ABSTRACT

IMPROVING THE COMPUTATION OF SIMULTANEOUS STOCHASTIC LINEAR EQUATIONS ESTIMATES

By

William Lewis Ruble

As a part of a much larger statistical package for a Control Data Corporation 6500 computer, the writer participated in the development of a set of simultaneous stochastic linear equations routines including direct least squares (DLS), two-stage least squares (2SLS), limited information single equation maximum likelihood (LIML), the Zellner-Aitken estimator (ZA), three-stage least squares (3SLS), linearized maximum likelihood (LML), subsystem maximum likelihood (SML), and full information maximum likelihood (FIML). This paper summarizes (1) computational approaches used, (2) many relationships between methods, (3) [to a very limited extent] some of the properties of these methods, and (4) forms of user control cards which may be used to specify and control the computation of problems. Computational techniques such as standardization of variables to reduce rounding error are noted.

Some of the computational approaches used and some of the relationship among estimators noted in this paper were derived by the writer. Of the computational approaches which are presented for the first time (as far as the writer is aware), the following are probably the most noteworthy:

(1) The use of direct orthogonalization in the calculations for many of the simultaneous stochastic linear equations methods. In addition to reducing rounding error, the use of direct orthogonalization eliminates some of the problems of multicollinearity among predetermined variables in the equation system. Also, the matrix of predetermined variables need not have full column rank.

(2) The development of a method for imposing arbitrary linear restrictions on coefficients which:

   (a) Allows the restrictions to be specified directly to the computer without prior solving out or conversion.

   (b) Provides a means of imposing arbitrary linear restrictions upon FIML and SML coefficients.

   (c) May be applied in essentially the same way to DLS, ZA, SML, FIML, LML, and 3SLS.

   (d) Is adapted to methods requiring iteration to a solution.

   (e) Allows redundant restrictions to be imposed on coefficients. The number of independent restrictions is calculated as a by-product of the computational procedure.

   (f) Detects inconsistent restrictions.

   (g) May be used to calculate restricted coefficients even though a unique solution for a method does not exist in the absence of the restrictions.

Relationships among methods which are shown for the first time in this paper include:

(1) For the special case of a system of equations in which only one jointly dependent variable occurs in each equation, the following computational procedures lead to the same coefficients:

   (a) FIML.

   (b) Iteratively applying ZA.

   (c) The Telser method of iteratively estimating each equation by DLS.

(2) For the general case in which more than one jointly dependent variable is permitted per equation and at least one equation is overidentified, the following computational procedures do not lead to the same coefficients:

   (a) FIML.

   (b) Iteratively applying 3SLS (I3SLS).

(3) Iteratively applying LIML (ILIML) leads to FIML estimates in the general case (multiple jointly dependent variables occurring in one or more stochastic equations).[1] The Telser method of iteratively estimating each equation by DLS may be considered a special case of ILIML; hence, IDLS is a maximum likelihood method. (A direct derivation of IDLS as a maximum likelihood method for the special case of one jointly dependent variable per equation is also given.)

In the derivation of the likelihood function for a system of equations for the application of FIML and for a subsystem of equations for the application of SML, identity equations are explicitly recognized. It is shown that the identity equations need not be used to eliminate jointly dependent variables from the stochastic equations

in order to express the likelihood function or to apply the FIML and SML estimation procedures.[2]

---

[1]The ILIML procedure was proposed to the writer by Professor Herman Rubin.

[2]T.J. Rothenberg and C.T. Leenders, "Efficient Estimation of Simultaneous Equation Systems", Econometrica, XXXII, No. 1-2 (January-April, 1964), 57-76 have already shown that it is unnecessary to use identity equations to eliminate jointly dependent variables from the stochastic equations; however, a slightly different approach to showing this is taken in this paper. Professor Herman Rubin informed the writer that it is unnecessary to use identity equations to eliminate jointly dependent variables for SML; however, the writer is not aware of any reference to this in the literature.

IMPROVING THE COMPUTATION OF SIMULTANEOUS STOCHASTIC LINEAR EQUATIONS ESTIMATES

By

William Lewis Ruble

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics

1968

ACKNOWLEDGEMENTS

The writer appreciates the help he has recieved from a large number of people during the early phases of projects leading to this manuscript as well as in the final preparation of this manuscript. Dr. William A.Cromarty served as the writer's academic advisor during his first year of graduate work at MSU and introduced the writer to the general area of simultaneous stochastic linear equations estimation. Professor Clifford Hildreth then served as the writer's academic advisor for three years during which time the writer formulated many of the computational procedures noted in this manuscript. Also, during this period, Professor Herman Rubin gave the writer a number of special lectures on the blackboard of his office in an effort to impart more understanding of the techniques involved and to suggest computational procedures not a part of the literature.

Professor Robert L. Gustafson served as chairman of the writer's thesis committee and spent a very large amount of time reading through two drafts of this manuscript and suggesting many substantive and editorial improvements which have been incorporated. The other members of the writer's thesis committee--Professors Lester B. Manderscheid, Jan Kmenta, Kenneth J. Arnold, and James H. Stapleton--also spent a large amount of time reading drafts of this manuscript and suggesting improvements which have been incorporated. Mrs. Janet Eyster read the manuscript in semi-final form and caught quite a few editorial and typographical errors.

Mrs. Marylyn Donaldson assisted the writer by calculating problems to test out a number of conjectures regarding special techniques to reduce rounding error. Mrs. Noralee Barnes did an exemplary job of typing this manuscript and Mrs. Barbara Bray, Mr. Tim Walters, and Mrs. Dona Smith assisted the writer to get the computer results in their final form for photographing and reproducing.

That it has been possible to accomplish this project at all is largely due to the support and encouragement the writer has received from Professor Lawrence L. Boger, Chairman of the Department of Agricultural Economics. The writer also appreciates the cooperation and assistance he has received during his employment in the Agricultural Experiment Station and the cooperation he has received from members of the MSU Computer Laboratory.

Finally, the writer greatly appreciates the patience, understanding, and encouragement of his wife, Margie, during the long duration of his graduate student career. Although a large number of people have made contributions, the writer assumes full responsibility for errors which may remain in the manuscript.

TABLE OF CONTENTS

PART I. SINGLE EQUATION METHODS

Chapter Page

LIST OF ABBREVIATIONS USED FOR ESTIMATORS

PART I

SINGLE EQUATION METHODS

CHAPTER I

INTRODUCTION

A. Background and Purpose

In 1963 the writer started development of a system of computer routines (presently called the AES STAT system) designed to calculate simultaneous stochastic linear equations estimates, direct least squares (including stepwise variations), analysis of variance and covariance, some "basic" statistics such as simple correlations, and the plotting of data and functions.[1] Emphasis from the start has been on developing only a few major routines which can compute by a number of methods on each routine. Additional flexibility has been obtained by incorporating considerable facility for the manipulation and transformation of data in the computer and for the manipulation of coefficients between methods of estimation (the estimated coefficients from one method often provide the starting coefficients for other methods). The parameters (instructions to the routines prepared by the user for the calculation of a particular problem) are of the same form for all of the AES STAT routines.[2]

---

[1] A computer routine is a set of instructions to a computer to accomplish a given calculation. The terms computer routine and computer program are used interchangeably.

[2] The general form of the control cards for the AES STAT system and some of the details on the development of the AES STAT system are given in chapter IX.

In the process of programming the simultaneous stochastic linear equations methods, the writer found it necessary to refer to many different articles and books for computational formulas and other specific aspects of the methods. Many of the computational approaches were desirable appraoches for computation on a hand calculator but were not well adapted for use on the computer. In adapting computational approaches for use on the computer, emphasis has been placed on reducing rounding error, increasing flexibility, and providing automatic decision branching in the computer in the solution of problems requiring iteration.

The purposes of this paper are to: (1) summarize computational formulas for a number of simultaneous stochastic linear equations methods, (2) summarize some of the relationships among these methods, and (3) to a limited extent summarize some of the properties of these methods.[1]

Some of the computational approaches used and some of the relationships among methods noted in this paper were derived by the writer. Of the computational approaches which

---

[1] Methods presented in this paper have already been incorporated into the AES STAT system with the following exceptions:

(1) The method of imposing arbitrary linear restrictions has not yet been implemented as it was derived by the writer in the process of writing this paper.
(2) The limited information subsystem maximum likelihood (SML) method has not been programmed, yet; however, it is planned for incorporation into the system shortly.
(3) Nagar's minimum second moment k (MSM) is not available in the system.
(4) The orthogonalization method described in appendix B is not available in the system.

are presented in this paper for the first time (at least as far as the writer is aware), the following are probably the most noteworthy:

(1) The use of direct orthogonalization in the calculations for the simultaneous stochastic linear equations methods. In addition to reducing rounding error, the use of direct orthogonalization eliminates some of the problems of multicollinearity among predetermined variables in the equation system. Also, the matrix of predetermined variables in the system need not have full column rank.

(2) The development of a method for imposing arbitrary linear restrictions on coefficients which:

   (a) Allows the restrictions to be specified directly to the computer without prior solving out or conversion.

   (b) Provides a means of imposing arbitrary linear restrictions directly upon full information maximum likelihood (FIML) coefficients.[1]

   (c) May be applied in essentially the same way to direct least squares (DLS), the Zellner-Aitken estimator (ZA), limited information subsystem maximum likelihood (SML), full information maximum likelihood (FIML), linearized maximum likelihood (LML), and three-stage least squares (3SLS).

---

[1] A summary of abbreviations used in this paper for estimators (e.g., FIML) follows the table of contents.

(d) Is adapted to methods requiring iteration to a solution.

(e) Allows redundant restrictions to be imposed on coefficients. The number of independent restrictions is calculated as a by-product of the computational procedure.

(f) Detects inconsistent restrictions.

(g) May be used to calculate restricted coefficients even though a unique solution for a method does not exist in the absence of the restrictions. (E.g., direct least squares may be applied directly to problems in which the matrix of explanatory variables has less than full column rank provided sufficient restrictions are placed on the coefficients. Thus, the step of eliminating linearly dependent explanatory variables from the equation before obtaining a solution is saved.)[1]

Relationships among methods which are shown for the first time in this paper include:

(1) For the special case of a system of equations in which only one jointly dependent variable occurs in each equation, the following computational procedures lead to the same coefficients:

---

[1]The use of the method of imposing restrictions suggested in this paper together with the method for computing 3SLS and ZA suggested in this paper (starting the computations with an identity matrix as the estimated disturbance variance-covariance matrix) provide an easily used procedure for imposing linear restrictions across equations on the coefficients used to obtain disturbance variance-covariance estimates in the calculation of 3SLS or ZA estimates. As a result, if sufficient restrictions on coefficients across equations occur, unique 3SLS or ZA estimates may exist even though unique two-stage least squares (2SLS) and DLS estimates do not exist.

(a) FIML.

