estimate of β is

$$\hat{\beta}_{2SLS} = [y'_2(P_X - P_{X_1})y_2]^{-1}[y'_2(P_X - P_{X_1})y_1] \tag{3.261}$$

and the LIML estimate is the vector satisfying

$$[Y'_w(P_x - P_{x_1})Y_w - \lambda_0 Y'_w\bar{P}_x Y_w](1 - \hat{\beta}')' = 0 \tag{3.262}$$

where λ_0 is the smallest root of

$$|Y'_w(P_x - P_{x_1})Y_w - Y'_w\bar{P}_x Y_w| = 0 \tag{3.263}$$

The asymptotic expansions of the joint density functions of 2SLS estimator are derived for the statistic

$$\hat{e}_{2S} = \begin{pmatrix} e_\beta \\ e_\gamma \end{pmatrix} = \sqrt{T}\begin{pmatrix} \hat{\beta}-\beta \\ \hat{\gamma}-\gamma \end{pmatrix} \text{ as } T \longrightarrow \infty. \tag{3.264}$$

(Note: a single structural equation is written:

$$\underset{T\times 1}{y_1} = \underset{T\times G_1}{y_2}\;\underset{G_1\times 1}{\beta} + \underset{T\times K_1}{Z_1}\;\underset{K_1\times 1}{\gamma} + \underset{T\times 1}{u}\;)$$

It is given by

$$f_{2S}(\xi) = \phi_{\sigma^2 Q}(\xi)\Big\{1 + \frac{1}{\sqrt{T}}(q'\xi)(G_1 + K_1 + 1 + L - \xi'(\sigma^2 Q)^{-1}\xi) +$$

$$\frac{1}{2T}\{L\sigma^2 tr(C_2 Q) - L^2\sigma^2 tr(C_1 Q) - \xi' C_2 \xi(G_1 + K_1 + 2 + L - \xi'(\sigma^2 Q)^{-1}\xi) +$$

$$\xi' C_1 \xi[(G_1 + K_1 + 1 + L - \xi'(\sigma^2 Q)^{-1}\xi)^2 + (G_1 + K_1 + 1 + 2L - 3\xi'(\sigma^2 Q)^{-1}\xi)]\}\Big\}$$

$$+ O(T^{-3/2}) \qquad (3.265)$$

For the LIML, we have

$$f_{LI}(\xi) = \phi_{\sigma^2 Q}(\xi) + \Big\{1 + \frac{1}{\sqrt{T}}(q'\xi)(G_1 + K_1 + 1 - \xi'(\sigma^2 Q)^{-1}\xi) + \frac{1}{2T}$$

$$\{-L\sigma^2 tr(C_2 Q) - \xi' C_2 \xi(G_1 + K_1 + 2 - L - \xi'(\sigma^2 Q)^{-1}\xi) + \xi' C_1 \xi\ [(G_1 + K_1 + 1$$

$$- \xi'(\sigma^2 Q)^{-1}\xi)^2 + G_1 + K_1 + 1 - 3\xi'(\sigma^2 Q)^{-1}\xi]\}\Big\} + O(T^{-3/2}) \qquad (3.266)$$

An integration of (3.265) with $G_1 = 1$ and with respect to the last K_1 elements of ξ gives the asymptotic expansion of the distribution of $\hat{e}_\beta$ identical to the result by Anderson and Sawa (1973) to term $O(T^{-1})$. A comparison of $\hat{\beta}_{2SLS}$ and $\hat{\beta}_{LIML}$ is made in terms of the mean square error and the concentration of probability around the true β for the general case of G_1. If $l \le 6$, the mean of $\hat{e}_{LI}\hat{e}'_{LI}$ in terms of the asymptotic expansions of the distribution $\hat{e}_{LI}\hat{e}'_{LI}$ up to $O(T^{-1})$ is at least equal to that of $\hat{e}_{2S}\hat{e}'_{2S}$ (i.e $AM_T(\hat{e}_{LI}\hat{e}'_{LI}) \ge AM_T(\hat{e}_{2S}\hat{e}'_{2S})$). For a comparison of the probability of concentration around the true value β, authors defined $\|x\| = \max\{|x_1|, \ldots, |x_{G_1+K_1}|\}$ and computed

