NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

ISBN 0-315-56033-9

BOOTSTRAPPING
GENERALIZED TWO STAGE LEAST SQUARE ESTIMATES
IN SIMULTANEOUS EQUATION MODEL
WITH BOTH FIXED AND RANDOM COEFFICIENTS

Shanshan Wang

A Thesis
in
The Department
of
Mathematics & Statistics

Presented in Partial Fulfilment of the Requirements
for the Degree of Master of Science at
Concordia University
Montreal, Quebec, Canada

April 1990

ABSTRACT

BOOTSTRAPPING GENERALIZED TWO STAGE LEAST SQUARE ESTIMATES
IN SIMULTANEOUS EQUATION MODEL WITH BOTH FIXED AND RANDOM COEFFICIENTS

Shanshan Wang

The bootstrap is a technique for estimating standard errors. the idea is to use Monte carlo simulation, based on a nonparametric estimate of the underlying error distribution. A simultaneous equation model with both fixed and random coefficients is fitted by generalized two-stage least-squares. It is shown that bootstrap approximation to the distribution of the estimates is asymptotically valid where the technical difficulties include simultaneity and random coefficients.

DEDICATION

This dissertation is dedicated to my father and my mother.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my thesis supervisor: Professor T. D. Dwivedi, for his genuine concern and his expert guidance. Thanks are also due to the professorial and administrative staff of Concordia University, who have assisted in my education.

Furthermore, I want to express my appreciation to my parents for their support and understanding.

I would also like to thank Dr.D.Morin-Wahhab for her encouragement and fellow graduate students Ms.F.du Burger, Mr.P.Chen and Mr.X.Hu for their useful discussion.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE

SIMULTANEOUS EQUATION AND ITS ESTIMATORS

1.1. Introduction.

Simultaneous equation model is one of the most important models in Econometrics, and is well developed by many economists and statisticians such as Thiel [1961], [1971] and Goldberger [1964].

In section 1.2 the model, structural form and reduced form are introduced. The identification problem is solved in section 1.3. Since we will mainly discuss the single equation in the system and the two stage least square (TSLS here and after) estimator in the following chapters, so we focus on the TSLS estimators in this chapter. The TSLS estimators and the generalized TSLS (GTSLS here and after) are derived in sections 1.4 and 1.5. In section 1.6, we discuss the bias, the moment matrix and some asymptotic properties of the TSLS estimators.

1.2. The Model, Structural Form and Reduced Form.

In this section, we present the simultaneous equation model, its structural form and its reduced form. First, we take a look at the model from both economic and statistical phenomena.

From economic phenomena, economic variables are generally decided by a group of equations respectively instead by one equation . The objective of such equation system is to describe a subset of its variable in terms of the other variables. The former variables are called endogenous, the latter exogenous The intuitive background of their distinction is that the values of variables are determined from the outside,that is in a way which is independent of the process described by the equation system, they are also called predetermined variables, whereas the values of the other endogenous variables are determined jointly and simultaneously by the equation of the system.

The statistical formalization of this idea is the assumption that the value of the exogenous variables are stochastically independent of the disturbances of the system. This assumption enables us to operate conditionally on the exogenous values, so that we may regard them as constants. The current endogenous variables are called jointly dependent variables. Meanwhile, either lagged or exogenous variables are called predetermined.

Now, we should design a systematic notation to a complete linear system in M jointly dependent variables and K predetermined variables. Write y_{tm} for the t-th value of the m-th dependent variable and x_{tk} for the corresponding value of the k-th determined variable. The number of observations is n, each consisting of M+K values, such as (y_{t1} y_{t2} ... y_{tM} x_{t1} x_{t2}... x_{tK}) where t=1, 2, ..., n.

The system consists of M structural equations and can be written as follows,

$$\beta_{11}y_{t1}+\beta_{21}y_{t2}+\ldots+\beta_{M1}y_{tM}+\gamma_{11}x_{t1}+\gamma_{21}x_{t2}+\ldots+\gamma_{K1}x_{tK}+\varepsilon_{t1}=0$$

$$\beta_{12}y_{t1}+\beta_{22}y_{t2}+\ldots+\beta_{M2}y_{tM}+\gamma_{12}x_{t1}+\gamma_{22}x_{t2}+\ldots+\gamma_{K2}x_{tK}+\varepsilon_{t2}=0$$

......

$$\beta_{1M}y_{t1}+\beta_{2M}y_{t2}+\ldots+\beta_{MM}y_{tM}+\gamma_{1M}x_{t1}+\gamma_{2M}x_{t2}+\ldots+\gamma_{KM}x_{tK}+\varepsilon_{tM}=0$$

$$t=1,\ 2,\ \ldots,\ n. \qquad (1.1)$$

The j-th equations will be,

$$\sum_{m=1}^{M} \beta_{mj}y_{tm} + \sum_{k=1}^{K} \gamma_{kj}x_{tk} + \varepsilon_{tj} = 0 \qquad \begin{array}{l} t=1,\ 2,\ \ldots,\ n \\ j=1,\ 2,\ \ldots,\ M \end{array} \qquad (1.2)$$

where ε_{tj} is the disturbance of the t-th observation in the j-th equation and the β's and γ's are parameters to be estimated. Using matrices, we can write,

$$X = [x_{tk}] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \ldots\ldots \\ x_{n1} & x_{n2} & \cdots & x_{nK} \end{bmatrix}_{n\times K} \qquad (1.3)$$

$$Y = [y_{tm}] = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1M} \\ y_{21} & y_{22} & \cdots & y_{2M} \\ \ldots\ldots \\ y_{n1} & y_{n2} & \cdots & y_{nM} \end{bmatrix}_{n\times M} \qquad (1.4)$$

$$E = [\varepsilon_{tm}] = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1M} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2M} \\ \cdots\cdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{nM} \end{bmatrix}_{n\times M} \tag{1.5}$$
$$= [\,\underset{\sim}{\varepsilon}_1 \;\; \underset{\sim}{\varepsilon}_2 \cdots \;\; \underset{\sim}{\varepsilon}_M\,]$$

$$B = [\beta_{ml}] = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1M} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2M} \\ \cdots\cdots \\ \beta_{M1} & \beta_{M2} & \cdots & \beta_{MM} \end{bmatrix}_{M\times M} \tag{1.6}$$

$$\Gamma = [\gamma_{km}] = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1M} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2M} \\ \cdots\cdots \\ \gamma_{K1} & \gamma_{K2} & \cdots & \gamma_{KM} \end{bmatrix}_{K\times M} \tag{1.7}$$

then we can write (1.1) in the following form:

$$YB + X\Gamma + E = 0 \; . \tag{1.8}$$

The matrix B is square because there are as many equations as jointly dependent variables. It is assumed to be nonsingular, so that the system can be solved for these variables,

$$\begin{aligned} Y &= -\, X\Gamma B^{-1} - EB^{-1} \\ &= X\Pi + V \end{aligned} \tag{1.9}$$

where

$$\Pi = -\Gamma B^{-1} \tag{1.10}$$

and

$$V = -EB^{-1}. \tag{1.11}$$

This is the reduced form for all n observations and all M jointly dependent variables.

The assumptions on the disturbance moments of the first and second order are as follows,

$$E(\underset{\sim}{\varepsilon}_j) = \underset{\sim}{0} \qquad j=1,\ 2,\ \dots,\ M, \tag{1.12}$$

the $\underset{\sim}{\varepsilon}_j$ here is the j-th column of the disturbance matrix E described in formula (1.5), and

$$\begin{aligned} E(\varepsilon_{tj}\varepsilon_{\eta m}) &= \sigma_{jm} && \text{when } t = \eta \\ &= 0 && \text{otherwise} \end{aligned} \tag{1.13}$$

then

$$E(\ \underset{\sim}{\varepsilon}_j\ \underset{\sim}{\varepsilon}'_m\ ) = \sigma_{jm}I \qquad j,m = 1,\ 2,\ \dots,\ M\ . \tag{1.14}$$

The covariance matrix of $[\varepsilon_{t1},\ \dots,\ \varepsilon_{tM}]$ for any value t is

$$\Sigma = [\sigma_{jm}] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M} \\ \cdots\cdots & & & \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM} \end{bmatrix}_{M\times M}, \tag{1.15}$$

therefore, Σ is symmetric and positive semidefinite.

Now, we can rewrite (1.11) as follows,

$$V = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1M} \\ v_{21} & v_{22} & \cdots & v_{2M} \\ \cdots\cdots \\ v_{n1} & v_{n2} & \cdots & v_{nM} \end{bmatrix}_{n\times M}$$

$$= -\begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1M} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2M} \\ \cdots\cdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{nM} \end{bmatrix}_{n\times M} B^{-1} \quad . \tag{1.16}$$

Then we can have

$$\underset{\sim}{v}_j = -B'^{-1}\underset{\sim}{\varepsilon}_j \qquad j=1,2,\ldots,M. \tag{1.17}$$

So that

$$E(\ \underset{\sim}{v}_j\ ) = -B'^{-1}E(\ \underset{\sim}{\varepsilon}_j) = 0 \tag{1.18}$$

or

$$E(V) = 0 \quad , \tag{1.19}$$

and

$$
\begin{aligned}
E\, \underset{\sim}{v}_j\, \underset{\sim}{v}'_j &= E\left[B'^{-1}\, \underset{\sim}{\varepsilon}_j\, \underset{\sim}{\varepsilon}'_j\, B^{-1} \right] \\
&= B'^{-1} E\left[\underset{\sim}{\varepsilon}_j\, \underset{\sim}{\varepsilon}'_j \right] B^{-1} \qquad j=1,2,\ldots,M, \qquad (1.20) \\
&= B'^{-1} \Sigma\, B^{-1} \\
&= \Omega
\end{aligned}
$$

say, where Ω is M×M nonnegative definite matrix or the contemporaneous covariance matrix of the disturbances in the different equations is the same for all t, that is

$$
\begin{aligned}
E(v_{tj} v_{\eta m}) &= \omega_{jm} \qquad \text{when } t = \eta \\
&= 0 \qquad \text{otherwise}
\end{aligned} \qquad (1.21)
$$

i.e., the reduced-form disturbance vector is temporally uncorre- llated. And for knowledge of the joint distribution of the dependent variables for all sets of values of the predetermined variables implies knowledge of the conditional expectations of this distribution, as a function of the predetermined variables. But this function is just the reduced-form coefficient matrix Π:

$$
E\left[y'_t \mid x'_t \right] = E\left[-x'_t \Pi + v'_t \mid x'_t \right] = -x'_t \Pi \, . \qquad (1.22)
$$

Furthermore, knowledge of the joint distribution of the dependent variables for all sets of values of predetermined variables implies knowledge of the conditional covariance matrix of this distribution. But this matrix is just the reduced form-disturbance covariance matrix

$$
\begin{aligned}
&E\left\{\left[y'_t - Ey'_t \mid x'_t \right]'\left[y'_t - Ey'_t \mid x'_t \right]\right\} \\
&= E\left[v_t v'_t \mid x'_t \right] \\
&= E\left[v_t v'_t \right] \qquad (1.23) \\
&= \Omega \quad .
\end{aligned}
$$

Similarly, higher moments of the joint distribution of the dependent variables may be identified as further parameters of the reduced-form. Thus the reduced-form parameters are always identified since they are uniquely deducible from the parameters of the joint distribution of the observations. Indeed it is easily seen that knowledge of the reduced form parameters implies knowledge of the conditional distribution of the predetermined variables. Therefore, we may conclude that a structural parameter is identified if and only if it can be uniquely deduced from the reduced form parameters. This introduces the identification problems.

1.3. Identification.

The notation (1.8) is compact and elegant, and it is very useful for the reduced form (1.9), but it is not really convenient when we want to estimate the parameters of our particular structure equation. The reason is that the parameter matrices B and Γ are "wasteful" in the sense that normally a large majority of their elements is known to be 1, 0, or -1. Therefore, we shall consider matters of notation prior to estimation. When attempting to estimate, however, we shall find that there is identification problem to be solved prior to the estimation.

We may believe that a structural parameter is identified if and only if it can be uniquely deduced from the reduced-form parameters. Now, the connections between the structural parameters Π and Ω are as follows,

$$\underset{K\times M}{\Pi} = -\underset{K\times M}{\Gamma}\ \underset{M\times M}{B^{-1}} \tag{1.24}$$

$$\Omega = B^{-1\prime}\Sigma B^{-1} \tag{1.25}$$

then, we can have,

$$\Pi^{*} A = 0 \tag{1.26}$$

where

$$\underset{K\times(M+K)}{\Pi^{*}} = (\underset{K\times M}{\Pi} \ \vdots\ \underset{K\times K}{I}) \tag{1.27}$$

$$A = \begin{bmatrix} B_{M\times M} \\ \Gamma_{K\times M} \end{bmatrix}_{(M+K)\times M} \tag{1.28}$$

and suppose that $\alpha_1, \alpha_2, \ldots, \alpha_M$ denote the columns of A

$$A = \left[\alpha_1, \alpha_2, \ldots, \alpha_M\right] \tag{1.29}$$

This implies that α_i represents the parameters of the i-th structural equation, and

$$\underset{K\times(M+K)}{\Pi^{*}} \underset{(M+K)\times 1}{\alpha_1} = 0 \tag{1.30}$$

provides us with K homogeneous linear equations in M+K variables. So the rank of Π^{*} cannot exceed K. This is the rank condition. Furthermore, let

$$\phi\ \alpha_1 = 0 \tag{1.31}$$

is a priori restriction on α_1 and rows in ϕ are equal to the number of restrictions on α_1. Suppose R restriction are imposed, the, ϕ is a $R\times(M+K)$ matrix and

$$\underset{(K+R)\times(M+K)}{\begin{bmatrix} \Pi^{*} \\ \phi \end{bmatrix}} \underset{(M+K)\times 1}{\alpha_1} = 0\ . \tag{1.32}$$

We have K+R linear homogeneous equations in M+K. This equation will have solutions if

$$\text{Rank} \begin{bmatrix} \Pi^{*} \\ \phi \end{bmatrix} < M+K\ . \tag{1.33}$$

If we normalize one element of α_1, we will have K+R homogeneous equations in M+K-1 variables and then $K+R \geq M+K-1$, or

$$R \geq M-1\ , \tag{1.34}$$

that is the number of prior restriction must be greater than the number of columns of Y minus 1. This is the order condition.

In practice, the most common type of a priori knowledge consists of "zero-restrictions". That is some coefficients being zero. Now, suppose m+1 of M jointly dependent and $K_1 < K$ predetermined variables have nonzero coefficients, and the structural equations have been normalized by dividing with nonzero coefficients of jointly dependent variables. We may write the equations then in the form

$$y_t = \sum_{i=1}^{m} \beta_i y_{ti} + \sum_{k=1}^{K_1} \gamma_k x_{tk} + \varepsilon_t \qquad t=1,2,\dots,n \tag{1.35}$$

and we assume that the equation (1.35) is one of a complete system of M stochastic linear equations in M jointly dependent and K predetermined variables. And this system can be solved for the jointly dependent variables. We also can write (1.35) in following form

$$y = Y_1\beta + X_1\gamma + \varepsilon \tag{1.36}$$

where

y is the n×1 vector whose coefficient is −1,

Y_1 is the n×m matrix whic non zero coefficients,

β is the m×1 vector of coefficients of Y_1,

X_1 is the n×K_1 matrix of observations on the included predetermined variables,

γ is the K_1×1 vector of coefficients of these included predetermined variables,

ε is the n×1 vector of disturbances in this structural equation, and $m<M-1$, $K_1<K$.