(b) Iteratively applying ZA.[1]

(c) The Telser method of iteratively estimating each equation by DLS.[2]

(2) For the general case in which more than one jointly dependent variable is permitted in each equation, the following computational procedures lead to the same coefficients:

(a) FIML .

(b) An estimation procedure in which limited information single equation maximum likelihood (LIML) estimation is used iteratively to estimate the coefficients of each equation.[3]

(3) For the general case in which more than one jointly dependent variable is permitted per equation and at least one equation in the system is over-identified, the following computational procedures do not lead to the same coefficients:

(a) FIML .

(b) Iterative 3SLS (I3SLS).

---

[1]The Zellner-Aitken (ZA) estimator is given in Zellner [1962] and in Chapter VII of this paper.

[2]The Telser estimator is given in Telser [1964] and in section VII.E of this paper.

[3]The iterative LIML computational procedure (ILIML) was proposed to the writer by Professor Herman Rubin in 1963 and the relationship of the method to FIML was shown to the writer by Professor Rubin at that time. Since DLS may be regarded as a particular case of LIML, the IDLS (Telser) method may be regarded as a particular case of the ILIML method.

In the derivation of the likelihood function for a system of equations for the application of FIML (chapter V) or for a subsystem of equations for the application of SML (chapter VI), identity equations are explicitly carried. It is shown in this paper that the identity equations need not be solved out to express the likelihood or to apply the FIML and SML estimation procedures.[1]

Detailed user descriptions have been developed for the AES STAT system except for the simultaneous stochastic linear equations portion. It is intended that this paper will serve as a basic reference to the computational procedures used in the simultaneous stochastic linear equations portion and that detailed user descriptions will soon be written for this portion as well.[2]

---

[1] Rothenberg and Leenders [1964] have already shown that it is unnecessary to eliminate jointly dependent variables to solve out identity equations for FIML; however, a slightly different approach to showing this is taken in this paper. Professor Herman Rubin informed the writer that it is unnecessary to solve out identity equations for SML; however, the writer is not aware of any reference to this in the literature.

[2] The form of the parameters to the system (control cards to the system to compute a given problem) and the form of the output is discussed and illustrated in part III of this paper.

The AES STAT system has been programmed on a Control Data Corporation (CDC) 3600 computer. Although 3600 FORTRAN is the primary language used, extensive use is made of COMPASS assembly language subroutines and features of the DRUM SCOPE executive system; hence, in its present form the AES STAT system is very difficult to convert to another computer system. Installation of a CDC 6500 computer at MSU is planned for late 1968 and the AES STAT system will then be converted to the CDC 6500 computer. In the process of conversion to the CDC 6500 computer, a number of assembly language subroutines will be replaced by FORTRAN subroutines thereby making conversion to another large scale computer system more feasible. In any event, the computational approaches suggested here, including especially the orthogonalization procedures, involve basic computational procedures which can be programmed for any computer as easily as less accurate but better known procedures.

Since this paper concentrates on computational procedures, properties of the estimators receive only cursory treatment; also many of the proofs are given in algorithmic form; that is, the computational method which is described provides the proof of the property claimed.

B. Some Asymptotic Properties of Estimators

In this paper, most of the properties noted for particular estimators are asymptotic properties rather than small sample properties.[1] The estimators are, however, used in estimation in which the number of observations in a given sample is finite and usually fairly small. Although it is hoped that the asymptotic properties mentioned give a guide to the comparable small sample properties, it must be realized that in particular cases a given ranking of estimators based on an asymptotic property may be reversed in samples of the size used in the usual application of the estimator. Also, asymptotic properties of an estimator may give a good guide to properties of that estimator for a sample of size of, say, 100 or larger, but a very poor guide for a sample size of, say, 10.

Since many readers are not familiar with many of the terms used in this paper, we will note some distinctions.[2] In what follows, let θ denote a parameter of a probability distribution

---

[1] In this paper, T denotes the number of observations in a sample. We say that an estimator has a given asymptotic property if there exists a sample size T_o such that for all $T > T_o$, the property holds within a given measure of closeness.

[2] See Goldberger [1964] for a more extensive treatment of the unbiased, asymptotically unbiased, and consistent properties.

and $\hat{\theta}$ an estimator of θ.[1] Let T denote sample size. θ is independent of T whereas $\hat{\theta}$ may not be independent of T. The properties of $\hat{\theta}$ for a given sample size of T are those which $\hat{\theta}$ has in repeated samples of sample size T -- not the estimated value of θ for a given sample (e.g., $\hat{\theta}$ may be the direct least squares estimator applied to a coefficient of a given equation--not the particular estimate obtained by applying direct least squares to a given sample). Some distinctions follow:

(1) $\hat{\theta}$ is an unbiased estimator of a parameter θ if $\mathcal{E}\hat{\theta} = \theta$ where $\mathcal{E}$ denotes expected value.[2]

(2) $\hat{\theta}$ is an asymptotically unbiased estimator of θ if $\lim_{T\to\infty} \mathcal{E}\hat{\theta} = \theta$.

Unbiased implies asymptotically unbiased, since if $\mathcal{E}\hat{\theta} = \theta$ for all T, $\lim_{T\to\infty} \mathcal{E}\hat{\theta} = \lim_{T\to\infty} \theta = \theta$.

Asymptotically unbiased does not imply unbiased. For example, let $\theta = 1$ and $\mathcal{E}\hat{\theta} = 1 + \frac{1}{T}$. Then $\lim_{T\to\infty} \mathcal{E}\hat{\theta} = \lim_{T\to\infty} (1 + \frac{1}{T}) = 1$ so that θ is asymptotically unbiased. $\hat{\theta}$ is not unbiased, however, since $\mathcal{E}\hat{\theta} = 1 + \frac{1}{T} \neq 1$.

---

[1] The properties given here derive from the properties of $\hat{\theta}$ as a random variable. It is more common in the statistical literature to use X_n in place of $\hat{\theta}$ so that the results are not limited to estimators and the location in the sequence is specifically noted. Here n would be T -- the sample size. A more complete notation would be the use of $\hat{\theta}_T$ in place of $\hat{\theta}$ to explicitly recognize sample size.

[2] More particularily, $\hat{\theta}$ is an unbiased estimator of a parameter $\theta \in \Theta$ if $\mathcal{E}\hat{\theta} = \theta$ for all $\theta \in \Theta$. For simplicity, in the remainder of this section, we will drop the mention of the class Θ to which θ belongs and the mention of the requirement that the defined property holds for all members of that class.

As T increases, $\hat{\theta}$ converges in probability to θ [i.e., converges stochastically--written $\operatorname{plim}_{T\to\infty} \hat{\theta} = \theta$ or equivalently as $\operatorname{plim}(\hat{\theta} - \theta) = 0$] if for every $\epsilon > 0$, $\lim_{T\to\infty} \text{Prob}[\text{abs}(\hat{\theta} - \theta) < \epsilon] = 1$, where Prob denotes the probability of the expression within the brackets.[1] Another way to say the same thing is that $\operatorname{plim}_{T\to\infty} \hat{\theta} = \theta$ if for any $\epsilon > 0$ and $\eta > 0$, however small, there is some $T'_{\epsilon,\eta}$, such that for all $T > T'_{\epsilon,\eta}$, $\text{Prob}[\text{abs}(\hat{\theta} - \theta) < \epsilon]$ exceeds $1 - \eta$.[2] We can also say that $\operatorname{plim}_{T\to\infty} \hat{\theta} = \theta$ if the probability distribution for $\hat{\theta}$ collapses about the single point θ as $T\to\infty$, i.e., if the mean square deviation of $\hat{\theta}$ from θ goes to zero as $T\to\infty$.

(3) An estimator whose probability limit is a finite parameter ($\operatorname{plim}_{T\to\infty} \hat{\theta} = \theta$) is said to be a consistent estimator of that parameter. Asymptotic unbiasedness does not imply consistency, since the probability distribution may not converge to a single point in the limit. As an example, suppose that for any T, $\hat{\theta}$ has the distribution:[4]

$$\text{Prob}(\hat{\theta} = 1) = 1/2$$

$$\text{Prob}(\hat{\theta} = 2) = 1/2$$

---

[1] abs denotes absolute value.

[2] Kendall and Stuart [1961], p. 3.

[3] The mean square deviation of $\hat{\theta}$ from θ is $\mathcal{E}(\hat{\theta} - \theta)^2 = \text{Var}(\hat{\theta}) + (\mathcal{E}\hat{\theta} - \theta)^2$.

[4] It is, of course, very unusual to define a distribution of $\hat{\theta}$ not containing θ as a parameter; however, using such distributions as examples permits construction of exceedingly simple examples.

Then $\mathcal{E}\hat{\theta} = (1/2)(2) + (1/2)(1) = 3/2$ and $\lim_{T\to\infty} \mathcal{E}\hat{\theta} = 3/2$ so that if $\theta = 3/2$ then $\hat{\theta}$ is both unbiased and asymptotically unbiased. $\hat{\theta}$ is not consistent since $\operatorname{plim}_{T\to\infty} \hat{\theta} \neq 3/2$. (The distribution does not concentrate on the point 3/2 as $T\to\infty$; in fact, in this example there is zero probability that $\hat{\theta} = 3/2$ even in an infinitely large sample.)

Consistency does not imply asymptotic unbiasedness. For example, let $\hat{\theta}$ have the distribution:[1]

$$\text{Prob}(\hat{\theta} = 0) = (T - 1)/T$$

$$\text{Prob}(\hat{\theta} = T^2) = 1/T$$

Then, since $\hat{\theta}$ concentrates at 0 as T becomes large, if $\theta = 0$, then $\hat{\theta}$ is a consistent estimator. On the other hand, $\hat{\theta}$ cannot be an asymptotically unbiased estimator of any θ since $\mathcal{E}\hat{\theta} = 0\cdot[(T - 1)/T] + T^2[1/T] = T$.

In discussing asymptotic properties it is often useful to use the "big O" and "little o" notation to give a magnitude or speed of convergence. In using this notation, it is important to distinguish between whether the magnitude is related to $\hat{\theta} - \theta$ (and therefore is an order of magnitude of consistency) or whether the magnitude is related to $\mathcal{E}\hat{\theta}$ (and therefore is an order of magnitude of asymptotic unbiasedness).

(4) Let $f(T)$ denote a positive valued function of T such as

---

[1]This example was suggested by Professor Kenneth J. Arnold.

$1/T$ or $1/T^2$. Then $\hat{\theta} - \theta$ is $O_p(f(T))$ [which is read $\hat{\theta} - \theta$ is "big O" of $f(T)$ in probability), or, (equivalently) $\hat{\theta} - \theta$ is of probability order $f(T)$ or (equivalently) $\hat{\theta} - \theta$ is of the <u>same</u> order of magnitude as $f(T)$ in probability as $T \to \infty$ if any of the following equivalent conditions hold:[1]

(a) If there exists a positive c independent of T such that

$$\operatorname{plim}_{T\to\infty}\left[\frac{1}{f(T)} \operatorname{abs}(\hat{\theta} - \theta)\right] \leq c.$$

(c can be a <u>very</u> large number and still be independent of T.)