$$P\Big\{\|(\sigma^2 Q)^{-1/2}\hat{e}_{LI}\| < w\Big\} - P\Big\{\|(\sigma^2 Q)^{-1/2}\hat{e}_{2S}\| < w\Big\} =$$

$$\int_{\|(\sigma^2 Q)^{-1/2}\xi\| < w} \cdots\cdots \int (f_{LI}(\xi) - f_{2S}(\xi)) \tag{3.267}$$

$$= \frac{L}{T}\left\{\Phi(w) - \Phi(-w)\right\}^{G_1+K_1} w\varphi(w)\sigma^2 tr(QC_1)D + O(T^{-3/2}) \tag{3.268}$$

where

$$\varphi(w) = \phi(w)/[\Phi(w) - \Phi(-w)] \tag{3.269}$$

$$D = L - 2(K_1 + G_1 - 1)(1 - 2w\varphi(w)) - 2\{tr(QC_1)\} \quad tr(QC_2) - 2w \tag{3.270}$$

and $\Phi(.)$ and $\phi(.)$ are the standard normal distribution and density functions. Let a_{min} and a_{max} be the smallest and the largest characteristic roots of Q_{11}. If $(a_{max}/a_{min})(L + 2)(trC_2/trC_1 + 1)^{-1} \leq 2$,

$$P\left\{\|(\sigma^2 Q)^{-1/2}\hat{e}_{LI}\| < w\right\} \leq P\left\{\|(\sigma^2 Q)^{-1/2}\hat{e}_{2S}\| < w\right\} + O(T^{-3/2})$$

If $w^2 \leq \frac{1}{2}(1 + 4 - 2K_1 - 2G_1) - (a_{max}/a_{min})(trC_2/trC_1 + 1)^{-1}$

$$P\left\{\|(\sigma^2 Q)^{-1/2}\hat{e}_{LI}\| < w\right\} \geq P\left\{\|(\sigma^2 Q)^{-1/2}\hat{e}_{2S}\| < w\right\} + O(T^{-3/2})$$

It follows that the small value of trC_2/trC_1 favors the LIML estimate, and the large value of trC_2/trC_1 favors the 2SLS estimate. One can write $(1 + trC_2/trC_1)^{-1} = E(u'v_2)E(v_2'u)/E(u'u)tr\{E(v_2'v_2)$ such that high correlation between u and y_2 favors LIML estimate.

Morimune and Tsukuda (1984) [59] derived asymptotic expansions of three alternative classes of structural coefficients for two parameter sequences: a sequence in which the non-centrality parameter increases while the sample size stays fixed and that in which the number of observations increases. The accuracy of approximations to small-sample

distribution are numerically examined with help of Monte Carlo studies. The authors also studied properties of the sum of squared residuals of an estimated structural equation.

Kunitomo (1986) [43'] proved that the LIML estimator in the simultaneous equation system is third-order asymptotically efficient when the number of excluded exogenous variables in a particular structural equation increases as the sample size increases. He argued that in the large sample asymptotic theory in econometrics, the LIML and 2SLS estimators are best asymptotically normal (BAN) estimators. Thus, these estimators can be modified in hope to improve BAN estimators in some sense. Given that K_2, in macroeconometric models, is fairly large even if there are only two endogenous variables in a particular structural equation, Kunitomo suggested the new large-K_2 asymptotics theory for large econometric models.

Kunitomo (1988) [44] studied the distributions of the test statistics for over-identifying restrictions in a system of simultaneous equations under the null and non-null hypotheses. The effects of the normality assumption for disturbances on the test statistics and their power functions based on their asymptotic expansions were investigated.

BIBLIOGRAPHY

[1] Amemiya, T. (1966b), "On the Use of Principal Components of Independent Variables in Two-Stage Least Squares Estimation". International Economic Review, 7: 283-303.