Now, in our original model (1.8) we let

$$Y = [\ y \mid Y_1 \mid Y_2] \tag{1.37}$$

and

$$X = [\ X_1 \mid X_2] \tag{1.38}$$

where

Y_2 is an $n\times(M-m-1)$ matrix which are jointly dependent variables with 0 coefficients.

X_2 is an $n\times K_2$ $(K_2=K-K_1)$matrix which are predetermined variables with 0 coefficients.

Then in the reduced form (1.9) we can have

$$y = X_1\pi^* + X_2\pi + v \qquad (.39)$$

$$Y_1 = X_1\Pi_1^* + X_2\Pi_1 + V_1 \tag{1.40}$$

$$Y_2 = X_1\Pi_2^* + X_2\Pi_2 + V_2 \tag{1.41}$$

which gives us

$$\Pi = \begin{bmatrix} \pi^* & \Pi_1^* & \Pi_2^* \\ K_1\times 1 & K_1\times m & K_1\times(M-m-1) \\ \pi & \Pi_1 & \Pi_2 \\ K_2\times 1 & K_2\times m & K_2\times(M-m-1) \end{bmatrix} \tag{1.42}$$

Then the reduced form is

$$[y \mid Y_1 \mid Y_2] = [X_1 \mid X_2] \begin{bmatrix} \pi^* & \Pi_1^* & \Pi_2^* \\ \pi & \Pi_1 & \Pi_2 \end{bmatrix} + [\, v \; V_1 \; V_2 \,] \; . \tag{1.43}$$

Suppose (1.35) is the first equation of the complete system, that is $\begin{bmatrix} 1 \\ -\beta \\ 0 \end{bmatrix}$ is the first column of B and $\begin{bmatrix} -\gamma \\ 0 \end{bmatrix}$ is the first column of Γ. Also from (1.10) and (1.11) we know that $\Pi B=-\Gamma$ and $VB=-E$. That is

$$\begin{bmatrix} \pi^* & \Pi_1^* & \Pi_2^* \\ \pi & \Pi_1 & \Pi_2 \end{bmatrix} \begin{bmatrix} 1 \\ -\beta \\ 0 \end{bmatrix} = -\begin{bmatrix} -\gamma \\ 0 \end{bmatrix} \tag{1.44}$$

and

$$[\, v \mid V_1 \mid V_2 \,] \begin{bmatrix} 1 \\ -\beta \\ 0 \end{bmatrix} = \varepsilon \tag{1.45}$$

then

$$\pi^* - \Pi_1^*\beta = \gamma \tag{1.46}$$

$$\pi - \Pi_1\beta = 0 \tag{1.47}$$

$$v - V_1\beta = \varepsilon \tag{1.48}$$

From (1.47)

$$\underset{K_2\times m}{\Pi}\ \underset{m\times 1}{\beta} = \underset{K_2\times 1}{\pi} \tag{1.49}$$

we can see, there are three situations,

1. if $K_2=m$ the equations will give unique solutions and we call it exact identification,

2. if $K_2>m$ the equations will have more than one solution, and we call it over identification,

3. if $K_2<m$ the equation will have no solution and we call it under identification.

1.4. Two Stage Least Squares.

The first method to estimate the coefficients is the two stage least squares which is developed by Theil [1953, 1961, pp 225-231, 334-344]. We rewrite the (1.36) and (1.40) here,

$$y = Y_1\beta + X_1\gamma + \varepsilon \tag{1.50}$$

and

$$\begin{aligned} Y_1 &= X_1\Pi_1^* + X_2\Pi_1 + V_1 \\ &= X\,\Pi_{X_1} + V_1 \end{aligned} \tag{1.51}$$

where $X=[X_1 \vdots X_2]$ and $\Pi_{X_1}=\begin{bmatrix} \Pi_1^* \\ \Pi_1 \end{bmatrix}$. Inserting (1.51) into (1.50) and rearranging, we can have

$$y = X\Pi_{X_1}\beta + X_1\gamma + (\varepsilon + V_1\beta) \tag{1.52}$$

Since the predetermined variables are contemporaneously uncorrelated with all disturbances (structural and reduced form), if we take the classical least-squares regression of y on $\bar{Y}_1=X\Pi_{X_1}$ and X_1 we would obtain consistent estimates of β and γ. This procedure is not available to us because we do not know Π_{X_1} and hence do not have observations on $\bar{Y}_1$. We can however, consistently estimate Π_{X_1} by $\hat{\Pi}_{X_1}$ and hence estimate $\bar{Y}_1$ by $X_1\hat{\Pi}_{X_1}$. These considerations here suggest the following two-stage procedure.

Stage one: Obtain the classical least-squares estimator $\hat{\Pi}_{X_1}$ of Π_{X_1} by regressing each column of Y_1 on X; this is the sub matrix of $\hat{\Pi}$. Where

$$\hat{\Pi} = (X'X)^{-1}X'Y \tag{1.53}$$

and

$$\hat{\Pi}_{X_1} = (X'X)^{-1}X'Y_1 \ . \tag{1.54}$$

We then obtain the calculated values in these regressions,

$$\hat{Y}_1 = X\hat{\Pi}_{X_1} \quad . \tag{1.55}$$

Stage two: Take the classical least-squares regression of y on $\hat{Y}_1$ and X_1. The resulting coefficients are the two stage least-squares estimators of β and γ.

Thus the two-stage least-squares estimator of $\begin{bmatrix} \beta \\ \gamma \end{bmatrix}$ is the $\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix}$ defined by the normal equations

$$\begin{bmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \hat{Y}_1'y \\ X_1'y \end{bmatrix} . \tag{1.56}$$

A system of $m+K_1$ equations in $m+K_1$ unknowns which will, in general, have a unique solution.

To establish the consistency of the two stage least-squares estimator we might write the equation as follows,

$$y = \hat{Y}_1\beta + X_1\gamma + [\varepsilon + (Y_1 - \hat{Y}_1)\beta\ ] \tag{1.57}$$

and show that $\hat{Y}_1$ and X_1 are independent of compound disturbance term in brackets, in the sense that the probability limits of the sample covariances are zero. Alternatively, we interpret (1.56) as the normal equations of instrumental variable estimation, which will allow us to draw on the results of (1.1) for standard errors as well. Suppose, then that we estimate (1.50) by instrumental variable method with calculated values of the right-hand dependent variables as instruments for the corresponding observed values and with the included predetermined variables such as their own instruments. Then the estimator of $\begin{bmatrix}\beta\\ \gamma\end{bmatrix}$ would be defined by

$$\begin{bmatrix} \hat{Y}_1'Y_1 & \hat{Y}_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} \hat{\beta}^* \\ \hat{\gamma}^* \end{bmatrix} = \begin{bmatrix} \hat{Y}_1'y \\ X_1'y \end{bmatrix} \tag{1.58}$$

and supposing that the instruments were legitimate, that this estimator would be consistent; that its asymptotic covariance matrix would be

$$\begin{aligned} n^{-1}\sigma^2 & \left[\operatorname{plim} n^{-1}\begin{pmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ \hat{X}_1'Y_1 & X_1'X_1 \end{pmatrix}\right]^{-1} \\ & \times\left[\operatorname{plim} n^{-1}\begin{pmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ \hat{X}_1'Y_1 & X_1'X_1 \end{pmatrix}\right] \\ & \times\left[\operatorname{plim} n^{-1}\begin{pmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ \hat{X}_1'Y_1 & X_1'X_1 \end{pmatrix}\right]^{-1} \end{aligned} \tag{1.59}$$

where σ^2 is the structural disturbance variance; and that this matrix would be consistently estimated by,

$$s^2 \begin{bmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1 \end{bmatrix}^{-1} \tag{1.60}$$

where s^2 is the sum of squared residuals divided by n. Now the instruments are legitimate and

$$\text{plim } n^{-1}X_1'\varepsilon = 0 \tag{1.61}$$

and since $\hat{\Pi}$ is consistent,

$$\text{plim } n^{-1}\hat{Y}_1'\varepsilon = \text{plim } n^{-1}\hat{\Pi}_{X_1}X_1'\varepsilon = \hat{\Pi}_{X_1}\text{plim } n^{-1}X_1'\varepsilon = 0 \ . \tag{1.62}$$

From the definition of $\hat{Y}_1$ in (1.55), we can have

$$\begin{aligned}\hat{Y}_1'\hat{Y}_1 &= Y_1'X\ (X'X)^{-1}X'X\ (X'X)^{-1}X'Y_1 \\ &= Y_1'X\ (X'X)^{-1}X'Y_1 \\ &= \hat{Y}_1'Y_1 \\ &= Y_1'\hat{Y}_1\end{aligned} \tag{1.63}$$

and

$$\begin{aligned}X_1'\hat{Y}_1 &= X_1'X\ (X'X)^{-1}X'Y_1 \\ &= (I\ 0)\begin{bmatrix}X_1' \\ X_2\end{bmatrix} Y_1 \\ &= X_1'Y_1\ .\end{aligned} \tag{1.64}$$

Since

$$\begin{aligned}X_1 &= (X_1\ X_2)\begin{bmatrix}I_{K_1} \\ 0_{K_2}\end{bmatrix} \\ &= X\begin{bmatrix}I \\ 0\end{bmatrix}\end{aligned} \tag{1.65}$$

and

$$X_1' = (I\ 0)\ X' . \tag{1.66}$$

Thus we see that TSLS is identical with an instrumental variable estimation which establishes its consistency. In addition, in (1.59) and (1.60) the middle matrix has as its inverse the third matrix so that these two matrices "cancel out". We also see that there is no need to compute explicitly the individual calculated values. Rather, the computationally efficient formula for TSLS estimations is given by

$$\begin{bmatrix} Y_1'X(X'X)^{-1}X'Y_1' & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} \hat\beta \\ \hat\gamma \end{bmatrix} = \begin{bmatrix} Y_1'X(X'X)^{-1}X'y \\ X_1'y \end{bmatrix} . \tag{1.67}$$

The asymtotic covariance matrix may be written as

$$\Sigma_{\hat\beta\hat\gamma,\hat\beta\hat\gamma} = n^{-1}\sigma^2 \begin{bmatrix} \Pi_{X_1}' \Sigma_{XX}\Pi_{X_1} & \Pi_{X_1}' \Sigma_{X1}' \\ \Sigma_{1X}\Pi_{X_1} & \Sigma_{11} \end{bmatrix}^{-1} \tag{1.68}$$

where Σ_{1X} and Σ_{11} are submatrices of Σ_{XX}=plim($n^{-1}X'X$). To see this, note that,

$$\begin{aligned} \text{Plim } n^{-1}X'Y_1 &= \text{Plim } n^{-1}X'(X\Pi_{X_1} + V_1) \\ &= \text{Plim } n^{-1}X'X\Pi_{X_1} + \text{Plim } n^{-1}X'V_1 \\ &= \Sigma_{XX}\Pi_{X_1} + 0 \\ &= \Sigma_{XX}\Pi_{X_1} . \end{aligned} \tag{1.69}$$

So that

$$\text{Plim } n^{-1}X_1'Y_1 = \text{Plim } n^{-1}X_1'(X\Pi_{X_1} + V_1)$$
$$= \text{Plim } n^{-1}X_1'X\Pi_{X_1} + \text{Plim } n^{-1}X_1'V_1$$
$$= \Sigma_{1X}\Pi_{X_1}. \tag{1.70}$$

The estimator of (1.68) may then be written as

$$S_{\hat\beta\hat\gamma,\hat\beta\hat\gamma} = s^2\begin{bmatrix} Y_1'X(X'X)^{-1}X'Y_1' & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1}. \tag{1.71}$$

It may also be seen that s^2 may be computed, without explicit computation of the residuals, by

$$s^2 = n^{-1}(y - Y_1\hat\beta - X_1\hat\gamma)'(y - Y_1\hat\beta - X_1\hat\gamma)$$

$$= n^{-1}(y'y - 2\hat\beta'Y_1'y - 2\hat\gamma'X_1'y + 2\hat\beta'Y_1'X_1\hat\gamma + \hat\beta'Y_1'Y_1\hat\beta + \hat\gamma'X_1'X_1\hat\gamma)$$

$$= n^{-1}(y'y + \hat\beta'Y_1'Y_1\hat\beta - 2\hat\beta'Y_1'y - \hat\gamma'X_1'X_1\hat\gamma) \tag{1.72}$$

using the second row of (1.67), namely, $\hat\gamma = (X_1'X_1)^{-1}(-X_1'Y_1\beta + X_1'y)$.

From the point of view of instrumental variable estimation, the calculated values of the right hand dependent variables are desirable instruments since they are contemporaneously, uncorrelated with the disturbance and yet are correlated with the observed values of the variables for which they serve as instruments.

Finally, we show that when the equation is just-identified, the TSLS estimators are identical with the indirect least squares estimators of (1.44), such as,

$$\begin{bmatrix} \pi^{*} & \vdots & \Pi^{*} & \vdots & \Pi_2^{*} \\ \pi & \vdots & \Pi & \vdots & \Pi_2 \end{bmatrix} \begin{bmatrix} 1 \\ -\hat{\beta} \\ 0 \end{bmatrix} = -\begin{bmatrix} -\hat{\gamma} \\ 0 \end{bmatrix} \qquad (1.73)$$

which in view of the definition of $\hat{\Pi}$ is

$$(X'X)^{-1}X'(y\ \ Y_1)\begin{bmatrix} -1 \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} -\hat{\gamma} \\ 0 \end{bmatrix}. \qquad (1.74)$$

If we premultiply (1.74) by $Y_1'X$ we find,

$$\begin{aligned} -Y_1'X(X'X)^{-1}X'y + Y_1'X(X'X)^{-1}X'Y_1\hat{\beta} &= Y_1'(X_1\ \ X_2)\begin{bmatrix} -\hat{\gamma} \\ 0 \end{bmatrix} \\ &= -Y_1'X_1\hat{\gamma} \end{aligned} \qquad (1.75)$$

If we premultiply (1.74) by $(X_1'X)$ we find,

$$\begin{aligned} -X_1'X(X'X)^{-1}X'y + X_1'X(X'X)^{-1}X'Y_1\hat{\beta} &= X_1'(X_1\ \ X_2)\begin{bmatrix} -\hat{\gamma} \\ 0 \end{bmatrix} \\ &= -X_1'X_1\hat{\gamma} \end{aligned} \qquad (1.76)$$

and notice that we can write $X_1' = (I\ \ 0)\begin{bmatrix} X_1' \\ X' \end{bmatrix} = (I\ \ 0)X$, then we have

$$-(I\ \ 0)\begin{bmatrix} X_1' \\ X_2' \end{bmatrix} y + (I\ \ 0)\begin{bmatrix} X_1' \\ X' \end{bmatrix} Y_1\hat{\beta} = -X_1'X_1\hat{\gamma} \qquad (1.77)$$

or

$$-X_1'y + X_1'Y_1\hat{\beta} = - X_1'X_1\hat{\gamma} \tag{1.78}$$

Now, (1.75) and (1.78) will be recognized as the first and second "rows" respectively of (1.67). Thus when they are defined the indirect least squares estimators satisfy the same (nonsingular) system (1.67) as do the TSLS estimators, so that they are identical with the latter.

The following assumptions will be made.

Assumption 1.1. Equation (1.36) is one of a complete system of $M(\geq m+1)$ linear stochastic equations in M jointly dependent variables and K predetermined variables. The reduced form of this system exists.

Assumption 1.2. The matrix Π_1, which is in (1.42) with the order of $K_2 \times m$, has rank m.

Assumption 1.3. The matrix X, which is of order $n \times K$ has rank K and consists of nonstochastic elements.