(b) For any $\delta > 0$, however small, there exists a positive constant c_δ independent of T such that

$$\lim_{T\to\infty} \operatorname{Prob}\left[\frac{1}{f(T)} \operatorname{abs}(\hat{\theta} - \theta) \leq c_\delta\right] \geq 1 - \delta$$

or (equivalently)

$$\lim_{T\to\infty} \operatorname{Prob}[\operatorname{abs}(\hat{\theta} - \theta) \leq c_\delta \cdot f(T)] \geq 1 - \delta.$$

(c) There exists a positive constant c independent of T such that for any positive η, however small, there exists T'_η such that for every $T > T'_\eta$,

$$\operatorname{Prob}\left[\frac{1}{f(T)} \operatorname{abs}(\hat{\theta} - \theta) < c\right] > 1 - \eta.$$

---

[1] Some of the "big O" and "little o" conditions which follow are given in Mann and Wald [1943b].

Most commonly $f(T)$ will be $T^{-1/2}$, T^{-1}, $T^{-3/2}$, or T^{-2} (i.e., $1/\sqrt{T}$, $1/T$, $1/T^{3/2}$, or $1/T^2$).

(5) $\hat{\theta} - \theta$ is $o_p(f(T))$ (which is read $\hat{\theta} - \theta$ is "little o" of $f(T)$ in probability), or (equivalently) $\hat{\theta} - \theta$ is of probability order <u>smaller</u> than $f(T)$, or (equivalently) $\hat{\theta} - \theta$ is of a <u>smaller</u> order of magnitude than $f(T)$ in probability as $T \to \infty$ if any of the following equivalent conditions hold:

(a) $\operatorname{plim}_{T\to\infty}\left[\frac{1}{f(T)} \operatorname{abs}(\hat{\theta} - \theta)\right] = 0$.

(b) For any positive ϵ and η, however small, there exists $T'_{\epsilon,\eta}$ such that for every $T > T'_{\epsilon,\eta}$,

$$\operatorname{Prob}\left[\frac{1}{f(T)} \operatorname{abs}(\hat{\theta} - \theta) < \epsilon\right] > 1 - \eta.$$

$\hat{\theta} - \theta$ is $o_p(f(T))$ implies that $\hat{\theta} - \theta$ is $O_p(f(T))$, but the reverse implication does not necessarily hold.

(6) To define order of magnitude of $\delta\hat{\theta}$ rather than $\hat{\theta}$, merely replace $\hat{\theta}$ by $\delta\hat{\theta}$ in (4) and (5) above. For example: $\delta\hat{\theta} - \theta$ is $O_p(f(T))$ [i.e., $\delta\hat{\theta} - \theta$ is of the <u>same</u> order of magnitude as $f(T)$ in probability as $T \to \infty$] if there exists a positive c independent of T such that

$$\operatorname{plim}_{T\to\infty}\left[\frac{1}{f(T)} \operatorname{abs}(\delta\hat{\theta} - \theta)\right] \leq c.$$

The order of magnitude of $f(T)$ gives the order of magnitude of stated convergence. Thus, if $\hat{\theta} - \theta$ is $O_p(T^{-2})$ then $\hat{\theta} - \theta$ is $O_p(T^{-3/2})$, $O_p(T^{-1})$, $O_p(T^{-1/2})$, etc., and $\hat{\theta}$ is consistent. ($\hat{\theta}$ is consistent if $\hat{\theta} - \theta$ is $O_p(T^{-n})$ with $n > 0$.)

On the other hand, $\hat{\theta}$ may be consistent but $\hat{\theta} - \theta$ not $O_p(T^{-1/2})$ or $\hat{\theta} - \theta$ may be $O_p(T^{-1/2})$ but not $O_p(T^{-2})$. Also, the order of magnitude of $\hat{\theta} - \theta$ does not imply anything regarding the order of asymptotic unbiasedness except for specified distributions.

Similarly if $\mathcal{E}\hat{\theta} - \theta$ is $O_p(T^{-2})$ then $\mathcal{E}\hat{\theta} - \theta$ is $O_p(T^{-3/2})$, $O_p(T^{-1})$, $O_p(T^{-1/2})$, etc., and $\hat{\theta}$ is asymptotically unbiased. ($\hat{\theta}$ is asymptotically unbiased if $\mathcal{E}\hat{\theta} - \theta$ is $O_p(T^{-n})$ with $n > 0$.) On the other hand, $\hat{\theta}$ may be asymptotically unbiased but $\mathcal{E}\hat{\theta} - \theta$ not $O_p(T^{-1/2})$ or $\mathcal{E}\hat{\theta} - \theta$ may be $O_p(T^{-1/2})$ but not $O_p(T^{-2})$. Also, the order of magnitude of $\mathcal{E}\hat{\theta} - \theta$ does not imply anything regarding the order of magnitude of consistency except for specified distributions.

The above properties are some of the properties referred to in this paper and do not, of course, include many properties of estimators which are important in estimation. The emphasis on asymptotic properties is dictated by the present state of our knowledge of the small sample properties of simultaneous stochastic equations estimators.

In the remainder of this paper, $\operatorname{plim}_{T\to\infty} A$ will be shortened to $\operatorname{plim} A$, i.e., the $T\to\infty$ will be understood.

C. Basic Model

1. System of structural equations

In part I, the estimation of a single equation from a complete system of equations will be considered. The complete system which consists of M stochastic equations and G - M identity equations (G - M may be zero) may be expressed as:[1,2,3]

$$(I.1)\qquad \underset{T\times G}{Y}\ \underset{G\times G}{\Gamma'} + \underset{T\times\Lambda}{X}\ \underset{\Lambda\times G}{B'} + \underset{T\times M}{[U}\ \vdots\ \underset{T\times(G-M)}{0]} = \underset{T\times G}{0}$$

or

$$(I.2)\qquad \underset{T\times(G+\Lambda)}{Z}\ \underset{(G+\Lambda)\times G}{\alpha'} + \underset{T\times G}{[U}\ \vdots\ 0] = \underset{T\times G}{0}\ .$$

The same model is also often written in transposed form as:

$$(I.3)\qquad \underset{G\times G}{\Gamma}\ \underset{G\times 1}{Y'_{[t]}} + \underset{G\times\Lambda}{B}\ \underset{\Lambda\times 1}{X'_{[t]}} + \begin{bmatrix} \underset{M\times 1}{U'_{[t]}} \\ \underset{(G-M)\times 1}{0'_{[t]}} \end{bmatrix} = \underset{G\times 1}{0'_{[t]}} \quad (t=1,\ldots,T)$$

---

[1] The dimensions of each matrix are listed below the matrix in many of the matrix equations given in this paper. In this paper, ' (prime) denotes transpose.

[2] An identity equation is an equation containing known coefficients and no disturbance (i.e., a disturbance vector of all zeros).

[3] The notation used in this paper was designed to meet the following requirements:

1) It should be consistent with the notation commonly used for direct least squares; in particular, the signs of the coefficients should not have to be reversed to make them comparable to direct least squares coefficients.

or

$$(I.4)\qquad \underset{G\times(G+\Lambda)}{\alpha}\ \underset{(G+\Lambda)\times 1}{Z'_{[t]}} + \begin{bmatrix} \underset{M\times 1}{U'_{[t]}} \\ \underset{(G-M)\times 1}{0'_{[t]}} \end{bmatrix} = \underset{G\times 1}{0'_{[t]}} \quad (t=1,\ldots,T)$$

where $[t]$ denotes the t^{th} observation and:

Y is a matrix of T sample observations taken on the G jointly dependent variables in the system.

X is a matrix of T sample observations taken on the Λ predetermined variables in the system.

$Z = [Y \vdots X]$ is a matrix of T sample observations taken on all $G + \Lambda$ variables in the system.

U is a $T\times M$ matrix containing T unobserved structural disturbances for each of the M equations containing disturbances.

Γ is the $G\times G$ matrix of population coefficients of the G jointly dependent variables. Each row of Γ (or column of Γ') contains the population coefficients corresponding to a

---

2) The coefficients for a structural equation are expressed as a row of the coefficient matrix for the system. Similarly, the coefficients for a reduced form equation are expressed as a row of the coefficient matrix for the reduced form.

3) An observation is a row in an observation matrix.

4) Identity equations are explicitly recognized in the notation.

5) Within the above limitations, the notation should be patterned after that of Theil [1961] and Zellner and Theil [1962], since this appears to be the most prevalently used simultaneous stochastic equations notation now appearing in the literature.

particular equation, and each column of Γ (or row of Γ') contains the population coefficients corresponding to a particular jointly dependent variable.

B is the $G \times \Lambda$ matrix of population coefficients of the Λ predetermined variables. Each row of B (or column of B') contains the population coefficients corresponding to a particular equation, and each column of B (or row of B') contains the population coefficients corresponding to a particular predetermined variable.

$\alpha = [\Gamma \vdots B]$ is the $G \times (G + \Lambda)$ matrix of population coefficients of all G equations for all of the $G + \Lambda$ variables in the system. The term "jointly dependent" will be treated as synonymous with the term "endogenous". Jointly dependent variables are random variables assumed to be contemporaneously correlated with the disturbances. They are assumed to be generated within (endogenous to) the system of equations.

Predetermined variables are the remaining variables in the system. They are either (1) exogenous variables, i.e., variables which are assumed to be generated outside the system of equations and therefore independent of the disturbances or (2) lagged values of jointly dependent variables which due to their lag are contemporaneously independent of the disturbances.[1] In some parts of this paper the predetermined variables will be assumed to consist

---

[1] Contemporaneously independent is defined in statistical assumption (3) of section I.C.3.

of "fixed" or non-stochastic variables only. In these cases, the set of predetermined variables must be restricted to exogenous variables only, since lagged values of jointly dependent variables are stochastic and not "fixed".

Klein's model I of the United States economy will be used to illustrate the above notation and some subsequent notation which will be introduced. Klein's model I is a complete system of equations (the number of jointly dependent variables equals the number of equations)[1] which may be written as an 8 equation system, the first 3 equations containing disturbances and the remaining 5 equations consisting of identity equations. These equations are:[2]

(I.5a) Consumption: $C = \alpha_0^{[1]} + \alpha_1^{[1]}P + \alpha_2^{[1]}W + \alpha_3^{[1]}P_{-1} + u_1$

(I.5b) Investment: $I = \alpha_0^{[2]} + \alpha_1^{[2]}P + \alpha_2^{[2]}P_{-1} + \alpha_3^{[2]}K_{-1} + u_2$

(I.5c) Private wage: $W_1 = \alpha_0^{[3]} + \alpha_1^{[3]}E + \alpha_2^{[3]}E_{-1} + \alpha_3^{[3]}t + u_3$

(I.5d) Product: $Y + R = C + I + G$

(I.5e) Income: $Y = P + W$

(I.5f) Capital: $K = K_{-1} + I$

(I.5g) Wages: $W = W_1 + W_2$

(I.5h) Private product: $E = Y + R - W_2$

---

[1] Also implied by the term "complete system" is the recognition that no lagged jointly dependent variable occurs in the system without the corresponding (non-lagged) jointly dependent variables also occurring in the system.