[2] Anderson, T.W. (1974), "An Asymptotic Expansion of Distribution of the Limited Information Maximum Likelihood Estimate of a Coefficient in a Simultaneous Equation System ". Journal of The American Statistical Association. 69, pp. 565-573.

[3] Anderson, T.W. (1976), "Estimation of Linear Functional Relationships: Approximate Distributions and Connections with Simultaneous Equations in Econometrics". Journal of The Royal Statistical Society, B. 38, pp. 1-36. 36.

[4] Anderson, T.W. (1977), "Asymptotic expansion of the distributions of estimates in simultaneous equations for alternative parameter sequences". Econometrica. 45, pp. 509-517.

[5] Anderson, T.W. (1982), "Some recent developments on the distributions of single equation estimators". Advances in Econometrics. Volume 1, pp. 109-122.

[6] Anderson, T.W, and H. Rubin, (1949). "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations". Annals of Mathematical Statistics, 20, pp. 46-63.

[6']Anderson, T.W, and H.Rubin, (1950). "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations". Annals of Mathematical Statistics, 21, pp. 570-582.

[7] Anderson, T.W. & Sawa, T (1973), "Distribution of estimates of coefficients of a single equation in a simultaneous system and their asymptotic expansions". Econometrica. 41, pp. 683-717.

[8] Anderson, T.W. & Sawa, Takamitsu (1979), "Evaluation of the distribution function of the two stage least squares estimate". Econometrica. 47, pp.163-182.

[9] Anderson, T.W. & Sawa, T. (1982), "Exact and Approximate Distributions of the Maximum Likelihood Estimator of a Slope Coefficient". Journal of The Royal Society (B). 44, pp. 52-62. 62.

[10] Anderson, T.W.;Morimune, K. & Sawa, T.(1983), "The numerical values of some key parameters in econometric models". Journal of Econometrics. 21, pp. 229-243.

[11] Basmann,R.L.(1957),"A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation". Econometrica, 25: 77-83.

[12] Basmann, R.L. (1959), "The Computation of Generalized Classical Estimates of Coefficients in a Structural Equation".

Econometrica, 27: 72-81.

[13] Basmann, R.L. (1960), "On the Asymptotic Distribution of Generalized Linear Estimators". Econometrica, 28: 97-108.

[14] Basmann, R.L. (1974), "Exact finite sample distribution for some econometric estimators and test statistics: a survey and appraisal". Frontiers of Quantitative Economics, Volume II, ed. M.D. Intriligator and D.A. Kendrich. 1974 North-Holland publishing Company. ny.

[15] Basmann, R.L. and D.H. Richardson (1973) "The exact finite sample distribution of a non-consistent structural variance estimator". Econometrica. Volume 41, pp. 41-58. 58.

[16] Basmann, R.L., Franklin Lee Brown, William S. Dawes, and Gregory K. Schoepfle (1971), "Exact Finite Sample Density Functions of GCL Estimators of Structural Coefficients in a Leading Exactly Identifiable Case". Journal of the American Statistical Association. Volume 66: 122-126.

[17] Bridger M. Mitchell (1974), "Fixed Point Estimation of Econometric Models". Australian Economic Papers. pp. 250-266.

[18] Brundy, J.M., and D.W. Jorgenson (1971), "Efficient Estimation of Simultaneous Equations by Instrumental Variables". Review of Economics and Statistics, 3: 207-24.

[19] Brundy, J.M., and D.W. Jorgenson (1973), "Consistent and Efficient Estimation of System of Simultaneous Equations", in P. Zarembka, Ed., Frontiers in Econometrics. New-York: Academic Press.

[20] Brundy, J.M., and D.W. Jorgenson (1974),"The Relative Efficiency of Instrumental Variables Estimators of Systems of Simultaneous Equations". Annals of Economics and Social Measurement, 3: 679-700.

[21] Carter, R.A.L. (1976), "The exact distribution of an instrumental variables estimator". International Economic Review. Vol. 17. pp.228-233. 33.