As to the n vectors of M disturbances corresponding to each of the M structural equations, we need,

Assumption 1.4. The n disturbance vectors are independent random drawings from the same M-dimensional normal parent with zero means.

Following all the discussion, we present the theorem 1.1 here.

THEOREM 1.1. (Theil [1971] pp497)

Suppose that the *assumptions 1.1 to 1.4* are satisfied, its disturbance variance σ^2 being positive. Then the TSLS estimator $\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix}$ in (1.56) is consistent for $\begin{bmatrix} \beta \\ \gamma \end{bmatrix}$ and $n^{-1/2}\left[\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} - \begin{bmatrix} \beta \\ \gamma \end{bmatrix}\right]$ has a normal distribution with zero mean vector and the following covariance matrix $\Sigma_{\hat{\beta}\hat{\gamma},\hat{\beta}\hat{\gamma}}$ which is described in (1.68). Also, the statistics s^2 in (1.72) is a consistent estimator of the variance σ^2.

1.5. Generalized Two Stage Least Squares.

A generalization can be obtained for the case of autocorrelated disturbances. We assume that first the disturbances of (1.35) have a finite and nonsingular covariance matrix, $E(\varepsilon\varepsilon')=\Sigma$. Secondly, for each pair t, $t'(=1, 2, \ldots, n)$ and for each pair i, $i'(=1, 2, \ldots, m)$ the parent reduced form disturbances v_{ti} and v'_{ti} corresponding to the right hand variable y_i and y'_i, respectively, of this equation has zero mean and satisfies

$$E(v_{ti}\ v'_{ti}) = \omega_{ii'}\sigma_{tt'} \tag{1.79}$$

where $\omega_{ii'}$ is independent of t and t', and $\sigma_{tt'}$ is the (t, t')-th element of Σ. This condition is weaker than the one which it replaces. Still it imposes rather strong restrictions on the probability structure of the disturbances, because the covariance matrices of those in the reduced form — at least in that part of the reduced form that corresponds to the dependent explanatory variables of the equation —

are supposed to be "proportional" to the covariance matrix of the disturbances in the structural equation under consideration. A sufficient condition under which this situation is reduced is that the equation system is linear and that the disturbance vectors of each pair of its equations have a covariance matrix equal to Σ except for a scalar. When viewed from this point, we can conclude that the assumption, although still rather restrictive, is certainly much weaker than the assumptions which it replaces. It allows us to take into account one of the important and almost common feature of disturbances in time series, viz, their positive first autocorrelation.

Suppose then that the reduced form is estimated according to Aitken's method of generalized least squares (Theil [1961]). This gives

$$Y_1 - V_1^* = X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y_1 \tag{1.80}$$

V^* being the matrix of resulting, estimated, reduced form disturbances. Consider also the estimation equations

$$\begin{bmatrix} Y_1'-V_1^{*\prime} \\ X_1' \end{bmatrix} \Sigma^{-1}y = \begin{bmatrix} Y_1'-V_1^{*\prime} \\ X_1' \end{bmatrix} \Sigma^{-1} \; (Y_1-V_1^* \; X_1) \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} \tag{1.81}$$

the last vector being an estimator of γ and β is an immediate generalization both of two stage least squares and of generalized least squares; so we shall call this method that of "generalized two stage least squares". Writing A for the product of the first three matrices in the right hand side of (1.81), such as

$$A = \begin{bmatrix} Y_1'-V_1^{*\prime} \\ X_1' \end{bmatrix} \Sigma^{-1} \; (Y_1 - V_1^* \; X_1) \tag{1.82}$$

$$\begin{aligned} \begin{bmatrix} \hat{\beta}^* \\ \hat{\gamma}^* \end{bmatrix} &= A^{-1} \begin{bmatrix} Y_1'-V_1^{*\prime} \\ X_1' \end{bmatrix} \Sigma^{-1} y \\ &= A^{-1} \begin{bmatrix} Y_1'-V_1^{*\prime} \\ X_1' \end{bmatrix} \Sigma^{-1} \left[(Y_1 - V_1^* \; X_1) \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \varepsilon \right] \\ &= A^{-1} \begin{bmatrix} Y_1'-V_1^{*\prime} \\ X_1' \end{bmatrix} \Sigma^{-1} (Y_1 - V_1^* \; X_1) \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + A^{-1} \begin{bmatrix} Y_1'-V_1^{*\prime} \\ X_1' \end{bmatrix} \Sigma^{-1} \varepsilon \\ &= A^{-1} A \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + A^{-1} \begin{bmatrix} Y_1'-V_1^{*\prime} \\ X_1' \end{bmatrix} \Sigma^{-1} \varepsilon \quad . \end{aligned} \tag{1.83}$$

Whence it follows that the estimator is asymptotically unbiased. Further, if e^* is the sampling error, we find for the asymptotic covariance matrix

$$\begin{aligned} \lim_{n \to \infty} E(n e^* e^{*\prime}) &= \operatorname*{plim}_{n \to \infty} (n A^{-1}) \\ &= \operatorname*{plim}_{n \to \infty} n \left[\begin{bmatrix} Y_1'-V_1^{*\prime} \\ X_1' \end{bmatrix} \Sigma^{-1} (Y_1 - V_1^* \; X_1) \right]^{-1} \end{aligned} \tag{1.84}$$

which is again an immediate generalization.

1.6. The Bias and Moment Matrix of The TSLS Estimators

We know for the equation (1.36) the TSLS is in (1.56) which can be written as

$$\begin{bmatrix} Y_1'-\hat{V}_1' \\ X_1' \end{bmatrix} y = \begin{bmatrix} Y_1'Y_1'-\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \tag{1.85}$$

where $\hat{V}_1 = Y_1 - X(X'X)^{-1}X'Y_1$ is the estimate of V_1 in (1.45) from the first stage.

We can write

$$V_1 = \varepsilon r' + W \tag{1.86}$$

which describes the (normally distributed reduced form disturbances as consisting of a part which is proportional to the corresponding disturbances of (1.36) (i.e. $\varepsilon r'$, r' being a column vector of m components) and a part (i.e. W) which is also normally distributed but independently of the ε vector. Consider the vector of covariances of the disturbances of and the right hand variables of equation (1.36):

$$q = \frac{1}{n} E\begin{bmatrix} Y_1'\varepsilon \\ X'\varepsilon \end{bmatrix} = \frac{1}{n}\begin{bmatrix} EV_1'\varepsilon \\ 0 \end{bmatrix} = \sigma^2 \begin{bmatrix} r \\ 0 \end{bmatrix} \tag{1.87}$$

where σ^2 is the variance of the disturbances of (1.36).

Further, we write

$$Q = \begin{bmatrix} Y_1'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \tag{1.88}$$

where $Q_{11} = (Y_1'M_1Y_1)^{-1}$,

$$Q_{12} = Q_{21} = -(X_1'X_1)^{-1}X_1'Y_1(Y_1'M_1Y_1)^{-1} ,$$

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

THEOREM 1.2. (Nagar [1959])

Under *assumptions 1.1 to 1.4*, the bias (to the order of n^{-1}) of the estimator $\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix}$ is given by

$$E(e) = (L-1)Qq \tag{1.89}$$

where e is the sampling error:

$$e = \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} - \begin{bmatrix} \beta \\ \gamma \end{bmatrix} \tag{1.90}$$

and $L=K-K_1-m$.

From (1.86) we can write for the moment matrix of V:

$$\frac{1}{n} E(V_1'V_1) = \sigma^2 rr' + \frac{1}{n} E(W'W) \tag{1.91}$$

and bordering these matrices with K_1 rows and K_1 columns of zeros; we obtain three square matrices of order $m+K_1$:

$$C_1 = \begin{bmatrix} \sigma^2 rr' & 0 \\ 0 & 0 \end{bmatrix} = \frac{1}{\sigma^2} q'q \quad , \tag{1.92}$$

$$C_2 = \begin{bmatrix} \frac{1}{n}E(W'W) & 0 \\ 0 & 0 \end{bmatrix} \tag{1.93}$$

and

$$\begin{aligned} C &= C_1 + C_2 \\ &= \begin{bmatrix} \sigma^2 rr' + \frac{1}{n}E(W'W) & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{n}E(V'V) & 0 \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{1.94}$$

THEOREM 1.3. (Nagar [1959])

Under the *assumptions* of theorem 1.2, the moment matrix, to the order of n^{-2}, of the estimator $\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix}$ around the parameter $\begin{bmatrix} \beta \\ \gamma \end{bmatrix}$ is given by:

$$E(ee') = \sigma^2 Q(I + D^*) \; ; \tag{1.96}$$

where D^* is a matrix of order n^{-1}

$$\begin{aligned} D^* = {} & \left[-(2L - 3)\mathrm{tr}(C_1Q) + \mathrm{tr}(C_2Q)\right]I \\ & + \left[(L-2)^2+2\right]C_1Q - (L-2)C_2Q \; . \end{aligned} \tag{1.97}$$

Then the estimators of Q, C, q and σ^2 are as follows;

$$\hat{Q} = \begin{bmatrix} \hat{Y}_1'\hat{Y}_1' & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1 \end{bmatrix}^{-1} = \begin{bmatrix} Y_1'PY_1' & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \tag{1.98}$$

$$\hat{Q}_{11} = \left[Y_1'(P - X_1(X_1X_1)^{-1}X_1)Y_1 \right] \tag{1.99}$$

$$\hat{C} = \begin{bmatrix} 1/nE(\hat{V}_1'\hat{V}_1) & 0 \\ 0 & 0 \end{bmatrix} \tag{1.100}$$

$$\hat{q} = \begin{bmatrix} \frac{1}{n}Y_1'(I-P)(y-Y_1\hat{\beta}-X_1\hat{\gamma}) \\ 0 \end{bmatrix} \tag{1.101}$$

where $P = X(X'X)^{-1}X'$.

In this chapter we studied the two-stage least-square estimates and the generalized two-stage least-square estimates of the simultaneous equation model and their asymptotic properties. More specifically, in this system we have to solve the identification problem. More complicated model will be discussed in the next chapter based on the discussions in this chapter.

CHAPTER TWO

SIMULTANEOUS EQUATIONS WITH BOTH FIXED AND RANDOM COEFFICIENTS AND THE GENERALIZED TWO STAGE LEAST SQUARE ESTIMATOR

2.1. Introduction.

As observed by Klein [1955, pp 212-216], it is unlikely that interindividual differences observed in a cross section sample can be explained by a simple regression equation with a few independent variables. In such situations, the coefficients can be treated as random to account for interindividual heterogeneity. This random coefficient regression model has been extensively investigated by Hildreth & Houck [1969], Swamy [1970, 1971, 1973], Swamy & Mehta [1975, 1977], Swamy & Tislay [1980], and Harville [1976, 1977]. They have proposed methods of estimating the mean parameters of such model.

Meanwhile, identification and estimation concerning simultaneous equation models with random parameters have been considered and some results have been derived by some authors such as Nerlove [1965], Zellner [1969], Kelejian [1974], and Raj, Srivastava and Ullah [1980] (R&S&U [1980] here and after).

The coefficients of simultaneous equation model discussed in previous chapter are fixed. Here we consider that model while the coefficients of exogenous variables are random and those of endogenous variables are fixed. We also show how to estimate consistently a single equation in the system and state some asymptotic properties.

2.2. Model Specification.

The model considered here is a simultaneous equation model in which the coefficients of endogenous variables are fixed as usual and the coefficients of exogenous variables are random. A generalized two stage least square estimator for a single equation in the model is then derived.

The single equation (1.35) in a complete system of M structural equations in N jointly dependent and K exogenous variables is rewritten here

$$y_t = \sum_{i=1}^{m} \beta_i y_{ti} + \sum_{k=1}^{K_1} \gamma_k x_{tk} + \varepsilon_t \qquad t=1,2,\ldots,n. \tag{2.1}$$

Now, the β_i's in (2.1) are fixed as usual and γ_k's are random, so (2.1) may be rewritten as follows,

$$y_t = \sum_{i=1}^{m} \beta_i y_{ti} + \sum_{k=1}^{K_1} \gamma_{k(t)} x_{tk} \qquad t=1,2,\ldots,n. \tag{2.2}$$

Assuming that $x_{t1}=1$, for $t=1,\ldots,n$, we assume the usual disturbance term in the equation (2.2) into the varying intercept term.

In more compact matrix notation the equation (2.2) can be written as:

$$y = Y_1\beta + \mathrm{Diag}\left(X'_{11}\ X'_{12}\ \cdots\ X'_{1n}\right)\left(\gamma'_{(1)}\ \gamma'_{(2)}\ \cdots\ \gamma'_{(n)}\right)'$$
$$= Y_1\beta + D_{X_1}\gamma \tag{2.3}$$

where y is a nx1 vector

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}_{n\times 1} . \tag{2.4}$$

Y_1 is a nxm matrix of observations on the other endogenous variables which appear in the equation, that is

$$Y_1 = \begin{bmatrix} y_{11}\ y_{12}\ \cdots\ y_{1m} \\ y_{21}\ y_{22}\ \cdots\ y_{2m} \\ \cdots\cdots \\ y_{n1}\ y_{n2}\ \cdots\ y_{nm} \end{bmatrix}_{n\times m} . \tag{2.5}$$

X_1 is $n\times K_1$ matrix of observations on K_1 exogenous variables which appear in the equation, such as

$$X_1 = \begin{bmatrix} x_{11}\ x_{12}\ \cdots\ x_{1K_1} \\ x_{21}\ x_{22}\ \cdots\ x_{2K_1} \\ \cdots\cdots \\ x_{n1}\ x_{n2}\ \cdots\ x_{nK_1} \end{bmatrix}_{n\times K_1} \tag{2.6}$$

while X'_{1t} is the t-th row of X_1, and β and $\gamma_{(t)}$ are coefficient vectors of dimensions $m\times 1$ and $K_1\times 1$ respectively, such as,

$$X_{1(t)} = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \cdot\cdot \\ x_{K_1 t} \end{bmatrix}_{K_1\times 1} \tag{2.7}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_m \end{bmatrix}_{m\times 1} \tag{2.8}$$

and

$$\gamma_{(t)} = \begin{bmatrix} \gamma_{1(t)} \\ \gamma_{2(t)} \\ \cdots \\ \gamma_{K_1(t)} \end{bmatrix}_{K_1\times 1} \qquad t=1,2,\ldots,n. \tag{2.9}$$

Furthermore D_{X_1} is of order $n\times nK_1$

$$D_{X_1} = \begin{bmatrix} X'_{11} & 0 & \ldots & 0 \\ 0 & X'_{12} & \ldots & 0 \\ \cdots\cdots & & & \\ 0 & 0 & \ldots & X'_{1n} \end{bmatrix}_{n\times nK_1} \tag{2.10}$$

and γ is of order $nK_1\times 1$

$$\gamma = \begin{bmatrix} \gamma_{(1)} \\ \gamma_{(2)} \\ \cdots \\ \gamma_{(n)} \end{bmatrix}_{nK_1 \times 1} . \qquad (2.11)$$

As the assumption at the beginning that the elements of β are fixed while the elements of γ in (2.11) are random and are independently and identically distributed about a mean vector $\bar{\gamma}$ such that

$$\gamma_{(t)} = \bar{\gamma} + \varepsilon_{(t)} \qquad t=1,2,\ldots,n \qquad (2.12)$$

where $\bar{\gamma}$ is a $K_1 \times 1$ vector with nonstochastic elements and ε is a $nK_1 \times 1$ discrepancy vector,

$$\bar{\gamma} = \begin{bmatrix} \bar{\gamma}_1 \\ \bar{\gamma}_2 \\ \cdots \\ \bar{\gamma}_{K_1} \end{bmatrix}_{K_1 \times 1} = \begin{bmatrix} \sum_{t=1}^{n} \gamma_{1(t)}/n \\ \sum_{t=1}^{n} \gamma_{2(t)}/n \\ \cdots \\ \sum_{t=1}^{n} \gamma_{K_1(t)}/n \end{bmatrix}_{K_1 \times 1} \qquad (2.13)$$

and

$$\varepsilon = \begin{bmatrix} \varepsilon_{(1)} \\ \cdots \\ \varepsilon_{(t)} \\ \cdots \\ \varepsilon_{(n)} \end{bmatrix}_{nK_1 \times 1} \qquad (2.14)$$

where

$$\varepsilon_{(t)} = \begin{bmatrix} \varepsilon_{1(t)} \\ \varepsilon_{2(t)} \\ \cdots \\ \varepsilon_{K_1(t)} \end{bmatrix}_{K_1 \times 1} . \tag{2.15}$$

Substituting (2.12) into (2.3), gives,

$$y = Y_1\beta + X_1\bar{\gamma} + w \tag{2.16}$$

and let $A=\left(Y_1 \mid X_1\right)$, $\delta=\begin{bmatrix} \beta \\ \hline \gamma \end{bmatrix}$ and $w=D_{X_1}\varepsilon$, this equation can be written as

$$y = A\delta + w \quad . \tag{2.17}$$

We assume that

Assumption 2.1.