[2] Although the notation has been changed slightly, the explanation of each equation has been taken almost verbatim from Goldberger [1964], pp. 303-304. Klein's model I is given in Klein [1950].

The first equation is a consumption function which describes consumption (C) linearly in terms of profits (P), profits lagged one year (P_{-1}), and the total wage bill (W).

The second equation is an investment equation describing net investment (I) linearly in terms of profits, lagged profits, and capital stock at the beginning of the year (K_{-1}).

The third equation is a demand for labor equation which describes the private wage bill linearly in terms of private product (E), private product lagged by one year (E_{-1}), and time (t) measured in calendar years.

The five identity equations complete the system. The additional variables in the identity equations are national income (Y), indirect taxes (R), government expenditure on goods and services (G), capital stock at the end of the year (K), the private wage bill (W_1), and the government wage bill (W_2).[1]

The variables C, P, W, I, W_1, E, Y, and K are designated as jointly dependent within the system and the variables P_{-1}, K_{-1}, E_{-1}, t, R, G, and W_2 are designated as predetermined to the system. It is convenient to consider one additional predetermined variable -- X_0, a variable which assumes the value 1.0 for all observations. Thus, X_0 is the variable whose coefficient

---

[1] G is also used in this paper to denote the number of jointly dependent variables in the system, and K is also used to denote the number of instrumental variables in the X_I matrix (a matrix which is defined further on); however, the particular uses of G and K should be clear from their contexts.

is $\alpha_0^{[1]}$ in the first equation, $\alpha_0^{[2]}$ in the second equation and $\alpha_0^{[3]}$ in the third equation.

As a digression we will note why certain of the variables defined by identity equations are considered to be jointly dependent rather than predetermined. As indicated in the wages equation, $W = W_1 + W_2$. In formulating the model, W_1 was designated as jointly dependent and W_2 as predetermined. Since W is composed of only one part not contemporaneously independent of the disturbance, W is also not contemporaneously independent of the disturbance and must be denoted jointly dependent. Similarly for K, since $K = I + K_{-1}$ with I designated as jointly dependent. Similarly also for E, since $E = Y + R - W_2$ and Y is jointly dependent.

If the annual data from 1921 through 1941 are used in the model, $T = 21$. Also, given the above designation of variables, $G = 8$, $\Lambda = 8$, and $M = 3$. Regarding each of the variables such as C as a 21×1 vector of observed values, the matrices of equations (I.1) and (I.2) may be defined as:

$$(I.6) \quad Z = [Y \vdots X] = [\underbrace{P, W, W_1, K, C, I, Y, E}_{\substack{\text{jointly dependent} \\ 21 \times 8}} \vdots \underbrace{X_0, P_{-1}, K_{-1}, E_{-1}, t, R, G, W_2}_{\substack{\text{predetermined} \\ 21 \times 8}}]$$

(I.7) $\underset{8\times16}{\alpha} = [\underset{8\times8}{\Gamma} \vdots \underset{8\times8}{B}] =$

	P	W	W_1	K	C	I	Y	E	
Eq.1	$\alpha_1^{[1]}$	$\alpha_2^{[1]}$	0	0	-1	0	0	0	
Eq.2	$\alpha_1^{[2]}$	0	0	0	0	-1	0	0	
Eq.3	0	0	-1	0	0	0	0	$\alpha_1^{[3]}$	
Eq.4	0	0	0	0	1	1	-1	0	
Eq.5	1	1	0	0	0	0	-1	0	
Eq.6	0	0	0	-1	0	1	0	0	
Eq.7	0	-1	1	0	0	0	0	0	
Eq.8	0	0	0	0	0	0	1	-1	

X_0	P_{-1}	K_{-1}	E_{-1}	t	R	G	W_2	
$\alpha_0^{[1]}$	$\alpha_3^{[1]}$	0	0	0	0	0	0	Eq.1
$\alpha_0^{[2]}$	$\alpha_2^{[2]}$	$\alpha_3^{[2]}$	0	0	0	0	0	Eq.2
$\alpha_0^{[3]}$	0	0	$\alpha_2^{[3]}$	$\alpha_3^{[3]}$	0	0	0	Eq.3
0	0	0	0	0	-1	1	0	Eq.4
0	0	0	0	0	0	0	0	Eq.5
1	0	1	0	0	0	0	0	Eq.6
0	0	0	0	0	0	0	1	Eq.7
0	0	0	0	0	1	0	-1	Eq.8

The part within the brackets constitutes the α matrix. The equation number corresponding to each row is given on each side of the matrix and the variable corresponding to each column is given above the matrix.

Notice that in the α matrix, a coefficient of -1 has been assigned for any variable listed to the left of the equality sign in equations (I.5a) through (I.5h). This is equivalent to assigning a coefficient of 1 in the equation and then transcribing the coefficient and variable to the same side of the equality sign as the remaining variables.

The coefficients of any equation may be multiplied by a scalar without changing the meaning of the equation. To avoid this indeterminacy we will follow the normalization rule given above, i.e., in each of the equations the coefficient of one of the jointly dependent variables will be assigned the value -1.[1]

Of all of the estimation procedures discussed in this paper, only in the case of limited information single equation

---

[1]Many normalization rules have been used in the past to avoid indeterminacy in each equation. One way is to specify that a coefficient corresponding to one jointly dependent variable in each equation be set to 1. A second way is to normalize each equation such that the resulting estimated disturbance variance-covariance matrix has 1's for diagonal elements. For our purpose here, however, we will find it most convenient to select a variable in each equation as the normalizeing variable and set its coefficient to -1.

If -1 is used as the normalizing coefficient, the resulting coefficients may be compared directly with coefficients from methods such as direct least squares (DLS) or two-stage least squares (2SLS); however, if 1 is used as the normalizing coefficient, the sign of each coefficient must be reversed before comparing the coefficient with a comparable coefficient from DLS or 2SLS. Since it is as easy to use -1 for normalizing as to use 1, it seems highly desirable to do so and avoid the reflection in coefficient signs.

maximum likelihood (LIML), limited information subsystem maximum likelihood (SML), linearized maximum likelihood (LML), and full information maximum likelihood (FIML) will it make no substantive difference in the estimated coefficients which jointly dependent variable is chosen as the normalizing variable. For the remainder of the procedures, selection of a different normalizing variable will make a substantive change in the resulting estimated coefficients.[1]

Only coefficients of the form $\alpha_i^{[j]}$ must be estimated. All of the remaining coefficients in the α matrix are assumed known.

---

[1] In the special case of a just-identified equation (see section II.D), it will also make no substantive difference for many methods since the same estimated coefficients are obtained by many of the single equation procedures as are obtained for LIML.

2. Single structural equation

In part I, estimation of only a single equation from among the M stochastic equations will be considered although for many of the procedures some of the information contained in the remaining $M - 1$ stochastic equations and the $G - M$ identity equations will be taken into account by the computational method. For most of the "single equation" estimating methods which are considered, account is taken of the structure of the particular equation to be estimated plus additional "instrumental variables." (The predetermined variables in the entire system including predetermined variables in identity equations usually comprise the instrumental variables.) No account is taken of jointly dependent variables inside the system of equations but outside the equation being estimated or of which particular equations the instrumental variables occur in (if the instrumental variables are predetermined variables in the system).

For ease of expressing computational formulas, some additional notation regarding a single equation from the system of equations will be recorded. The μ^{th} equation may be written separately as:

$$(I.8) \qquad \underset{T\times 1}{y_\mu} = \underset{T\times m_\mu}{Y_\mu}\ \underset{m_\mu\times 1}{\gamma_\mu} + \underset{T\times \ell_\mu}{X_\mu}\ \underset{\ell_\mu\times 1}{\beta_\mu} + \underset{T\times 1}{u_\mu}$$

or

$$(I.9) \qquad y_{\mu} = [Y_{\mu} \vdots X_{\mu}] \begin{bmatrix} \gamma_{\mu} \\ \beta_{\mu} \end{bmatrix} + u_{\mu}$$

or

$$(I.10) \qquad \underset{T\times 1}{y_{\mu}} = \underset{T\times n_{\mu}}{Z_{\mu}} \; \underset{n_{\mu}\times 1}{\delta_{\mu}} + \underset{T\times 1}{u_{\mu}}$$

where:

y_{μ} is the vector of T sample observations taken on the normalizing jointly dependent variable (i.e., the jointly dependent variable assigned a coefficient of -1) for the μ^{th} equation.

Y_{μ} is the $T\times m_{\mu}$ matrix of T sample observations taken on the remaining m_{μ} jointly dependent variables in the μ^{th} equation. We will refer to jointly dependent variables other than the normalizing jointly dependent variable as "explanatory" jointly dependent variables.

X_{μ} is the $T\times \ell_{\mu}$ matrix of T sample observations taken on the ℓ_{μ} predetermined variables in the μ^{th} equation.

Z_{μ} $= [Y_{\mu} \vdots X_{\mu}]$ is the $T\times n_{\mu}$ matrix of T sample observations taken on all n_{μ} $(= m_{\mu} + \ell_{\mu})$ explanatory variables in the equation; that is, on all variables in the equation except the normalizing variable, y_{μ}.

u_{μ} is the vector of T unobserved structural disturbances.

γ_{μ} is the $m_{\mu}\times 1$ vector of population coefficients of the m_{μ} explanatory jointly dependent variables in the μ^{th} equation. (The elements of γ_{μ} may be obtained from the

μ^{th} row of Γ by deleting the normalizing coefficient and the coefficients which are known to be zero.)

β_μ is the $\ell_\mu \times 1$ vector of population coefficients of the ℓ_μ predetermined variables in the μ^{th} equation. (The elements of β_μ are the non-zero elements of the μ^{th} row of B.)

$\delta_\mu = \begin{bmatrix} \gamma_\mu \\ \beta_\mu \end{bmatrix}$ is the $n_\mu \times 1$ vector of population coefficients of the n_μ $(= m_\mu + \ell_\mu)$ explanatory variables in the equation. (The elements of δ_μ may be obtained from the μ^{th} row of α be deleting the normalizing coefficient and the coefficients which are known to be zero.)

Sometimes it will be desirable that the normalizing variable be included as part of an observation matrix or the normalizing coefficient as part of a coefficient vector. To do this the following additional notation is used:

$_+Y_\mu = [y_\mu \vdots Y_\mu]$ is the $T \times (m_\mu + 1)$ matrix of T sample observations taken on all $m_\mu + 1$ jointly dependent variables in the μ^{th} equation.

$_+Z_\mu = [y_\mu \vdots Z_\mu] = [y_\mu \vdots Y_\mu \vdots X_\mu] = [_+Y_\mu \vdots X_\mu]$ is the $T \times (n_\mu + 1)$ matrix of T sample observations taken on all $n_\mu + 1$ variables in the μ^{th} equation.