[22] Dhrymes, P.J. (1969), "Efficient Estimation of Distributed Lags with Autocorrelated Errors". International Economic Review, 10, pp. 47-67.

[23] Dhrymes, P.J. (1970), Econometrics: Statistical Foundations and Applications. New-York: Harper & Row.

[24] Dhrymes, P.J., and H. Erlat (1974), "Asymptotic Properties of Full Information Estimation in Dynamic Autoregressive Simultaneous Equation Model". Journal of Econometrics, 2: 247-60.

[25] Duesenberry, J. S., O. Eckstein, and G. Fromm (1960), "Simulation of the United States Economy in Recession". Econometrica, 28: 749-809

[26] Duesenberry, J. S., G. Fromm, L.R. Klein, and E. Kuh, Eds. (1965), The Brookings Quaterly Econometric Model of United States. Chicago: Rand-McNally & Company; Amsterdam: North-Holland Publishing Co.

[27] Dutta, M., and E. Lyttkens (1974), "Iterative Instrumental Variables Method and Estimation of a Large Simultaneous System".Journal of the American Statistical Association, 69: 977-86.

[28] Dwivedi, T.D. (1987), A monograph on Simultaneous Equations. Concordia University. Department of Mathematics and Statistics.(memeograph)

[29] Eckstein, O., E. W. Green, and A. Sinai (1974), "The Data Resources Model: Uses, Structure, and Analysis of the U.S. Economy". International Economic Review, 15: 595-615.

[30] Evans, M. K., and L.R. Klein (1967), The Wharton Econometric Forecasting Model. Philadelphia: Economics Research Unit, Wharton School, University of Pennsylvania.

[31] Evans, M. K., and L.R. Klein (1967), The Wharton Econometric Forecasting Model, 2nd Enlarged Ed. Philadelphia: Economics Research Unit, Wharton School, University of Pennsylvania.

[32] Fair,R. C.(1971), A Short-Run Forecasting Model of the United States Economy. Lexington: Heath Lexington Books.

[33] Fair, R.C. (1972), " Efficient Estimation of Simultaneous Equations with Autoregressive Errors by Instrumental Variables". Review of Economics and Statistics, 54:444-9.

[34] Fair, R. C. (1974), "An Evaluation of a Short-Run Forecasting Model". International Economic Review, 15: 285-303.

[35] Fromm, G., and L. R. Klein, Eds. (1975), The Brookings Model: Perspective and Recent Developments. Amsterdam: North-Holland Publishing Co.

[36] Fujikoshi, Y.; Morimune, K.; Kunitomo, N. & Taniguchi, M (1982), "Asymptotic expansions of the distributions of estimates of coefficients in a simultaneous equation system". Journal of Econometrics. 18, pp. 191-205.

[37] Grant H. Hillier (1985), "On the joint and marginal densities of instrumental variable estimators in a general structural equation". Econometric Theory, 1: 53-72.

[38] Haavelmo, T. (1943), "The Statistical Implications of a System of Simultaneous Equations". Econometrica, 11: 1-12. 12.

[39] Haavelmo, T. (1947), "Methods of Measuring the Propensity to Consume". Journal of the American Statistical Association, 42: 105-22.

[40] Hymans, S. H., and H. T. Shapiro (1974). "The Structure and Properties of the MQEM of the U.S. Economy". International Economic Review.

[41] Klein, L. R. (1950), Economic Fluctuations in the United States, 1921-1941. Cowles Commission Monograph No. 11. New-York: John Wiley & Sons, Inc.

[42] Klein, L. R., and A. S. Goldberg (1955), An econometric of The United States, 1929-1952. Amsterdam: North-Holland Publishing Co.

[43] Kunitomo, Naoto (1980), "Asymptotic Expansions of the Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations". Journal of The American Statitical Association. 75, pp. 693-700.

[43'] Kunitomo, Naoto (1986), "A Third Order Optimum Property of the M L Estimator in a Linear Functional Relationship Model and Simultaneous Equation System in Econometrics". Ann. Inst. Statist. Math. 39 (1987), Part A, 575-591.