(i) $\quad E\varepsilon_{(t)} = 0 \qquad t=1,2,\ldots n.$

(ii) $$E\left(\varepsilon_{(t)}\varepsilon_{(t')}\right) = \Delta = \begin{bmatrix} \theta_1 & 0 & \ldots & 0 \\ 0 & \theta_2 & \ldots & 0 \\ \ldots & \ldots & & \\ 0 & 0 & & \theta_{K_1} \end{bmatrix}_{K_1 \times K_1}$$

where $\varepsilon_{j(t)}$ is the j-th element of $\varepsilon_{(t)}$, $j=1,2,\ldots,K_1$.

(iii) The elements of D_{X_1} are independent of the elements of $\varepsilon_{(t)}$.

Clearly, under Assumption 2.1

$$E(w) = E(D_{X_1}\varepsilon) = D_{X_1}E(\varepsilon) = 0 \tag{2.18}$$

and

$$
\begin{aligned}
Eww' &= ED_{X_1}\varepsilon\varepsilon' D'_{X_1} \\
&= D_{X_1} E\varepsilon\varepsilon' D'_{X_1} \\
&= D_{X_1} \Theta\, D_{X_1} \\
&= \Omega
\end{aligned}
\tag{2.19}
$$

where

$$
\Theta = \begin{bmatrix} \Delta & 0 & \dots & 0 \\ 0 & \Delta & \dots & 0 \\ \cdots\cdots & & & \\ 0 & 0 & & \Delta \end{bmatrix}_{nK_1 \times nK_1}
\tag{2.20}
$$

0's are $K_1 \times K_1$ null matrices.

$$
\Omega = \begin{bmatrix} \Omega_1 & 0 & \dots & 0 \\ 0 & \Omega_2 & \dots & 0 \\ \cdots\cdots & & & \\ 0 & 0 & & \Omega_n \end{bmatrix}_{n \times n}
$$

$$
= \begin{bmatrix} X_{1(1)}\Delta X'_{1(1)} & 0 & \dots & 0 \\ 0 & X_{1(2)}\Delta X'_{1(2)} & \dots & 0 \\ \cdots\cdots\cdots & & & \\ 0 & 0 & & X_{1(n)}\Delta X'_{1(n)} \end{bmatrix}
$$

$$= \begin{bmatrix} \sum_{j=1}^{K_1} x_{1j}^2\theta_j & 0 & \cdots & 0 \\ 0 & \sum_{j=1}^{K_1} x_{2j}^2\theta_j & \cdots & 0 \\ \cdots\cdots & & & \\ 0 & & & \sum_{j=1}^{K_1} x_{2j}^2\theta_j \end{bmatrix} \tag{2.21}$$

$$\Omega_t = \sum_{j=1}^{K_1} x_{tj}^2\theta_j \qquad t=1,\ 2,\ \ldots,\ n. \tag{2.22}$$

Now a necessary and sufficient condition for the identifiability of β and $\bar{\gamma}$ in equation (2.21) is exactly as it would be in the fixed coefficients version of (2.1). Then we can have another assumption as follows.

Assumption 2.2. Equation (2.21) is identifiable.

Assumption 2.3. The rank of matrix X is $K<n$.

Then, we can discuss the estimaes of coeffcient of the model.

2.3. Consistent Estimators.

The Two Stage Least Square (TSLS) estimator of δ which does not take into account the heteroscedasticity of w as described in (1.67) is

$$\hat{\delta} = \left[\begin{pmatrix} \hat{Y}_1' \\ X_1' \end{pmatrix} (\hat{Y}_1 \ \ X_1)\right]^{-1} (\hat{Y}_1 \ \ X_1)y$$

$$= \left[\begin{pmatrix} Y_1'X(X'X)^{-1}X' \\ X_1' \end{pmatrix} \left(X(X'X)^{-1}X'Y_1 \ \ X_1\right)\right]^{-1} \left(Y_1'X(X'X)^{-1}X' \ \ X_1\right)y$$

$$= \begin{bmatrix} Y_1'X(X'X)^{-1}X'Y_1 & Y_1'X(X'X)^{-1}X'X_1 \\ X_1'X(X'X)^{-1}X'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1'X(X'X)^{-1}X' & X_1 \end{bmatrix} y$$

$$= \left[\begin{bmatrix} Y_1' \\ X_1' \end{bmatrix} (X(X'X)^{-1}X)(Y_1 \ X_1)\right]^{-1} \left[(Y_1 \ X_1)(X(X'X)^{-1}X')\right] y$$

$$= \left(A'PA\right)^{-1}A'Py \tag{2.22}$$

where $P=X\left(X'X\right)^{-1}X'$ and $X_1'P'=(I\ 0)X'P=(I\ 0)X'X(X'X)^{-1}X'=X_1'$.

From Chapter 1, we know that under general conditions the TSLS estimator $\hat{\delta}$ is consistent and the second order moment matrix of the asymptotic distribution of $\sqrt{n}\,(\hat{\delta}-\delta)$ is given by

$$\underset{n\to\infty}{\text{Plim}}\ n\left(A'PA\right)^{-1}AP\Omega PA\left(A'PA\right)^{-1} . \tag{2.23}$$

When $\Omega=\sigma^2 I_n$, we easily verify the conventional result of (2.23) as

$$\begin{aligned} &\underset{n\to\infty}{\text{Plim}}\ n\left(A'PA\right)^{-1}AP\sigma I_n PA\left(A'PA\right)^{-1} \\ &=\sigma^2 \underset{n\to\infty}{\text{Plim}}\ n\left(A'PA\right)^{-1}. \end{aligned} \tag{2.24}$$

Now, since $E(ww') = \Omega \neq \sigma^2 I_n$, an efficient TSLS estimator of δ in (2.17) can be obtained as follows. Given that Ω is positive definite, we can apply the generalized TSLS in Chapter 1 equation (1.81) to (2.17)

$$\hat{Y}_1 = X(X'\Omega^{-1}X)^{-1}X\Omega^{-1}Y_1 \tag{2.25}$$

then

$$y = (\hat{Y}_1 \; X_1)\delta + w \tag{2.26}$$

and

$$\begin{pmatrix} \hat{Y}_1' \\ X_1' \end{pmatrix}\Omega^{-1}y = \begin{pmatrix} \hat{Y}_1' \\ X_1' \end{pmatrix}\Omega^{-1}(\hat{Y}_1 \; X_1)\delta + \begin{pmatrix} \hat{Y}_1' \\ X_1' \end{pmatrix}\Omega^{-1}w , \tag{2.27}$$

so we can have

$$\hat{\delta} = \left[\begin{pmatrix} \hat{Y}_1' \\ X_1' \end{pmatrix}\Omega^{-1}(\hat{Y}_1 \; X_1)\right]^{-1}\begin{pmatrix} \hat{Y}_1' \\ X_1' \end{pmatrix}\Omega^{-1}y . \tag{2.28}$$

The first three matrices in (2.28) can be written as

$$\begin{aligned}
&\begin{pmatrix} \hat{Y}_1' \\ X_1' \end{pmatrix}\Omega^{-1}(\hat{Y}_1 \; X_1) \\
&= \begin{bmatrix} Y_1'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X' \\ X_1' \end{bmatrix}\Omega^{-1}\left(Y_1'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}Y_1 \;\vdots\; X_1\right) \\
&= \begin{bmatrix} Y_1'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y_1 & Y_1'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X_1 \\ X_1'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y_1 & X_1'\Omega^{-1}X_1 \end{bmatrix} \\
&= \begin{bmatrix} Y_1'RY_1 & Y_1'RX_1 \\ X_1'RY_1 & X_1'RX_1 \end{bmatrix} \\
&= A'RA
\end{aligned} \tag{2.29}$$

where

$$R = \Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} \tag{2.30}$$

and

$$\begin{aligned}
&X_1'\Omega^{-1}X_1 \\
&= \begin{pmatrix} I & 0 \end{pmatrix} X'\Omega^{-1}X\begin{pmatrix} I \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} I & 0 \end{pmatrix} X'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X\begin{pmatrix} I \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} I & 0 \end{pmatrix} X'RX\begin{pmatrix} I \\ 0 \end{pmatrix} \\
&= X_1'RX_1.
\end{aligned} \tag{2.31}$$

Furthermore,

$$\begin{aligned}
&\begin{pmatrix} \hat{Y}_1' \\ X_1' \end{pmatrix} \Omega^{-1} \\
&= \begin{bmatrix} Y_1'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} \\ \begin{pmatrix} I & 0 \end{pmatrix}X'\Omega^{-1} \end{bmatrix} \\
&= \begin{bmatrix} Y_1'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} \\ \begin{pmatrix} I & 0 \end{pmatrix}X'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} \end{bmatrix} \\
&= \begin{bmatrix} Y_1'R \\ X_1'R \end{bmatrix} \\
&= A'R\ .
\end{aligned} \tag{2.32}$$

All these would lead to the generalized TSLS estimate of δ as

$$\tilde{\delta} = (A'RA)^{-1}A'Ry\ . \tag{2.33}$$

Without knowledge of θ_j's it is natural to consider the possibility of estimating θ_j's and substituting the estimators for θ_j's involved in (2.33). Thus, to estimate θ_j's are both of interest in themselves and as aids in obtaining improved estimates of δ. For this purpose some consistent estimators of δ, θ_j's are proposed here.

First let,

$$y^{*} = y - Y_1\beta = X_1\bar{\gamma} + w \ . \tag{2.34}$$

Replacing β by its consistent and approximate unbiased TSLS estimator $\hat{\beta}^{*}$ as shown in chapter 1,

$$\hat{\beta}^{*} = \hat{\beta} - (K-K_1-m-1)\hat{Q}_{11}\hat{q}_1 \tag{2.35}$$

where β consists of the first m elements of δ in (2.22), and $\hat{Q}_{11}$ and $\hat{q}_1$ as in (1.98) and (1.100). Then the estimator of y^{*} in (2.34) is

$$\hat{y}^{*} = y - Y_1\hat{\beta}^{*} \tag{2.36}$$

using $\hat{y}^{*}$ instead of y^{*} used in (2.34) we obtain

$$\hat{y}^{*} = X_1\bar{\gamma} + w, \tag{2.37}$$

then,

$$\hat{\bar{\gamma}}^{*} = (X_1'X_1)^{-1}X_1'\hat{y}^{*} \tag{2.38}$$

and the vector of the residuals from the least square regression of $\hat{y}^*$ on X_1 is

$$\begin{aligned}\hat{w}^* &= \hat{y}^* - X_1\hat{\bar{\gamma}}^* \\ &= \hat{y}^* - X_1(X_1'X_1)^{-1}X_1'\hat{y}^* \\ &= M_1\hat{y}^* \\ &= M_1 w \qquad (2.39)\end{aligned}$$

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

Secondly, we define the Hadamard products "*" which is to multiply corresponding elements in two matrices such as

$$\dot{X}_1 = X_1 * X_1 = \begin{bmatrix} x_{11}^2 & x_{12}^2 & \cdots & x_{1K_1}^2 \\ x_{21}^2 & x_{22}^2 & \cdots & x_{2K_1}^2 \\ \cdots\cdots \\ x_{n1}^2 & x_{n2}^2 & \cdots & x_{nK_1}^2 \end{bmatrix}. \qquad (2.40)$$

We rewrite (2.20) by using $\dot{X}_{1t}$, the t-th row of $\dot{X}_1$, as

$$\Omega_t = \sum_{j=1}^{K_1} x_{tj}^2\theta_j = \dot{X}_{1t}\bar{\theta} \qquad t=1,2,\ldots,n, \qquad (2.41)$$

where

$$\bar{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \cdot\cdot \\ \theta_{K_1} \end{bmatrix}. \qquad (2.42)$$

Then the variance matrix of $\hat{w}^*$ is

$$E(\hat{w}^*\hat{w}^{*\prime}) = E(M_1ww'M_1)$$
$$= M_1E(ww')M_1$$
$$= M_1\Omega M_1 \tag{2.43}$$

from which

$$E\,\hat{\dot{w}}^* = \dot{M}_1\bar{\Omega} \tag{2.44}$$

where $\hat{\dot{w}}^*_t = \hat{w}^{*2}_t$, $\dot{m}_{st} = m^2_{st}$, for s, t=1,2,...,n, and

$$\bar{\Omega} = \begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \cdots \\ \Omega_n \end{bmatrix} = \begin{bmatrix} \dot{X}_{11}\theta \\ \dot{X}_{12}\theta \\ \cdots \\ \dot{X}_{1n}\theta \end{bmatrix} = \dot{X}_1\bar{\theta} \ . \tag{2.45}$$

Let $u = \hat{\dot{w}}^* - E\hat{\dot{w}}^* = \hat{\dot{w}}^* - \dot{M}_1\dot{X}_1\bar{\theta}$, then

$$\hat{\dot{w}}^* = \dot{M}_1\dot{X}_1\bar{\theta} + u \ . \tag{2.46}$$

Since $\dot{M}_1\dot{X}_1$ is a known function of X_1 which is also a known matrix, then (2.46) is a linear model and $\dot{M}_1\dot{X}_1$ is rank of K_1. Applying the least squares method to (2.46), an operational version of the estimator of $\bar{\theta}$ yields as

$$\hat{\theta}^* = \left[(\dot{M}_1\dot{X}_1)\dot{M}_1\dot{X}_1\right]^{-1}(\dot{M}_1\dot{X}_1)'\hat{\dot{w}}^* \ . \tag{2.47}$$

It has one clearly undesirable feature as an estimator of variances; namely, some elements of $\hat{\theta}^*$ may be negative. The simplest remedy would be to use the alternative estimator $\hat{\theta}^{*M}$ defined by

$$\hat{\theta}_k^{*M} = \max\{\hat{\theta}^*, 0\} \qquad k=1,2,\ldots,K_1 \tag{2.48}$$

In order to show the consistency of $\hat{\theta}^*$, first we observe that

$$\begin{aligned}\hat{w}^* &= M_1\hat{y}^* \\ &= M_1\left(y - Y_1\beta^*\right) \\ &= M_1\left(y - Y_1\hat{\beta} + (K-K_1-m-1)Y_1\hat{Q}_{11}\hat{q}_1\right) \\ &= M_1\left(A\delta + w - A\hat{\delta} + (K-K_1-m-1)Y_1\hat{Q}_{11}\hat{q}_1\right) \\ &= M_1\left(w - A(\hat{\delta} - \delta) + (K-K_1-m-1)Y_1\hat{Q}_{11}\hat{q}_1\right) \\ &= M_1\left[w - A(\hat{\delta} - \delta) + (K-K_1-m-1)Y_1\hat{Q}\,\hat{q}\right].\end{aligned} \tag{2.49}$$

Let $\nu=\hat{\delta}-\delta$ and $D=(K-K_1-m-1)Y_1\hat{Q}\hat{q}$, (2.49) can be written as

$$\hat{w}^* = M_1\left(w - A\nu + D\right) \tag{2.50}$$

and

$$\hat{\dot{w}}^* = \dot{M}_1\left(\dot{w} + \dot{A}\dot{\nu} + \dot{D} + \dot{G}\right) \tag{2.51}$$

where $\dot{G}$ represents those crossing terms of Hadamard product.