$_+\gamma_\mu = \begin{bmatrix} -1 \\ \gamma_\mu \end{bmatrix}$ is the $(m_\mu + 1) \times 1$ vector of population coefficients of all $m_\mu + 1$ jointly dependent variables in the μ^{th} equation.

$+\delta_\mu = \begin{bmatrix} -1 \\ \delta_\mu \end{bmatrix} = \begin{bmatrix} -1 \\ \gamma_\mu \\ \beta_\mu \end{bmatrix} = \begin{bmatrix} +\gamma_\mu \\ \beta_\mu \end{bmatrix}$ is the $(n_\mu + 1)\times 1$ vector of population coefficients of all $n_\mu + 1$ variables in the μ^{th} equation.

Using the above additional notation, the μ^{th} equation may also be written as:

$$(I.11) \qquad \underset{T\times(m_\mu+1)}{+Y_\mu} \quad \underset{(m_\mu+1)\times 1}{+\gamma_\mu} + \underset{T\times \ell_\mu}{X_\mu} \quad \underset{\ell_\mu \times 1}{\beta_\mu} + \underset{T\times 1}{u_\mu} = \underset{T\times 1}{0}$$

or

$$(I.12) \qquad \underset{T\times(n_\mu+1)}{+Z_\mu} \quad \underset{(n_\mu+1)\times 1}{+\delta_\mu} + \underset{T\times 1}{u_\mu} = \underset{T\times 1}{0} \; .$$

The estimation of a particular equation will often be accomplished by essentially a two stage procedure in which the jointly dependent variables in the equation are first adjusted by a set of "instrumental variables" (we will refer to a set of instrumental variables as a set of "instruments") which consist of the predetermined variables in the equation plus additional instruments--usually the additional predetermined variables in the system.[1] The coefficients are then estimated from the adjusted jointly dependent variables and the predetermined variables in the equation. We will refer to a matrix of instruments as X_I. (The I denotes instruments.) Thus,

---

[1] The use of the term instruments follows Fisher [1965].

$$(I.13) \qquad \underset{T\times K}{X_I} = \underset{T\times \ell_\mu}{[\, X_\mu} \;\vdots\; \underset{T\times (K-\ell_\mu)}{X^{**}} \,] .$$

Strictly speaking, X_I, X^{**}, and K should be written as X_{I_μ}, X^{**}_μ, and K_μ, respectively since they may be different for each equation, μ; however, since the equation referred to will be clear from the context, we will simplify the notation by writing X_{I_μ} as X_I, X^{**}_μ as X^{**}, and K_μ as K.[1]

During discussion of the single equation techniques (part I), we will drop the subscript μ when such will not be confusing. Thus, y_μ, Y_μ, Z_μ, u_μ, ${}_+Y_\mu$, ${}_+Z_\mu$, and X^{**}_μ will usually be shortened to y, Y, Z, u, ${}_+Y$, ${}_+Z$, and X^{**} respectively. Also, γ_μ, β_μ, δ_μ, ${}_+\gamma_\mu$, and ${}_+\delta_\mu$ will usually be shortened to γ, β, δ, ${}_+\gamma$, and ${}_+\delta$ respectively, and m_μ, ℓ_μ, n_μ, and K_μ will usually be shortened to m, ℓ, n, and K, respectively.

The consumption equation (I.5a) of Klein's model I will be used to illustrate the above notation. Since the consumption equation is the first equation, μ will be 1 and:

$$y_1 = C, \quad Y_1 = [P, W], \quad X_1 = [X_0, P_{-1}], \quad Z_1 = [P, W, X_0, P_{-1}],$$

$${}_+Y_1 = [C, P, W], \quad {}_+Z_1 = [C, P, W, X_0, P_{-1}],$$

$$X^{**} = [K_{-1}, E_{-1}, t, R, G, W_2] .$$

---

[1] In 3SLS estimation X_I is again used to denote a set of instruments but in this case X_I includes as instruments the predetermined variables in all equations of the subsystem being estimated plus additional instruments desired. For 3SLS estimation the same X_I matrix is used by all equations in the subsystem being estimated.

If all of the predetermined variables in the system are used as instrumental variables for adjusting the jointly dependent variables, then X_I coincides with X except possibly for a renumbering of variables, i.e.,

$$X_I = [X_1 \vdots X^{**}] = [X_0, P_{-1} \vdots K_{-1}, E_{-1}, t, R, G, W_2] .$$

The population coefficient vectors are:

$$\gamma_1 = \begin{bmatrix} \alpha_1^{[1]} \\ \alpha_2^{[1]} \end{bmatrix}, \qquad \beta_1 = \begin{bmatrix} \alpha_0^{[1]} \\ \alpha_3^{[1]} \end{bmatrix}, \qquad \delta_1 = \begin{bmatrix} \alpha_1^{[1]} \\ \alpha_2^{[1]} \\ \alpha_0^{[1]} \\ \alpha_3^{[1]} \end{bmatrix}, \qquad {}^{+}\gamma_1 = \begin{bmatrix} -1 \\ \alpha_1^{[1]} \\ \alpha_2^{[1]} \end{bmatrix},$$

$$ {}^{+}\delta_1 = \begin{bmatrix} -1 \\ \alpha_1^{[1]} \\ \alpha_2^{[1]} \\ \alpha_0^{[1]} \\ \alpha_3^{[1]} \end{bmatrix}$$

Thus, $m_1 = 2$, $\ell_1 = 2$, $n_1 = 4$, and $K = 8$.

Equations (I.8) through (I.12) become:

$$\text{(I.8')} \qquad C = [P, W] \begin{bmatrix} \alpha_1^{[1]} \\ \alpha_2^{[1]} \end{bmatrix} + [X_0, P_{-1}] \begin{bmatrix} \alpha_0^{[1]} \\ \alpha_3^{[1]} \end{bmatrix} + u_1$$

(I.9') and (I.10')
$$C = [P,\ W \vdots X_0,\ P_{-1}] \begin{bmatrix} \alpha_1^{[1]} \\ \alpha_2^{[1]} \\ \alpha_0^{[1]} \\ \alpha_3^{[1]} \end{bmatrix} + u_1$$

(I.11')
$$[C \vdots P,\ W] \begin{bmatrix} -1 \\ \alpha_1^{[1]} \\ \alpha_2^{[1]} \end{bmatrix} + [X_0,\ P_{-1}] \begin{bmatrix} \alpha_0^{[1]} \\ \alpha_3^{[1]} \end{bmatrix} + u_1 = 0$$

(I.12')
$$[C \vdots P,\ W \vdots X_0,\ P_{-1}] \begin{bmatrix} -1 \\ \alpha_1^{[1]} \\ \alpha_2^{[1]} \\ \alpha_0^{[1]} \\ \alpha_3^{[1]} \end{bmatrix} + u_1 = 0$$

3. Statistical assumptions

Following are the statistical assumptions which will be made in this paper unless noted otherwise:

(1) If estimating the μ^{th} structural equation by a single equation method, the μ^{th} equation is identifiable by the a priori restrictions on the values of the coefficients in the equation. If estimating by a multiple equations method, all equations in the subset of equations being estimated are identifiable by the a priori restrictions on the values of the coefficients in the subsystem.[1]

In the early chapters on single equation methods (chapters II

---

[1] See Goldberger [1964], pp. 306-318, Johnston [1963], pp. 240-252, or Koopmans and Hood [1953], pp. 135-142 for a discussion of identification. In addition to the usual order conditions imposed on a single equation, (i.e., the usual counting rule $K^{**}_{\mu} > m_{\mu}$ given below), the assumption here is that observationally equivalent structures for the subsystem being estimated do not occur. Johnston [1963], p. 252 states: "If $K^{**}_{\mu} \geq m_{\mu}$, the parameters of a relation are identifiable. Our practical estimation procedure may then be influenced by whether $K^{**}_{\mu} = m_{\mu}$ or $K^{**}_{\mu} > m_{\mu}$. In the former case rk $\hat{\pi}_{\Delta,**}$ will, apart from a freakish statistical accident, be equal to m_{μ} ..." (Johnston used K^{**} in place of K^{**}_{μ}, $G^{\Delta} - 1$ in place of m_{μ}, and ρ in place of rk.) Unfortunately, Johnston's statement seems stronger than is warranted. rk $\hat{\pi}_{\Delta,**} < m_{\mu}$ has probably occurred much more often than is generally realized, but has not been apparent when estimating by single equation methods; however, such a situation is more likely to become apparent when estimating by multiple equation techniques. Koopman, Rubin, and Leipnik [1950], pp. 78-80 present an additional examination (in addition to the usual counting rule for a single equation) which may be readily performed to detect observationally equivalent structures.

and III) only a priori restrictions that certain coefficients are zero are used; therefore, during this part, identifiability implies $n_\mu \leq \text{rk } X \leq \Lambda$ and the ${}_+Z_\mu$ matrix has full column rank.[1] (rk X is used to denote rank of X.) In the last chapter on single equation methods (chapter IV), more general linear restrictions are considered. For that chapter, n_μ may be greater than Λ and ${}_+Z_\mu$ need not have full column rank.

Except for showing some relationships, in no part of this paper will the usual assumption that X has full column rank (i.e., the assumption that rk X = Λ) be made, since the computational procedures given in this paper automatically handle the more general situation of rk X < Λ.

(2) The $T^{X}M$ matrix of disturbances of the first M equations,

$$U = [u_1 \cdots u_M] = \begin{bmatrix} U_{[1]} \\ \vdots \\ U_{[T]} \end{bmatrix} = \begin{bmatrix} u_{11} & \cdots & u_{1M} \\ \vdots & & \vdots \\ u_{T1} & \cdots & u_{TM} \end{bmatrix}$$

has a multivariate distribution with $\mathcal{E}U = 0$, $\mathcal{E}U'_{[t]}U_{[t]} = \Sigma$ for all t and $\mathcal{E}U'_{[t]}U_{[t']} = 0$ for $t \neq t'$, Σ being an $M^{X}M$ positive semi-definite matrix. Thus, Σ is the population variance-covariance matrix of disturbances. When estimating

---

[1] $n_\mu \leq \Lambda$ is equivalent to $m_\mu \leq K^{**}_\mu$ since $n_\mu = m_\mu + \ell_\mu$ and $\Lambda = \ell_\mu + K^{**}_\mu$. (Goldberger [1964], pp. 306-318 and Johnston [1963], pp. 240-252 use the latter order condition.) $n_\mu \leq \text{rk } X$ is imposed since we are permitting X to have less than full column rank in this paper. rk X $\leq \Lambda$ always holds since the rank of any matrix is less than or equal to the number of columns (and rows) in the matrix.

a subset of structural equations (including the entire set of equations) by a multiple equations technique, the stronger assumption that Σ is positive definite will be required so that the determinant of Σ will be greater than zero and the inverse of Σ will exist. The restriction $\mathcal{E}U'_{[t]}U_{[t]} = \Sigma$ for all t implies that the disturbance variance-covariance matrix is constant for all observations. The restriction $\mathcal{E}U'_{[t]}U_{[t']} = 0$ for $t \neq t'$ implies that there exists no serial correlation between observations of the disturbance elements. The μ^{th} diagonal element of Σ will be denoted σ_μ^2, i.e., Var $u_\mu = \sigma_\mu^2$. For notational simplicity σ_μ^2 will be written simply as σ^2 during discussion of the single equation methods. The above assumptions regarding U imply that under general conditions $\text{plim}(1/T)U'U = \Sigma$.[1]

(3) The $T^{X}\Lambda$ matrix of predetermined variables, X, has the property $\text{plim}(1/T)X'X = \mathcal{E}(1/T)X'X = \Omega_{XX}$, a finite positive semi-definite matrix. Also, the variables in U are contemporaneously independent of the variables in X; that is, $U_{[t]}$ is statistically independent of $X_{[t']}$ for all t, t' with $t \geq t'$.[2] This implies that $\text{plim}(1/T)X'U = 0$.[3]

(4) $\det \Gamma \neq 0$, hence Γ^{-1} exists. (det denotes determinant.)