[44] Kunitomo, Naoto (1988), "Approximate distributions and power of test statistics for overidentifying restrictions in a system of simultaneous equations". Econometric Theory, 4, 1988, 248-274.

[45] Liebenberg, M.,G. Green, and A. Hirsch (1971), "The OBE 1970 Quaterly Econometric Model".Washington, D.C.: U.S Department of

Commerce, Econometrics Branch.

[46] Liu, T C. (1963), "An Exploratory Quaterly Econometric Model of Effective Demand in the Post-War U. S. Economy". Econometrica 31: 301-48.

[47] Liu, T. C., and E. C. Hwa (1974), "Structure and Applications ofof a Monthly Econometric Model of The U. S. Economy". International Economic Review, 15: 328-65.

[48] Madansky, A (1964), "On the Efficiency of Three Stage Least Squares Estimation". Econometrica, 32: 51-6

[49] Madansky, A (1976), Foundation of Econometrics. Amsterdam: North-Holland Publishing Co.

[50] Malinvaud, E. (1970), Statistical Methods of Econometrics, 2nd Rev. Ed. Amsterdam: North-Holland Publishing Co.

[51] Mariano, Roberto S. (1973), "Approximations to the Distribution functions of the ordinary least squares and two stage least squares estimators in the case of two included endogenous variables". Econometrica. 41, pp. 67-77

[52] Mariano, Roberto S. (1973), "Approximations to the Distribution of Theil's K-Class Estimators". Econometrica. 41, pp. 715-721. 21.

[53] Mariano, Roberto S. (1982), "Analytical small-sample distribution theory in econometrics: the simultaneous-equations case". International Economic Review. Vol.23, pp. 503-533. 33.

[54] Mariano, Roberto S, and Takamitsu Sawa (1972), "The Exact Finite-Sample Distribution of the Limited Information Maximium Likelihood Estimator in the Case of Two Included Endogenous Variables". Journal of the American Statistical Association. Volume 67: 159-163.

[55] McCarthy, M. D. (1972), The Wharton Quaterly Econometric Forecasting Model, Mark III. Philadelphia: Economics Research Unit, University of Pennsylvania.

[56] McCarthy, Michael D. (1987), "Exact finite sample test statistics for restricted reduced forms: A generalization of "Student's" result". International Economic Review. Vol. 28.

[57] Mitchell, B. (1971), "Estimation of Large Econometric Models by Principal Component and Instrumental Variable Methods". Review of Econometrics and Statistics, 53: 140-6.

[57']Morimune, Kimio (1978), "Improving the Limited Information Maximum Likelihood Estimator When the Disturbances Are Small". Journal of the American Statistical Association, 73, 867-871.

[58] Morimune, Kimio (1981), "Asymptotic Expansions of the Distribution

of an Improved Limited information Maximum Likelihood Estimator". Journal of The American Statistical Association. Volume 76, pp. 476-478.

[58']Morimune, K. & Kunitomo, N. (1980), "Improving the Maximum Likelihood Estimate in Linear Functional Relationships for Alternative Parameter Sequences". Journal of the American Statistical Association, March 1980, Volume 75, Number 369.

[59] Morimune, K & Tsukuda, Y. (1984), "Asymptotic expansions of the distributions of the structural variance estimators in a simultaneous equations system".Journal of Econometrics, pp. 279-292.

[60] Nagar, A.L. (1962), "Double k-Class Estimators of Parameters in Simultaneous Equations and Their Small Sample Properties". International Economic Review, 3, pp. 168-188.

[61] Nerlove, M. (1966), "A tabular Survey of Macroeconometric Models". International Economic Review, 7: 127-75.

[62] Oi, W.Y. (1969), "On the Relationship among Different Members of the k-Claas". International Economic Review, 10, pp. 36-46 -46

[63] Phillips, P.C.B. and A. Holly (1979), "A saddlepoint approximation to the distribution of the k-class estimator of a coefficient in a simultaneous system". Econometrica, Vol. 47, pp. 1527-1546.