Substituting (2.51) in (2.47), we can verify that

$$
\begin{aligned}
\hat{\theta}^{*} - \bar{\theta} &= \left[(\dot{M}_1\dot{X}_1)'\dot{M}_1\dot{X}_1\right]^{-1}(\dot{M}_1\dot{X}_1)'\dot{M}_1\left(\dot{w} + \dot{A}\dot{\nu} + \dot{D} + \dot{G}\right) - \bar{\theta} \\
&= \left[(\dot{M}_1\dot{X}_1)'\dot{M}_1\dot{X}_1\right]^{-1}(\dot{M}_1\dot{X}_1)'\dot{M}_1\left(\dot{X}_1\bar{\theta} + \dot{A}\dot{\nu} + \dot{D} + \dot{G}\right) - \bar{\theta} \\
&= \left[(\dot{M}_1\dot{X}_1)'\dot{M}_1\dot{X}_1\right]^{-1}(\dot{M}_1\dot{X}_1)'\dot{M}_1\left(\dot{A}\dot{\nu} + \dot{D} + \dot{G}\right) \\
&= \eta. \qquad (2.52)
\end{aligned}
$$

From chapter 1 we see that the elements of η are of order $O_p(n^{-g})$, $g \ge 1/2$.

Now a consistent estimator $\hat{\Omega}^{*}$ of Ω can be constructed by replacing the unknown $\bar{\theta}$ in (2.21) by its estimates $\hat{\theta}^{*}$ in (2.47). And using (2.52), we find that

$$\hat{\Omega}^{*} - \Omega = O_p(n^{-g}), \qquad g \ge 1/2, \qquad (2.53)$$

whence $\hat{\Omega}^{*}$ is a consistent estimator of Ω.

GTSLSR estimator (R&S&U [1980])

By using $\hat{\theta}^{*}$, the estimator of $\bar{\theta}$ in (2.47), to construct the $\hat{\Omega}^{*}$, the estimator of δ in (2.33) is named the Generalized Two Stage Least Square estimator to simultaneous equation with both Fixed and Random coefficients (GTSLSR here and after) such as,

$$\tilde{\delta}^{*} = \left(A'\hat{R}^{*}A\right)^{-1}A'\hat{R}^{*}y \qquad (2.54)$$

where $\hat{R}^{*}$ is the estimator of R^{*} in (2.30), such as

$$\hat{R}^{*} = \hat{\Omega}^{*-1}X\left(X'\hat{\Omega}^{*-1}X\right)^{-1}X'\hat{\Omega}^{*-1} \qquad (2.55)$$

provide $\hat{\Omega}^{*-1}$ exists.

GTSLSRM estimator (R&S&U [1980])

Using $\hat{\theta}^{*M}$, the modified estimator of $\bar{\theta}$ in (2.48), we may obtain another estimator $\hat{\Omega}^{*M}$ instead of Ω used in (2.33), which gives another estimator of δ which may be named the GTSLSRM (here and after estimator where M stands for Modified. The GTSLSRM will be the same as GTSLSR if none of elements of $\hat{\theta}^{*}$ is negative. Which is

$$\tilde{\delta}^{*M} = \left(A'\hat{R}^{*M}A\right)^{-1}A'\hat{R}^{*M}y \qquad (2.56)$$

where $\hat{R}^{*M}$ is the estimator of R^{*} in (2.30), that is

$$\hat{R}^{*M} = \hat{\Omega}^{*M-1}X(X'\hat{\Omega}^{*M-1}X)^{-1}X'\hat{\Omega}^{*M-1} \qquad (2.57)$$

given $\hat{\Omega}^{*M-1}$ exists.

2.4. Asymptotic Properties.

In this section we shall show that the TSLS estimator of the parameter vector δ of (2.17) is asymptotically normally distributed. We shall also show that the limiting distribution of $\sqrt{n}(\tilde{\delta}^{*} - \delta)$ is the same as that of $\sqrt{n}(\tilde{\delta} - \delta)$. The lemma, the theorem and the proof about the properties in R&S&U [1980] are presented.

We set another two assumptions here by following Fuller and Battese (1973, p629) and Swamy and Mehta (1977).

Assumption 2.4.

(i) The elements of R in (2.30) are functions of the parameter vector θ such that the matrices $(\frac{\partial R}{\partial \theta_i})$, $i=1,2,\ldots,K_1$, where θ_i is the i-th element of θ, are continuous functions of θ in an opened sphere of θ°, the true value of θ.

(ii) The matrices A and R are such that

$$\underset{n\to\infty}{\text{Plim}} \frac{1}{n} A'RA = M(\theta) \tag{2.58}$$

is a finite matrix, $M^{-1}(\theta)$ exists for all θ in S, and

$$\underset{n\to\infty}{\text{Plim}} \frac{1}{n} A'\left(\frac{\partial R}{\partial \theta_i}\right)A = H_i(\theta) \tag{2.59}$$

is a finite matrix whose elements are continuous function of θ, $i=1,2,\ldots,K_1$;

(iii)an estimator $\hat{R}^{*}=R(\hat{\theta}^{*})$ for $R=R(\theta^{\circ})$ is available and $\hat{\theta}^{*}$ satisfies the condition

$$\hat{\theta}^{*} = \theta^{\circ} + O_p(N^{-g}) \qquad g>0. \tag{2.60}$$

Assumption 2.5.

(i) Let f'_t be the t-th row of $\Omega^{-1/2}$. Then the elements $f'_t w_1 = f'_t D_{X_1} \varepsilon_1$, $t=1,2,\ldots,n$, are independent and $f'_t v$ has distribution function $F_t(f'w)$, $t=1,2,\ldots,n$, such that

$$\sup_{t=1,2,\ldots n} \int_{|f'w|>c} (f'w)^2 dF_r(f'w) \longrightarrow 0 \qquad \text{as } c \longrightarrow \infty \tag{2.61}$$

(ii) $\operatorname{Plim}_{n\to\infty} \max_{1\le t\le n} (x^{*2}_{1jt}/n)=0$, $j=1,2,\ldots,K$, where x^{*2}_{1jt} is the (t,j)-th element of $\Omega^{-1/2}X$;

(iii) $\operatorname{Plim}_{n\to\infty} (X'\Omega^{-1}X/n)$ is finite and nonsingular, and $\operatorname{Plim}_{n\to\infty}(A'\Omega^{-1}X/n)$ is finite.

Lemma. (R&S&U [1980])

If Assumptions 2.1, 2.2, 2.3, and 2.5 are satisfied, then the limiting distribution of $\sqrt{n}(\tilde{\delta} - \delta)$ is normal with mean vector 0 and variance-covariance matrix $\operatorname{Plim}_{n\to\infty}(A'\Omega^{-1}A)^{-1}$.

Proof.

Write

$$\sqrt{n}(\tilde{\delta} - \delta)$$
$$= \left[\frac{A'\Omega^{-1}X}{n} \left(\frac{X'\Omega^{-1}X}{n}\right)^{-1} \frac{X'\Omega^{-1}A}{n}\right]^{-1} \frac{A'\Omega^{-1}X}{n} \left(\frac{X'\Omega^{-1}X}{n}\right)^{-1} \frac{X'\Omega^{-1}w}{\sqrt{n}}. \tag{2.62}$$

By Assumption 2.5 (iii)

$$\operatorname*{Plim}_{n\to\infty} \left(\frac{A'\Omega^{-1}X}{n}\left(\frac{X'\Omega^{-1}X}{n}\right)^{-1}\frac{X'\Omega^{-1}A}{n}\right)^{-1}\frac{A'\Omega^{-1}X}{n}\left(\frac{X'\Omega^{-1}X}{n}\right)^{-1}\frac{X'\Omega^{-1}w}{\sqrt{n}} = \bar{G}. \tag{2.63}$$

a finite matrix.

It follows from Anderson's [1971, pp. 23-25, 585] Theorem 2.6.1 that under Assumptions 2.5(i) and (ii), $X'\Omega^{-1}w/\sqrt{n}$ converges in distribution to normal with mean vector 0 and variance covariance matrix $\operatorname*{Plim}_{n\to\infty} X'\Omega^{-1}X/n$.

Now applying the limit theorem (x)b in Rao [1973,2c.4,p.122], we have the result of the Lemma.

Theorem 2.1. (R&S&U [1980])

Suppose that Assumptions 2.1, 2.2, 2.3, and 2.4 are true. Then the limiting distribution of $\sqrt{n}(\tilde{\delta}^{*} - \delta)$ is the same as that of $\sqrt{n}(\tilde{\delta} - \delta)$.

Proof.

Write

$$\tilde{\delta}^{*} - \delta = \left(A'R(\hat{\theta}^{*})A\right)^{-1}A'R(\hat{\theta}^{*})w \tag{2.64}$$

where $R(\hat{\theta}^{*}) = \hat{\Omega}^{*-1}X\left(X'\hat{\Omega}^{*-1}X\right)^{-1}X'\hat{\Omega}^{*-1}$.

By a Taylor's expansion with remainder about the true parameter $\bar{\theta}^{\circ}$, we obtain,

$$\left(A'R(\hat{\theta}^{*})A\right)^{-1}A'R(\hat{\theta}^{*})w = \left[\frac{1}{n}A'R(\theta^{\circ})A\right]^{-1}\left[\frac{1}{n}A'R(\theta^{\circ})w\right]$$

$$+\sum_{i=1}^{K_1}\left\{\left[\frac{1}{n}A'R(\theta^{1})A\right]^{-1}\left[\frac{1}{n}\; A'\left[\frac{\partial}{\partial_{\theta t}}R(\theta)\right]_{\theta=\theta 1}w\right]\right.$$

$$-\left[\frac{1}{n}A'R(\theta^{1})A\right]^{-1}\left[\frac{1}{n}\; A'\left[\frac{\partial}{\partial_{\theta t}}R(\theta)\right]_{\theta=\theta 1}A\right]$$

$$\left.\cdot\left[\frac{1}{n}A'R(\theta^{1})A\right]^{-1}\left[\frac{1}{n}\; A'R(\theta^{1})w\right]\right\}\left(\tilde{\theta}_i^{*}-\theta_i^{\circ}\right) \qquad (2.65)$$

where θ^1 is between θ° and θ. By Assumption 2.4 and limit theorem (x)a in Rao [1973,2c.4,pp.122], it follows then

$$\tilde{\delta}^{*}-\delta = \left(\tilde{\delta}^{*}-\delta\right) + O_p\left(n^{-(1/2+g)}\right) \qquad (2.66)$$

or

$$\sqrt{n}\left(\tilde{\delta}^{*}-\delta\right) = \sqrt{n}\left(\tilde{\delta}^{*}-\delta\right) + O_p\left(n^{+g}\right) \qquad (2.67)$$

which, in view of the theorem (x)d in Rao [1973,2c.4,p.122], gives the result of the Theorem.

The problem of estimation of simultaneous equation models with fixed and random coefficients was studied and some useful estimators were discussed. An efficient single equation method of estimating the structural coefficients of endogenous variables and the means of structural coefficients of exogenous variables was shown and its asymptotic properties was established. A reformulated Klein Model I and corresponding GTSLSR estimates of the coefficients will be shown in the section 3.6.

CHAPTER THREE

BOOTSTRAPPING GTSLSR ESTIMATORS IN A SIMULTANEOUS EQUATION MODEL WITH BOTH FIXED AND RANDOM COEFFICIENTS

3.1 Introduction.

In this chapter, we consider the bootstrapping GTSLSR estimators of simultaneous equation model with both fixed and random coefficients (FRSE model here and after) which we have studied in chapter two. As discussed in chapter two, existing methods are largely asymptotic, and may not apply with finite samples. We use "the bootstrap", a computer-based methodology, to check the accuracy of the asymptotic results and to make alternative estimates of the standard errors that are more reliable.

The bootstrap methodology will be introduced in section 3.2. In section 3.3, we will give a brief review of FRSE model and some kind of notations of the consistent GTSLSR estimators. Then, the large sample properties and the finite sample properties of bootstrap GTSLSR estimators will be studied in section 3.4 and 3.5 respectively. An experiment based on bootstrapping Klein model I will be shown in the last section and some empirical results will be presented there too.

3.2. Bootstrap

Bootstrap is a relatively new statistical technique which was first introduced by Efron (1979). This nonparametric method resamples the original observations in a suitable way in order to construct "pseudo data" on which the estimator of interest is exercised. Efron (1982) considered some general problems, such as median, bias, and regression model.

Asymptotic properties of the bootstrap were first studied by Bickel and Freedman [1981] (B&F here and after) and Singh [1981]. Freedman [1981] developed some asymptotic properties for the application of bootstrap to the regression model.

Freedman and Peters [1984a] (F&P [1984a] here and after) discussed estimating standard errors for regression coefficients obtained by constrained generalized least squares with an estimated covariance matrix. In Freedman and Peter [1984b] (F&P [1984b] here and after), they applied the bootstrap to an econometrics model which is fitted by three stage least squares. In his paper "on Bootstrapping Two-stage Least Squares Estimates in Stationary Linear Models" (1984), Freedman showed that for large samples the bootstrap will give the right answers even in the presence of heteroscedastic errors.

Bootstrap provides a very useful tool for assessing statistical accuracy. Consider for example estimation of the variance of a statistic T under a true distribution F, i.e. $VAR_F T$, the bootstrap works by replacing the unknown distribution F by the empirical distribution function $\hat{F}$ and the bootstrap estimate of $VAR_{\hat{F}}T$. Unless T is very simple, this cannot be computed analytically, and hence must be approximated by a Monte Carlo simulation. To do this, we sample n times with replacement from the original data (n is the sample size), then evaluate the statistic of interest for this "bootstrap sample". This process is repeated B times, where B is typically 100 to 1000. the Monte Carlo estimate of $VAR_{\hat{F}}T$ is the sample variance of the B bootstrap values of T. We can see that sampling with replacement from the data is equivalent to sampling from $\hat{F}$. That is, in bootstrapping, only the observed data are required and no other extraneous data are needed.

Suppose that our data consist of random sample $X_1, X_2, \ldots, X_n$ from an unknown distribution F. The statistic of interest is some symmetric function $T(X_1, X_2, \ldots, X_n)$. An estimate of functional $Q(T,F,X)$ is required. Using $\hat{F}$ to represent the empirical distribution function, the bootstrap estimate is defined as $Q(T,\hat{F},X)$. Usually, we cannot compute this analytically, so we estimate it through a Monte Carlo simulation. This can be done by writing $Q(T,\hat{F},X)$ in terms of quantities of the form $E_{\hat{F}}R$ and estimating each quantity by a Monte Carlo estimate of expectation. For example, in the case of $Q(T,F,X)=VAR_F T$, we write $VAR_{\hat{F}}T=E_{\hat{F}}T^2-(E_{\hat{F}}T)^2$. We draw B bootstrap samples which are size n drawn with replacement from X_1, X_2, ..., X_n and compute the bootstrap values $T_1^B, T_2^B, \ldots, T_B^B$. The Monte Carlo estimate of $VAR_F T$ is $\sum_{b=1}^{B} T_b^{B2}/B-(\sum_{b=1}^{B} T_b^B/B)^2$.