---

[1] Goldberger [1964], p. 300. As noted earlier plim denotes $\underset{T\to\infty}{\text{plim}}$.

[2] This assumption holds by definition when X includes only exogenous variables. It allows inclusion of lagged jointly dependent variables in X provided there is no serial correlation in the disturbances.

[3] See Christ [1966], pp. 377, 378 and footnote 70 of p. 439.

(5) In some of the computational methods, jointly dependent variables will be adjusted by a $T \times K$ matrix of instrumental variables which we will denote X_I (the subscript I denoting instruments). These instrumental variables will be assumed to have essentially the same properties as the predetermined variables in the system. (Usually X_I consists of the set of predetermined variables in the system.) In particular we will assume that $\text{plim}(1/T)X_I'U = 0$ (where 0 is a $K \times M$ matrix) and that $\text{plim}(1/T)X_I'X_I = \mathcal{E}(1/T)X_I'X_I = \Omega_{X_I X_I}$, a $K \times K$ positive semi-definite matrix. (We will not, in general, assume that X_I has full column rank.) We will also assume that $\text{plim}(1/T)X_I'X = \mathcal{E}(1/T)X_I'X = \Omega_{X_I X}$, a $K \times \Lambda$ matrix.

4. Reduced form equations

Often after estimating the coefficients of the structural equations in a system, it is desired that these be used for predictive purposes; however, (1) multiple jointly dependent variables may occur in the equation (therby requiring that values must be assumed for the "explanatory" dependent variables as well as the predetermined variables), (2) it may not be obvious which equation to use in the prediction of a particular jointly dependent variable, and (3) a given equation will not reflect the repercussion of the assumed levels of all predetermined variables in the system. As a result, it is often desired that the structural equations be solved for a set of "reduced form" equations, each reduced form equation containing 1 jointly dependent variable and all of the predetermined variables in the system (except that the *a priori* restrictions on the structural equations may imply that certain of the coefficients of the reduced form equations are to be zero, also).

An additional reason for calculating reduced form coefficients comes from the calculation of elasticities. Certain direct elasticities between variables should be based on the coefficients of the structural equations; however, in many cases, when specifying elasticities between two variables, the relationship between these variables after taking account of all repercussions in the system is desired. Such elasticities should be based on the reduced form coefficients.

The reduced form equations may be derived by premultiplying (I.3) by Γ^{-1} or by postmultiplying (I.3) by $(\Gamma^{-1})'$. Premultiplying (I.3) by Γ^{-1} we have:

$$(I.14)\quad \underset{G\times G}{\Gamma^{-1}}\ \underset{G\times G}{\Gamma}\ \underset{G\times 1}{Y'_{[t]}} + \underset{G\times G}{\Gamma^{-1}}\ \underset{G\times\Lambda}{B}\ \underset{\Lambda\times 1}{X'_{[t]}} + \underset{G\times G}{\Gamma^{-1}} \begin{bmatrix} \underset{M\times 1}{U'_{[t]}} \\ \\ \underset{(G-M)\times 1}{0'_{[t]}} \end{bmatrix} = \underset{G\times 1}{0'_{[t]}}$$

or

$$(I.15)\quad \underset{G\times 1}{Y'_{[t]}} = \underset{G\times G}{-\Gamma^{-1}}\ \underset{G\times\Lambda}{B}\ \underset{\Lambda\times 1}{X'_{[t]}} - \underset{G\times G}{\Gamma^{-1}} \underset{G\times 1}{[U_{[t]} \vdots 0_{[t]}]'}$$

or

$$(I.16)\quad \underset{G\times 1}{Y'_{[t]}} = \underset{G\times\Lambda}{\Pi}\ \underset{\Lambda\times 1}{X'_{[t]}} + \underset{G\times 1}{V'_{[t]}} .$$

In terms of the entire observation matrices for Y, X, and V, (I.16) may also be written as:

$$(I.17)\quad \underset{G\times T}{Y'} = \underset{G\times\Lambda}{\Pi}\ \underset{\Lambda\times T}{X'} + \underset{G\times T}{V'}$$

or

$$(I.18)\quad \underset{T\times G}{Y} = \underset{T\times\Lambda}{X}\ \underset{\Lambda\times G}{\Pi'} + \underset{T\times G}{V}$$

where in (I.16) through (I.18)

$$(I.19)\quad \Pi = -\Gamma^{-1}B$$

is the $G^{X}\Lambda$ matrix of coefficients of the reduced form equations, in which each row of Π gives the coefficients corresponding to the predetermined variables of a single reduced form equation. Each equation contains only one jointly dependent variable and this variable is written to the left of the equality sign.

$$V = -[U \vdots 0](\Gamma^{-1})' = [v_1 \cdots v_G] = \begin{bmatrix} V_{[1]} \\ \vdots \\ V_{[T]} \end{bmatrix} = \begin{bmatrix} v_{11} & \cdots & v_{1G} \\ \vdots & & \vdots \\ v_{T1} & \cdots & v_{TG} \end{bmatrix}$$

is the $T^{X}G$ matrix of reduced form disturbances.

Some of the statistical characteristics of the reduced form matrices which follow from their relationships to the structural equations and the statistical assumptions regarding the structural equations are:

$$(I.20) \qquad \mathcal{E}V = \mathcal{E}\left[[-U \vdots 0](\Gamma^{-1})'\right] = [-\mathcal{E}U \vdots 0](\Gamma^{-1})' = [0 \vdots 0] = 0 \quad .$$

$$(I.21) \qquad \mathcal{E}V'_{[t]}V_{[t']} = \mathcal{E}\begin{bmatrix} v_{t1} \\ v_{tG} \end{bmatrix}[v_{t'1} \cdots v_{t'G}]$$

$$= \mathcal{E}(\Gamma^{-1})'[U_{[t]} \vdots 0_{[t]}]'[U_{[t']} \vdots 0_{[t']}]\Gamma^{-1}$$

$$= (\Gamma^{-1})'\begin{bmatrix} \mathcal{E}U'_{[t]}U_{[t']} & 0 \\ 0 & 0 \end{bmatrix}\Gamma^{-1} \quad .$$

Thus,

$$(I.22) \qquad \mathcal{E}V'_{[t]}V_{[t]} = (\Gamma^{-1})'\begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}\Gamma^{-1} \overset{\text{def}}{=}^{1} \text{a } G^{X}G \text{ positive semi-}$$

$^{1}\overset{\text{def}}{=}$ denotes that we are defining the matrix $(\Gamma^{-1})'\begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}\Gamma^{-1}$ as Ω_{VV} .

definite matrix which is fixed for all t. Under general conditions, $\text{plim}(1/T)V'V = \Omega_{VV}$.

$$(I.24) \qquad \delta V'_{[t]} V_{[t']} = (\Gamma^{-1})' \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \Gamma^{-1} = 0 \quad \text{for} \quad t \neq t' \ .$$

$$(I.25) \qquad \text{plim}(1/T)X'V = -\text{plim}(1/T)X'[U \vdots 0](\Gamma^{-1})'$$

$$= -[\text{plim}(1/T)X'U \vdots 0](\Gamma^{-1})' = [0 \vdots 0](\Gamma^{-1})' = 0 \ .$$

$$(I.26) \qquad \text{plim}(1/T)X_I'V = -\text{plim}(1/T)X_I'[U \vdots 0](\Gamma^{-1})' = [0 \vdots 0](\Gamma^{-1})' = 0 \ .$$

D. Orthogonality Relationships

1. Notation expressing orthogonality relationships

We can shorten the expression of many formulas which are given in this paper while giving a better idea of the computational use of these formulas by introducing some additional notation at the outset.

Two variables are said to be orthogonal if their sum of cross-products is zero; that is z_1 is orthogonal to z_2 if $z_1'z_2 = 0$. The extension of the concept of orthogonality to matrices whose columns are variables is quite straightforward. Let Z_1 be a $T^{X}N_1$ matrix of variables and Z_2 be a $T^{X}N_2$ matrix of variables. Then the columns of Z_1 are said to be orthogonal to the columns of Z_2 if the matrix $Z_1'Z_2 = 0$ where 0 is an $N_1{}^{X}N_2$ matrix. (The ij^{th} element of the matrix 0 is the sum of cross-products of the i^{th} variable of Z_1 and the j^{th} variable of Z_2. The ij^{th} element is, of course, zero since the entire matrix of sums of cross-products is zero.)

In this paper each column in a matrix of variables is a variable; thus, for this paper we will shorten our definition of orthogonality between the variables in two matrices to: Z_1 is orthogonal to Z_2 if $Z_1'Z_2 = 0$.

It is often convenient to divide a vector into two components -- (1) the part of the vector within the space spanned by (the variables of) another matrix and (2) the part of the vector orthogonal to the other matrix (i.e., outside the space spanned

by the other matrix). Thus, given a $T^{X}1$ vector, y, and a $T^{X}N_1$ matrix, X, y may be separated into:

(I.27)
$$\underset{T^{X}1}{y} = \underset{T^{X}1}{y_{\parallel X}} + \underset{T^{X}1}{y_{\perp X}}$$

where:

$y_{\parallel X}$ is the part of y which is in the space spanned by X.[1]

$y_{\perp X}$ is the part of y which is orthogonal to X. ($y_{\perp X}$ is the part of y outside the space spanned by the variables in X.)[2]

The $\perp$ (perpendicular) symbol is used to denote orthogonality, because if two vectors, y and x, are geometrically at right angles to each other, then $y'x = 0$ (i.e., the vectors are orthogonal). The use of the $\parallel$ (parallel) symbol to represent "within the space spanned by" may be justified by a similar geometrical argument.