[64] Phillips, P.C.B. (1980a), "The exact distribution of instrumental variable estimators containing in an equation containing n + 1 endogenous variables". Econometrica. 48, pp. 861-873.

[65] Phillips, P.C.B. (1980b), "Finite Sample Theory and the Distributions of Alternative Estimators of the Marginal Propensity to consume". Review of Economic Studies. XLVII, pp. 183-224.

[66] Phillips, P.C.B. (1983a), "Marginal densities of instrumental variable estimators in the general single equation case". Advances in Econometrics. Volume 2, pp. 1-24.

[67] Phillips, P.C.B.(1983b) "Exact small sample theory in the simultaneous equations model". Handbook of Econometrics, Volume I, Edited by Z. Griliches and M.D. Intriligator. North-Holland Publishing Company, 1983.

[68] Phillips, P.C.B.(1984a), "The exact distribution of LIML: I". International Economic Review, 25: 249-261.

[69] Phillips, P.C.B. (1984b), "The exact distribution of exogenous variable coefficient estimators". Journal of Econometrics. 26 (1984) 387-398. North-Holland.

[70] Phillips, P.C.B. (1985b), "The exact distribution of LIML: II". International Economic Review, 27: 21-36.

[71] Phillips, P.C.B. (1986), "The distribution of FIML in the leading case". International Economic Review, 27: 239-243.

[72] Preston, R. S. (1972), The Wharton Annual and Industry Forecasting Model. Philadelphia: Economics Research Unit, Wharton School, University of Pennsylvania.

[73] Preston, M. (1975), "THe Wharton Long Term Model: Input-Output Within the Context of a Macro-Forecasting Model". International Economic Review, 16: 3-19.

[74] Rasche, R. H., and H. Shapiro (1968), "The FRB-MIT Econometric Model: Its Special Features and Implications for Stabilization Policies". American Economic Review, 58: 123-49.

[75] Richardson, David H. (1963), "The Exact Distribution of a Structural Coefficient Estimator". American Statistical Association Journal, 63: 1214-1226.

[76] Richardson, David H. (1986), "Exact finite sample distribution theory: some results for equations containing three endogenous variables". Advances in Econometrics, vol. 5, pp. 143-170.

[77] Richardson, D.H. & Rohr, R.J. (1983), "An experimental study of two-stage least squares distributions in equations containing three endogenous variables". Advances in Econometrics. Volume 2, pp. 25-50.

[78] Rothenberg, T.J. (1974), Efficient Estimation with A Priori Information. New Haven: Yale University Press.

[79] Sargan, J.D. (1958), "The Estimation of Economic Relationships Using Instrumental Variables". Econometrica, 26: 393-415. 15.

[80] Sargan, J.D. (1961), "The Maximum Likelihood Estimation of Economic Relationships with Autoregressive Residuals". Econometrica, 29: 414-26.

[81] Sargan, J.D. (1964), "Three-Stage Least Squares and Full Information Maximum Likelihood Estimates". Econometrica, 32: 77-81.

[82] Sargan, J.D. (1976), "Econometric Estimators and the Edgeworth Approximations". Econometrica, 44, pp. 421-428.

[83] Sawa, Takamitsu (1969), "The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators". American Statistical Association Journal. 64: 923-937. 37.

[84] Theil, H. (1961), Economic Forecasts and Policy. Second edition. Amsterdam: North-Holland Publishing Company.

[85] Theil, H. (1971), Principles of Econometrics. New-York: John Wiley & Sons, Inc.

[86] Thomas J. Rothenberg (1984), "Approximating the distribution of

econometric estimators and test statistics". Handbook of Econometrics. Volume II, pp. 881-935. 35.

[87] Tinbergen, J. (1939), Statistical Testing of Business Cycle Theories. Vol. 1: A Method and Its Application to Investmentent Activity. Vol. 2: Business Cycles in the United States of America, 1919-1932. Geneva: League of Nations.

[88] Zellner, A., and H.Theil (1962), "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations". Econometrica, 30: 54-78.