In the regression case the bootstrap is useful for investigations when mathematical analysis can give only asymptotic results. More particularly, in the model discussed in chapter two, when the simultaneity and random coefficients involve, the bootstrap will be applied to check the accuracy of the asymptotic and to make alternative estimates of the standard errors that are more reliable.

3.3. The Model and the Consistent GTSLSR Estimators.

The model discussed in the second chapter is

$$\begin{aligned} y &= Y_1\beta + X_1\bar{\gamma} + D_{X_1}\varepsilon \\ &= A\delta + w \end{aligned} \tag{3.1}$$

where $E\varepsilon=0$, $E\varepsilon\varepsilon'=\Theta$, $Ew=0$ and $Eww'=\Omega$.

Whatever the variance-covariance matrix Ω may be, it is possible to linearly transform the errors w by a matrix H, so that

$$Cov(Hw) = I \tag{3.2}$$

Since Ω is the variance-covariance matrix of w, it is positive definite and symmetric, therefore there exists an invertible matrix H, so that

$$H\Omega H' = I \tag{3.3}$$

It is easily to verify that

$$H'H = \Omega^{-1} . \tag{3.4}$$

So

$$E(Hww'H') = HE(ww')H' = H\Omega H' = I \ . \tag{3.5}$$

Premultiply 3.1 by H, then

$$Hy = HA\delta + Hw \tag{3.6}$$

Now the design matrix X'H' is orthogonal to the Hw in equation (3.6), also

$$E(X'H'Hw) = E(X\Omega^{-1}w) = 0 \tag{3.7}$$

Premultiply 3.4 by X'H' and use the assumed orthogonality,

$$X'\Omega^{-1}y = X'\Omega^{-1}A\delta \tag{3.8}$$

To get the standard theory in the present setting, let

$$Q = E(X'\Omega^{-1}y) \tag{3.9}$$

$$R = E(X'\Omega^{-1}A) \tag{3.10}$$

and

$$S = E(X'\Omega^{-1}X) \ . \tag{3.11}$$

Take the expectations of (3.5) and use (3.4), we have

$$Q = R\delta \ . \tag{3.12}$$

Assumption 3.1.

(i) (y, A, X) is a random observable vector of dimension 1+L+K where $L=m+K_1$.

(ii) $E\left[\ |y,\ A,\ X|^2\ \right] < \infty.$

(iii) $E\left[\ |y,\ A,\ X|^4\ \right] < \infty.$

where $|\ |$ represents the Euclidean norm.

Assumption 3.2. System (3.9) is identified, i.e. $K \geq m+K_1$, rank(R) is $m+K_1$ and S is invertible.

From chapter two we know the consistent GTSLSR estimators of δ is

$$\tilde{\delta} = \left(A'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}A\right)^{-1}A'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \ . \tag{3.13}$$

Using (3.9)-(3.11), we can have

$$\tilde{\delta} = \left(R'S^{-1}R\right)^{-1}R'S^{-1}Q \ . \tag{3.14}$$

We shall use the notations formulated here in the next two sections.

3.4. Large Sample Properties of the Bootstrap GTSLSR.

Here, we show that the bootstrap gives the right answers with large samples, so that the bootstrap is at least as sound as the conventional asymptotics.

Following the discussion in section 3.3, the data are modeled as a sample of size n from this structure. More particularly,

(i) The vectors (y_i, A_i, X_i, w_i) for $i=1,2,\ldots,n$, are independent with the common (unknown) distribution as (y, A, X, w) in 2+L+K dimensional space,

(ii) $\varepsilon_{(i)} \sim N(0,\Theta)$,

(iii) $X_i'H'$ is orthogonal to Hw_i in the sense $E(X_i\Omega^{-1}w_i)=0$.

The quantities Q, R, and S can be calculated as

$$Q_n = \frac{1}{n} \sum_{i=1}^{n} X_i'\Omega^{-1}y_i \tag{3.15}$$

$$R_n = \frac{1}{n} \sum_{i=1}^{n} X_i'\Omega^{-1}A_i \tag{3.16}$$

$$S_n = \frac{1}{n} \sum_{i=1}^{n} X_i'\Omega^{-1}X_i \tag{3.17}$$

and one more,

$$\Delta_n = \frac{1}{n} \sum_{i=1}^{n} X_i'\Omega^{-1}w_i \;. \tag{3.18}$$

The model can be written as

$$Q_n = R_n\delta + \Delta_n \tag{3.19}$$

and the GTSLSR estimator is

$$\tilde{\delta}_n = \left(R_n' S_n^{-1} R_n\right)^{-1} R_n' S_n^{-1} Q_n \ . \tag{3.20}$$

Since $Q_n \longrightarrow Q$, $R_n \longrightarrow R$, and $S_n \longrightarrow S$ by the law of large numbers, and from the lemma in chapter two, we know

$$\sqrt{n}\left(\tilde{\delta}_n - \delta\right) = \left(R_n' S_n^{-1} R_n\right)^{-1} R_n' S_n^{-1} \left(\sqrt{n}\ \Delta_n\right) \tag{3.21}$$

is normal with mean vector 0 and variance-covariance matrix $\operatorname*{Plim}_{n\to\infty}\left(R_n' S_n^{-1} R_n\right)^{-1}$. The $\sqrt{n}\ \Delta_n$ satisfies the central limit theorem in K_1-dimensional space.

The residuals from the fit are

$$\begin{aligned} \hat{w}_i &= y_i - A_i \tilde{\delta}_n \\ &= y_i - A_i \left(R_n' S_n^{-1} R_n\right)^{-1} R_n' S_n^{-1} Q_n \ . \end{aligned} \tag{3.22}$$

As data, the $H\hat{w}$ will not in general be exactly orthogonal to the $X_i H$. To keep this property, let $\tilde{w}_i$ be part of the $\hat{w}_i$:

$$\tilde{w}_i = \hat{w}_i - X_i \tilde{b}_n \qquad i=1,\ldots,n, \tag{3.23}$$

where $\tilde{b}_n = S_n^{-1}\left(Q_n - R_n \tilde{\delta}_n\right)$. So that when we use the vector sign in (3.23) and multiply it on the left by $X'\Omega^{-1}$, we can have

$$
\begin{aligned}
X'\Omega^{-1}\tilde{w} &= X'\Omega^{-1}\hat{w} - X'\Omega^{-1}XS_n^{-1}\left(Q_n - R_n\tilde{\delta}_n\right) \\
&= X'\Omega^{-1}\left(y_n - A_n\tilde{\delta}_n\right) - Q_n + R_n\tilde{\delta}_n \\
&= X'\Omega^{-1}y_n - X'\Omega^{-1}A_n\tilde{\delta}_n - Q_n + R_n\tilde{\delta}_n \\
&= 0.
\end{aligned} \tag{3.24}
$$

This means $H\tilde{w}$ is orthogonal to the $X'H'$.

Bootstrap method can now be used to generate the "pseudo data". In (3.1), $A=(Y_1, X_1)$ and $w=D_{X_1}\varepsilon$, where Y_1 and X_1 are related to w. So it is inappropriate to resample the residuals, for the dependence. Instead, it is necessary to resample the vectors. More specifically, let $\tilde{\mu}_n$ be the empirical distribution of the vector $(A_i, X_i, \tilde{w}_i)$ for i=1,2,...,n. Thus $\tilde{\mu}_n$ is a probability on R^{1+L+K}, putting mass 1/n at each vector $(A_i, X_i, \tilde{w}_i)$. Given (y_i, A_i, X_i) for i=1,2,...,n, let $(A_j^B, X_j^B, \tilde{w}_j^B)$ be independent, with common distribution $\tilde{\mu}_n$, for j=1,...,n. Informally, data from a small sample can be used to judge the likely performance of a large sample. Resampling the data this way, any relationship there may be between instruments and disturbances can be preserved. Let

$$
y_j^B = A_j^B\tilde{\delta}_n + \tilde{w}_j^B . \tag{3.25}
$$

The data with B superscription can be used to get some bootstrap results as follows,

$$
\Omega^B = D_{X_1^B} \otimes D_{X_1^B} , \tag{3.26}
$$

where the Θ is as defined in (2.20) and the X_1^B is the last K_1 columns in the matrix A^B,

$$Q_n^B = \frac{1}{n} \sum_{j=1}^{n} X_j^{B\prime} \Omega^{B-1} y_j^B , \tag{3.27}$$

$$R_n^B = \frac{1}{n} \sum_{j=1}^{n} X_j^{B\prime} \Omega^{B-1} A_j^B , \tag{3.28}$$

$$S_n^B = \frac{1}{n} \sum_{j=1}^{n} X_j^{B\prime} \Omega^{B-1} X_j^B \tag{3.29}$$

and

$$\Delta_n^B = \frac{1}{n} \sum_{j=1}^{n} X_j^{B\prime} \Omega^{B-1} \tilde{w}_j^B . \tag{3.30}$$

The bootstrap estimator of GTSLSR is

$$\begin{aligned}
\tilde{\delta}_n^B &= \left(R_n^{B\prime} S_n^{B-1} R_n^B\right)^{-1} P_n^{B\prime} S_n^{B-1} Q_n^B \\
&= \left(R_n^{B\prime} S_n^{B-1} R_n^B\right)^{-1} R_n^{B\prime} S_n^{B-1} \frac{1}{n} \sum_{j=1}^{n} X_j^{B\prime} \Omega^{B-1} \left(A_j^B \tilde{\delta}_n + \tilde{w}_j^B\right) \\
&= \left(R_n^{B\prime} S_n^{B-1} R_n^B\right)^{-1} R_n^{B\prime} S_n^{B-1} R_n^B \tilde{\delta}_n^B + \left(R_n^{B\prime} S_n^{B-1} R_n^B\right)^{-1} R_n^{B\prime} S_n^{B-1} \Delta_n^B \\
&= \tilde{\delta}_n + \left(R_n^{B\prime} S_n^{B-1} R_n^B\right)^{-1} R_n^{B\prime} S_n^{B-1} \Delta_n^B.
\end{aligned} \tag{3.31}$$

The bootstrap principle is that the error structure of the estimates with B superscription, which can be computed directly from the data, approximates that in the original estimates. It will be shown that the conditional law of $\sqrt{n}\,\Delta_n^B$ must be close to the unconditional law of $\sqrt{n}\,\Delta_n$, i.e. the bootstrap approximation is valid.

We can now use lemmas from B&F [1981] and lemmas and the theorem in Freedman [1984]. The only difference in our estimates from Freedman [1984] is that we use GTSLSR estimators for the coefficients and suppose the variance-covariance θ of ε is known. But it does not change the dimension of the space. So Freedman's theorem can be used here. Following are the relative lemmas and the theorem.

First, the definition of "Mallows metrics" discussed in section 8 of Freedman [1981] is given here

Definition.

Let d_l^p be the Mallows metric for probabilities in R^p, relative to the Euclidean norm $|\ |$. Thus, if μ and ν are probabilities in R^p, $d_l^p(\mu,\nu)$ is the infimum of $E\left(|U - V|^l\right)^{1/l}$ over all pairs of random vectors U and V, where U has law μ and V has law ν. Abbreviate d_l for d_l^1. Only l=1 or 2 are of present interest. Let $1 \le p \le \infty$; only p=1 or 2 are of present interest.

Let B be a separable Banach space with $\|\cdot\|$. Let $\Gamma_p=\Gamma_p(B)$ be the set of probabilities γ on the Borel σ-field of B, such that $\int\|x\|^p\gamma(dx)<\infty$. For α and β in Γ_p, let $d_p(\alpha,\beta)$ be the infimum of $E\left(\|U - V\|^p\right)^{1/p}$ over pairs of B-valued random variables X and Y, where X has law α and Y has law β.

lemma 3.1

(a) The infimum is attained.

(b) d_p is a metric on Γ_p.

See proof of lemma 8.1 in B&F [1981].

lemma 3.2

Let ν_n, ν be probabilities in R^j. Let $\alpha \geq 1$, and suppose the Mallows metric $d_p(\nu_n, \nu) \longrightarrow 0$. Let M_n be a linear map from R^j to R^k, also equipped with the Euclidean norm. Suppose $M_n \longrightarrow M$. Then $d_\alpha^k(\nu_n M_n^{-1}, \nu M^{-1}) \longrightarrow 0$.

Proof. Construct U_n and U with distribution ν_n and ν respectively, and

$$E\left(|U_n - U|^\alpha\right)^{1/\alpha} = d_\alpha^j(\nu_n, \nu) \tag{3.32}$$

which is from lemma 3.1. The $\|\cdot\|$ is a operator norm, so $|M_n u| \leq \|M_n\| \cdot |u|$. Then

$$\begin{aligned} d_\alpha^k(\nu_n M_n^{-1}, \nu M^{-1}) &\leq E\left(|M_n U_n - MU|^\alpha\right)^{1/\alpha} \\ &= E\left(|M_n U_n - M_n U + M_n U - MU|^\alpha\right)^{1/\alpha} \\ &\leq E\left(|M_n(U_n - U)|^\alpha\right)^{1/\alpha} + E\left(|(M_n - M)U|^\alpha\right)^{1/\alpha} \\ &\leq \|M_n\| \cdot E\left(|U_n - U|^\alpha\right)^{1/\alpha} + \|M_n - M\| \cdot E\left(|U|^\alpha\right)^{1/\alpha} \\ &\longrightarrow 0. \end{aligned} \tag{3.33}$$

Lemma 3.3

Let X_i be independent B-valued random variables, with common distribution $\mu \in \Gamma_p$. Let μ_n be the empirical distribution of $X_1, \ldots, X_n$.

Then $d_p(\nu_n, \nu) \longrightarrow 0$ a.e.

See proof of lemma 8.4 in B&F [1981].

lemma 3.4

Let μ_n be the empirical distribution of (y_i, A_i, X_i) for $1 \le i \le n$. Let μ be the common theoretical distribution of (y_i, A_i, X_i). Then $d_4^{1+L+K}(\mu_n, \mu) \longrightarrow 0$ a.e.

Proof. Same as lemma 3.3.

Lemma 3.5

Let $\tilde{\mu}_n$ be the empirical distribution of $(A_i, X_i, \tilde{w}_i)$ for $1 \le i \le n$. Let $\tilde{\mu}$ be the common theoretical distribution of (A_i, X_i, w_i). Then $d_4^{1+L+K}(\tilde{\mu}_n, \tilde{\mu}) \longrightarrow 0$ a.e.

Proof. This follows from Lemma 3.2 and 3.4, because $\tilde{\mu}_n$ is the image of μ_n under the linear mapping L_n:

$$L_n(y, a, x) = (a, x, y - a\tilde{\delta}_n - \tilde{b}_n' x) \tag{3.34}$$

and $\tilde{\delta}_n \longrightarrow \delta$, $\tilde{b}_n \longrightarrow 0$ a.e. So, $\tilde{\mu}_n$ tends to the image of μ under the linear mapping

$$L_n(y, a, x) = (a, x, y - a\delta) . \tag{3.35}$$

This is $\tilde{\mu}$.