Extension of this notation to matrices of variables is straightforward. Thus, a $T^{X}G$ matrix of variables Y ($Y = [y_1 \ldots y_G]$) may be partitioned as:

(I.28)
$$\underset{T^{X}G}{Y} = \underset{T^{X}G}{y_{\parallel X}} + \underset{T^{X}G}{Y_{\perp X}}$$

where:

$Y_{\parallel X}$ is the part of Y in the space spanned by X.

$Y_{\parallel X} = \{[y_1]_{\parallel X} \ldots [y_G]_{\parallel X}\}$; that is, $Y_{\parallel X}$ is merely the matrix obtained by calculating the part of each variable in Y which is in the space spanned by the variables in X.

---

[1] $y_{\parallel X}$ is the projection of y onto the space spanned by the variables in X.

[2] $y_{\perp X}$ is the projection of y onto the space orthogonal to X.

$Y_{\perp X}$ is the part of Y which is orthogonal to X.

$Y_{\perp X} = \{[y_1]_{\perp X} \cdots [y_G]_{\perp X}\}$; that is, $Y_{\perp X}$ is merely the matrix obtained by calculating the part of each variable in Y which is orthogonal to X.

The partitioning of Y into $Y_{\parallel X}$ and $Y_{\perp X}$ may be accomplished by an application of least squares. Let us estimate the equation $y = X\pi + v$ by least squares. (In the notation previously given, this equation would be a reduced form equation.) The usual least squares vector of predicted (estimated) values of y (i.e., $\hat{y}$) is $y_{\parallel X}$, i.e., the part of y in the space spanned by X, and the usual least squares vector of residuals (i.e., $\hat{v}$) is $y_{\perp X}$ -- the part of y orthogonal to X.

To show this we note that (assuming that X has full column rank) the usual least squares solution for the vector of estimated coefficients is given by:

$$\hat{\pi} = (X'X)^{-1}X'y \; ; \tag{I.29}$$

the $T \times 1$ vector of predicted values for y is given by

$$y_{\parallel X} = \hat{y} = X\hat{\pi} = X(X'X)^{-1}X'y \; ; \tag{I.30}$$

and the $T \times 1$ vector of residuals is given by

$$y_{\perp X} = \hat{v} = y - \hat{y} = y - X\hat{\pi} = y - X(X'X)^{-1}X'y \; . \tag{I.31}$$

Notice that in the usual least squares calculations, y is divided into a part $\hat{y}$ (or $y_{\parallel X}$) within the space spanned by X and a part $\hat{v}$ (or $y_{\perp X}$) orthogonal to X. That $\hat{v}$ is orthogonal

to X is easily demonstrated:

$$(I.32)\quad X'\hat{v} = X'(y - X(X'X)^{-1}X'y) = X'y - X'X(X'X)^{-1}X'y = X'y - X'y = 0 .$$

Since $Y_{\|X}$ and $Y_{\perp X}$ were defined as the parts of each variable in Y in the space spanned by X and orthogonal to X, respectively, the calculation of $Y_{\|X}$ and $Y_{\perp X}$ may also be regarded as least squares calculations. Let us consider G equations with separate dependent variables, $y_1 \ldots y_G$, but the same independent variables, $x_1 \ldots x_\Lambda$ for each equation. Let $Y = [y_1 \ldots y_G]$ and $X = [x_1 \ldots x_\Lambda]$ as before. The least squares solutions for the coefficients of the G equations are given by the matrix

$$(I.33)\quad \hat{\Pi}' = [(X'X)^{-1}X'y_1 \ldots (X'X)^{-1}X'y_G] = [X'X]^{-1}X'Y \ ;$$

the $T^{X}G$ matrix of predicted values for Y is given by:

$$\begin{aligned}(I.34)\quad Y_{\|X} = \hat{Y} &= [\hat{y}_1 \ldots \hat{y}_G] = [X\hat{\pi}_1 \ldots X\hat{\pi}_G] \\ &= [X(X'X)^{-1}X'y_1 \ldots X(X'X)^{-1}X'y_M] \\ &= X(X'X)^{-1}X'[y_1 \ldots y_M] = X(X'X)^{-1}X'Y \ ;\end{aligned}$$

and the $T^{X}G$ matrix of residuals for Y is given by:

$$\begin{aligned}(I.35)\quad Y_{\perp X} = \hat{U} &= [\hat{u}_1 \ldots \hat{u}_G] = [y_1 - \hat{y}_1 \cdots y_G - \hat{y}_G] \\ &= [y_1 - X(X'X)^{-1}X'y_1 \ldots y_M - X(X'X)^{-1}X'y_M] \\ &= [y_1 \ldots y_M] - X(X'X)^{-1}X'[y_1 \ldots y_M] = Y - X(X'X)^{-1}X'Y .\end{aligned}$$

If X does not have full column rank, the X'X matrix will

be singular, the inverse of $X'X$ will not exist and unique least squares coefficients will not exist for any of the G equations.[1] Even though the least squares coefficients are not unique, the least squares predicted values for Y and the residuals for each equation are unique and can be readily calculated. This illustrates a desirable characteristic of $Y_{\|X}$ and $Y_{\perp X}$--even though X is not of full column rank, $Y_{\|X}$ and $Y_{\perp X}$ are unique and can be readily calculated.[2]

---

[1] A set of least squares coefficients can be obtained by putting enough restrictions on the coefficients of each equation; e.g., by setting certain of the coefficients to zero thereby, in effect, omitting variables from the X matrix.

[2] A direct orthogonalization procedure such as the Gram-Schmidt orthogonalization procedure is probably the most accurate method available for the calculation of residuals ($Y_{\perp X}$), and predicted values of Y ($Y_{\|X}$) may be calculated as $Y - Y_{\perp X}$; however, if very many observations occur it will be considerably more efficient to obtain a set of least squares coefficients for each column of Y and then use these coefficients to calculate $Y_{\|X}$ and $Y_{\perp X}$. This may be done by selecting variables from the X matrix in such a way that the subset of variables selected have full column rank which is the same rank as the original X matrix and then using this smaller set of variables in place of the original X matrix in the calculation of least squares coefficients, predicted values of Y ($Y_{\|X}$) and residuals ($Y_{\perp X}$). The selection of a subset of variables having full column rank and having the same rank as X may be built into the inversion routine in the manner noted in section I.D.2.

Although less computer time will generally be required if a set of least squares coefficients are calculated as indicated above and $Y_{\|X}$ and $Y_{\perp X}$ are calculated from these coefficients, it is desirable that matrices of the form $Y'_{\|X}Y_{\|X}$, $Y'_{\perp X}Y_{\perp X}$, $Y'_{\|X}Z_1$, or $Y'_{\perp X}Z_2$ be calculated by direct orthogonalization, since matrices of this form may be calculated directly from moment matrices instead of from the observation matrix. (A method for doing so is given in section I.D.2.)

In this paper, we will make extensive use of matrices of the form $Y'_{\|X}Y_{\|X}$ and $Y'_{\perp X}Y_{\perp X}$ which we will denote as $[Y'Y]_{\|X}$ and $[Y'Y]_{\perp X}$, respectively. Thus, $[Y'Y]_{\|X}$ is the moment matrix (i.e., the sums of squares and cross-products matrix) of the part of Y in the space spanned by the columns of X and $[Y'Y]_{\perp X}$ is the moment matrix of the part of Y orthogonal to X.

If X has full column rank, then $[Y'Y]_{\|X}$ may be expressed as:

$$(I.36)\quad [Y'Y]_{\|X} = Y'_{\|X}Y_{\|X} = [Y'X(X'X)^{-1}X'][X(X'X)^{-1}X'Y] = Y'X(X'X)^{-1}X'Y .$$

and $[Y'Y]_{\perp X}$ may be expressed as:

$$\begin{aligned}(I.37)\quad [Y'Y]_{\perp X} &= Y'_{\perp X}Y_{\perp X} = [Y' - Y'X(X'X)^{-1}X'][Y - X(X'X)^{-1}X'Y] \\ &= Y'Y - 2Y'X(X'X)^{-1}X'Y + Y'X(X'X)^{-1}X'X(X'X)^{-1}X'Y \\ &= Y'Y - Y'X(X'X)^{-1}X'Y = Y'Y - [Y'Y]_{\|X} .\end{aligned}$$

Although $Y'X(X'X)^{-1}X'Y$ is the usual computational formula given for $[Y'Y]_{\|X}$ and $Y'Y - Y'X(X'X)^{-1}X'Y$ is the usual computational formula given for $[Y'Y]_{\perp X}$, these formulas will not be used in this paper, except possibly for the purpose of a derivation. Instead, the $[Y'Y]_{\perp X}$ matrix will be calculated by direct orthogonalization from the $Y'Y$, $Y'X$, and $X'X$ matrices by a computational scheme given in section I.D.2; hence, the use of the form $[Y'Y]_{\perp X}$ rather than $Y'_{\perp X}Y_{\perp X}$. Calculation by this direct orthogonalization method has the advantages of being more accurate, requiring less computer time, and requiring less computer storage.

Also, calculation by direct orthogonalization has the additional advantage that $[Y'Y]_{\perp X}$ is easily calculated when X has less than full column rank. ($[Y'Y]_{\perp X}$ is unique even though X has less than full column rank.) The $[Y'Y]_{\parallel X}$ matrix will be calculated as $Y'Y - [Y'Y]_{\perp X}$ rather than by the computational formula $Y'X(X'X)^{-1}X'Y$ for the same reasons that $[Y'Y]_{\perp X}$ is calculated by direct orthogonalization.

More general matrices of the form $[Z_1'Z_2]_{\parallel Z_3}$ and $[Z_1'Z_2]_{\perp Z_3}$ will also be calculated, where Z_1 is a $T \times N_1$ matrix of variables, Z_2 is a $T \times N_2$ matrix of variables and Z_3 is a $T \times N_3$ matrix of variables, the variables in any of these matrices being jointly dependent, predetermined, or some jointly dependent and others predetermined within the same matrix. (Some of the variables in one matrix may be repeated in the other two.) Although $[Z_1'Z_2]_{\perp Z_3} = [Z_1]_{\perp Z_3}'[Z_2]_{\perp Z_3}$, $[Z_1'Z_2]_{\perp Z_3}$ will not be computed in this way but will instead be extracted from the moment matrix of the part of $Z = [Z_1 \vdots Z_2]$ orthogonal to Z_3, i.e., $[Z_1'Z_2]_{\perp Z_3}$ may be extracted as the upper right hand block of:

$$[Z'Z]_{\perp Z_3} = \begin{bmatrix} Z_1'Z_1 & Z_1'Z_2 \\ Z_2'Z_1 & Z_2'Z_2 \end{bmatrix}_{\perp Z_3} = \begin{bmatrix} [Z_1'Z_1]_{\perp Z_3} & [Z_1'Z_2]_{\perp Z_3} \\ [Z_2'Z_1]_{\perp Z_3} & [Z_2'Z_2]_{\perp Z_3} \end{bmatrix}$$

which is computed directly from the $Z'Z$, $Z'Z_3$ and $Z_3'Z_3$ matrices. Any variables common to Z_1 and Z_2 need not be repeated in Z. The row and column of $[Z'Z]_{\perp Z_3}$ corresponding to any variable of Z_1 or Z_2 which also occurs in Z_3 will be

zero, since no part of that variable is orthogonal to Z_3.