Lemma 3.6

Let U_j be independent; likewise for V_j; assume the laws are in Γ_p. Then

$$d_p\Big(\sum_{j=1}^{n} U_j,\ \sum_{j=1}^{n} V_j\Big) \le \sum_{j=1}^{n} d_p\big(U_j,\ V_j\big)\ . \tag{3.36}$$

See proof of lemma 8.6 in B&F [1981].

Lemma 3.7

Suppose B is a Hilbert space with inner product $\langle\cdot,\cdot\rangle$, and $p=2$. Suppose the U_j are independent, likewise for V_j; assume the laws are in Γ_2, and $E(U_j)=E(V_j)$. Then

$$d_2\Big(\sum_{j=1}^{n} U_j\ \sum_{j=1}^{n} V_j\Big)^2 \le \sum_{j=1}^{n} d_2\big(U_j\ V_j\big)^2 \tag{3.37}$$

See proof of lemma 8.7 in B&F [1981].

Theorem 3.1. (B&F [1984])

Along almost all sample sequences, as $n \to \infty$, conditionally on the data:

(a) $Q_n^B \to Q$ and $R_n^B \to R$ and $S_n^B \to S$ in conditional probability.

(b) the conditional law of $\sqrt{n}\,\Delta_n^B$ has the same limit as the unconditional law of $\sqrt{n}\,\Delta_n$.

Proof.

(a) Let $U_j = X_j^{B\prime}\Omega^{-1}y_j^B/n$ and $V_j = X_j'\Omega^{-1}y_j/n$, and their laws be in Γ_1. The

U_j's are independent, so are V_j's for $j=1,2,\ldots n$. Let

$$E\left(|U_j - V_j| \right) = d(U_j\ V_j) \tag{3.38}$$

and $d(U_j\ V_j) \longrightarrow 0$ as $n \longrightarrow \infty$. Using Lemma 3.6, then

$$d_1\left(\sum_{j=1}^{n} U_j\ \sum_{j=1}^{n} V_j \right) \le \sum_{j=1}^{n} d_1\left(U_j\ V_j\right) \tag{3.39}$$

i.e. $Q_n^B \longrightarrow Q_n$ as $n \longrightarrow \infty$. From the large sample theory, we know that $Q_n \longrightarrow Q$. Then

$$\begin{aligned} E\left(Q_n^B - Q \right) &= E\left(Q_n^B - Q_n + Q_n - Q \right) \\ &\le E\left(Q_n^B - Q_n \right) + E\left(Q_n - Q \right) \end{aligned} \tag{3.40}$$

So $Q_n^B \longrightarrow Q$ in conditionally probability.

(b) Let $U_j = X_j^{B\prime}\Omega^{-1}\tilde{w}_j^B/n$ be independently, likewise $V_j = X_j'\Omega_j^{-1}w_j/n$. Indeed, $E\left(U_j\right) = E\left(X_j^{B\prime}\Omega^{-1}\tilde{w}_j^B \right) = 0$ due to the orthogonality, and by lemma 3.5, its conditional law is close in d_2^K to the unconditional law of $X_j'\Omega^{-1}w_j$. Now suppose the assumption 3.1 is true, using lemma 3.7 we can have

$$d_2\left(\sum_{j=1}^{n} U_j,\ \sum_{j=1}^{n} V_j \right) \le \sum_{j=1}^{n} d_2\left(U_j,\ V_j\right) . \tag{3.41}$$

That is as $n \rightarrow \infty$, given the data, the d_2^K-distance between the conditional law of $\sqrt{n}\,\Delta_n^B$ and the unconditional law of $\sqrt{n}\,\Delta_n$ tends to 0.

Corollary 3.1. (B&F [1984])

The conditional law of $\sqrt{n}\left(\tilde{\delta}_n^B - \tilde{\delta}_n\right)$ and unconditional law of $\sqrt{n}\left(\tilde{\delta}_n - \delta\right)$ have same limit.

From (3.31), we have

$$\sqrt{n}\left(\tilde{\delta}_n^B - \tilde{\delta}_n\right) = \left(R_n^{B\prime} S_n^{B-1} R_n^B\right)^{-1} R_n^{B\prime} S_n^{B-1} \sqrt{n}\,\Delta_n^B \qquad (3.42)$$

is a continuous function of $\sqrt{n}\,\Delta_n^B$. An application of Slutzky's theory indicates that the conditional law of $\sqrt{n}\left(\tilde{\delta}_n^B - \tilde{\delta}_n\right)$ is the function of conditional law of $\sqrt{n}\,\Delta_n^B$. Similarly, the conditional law of

$$\sqrt{n}\left(\tilde{\delta}_n - \tilde{\delta}\right) = \left(R_n' S_n^{-1} R_n\right)^{-1} R_n' S_n^{-1} \sqrt{n}\,\Delta_n \qquad (3.43)$$

is just a function of unconditional law of $\sqrt{n}\,\Delta_n$.

The corollary now is followed directly by using theorem 3.1.

3.5. Finite Sample Properties of the Bootstrap GTSLSR Estimators.

As we stated in section 3.2, the principle of bootstrap is to resample the pseudo data from the original data with replacement B times and each time to construct a statistics of interest, we then use the mean and standard error of the bootstrap statistics to approximate the real distribution of the statistics. We are showing here, when n is finite, the bootstrap actually outperforms the conventional asymptotics.

First, an algorithm is given to obtain $\delta^B_{(b)}$ for $b = 1, 2, \ldots, B$, where B is usually between 100 to 1000.

Algorithm 3.1.

Step 1. Obtain $\tilde{\delta}$ in (3.17) and $\tilde{w}$ in (3.20) respectively.

Step 2. Fit a nonparametric MLE of F,

$\hat{F}$: mass $1/n$ at $(A_i, X_i, \tilde{w}_i)$, $i=1, 2, \ldots, n$.

Step 3. Choose a random seed. Then set this value in the UNIFORM function which is a scalar function returning one or more pseudo random numbers with a uniform distribution over 0 to 1 in SAS IML procedure.

Step 4. Generate a random integer $1<j<n$ by using the number obtained in step 3, draw a bootstrap sample $(A_j, X_j, \tilde{w}_j)$ with replacement from $(A, X, \tilde{w})$.

Step 5. Repeat step 3 n times to construct $(A^B_{(b)}, X^B_{(b)}, \tilde{w}^B_{(b)})$.

Step 6. Reconstruct the linear responses as

$$y^B_{(b)} = A^B_{(b)}\tilde{\delta} + \tilde{w}^B_{(b)} \quad . \tag{3.44}$$

Step 7. Compute

$$Q^B_{(b)} = \frac{1}{n} X^{B\prime}_{(b)}\Omega^{-1}y^B_{(b)} \quad , \tag{3.45}$$

$$R^B_{(b)} = \frac{1}{n} X^{B\prime}_{(b)}\Omega^{-1}A^B_{(b)} \quad , \tag{3.46}$$

$$S^B_{(b)} = \frac{1}{n} X^{B\prime}_{(b)}\Omega^{-1}X^B_{(b)} \quad , \tag{3.47}$$

$$\Delta^B_{(b)} = \frac{1}{n} X^{B\prime}_{(b)}\Omega^{-1}\tilde{w}^B_{(b)} \quad , \tag{3.48}$$

and

$$\tilde{\delta}^B_{(b)} = \left[R^{B\prime}_{(b)}S^{B-1}_{(b)}R^B_{(b)}\right]^{-1}R^{B\prime}_{(b)}S^{B\;-1}_{(b)}Q^B_{(b)} \quad . \tag{3.49}$$

Step 8. Repeat steps 3 to 7 for $b=1,2,\ldots,B$.

The following theorems concern the finite sample properties of the bootstrap estimates $\tilde{\delta}^{B}_{(b)}$ for b=1,2,...,B.

Theorem 3.2.

Let n be finite and

$$\tilde{\delta}^{B} = \left[\frac{1}{B}\sum_{b=1}^{B}\tilde{\delta}^{B}_{(b)}\right] . \tag{3.50}$$

Then

$$\tilde{\delta}^{B} \xrightarrow[B]{} \tilde{\delta} \tag{3.51}$$

in conditional probability.

Proof.

Substitute (3.35) into (3.36),

$$Q^{B}_{(b)} = \frac{1}{n} X^{B\prime}_{(b)}\Omega^{-1}\left[A^{B}_{(b)}\tilde{\delta} + \tilde{W}^{B}_{(b)}\right] = R^{B}_{(b)}\tilde{\delta} + \Delta^{B}_{(b)} \tag{3.52}$$

then

$$\tilde{\delta}^{B}_{(b)} = \tilde{\delta} + \left(R^{B\prime}_{(b)}S^{B-1}_{(b)}R^{B}_{(b)}\right)^{-1}R^{B\prime}_{(b)}S^{B\ -1}_{(b)}\Delta^{B}_{(b)} . \tag{3.53}$$

So

$$\tilde{\delta}^{B} - \tilde{\delta} = \frac{1}{B}\sum_{b=1}^{B}\left(R^{B\prime}_{(b)}S^{B-1}_{(b)}R^{B}_{(b)}\right)^{-1}R^{B\prime}_{(b)}S^{B\ -1}_{(b)}\Delta^{B}_{(b)} \tag{3.55}$$

Upon the application of large sample theory and the assumed orthogonality. $\tilde{\delta}^B - \tilde{\delta} \xrightarrow[B]{} 0$ in probability follows directly.

Theorem 3.3.

Let n be finite and

$$SD(\tilde{\delta}^B) = \frac{1}{B} \sum_{b=1}^{B} \left[(\tilde{\delta}^B_{(b)} - \tilde{\delta}^B)(\tilde{\delta}^B_{(b)} - \tilde{\delta}^B)' \right] \tag{3.56}$$

Then

$$SD(\tilde{\delta}^B) \xrightarrow[B]{} Cov(\tilde{\delta}) \tag{3.57}$$

in conditional probability.

Proof.

From (3.53)

$$\begin{aligned} \tilde{\delta}^B_{(b)} - \tilde{\delta}^B &= \tilde{\delta} - \left(R^{B\prime}_{(b)} S^{B-1}_{(b)} R^B_{(b)}\right)^{-1} R^{B\prime}_{(b)} S^{B\,-1}_{(b)} \Delta^B_{(b)} - \tilde{\delta}^B \\ &= \tilde{\delta} - \tilde{\delta}^B - \left(R^{B\prime}_{(b)} S^{B-1}_{(b)} R^B_{(b)}\right)^{-1} R^{B\prime}_{(b)} S^{B\,-1}_{(b)} \Delta^B_{(b)} \end{aligned} \tag{5.58}$$

$$\begin{aligned} &\frac{1}{B} \sum_{b=1}^{B} \left[(\tilde{\delta}^B_{(b)} - \tilde{\delta}^B)(\tilde{\delta}^B_{(b)} - \tilde{\delta}^B)' \right] \\ &= \frac{1}{B} \sum_{b=1}^{B} \Big[(\tilde{\delta} - \tilde{\delta}^B)(\tilde{\delta} - \tilde{\delta}^B)' \\ &\quad + 2(\tilde{\delta} - \tilde{\delta}^B) \left(R^{B\prime}_{(b)} S^{B-1}_{(b)} R^B_{(b)}\right)^{-1} R^{B\prime}_{(b)} S^{B\,-1}_{(b)} \Delta^B_{(b)} \\ &\quad \left(R^{B\prime}_{(b)} S^{B-1}_{(b)} R^B_{(b)}\right)^{-1} R^{B\prime}_{(b)} S^{B\,-1}_{(b)} \Delta^B_{(b)} \Delta^{B\prime}_{(b)} S^{B\,-1}_{(b)} R^B_{(b)} \left(R^{B\prime}_{(b)} S^{B-1}_{(b)} R^B_{(b)}\right)^{-1} \Big] \end{aligned} \tag{3.59}$$

The theorem 3.2 indicates that the first two terms in (3.58) tend to 0 when B goes to infinite. Since the $(A^{B}_{(b)}, X^{B}_{(b)}, \tilde{W}^{B}_{(b)})$ are sampled randomly from $\hat{F}$, as $B \longrightarrow \infty$, the third term in (3.58) is nothing but the estimate of variance covariance of $\tilde{\delta}$.

In section 3.3 and this section only the properties of $\tilde{\delta}$ are discussed but not the $\tilde{\delta}^{*}$. In practice, the variance covariance of w, i.e. Ω, is almost impossible to know, so $\tilde{\delta}$ is not available most of time. But theorem 2.1 indicates that $\sqrt{n}\left(\tilde{\delta}^{*} - \delta\right)$ has the same limiting distribution as that of $\sqrt{n}\left(\tilde{\delta} - \delta\right)$. So the $\tilde{\delta}^{*}$ may be bootstrapped and the same properties as $\tilde{\delta}$ may be obtained. We will show some empirical results in section 3.5.

3.6. Bootstrapping the Klein Model I.

Klein Model I will be first introduced in this section. The TSLS, GTSLSR, and GTSLSRM estimates of coefficients will be calculated and shown in table 4.1. In the last section, we bootstrap the Klein Model I to get some empirical results for our studies in the two previous sections. To do this two algorithms are constructed. The mean and variance of bootstrap GTSLSR estimators are shown in table 4.2 and 4.3.

I. The Klein Model I and The Estimates.

The model we will discussed here is one of the models constructed by L.R. Klein [1950], the so called Klein Model I. It is eight-equation system based on annual data for the United States in the period of time between the two wold wars. Furthermore, it is dynamic in the sense that it is formulated in terms of variables belonging to different points or periods of time. It is formulated as follows:

$$C = \beta_{C1}P + \beta_{C2}W + \gamma_{C1}P_{-1} + \gamma_{C0} + \varepsilon_1 \qquad (3.60)$$

$$I = \beta_{I1}P + \gamma_{I1}K_{-1} + \gamma_{I2}P_{-1} + \gamma_{I0} + \varepsilon_2 \qquad (3.61)$$

$$W^* = \beta_{W^*1}E + \gamma_{W^*1}E_{-1} + \gamma_{W^*2}YR + \gamma_{W^*0} + \varepsilon_3 \qquad (3.62)$$

$$E = Y + T - W^{**} \qquad (3.63)$$

$$Y = C + I + G - T \qquad (3.64)$$

$$P = Y - W^* - W^{**} \qquad (3.65)$$

$$K = I + K_{-1} \qquad (3.66)$$

$$W = W^* + W^{**} \qquad (3.67)$$

Where the endogenous variables are

C = consumption,

I = investment,

W^*= private wage bill,

E = private product

Y = national income, P = profits,

K = end-of-year capital stock,

W = total wage bill,

and the exogenous and lagged variables are

1 = unity,

W^{**}= government wage bill,

T = indirect taxes,

G = government expenditures,

YR= time,

K_{-1} = the lagged variable of K,

P_{-1} = the lagged variable of P,

E_{-1} = the lagged variable of E.

We will treat the lagged endogenous variables K_{-1}, P_{-1}, and E_{-1} as the exogenous variables.

The structural form of this system is

$$BY + X\Gamma + E = 0 \tag{3.68}$$

where

$$Y = \left(C \ I \ W^{*} \ E \ Y \ P \ K \ W \right), \tag{3.69}$$

and

$$X = \left(K_{-1} \ P_{-1} \ X_{-1} \ W^{**} \ G \ T \ YR \ 1 \right). \tag{3.70}$$

The coefficients of the structural form are

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & -1 \\ C & 0 & -\beta_{W^{*}1} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ -\beta_{C1} & -\beta_{I1} & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -\beta_{C2} & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.71}$$

and

$$\Gamma = \begin{bmatrix} 0 & -\gamma_{I1} & 0 & 0 & 0 & 0 & -1 & 0 \\ -\gamma_{C1} & -\gamma_{I2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\gamma_{W^*1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\gamma_{W^*2} & 0 & 0 & 0 & 0 & 0 \\ -\gamma_{C0} & -\gamma_{I0} & -\gamma_{W^*0} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} . \tag{3.72}$$

The first three single equations in the system (3.48)-(3.55) which has fixed coefficients can be written as

$$y_d = A_d \delta_d + \varepsilon_d \qquad d = C,\ I,\ W^*, \tag{3.73}$$

and the TSLS estimators are

$$\hat{\delta}_d = \left(A_d' P A_d\right)^{-1} A_d' P y_d \qquad d = C,\ I,\ W^*, \tag{3.74}$$

where $P=X(X'X)^{-1}X'$. These estimators are given in the table 4.1 column (1) . The standard errors of the estimators are calculated from the matrix $\hat{\sigma}_d^2\left(A_d' P A_d\right)^{-1}$, where

$$\hat{\sigma}_d^2 = \frac{1}{n} \left(y_d - A_d \hat{\delta}_d\right)' \left(y_d - A_d \hat{\delta}_d\right) = \frac{1}{n} \hat{\varepsilon}_d' \hat{\varepsilon}_d \tag{3.75}$$

Then, another kind of standard errors is calculated from the matrix $\left(A_d' P A_d\right)^{-1} A_d' P \hat{\Omega}_{Md}^* P A_d \left(A_d' P A_d\right)^{-1}$, here the $\hat{\Omega}_{Md}^*$ is the modified variance covariance matrix Ω. These standard errors corresponding to the TSLS are listed in column (2). If the model (3.58)-(4.65) with random coefficients is true, a comparison of standard errors in column (1) and (2) tells that the conventional estimators of the variance of TSLS may be underestimators.

Now, suppose that the coefficients of exogenous variables in the Klein model I are random while those of endogenous variables there are fixed. The first three single equations in the system (3.48)-(3.55) can be written as

$$y_d = A_d\delta_d + w_d \qquad d = C,\ I,\ W^*, \tag{3.76}$$

the GTSLSR estimators are

$$\tilde{\delta}^*_d = \left(A_d'\hat{R}^*_d A_d\right)^{-1}A_d'\hat{R}^*_d y_d \qquad d = C,\ I,\ W^*, \tag{3.77}$$

where $\hat{R}^*_d = \hat{\Omega}^{*-1}_d X(X'\hat{\Omega}^{*-1}_d X)^{-1}X'\hat{\Omega}^{*-1}_d$ is the one defined in (2.55) and

$$\tilde{w}^*_d = y_d - A_d\tilde{\delta}^*_d \qquad d = C,\ I,\ W^*. \tag{3.78}$$

The $\tilde{\delta}^*_d$ are given in the table 4.1 column (3). The standard errors of the estimators are also presented there. The column (4) in table 4.1 is the GTSLSR estimators $\tilde{\delta}^*_d$ and their standard errors, the only difference is here we use $\hat{\Omega}^*_{Md}$ The modified Ω.

TABLE 3.1

COEFFICEINT OR MEAN COEFFICIENT ESTIMATES

AND THEIR STANDARD ERRORS

EQUT	VARI	Coefficient or mean coefficient estimates (Standard Errors)			
		(1) TSLS	(2) TSLS	(3) GTSLSR	(4) GTSLSRM
C	P β_{C1}	0.0173 (0.1180)	0.0173 (0.1308)	0.0206 (0.1300)	0.0206 (0.1300)
	W β_{C2}	0.8102 (0.0402)	0.8102 (0.0447)	0.8125 (0.0447)	0.8125 (0.0447)
	P_{-1} γ_{C1}	0.2162 (0.1073)	0.2162 (0.1192)	0.2115 (0.1184)	0.2115 (0.1184)
	INT γ_{C0}	16.5548 (1.3208)	16.5548 (1.1563)	16.4798 (1.4569)	16.4798 (1.4569)
I	P β_{I1}	0.1502 (0.1732)	0.1502 (0.2707)	0.1251 (0.2233)	0.1403 (0.2764)
	K_{-1} γ_{I1}	-0.1578 (0.0361)	-0.1578 (0.0548)	-0.1571 (0.0421)	-1.5624 (0.0539)
	P_{-1} γ_{I2}	0.6159 (0.1628)	0.6159 (0.2555)	0.6422 (0.2104)	0.6248 (0.2690)
	INT γ_{I0}	20.2782 (7.5427)	20.2782 (11.3435)	20.1388 (8.7283)	19.8946 (11.2188)
W	F β_{W^*1}	0.4389 (0.0356)	0.4389 (0.0371)	0.4288 (0.0317)	0.4288 (0.0317)
	E_{-1} γ_{W^*1}	0.1467 (0.0388)	0.1467 (0.0380)	0.1528 (0.0334)	0.1528 (0.0334)
	YR γ_{W^*1}	0.1304 (0.0291)	0.1304 (0.0323)	0.1324 (0.0314)	0.1324 (0.0314)
	INT γ_{W^*0}	1.5003 (1.1478)	1.5003 (1.1527)	1.8343 (1.0413)	1.8343 (1.0413)

II. Bootstrapping the Klein Model I.

What we discussed in the section 3.5 will be executed to the Klein Model I. Since Ω, the variance covariance matrix of w, is not available in Klein model. $\hat{\Omega}^{*}_{d}$, the estimate of Ω, is used here. The algorithm 3.2 is given here to show the steps of the experiment. We omit the subscript d to simplify the formula in the algorithms. The same algorithm will execute to each of the three equations.

Algorithm 3.2.

Step 1. Obtain $\hat{\theta}^{*}$ defined in (2.47) to get $\hat{\Omega}^{*}_{d}$. Obtain also $\tilde{\delta}^{*}$ in (3.75) and $\tilde{w}^{*}$ in (3.76) respectively.

Step 2. Fit a nonparametric MLE of F,
$\hat{F}$: mass 1/?1 at $(A_i, X_i, \tilde{w}_i)$, i=1, 2, ..., n.

Step 3. Use 0 as the seed in the 'JNIFORM function which is a scalar function returning one or more pseudo random numbers with a uniform distribution over 0 to 1 in SAS IML procedure.

Step 4. Generate a random integer 1<j<21 by using the number obtained in step 3, draw a bootstrap sample $(A_j, X_j, \tilde{w}_j)$ with replacement from $(A, X, \tilde{w})$.

Step 5. Repeat step 3 21 times to con ruct $(A^{B}_{(b)}, X^{B}_{(b)}, \tilde{w}^{B}_{(b)})$.

Step 6. Reconstruct the linear responses as

$$y^{B}_{(b)} = A^{B}_{(b)}\tilde{\delta}^{*} + \tilde{w}^{*B}_{(b)} . \tag{3.77}$$

Step 7. Compute

$$\tilde{\delta}^{*}_{(b)} = \left(A^{B\prime}_{(b)}\hat{R}^{B-1}_{(b)}A^{B}_{(b)}\right)^{-1}A^{B\prime}_{(b)}\hat{R}^{B\,-1}_{(b)}y^{B}_{(b)} . \tag{3.78}$$

Step 8. Repeat steps 3 to 7 for b=1,2,...,B.

TABLE 3.2

MEANS AND STANDARD ERRORS OF BOOTSTRAP ESTIMATES

(200 REPLICATIONS)

EQUT	VARI	(1) GTSLSR SE	(2) BOOTSTRAP MEAN (3.50)	(3) BOOTSTTRAP SD (3.56)	(4) t (3.79)
C	P β_{C1}	0.0206 (0.1323)	0.06987	0.10834	-6.4314
	W β_{C2}	0.8125 (0.0456)	0.78351	0.17172	0.2388
	P_{-1} $\bar{\gamma}_{C1}$	0.2116 (0.1204)	0.22991	0.03333	-7.8114
	INT $\bar{\gamma}_{C0}$	16.4798 (1.4819)	13.17201	5.47957	-0.3646
I	P β_{I1}	0.1251 (0.2712)	0.34457	0.34457	-6.5186
	K_{-1} $\bar{\gamma}_{I1}$	-0.1571 (0.0511)	-0.12242	0.08773	-0.5590
	P_{-1} $\bar{\gamma}_{I2}$	0.6422 (0.2555)	0.45504	0.37271	-0.8184
	INT $\bar{\gamma}_{I0}$	20.1388 (10.5963)	12.55243	19.19145	5.5903
W	E β_{W^*1}	0.4288 (0.0221)	0.43744	0.16984	-0.7194
	E_{-1} $\bar{\gamma}_{W^*1}$	0.1528 (0.0233)	0.14352	0.12741	1.0301
	YF. $\bar{\gamma}_{W^*1}$	0.1324 (0.0219)	0.14167	0.09816	-1.3356
	INT $\bar{\gamma}_{W^*0}$	1.5003 (1.1478)	3.24648	17.50562	-1.1224

The computations described in the algorithm 3.2 were performed by a SAS IML program. The procedure was repeated 200 times. Column 3 and 4 in Table 4.2 show, for each parameter in the original model, the sample mean and sample standard deviation defined in (3.50) and (3.56) respectively. These SD's are the bootstrap estimates of variability in the parameter estimates. The t statistic in column 4 is defined as

$$t = (\text{ESTIMATE} - \text{MEAN})/(\text{SD}/\sqrt{B}) \ . \tag{3.79}$$

For example, the t statistic of β_{C2}, the coefficient of W in consumption equation, is

$$(0.8125 - 0.78351)/(0.17172/\sqrt{200}\ ) = 0.2388 \tag{3.80}$$

Compare to $t_{\infty, 0.05} = 1.645$, the small sample bias is statistically significant in four of the estimates; however, the practical significance may be small because the means are so close to the assumed values. On the whole the GTSLSR is performing well: bias in the coefficient estimates is small in practical terms. Much of the bias may be due to the impact of fitting, in making the residuals smaller than the disturbance terms. On the whole, the conventional formulas seem to be doing very well.

The shapes of the bootstrap distributions may be interesting. The coefficient estimate like β_{W^*1} is close to normally distributed, as may be anticipated. with a discrete error distribution, the bootstrap distribution of coefficient estimators will usually have moments of all orders. With a continuous distribution like the normal, moments may not exist. Still, there is good evidence to show that the bootstrap distribution can be well approximated by a normal distribution whose first and second moments can be estimated by the method indicated here.

The use of the bootstrap to attach standard errors to GTSLSR estimators in the FRSE model was demonstrates in this chapter. Instead to resample only the residuals, we resampled the vectors $(A_i, X_i, \tilde{w}_i^*)$ for $i=1,2,...n$, to show the relationship between the A, X and w. The asymptotic formulas for coefficient standard errors performed reasonably well.

REFERENCES

ANDERSON, T. W., [1971] *The statistical Analysis of Time Series* New York: Wiley.

BALDEV RAJ, V.K. SRIVASTAVA AND A. ULLAH [1980], "Generalized Two Stage Least Squares Estimators for a Structural Equation with Both Fixed and Random Coefficients." *International Economic Review*, 21 pp171-183.

BICKEL P.J. AND FREEDMAN, D.A. [1981], "Some Asymptotic Theory for the Bootstrap." *Annals of Siatistics*, 9 pp1197-1217.

DWIDEDI T.D. AND SRIVASTAVA. V.K. [1984], "Exact Sample Properties of Double K-Class Estimator." *Journal of Econometrics;* 25 pp263-284.

DWIVEDI T.D. [1987], "A Monograph on Simultaneous Equation Model" *Concordia University, Department of Mathematics and Statistics, (Mimeograph).*

EFRON, B. [1979], "Bootstrap Methods: Another Look at the Jackknife." *Annals of Statistics*, 7, pp1-26.

—— [1982], *The Jackknife, the bootstrap, the Other Resampling Plans*, CBMS-NSF Regional Conference Series in applied mathematics, Monograph 38, Philadelphia: Society for Industrial and applied Mathematics.

FREEDMAN, D.A. [1981], "Bootstrapping Regression Models." *Annals of Statistics*, 9, pp1218-1228.228.

—— [1984] , "On Bootstrapping Two Stage Least Squares Estimates in Stationary Linear Models." *the Annals of Statistics*, 12 pp827-842.

——, AND PETERS, S.C. [1984a] , "Bootstrapping a Regression Equation: Some Empirical Results." *Journal of the Statistical Association*, 79, pp97-106.

——, AND —— [1984b], "Bootstrapping an Econometric Model: some Empirical results. *Journal of Business and Economic Statistics*. 2 pp150-158.

FROEHLICH, B.R. [1973], "Small Sample Properties of various Consistent Variance Estimators Proposed for a Random Coefficient Regression Model." *Journal of the American statistical Association*, 68, pp626-632.

GOLDERBERGER A. S. [1964], *Econometric Theory*. New York: John Wiley.

HILDRETH, C. AND J. HOUCK [1968], "Some Estimates for a Linear Model with Random Coefficients." *Journal of the American Statistical association*, 63 (June), pp584-595.

KELEJIAN, J. B. [1971], "Comparison of k-Class Estimators when the Disturbances are Small," *Econometrica*, 39 (July), pp723-737.

KELEJIAN, H. H. [1974], "Random Parameters in a Simultaneous equation framework: Identification and Estimation." *Econometrica*, 42 (May), pp517-527.

KLEIN, L. [1950], *Economic Fluctuations in the United States*. pp1921-1941 (New York: Wiley).

KHENTA J. [1971], *Elements of Econometrics*. New York: Macmillan.

NAGAR, A. L. [1959], "The Bias and Moment Matrix of the General K-Class Estimators of the Parameters in Simultaneous Equations." *Econometrica*, 27 (October), pp575-595.

RAJ, BALDEV [1975], "Linear Regression with Random Coefficients: The Finite Sample and Convergence Properties." *Journal of the American Statistical Association*, 70 (March), pp127-137.

RAO, C. R. [1973], *Linear Statistical Inference and Its Applications*, 2nd edition (New York: Wiley).

ROTHENGER, T. AND C. LEENDERS [1964], "Efficient Estimation of simultaneous Equation Systems." *Econometrica*, 32 (January-April), 57-76.

SAWA, T. [1969], "The Exact Sampling Distribution of Ordinary Least Squares and Two Stage Least Squares Estimators." *Journal of the American Statistical Association*, 64 'September) pp923-937.

SINGH, K. [1981], "On Asymptotic Accuracy of Efron's bootstrap." *The Annals of Statistics*. 9 pp1187-1195.

SWAMY, P. A. V. B. [1970], "Efficient Inference in a Random Coefficient Regression Model." *Econometrica*, 38 (March), pp311-323.

——— AND J. HOLMES, [1971] "The Use of Undersized Samples in the Estimation of Simultaneous Equation systems." *Econometrica*, 39 (May), pp455-459.

——— AND J. S. MEHTA [1975], "Bayesian and Non-bayesian Analysis of Switching Regressions and of Random coefficient Regression Models." *Journal of the American Statistical Association*, 70 (September), pp593-602.

——— AND ——— [1977], "Estimation of Linear Models with time and Cross-Varying Co cients." *Journal of the American Statistical Association*, 72 (December), pp890-898.

THEIL, H. [1961], *Economic Forecasts and Policy* (Amsterdam: North-Holland).

——— [1971], *Principles of Econometrics*, New York: John Wiley.

WANG S. AND DWIVEDI T.D. [1989] "Bootstrapping an econometric model with Random Coefficients." *Presented during Statistical Society of Canada meeting, Ottawa.*

WANG S. AND DWIVEDI. T.D. [1990] "Bootstrapping a Single Equation of a Simultaneous Equation Model." *Presented during Interface'90, East Lansing, Michigan.*

WONNACOTT R.J. AND WONNACOTT T.H. [1970], *Econometrics* New York: John Wiley,