$[Z_1'Z_2]_{\|Z_3}$ is calculated simply as $Z_1'Z_2 - [Z_1'Z_2]_{\perp Z_3}$.

If the reader has difficulty with the concept of orthogonality, he may regard a matrix $[Z_1'Z_2]_{\perp Z_3}$ as merely a matrix that is calculated from the matrix $Z_1'Z_2$ (and has the same number of rows and columns as $Z_1'Z_2$) through use of the matrices $Z_3'Z_3$ and $Z_3'Z$ (where Z contains all variables in Z_1 and Z_2) by a standard computational procedure given in section I.D.2 (which we will call the direct orthogonalization procedure). The matrix $[Z_1'Z_2]_{\|Z_3}$ may be regarded (and calculated) as $Z_1'Z_2 - [Z_1'Z_2]_{\perp Z_3}$.[1]

The remainder of this section is devoted to deriving some fundamental relationships which will be useful in verifying various results throughout this paper; however, these relationships are not required to apply the formulas which are given in this paper.

---

[1] Some readers will find it of interest to remember that $[Z_1'Z_2]_{\|Z_3}$ and $[Z_1'Z_2]_{\perp Z_3}$ can be calculated by direct least squares by using the variables in Z_1 and Z_2 as dependent variables and the variables in Z_3 as independent variables (the maximum number of linearly independent variables in Z_3 being used as the set of independent variables if Z_3 has less than full column rank). The matrices $[Z_1]_{\|Z_3}$ and $[Z_2]_{\|Z_3}$ are the matrices of predicted values of the variables in Z_1 and Z_2, respectively; the matrices $[Z_1]_{\perp Z_3}$ and $[Z_2]_{\perp Z_3}$ are the matrices of residuals of the variables in Z_1 and Z_2, respectively; $[Z_1'Z_2]_{\|Z_3} = [Z_1]_{\|Z_3}'[Z_2]_{\|Z_3}$; and $[Z_1'Z_2]_{\perp Z_3} = [Z_1]_{\perp Z_3}'[Z_2]_{\perp Z_3}$. The use of the direct orthogonalization procedure merely saves computer time and provides a more accurate calculation of the desired matrices.

In the remainder of this section let Z_1, Z_2, Z_3, and Z_4 be any $T^{X}N_1$, $T^{X}N_2$, $T^{X}N_3$, and $T^{X}N_4$ matrices of variables, respectively, the variables in any of these matrices being jointly dependent, predetermined, or some jointly dependent and some predetermined. Variables in any of these matrices may also occur in any of the other matrices. We will assume only that Z_1, Z_2, Z_3, and Z_4 have rank N_1^*, N_2^*, N_3^*, and N_4^* respectively, i.e., we will assume that any of the matrices may have less than full column rank. In showing algebraically that each of the claimed relationships hold, we will often use matrices Z_1^*, Z_2^*, Z_3^*, and Z_4^* where Z_1^* is a $T^{X}N_1^*$ matrix of variables extracted from the Z_1 matrix such that Z_1^* has full column rank and every variable in Z_1 can be expressed as a linear combination of variables in Z_1^*. (It is always possible to extract such a matrix. A method for doing so is given in section I.D.2.) Z_2^*, Z_3^*, and Z_4^* are constructed from Z_2, Z_3, and Z_4 in the same fashion. Following are some additional relationships which will be helpful in deriving computational formulas.

(1) If x is a variable in the matrix Z_3 or if x is in the space of Z_3 (i.e., x may be expressed as a linear combination of the columns of Z_3), then

$$[x]_{\|Z_3} = x \; ; \tag{I.38}$$

hence

$$[x]'_{\|Z_3} Z_2 = x'Z_2 \tag{I.39}$$

or if X_1 is a matrix of variables which are also contained in Z_3, then:

(I.40) $[X_1]'_{\|Z_3} Z_2 = X_1'Z_2$

and in particular,

(I.41) $[X_1'X_1]_{\|Z_3} = [X_1]'_{\|Z_3}[X_1]_{\|Z_3} = X_1'X_1$.

Also, (continuing to let X_1 be a submatrix of Z_3 or at least in the space of Z_3):

(I.42) $[X_1]_{\perp Z_3} = 0$ [where 0 is a T^X(number of variables in X_1) matrix]; therefore, the matrix of sums of cross-products of $[X_1]_{\perp Z_3}$ with any other matrix of variables is zero, i.e.,

(I.43) $[X_1]'_{\perp Z_3} Z_2 = 0'Z_2 = 0$

and, in particular,

(I.44) $[X_1'X_1]_{\perp Z_3} = [X_1]'_{\perp Z_3}[X_1]_{\perp Z_3} = 0$

and

(I.45) $[X_1'Z_2]_{\perp Z_3} = [X_1]'_{\perp Z_3}[Z_2]_{\perp Z_3} = 0'[Z_2]_{\perp Z_3} = 0$.

(2) (I.46) $[Z_1]_{\|Z_3} = Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}Z_1$

where $Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}$ is called the projection matrix for the space Z_3 .[1] (A matrix P is a projection matrix for a space Z_3 if $PZ_1 = [Z_1]_{\|Z_3}$ for any matrix of variables, Z_1 .)

---

[1] Z_3^* is defined on page 47.

(3) (I.47) $[Z_1]_{\perp Z_3} = Z_1 - [Z_1]_{\| Z_3} = Z_1 - Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}Z_1$

$= [I - Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}]Z_1$

where $I - Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^*$ is called the projection matrix for the space orthogonal to Z_3. (A matrix P is a projection matrix for the space orthogonal to Z_3 if $PZ_1 = [Z_1]_{\perp Z_3}$ for any matrix of variables, Z_1.)

(4) $Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}$ is symmetric and idempotent. (A matrix, P, is symmetric if $P = P'$ and idempotent if $PP = P$.) That $Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}$ is symmetric and idempotent is easily verified:[1]

(I.48) $(Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime})' = (Z_1^{*\prime})'(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime} = Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}$

and

(I.49) $[Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}][Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}] = Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}$,

since the part underlined in (I.49) is an identity matrix.

(5) $I - Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}$ is symmetric and idempotent. Again this is easily verified:

(I.50) $[I - Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}]' = I - Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}$

and

(I.51) $[I - Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}][I - Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}] =$

$I - Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime} - Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime} + Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}$

$= I - Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}$.

---

[1] For any three matrices A, B, and C of compatible dimensions, $(ABC)' = C'B'A'$. Also, for any matrix A, $(A')' = A$.

(6) The projection matrices $Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}$ and $I - Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}$ are mutually orthogonal, since

(I.52) $$[Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}][I - Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime}] = Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime} - Z_1^*(Z_1^{*\prime}Z_1^*)^{-1}Z_1^{*\prime} = 0 \quad .$$

(7) $[Z_1]'_{\|Z_3}[Z_2]_{\perp Z_3} = 0$ (where 0 is an $N_1 \times N_2$ matrix), since the variables in $[Z_1]_{\|Z_3}$ may be expressed as a linear combination of the variables in Z_3 and the variables in $[Z_2]_{\perp Z_3}$ are orthogonal to the variables in Z_3. In particular, from (I.46), (I.47), and (I.52) we have that:

(I.53) $$[Z_1]'_{\|Z_3}[Z_1]_{\perp Z_3} = Z_1'[Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}][I - Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}]Z_1 = Z_1'0Z_1 = 0 \quad .$$

(8) (I.54) $$Z_1'[Z_2]_{\|Z_3} = [Z_1]'_{\|Z_3}[Z_2]_{\|Z_3} = [Z_1'Z_2]_{\|Z_3} = [Z_1]'_{\|Z_3}Z_2 \quad .$$

This comes from the idempotency of $Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}$ [see (I.49)] as follows:

(I.55) $$[Z_1]'_{\|Z_3}[Z_2]_{\|Z_3} = [Z_1'Z_3^*(Z_3^{*\prime}Z^*{}_3)^{-1}Z_3^{*\prime}][Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}Z_2] = Z_1'Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}Z_2 = [Z_1]'_{\|Z_3}Z_2 \text{ or } Z_1'[Z_2]_{\|Z_3} \quad .$$

(9) Similarly,

(I.56) $$Z_1'[Z_2]_{\perp Z_3} = [Z_1]'_{\perp Z_3}[Z_2]_{\perp Z_3} = [Z_1'Z_2]_{\perp Z_3} = [Z_1]'_{\perp Z_3}Z_2 \quad .$$

This comes from the idempotency of $I - Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}$ [see (I.51)] as follows:

$$(I.57)\quad [Z_1'Z_2]_{\perp Z_3} = [Z_1]'_{\perp Z_3}[Z_2]_{\perp Z_3}$$
$$= Z_1'(I - Z_3^*(Z_3^{*'}Z_3^*)^{-1}Z_3^{*'})(I - Z_3^*(Z_3^{*'}Z^*{}_3)^{-1}Z_3^{*'})Z_2$$
$$= Z_1'(I - Z_3^*(Z_3^{*'}Z_3^*)^{-1}Z_3^{*'})Z_2 = [Z_1]'_{\perp Z_3}Z_2 \text{ or } Z_1'[Z_2]_{\perp Z_3} .$$

(10) Let A_1 be any $N_1 \times p_1$ matrix. Then using (I.46) we obtain:

$$(I.58)\quad [Z_1A_1]_{\parallel Z_3} = Z_3^*(Z_3^{*'}Z_3^*)^{-1}Z_3^{*'}Z_1A_1 = [Z_1]_{\parallel Z_3}A_1 \ ;$$

and using (I.47) we obtain:

$$(I.59)\quad [Z_1A_1]_{\perp Z_3} = (I - Z_3^*(Z_3^{*'}Z_3^*)^{-1}Z_3^{*'})Z_1A_1 = [Z_1]_{\perp Z_3}A_1 \ .$$

(11) Thus, letting A_1 be any $N_1 \times p_1$ matrix and A_2 be any $N_2 \times p_2$ matrix we obtain:

$$(I.60)\quad [A_1'Z_1'Z_2A_2]_{\parallel Z_3} = [Z_1A_1]'_{\parallel Z_3}[Z_2A_2]_{\parallel Z_3} = A_1'[Z_1]'_{\parallel Z_3}[Z_2]_{\parallel Z_3}A_2$$
$$= A_1'[Z_1'Z_2]_{\parallel Z_3}A_2$$

and

$$(I.61)\quad [A_1'Z_1'Z_2A_2]_{\perp Z_3} = [Z_1A_1]'_{\perp Z_3}[Z_2A_2]_{\perp Z_3} = A_1'[Z_1]'_{\perp Z_3}[Z_2]_{\perp Z_3}A_2$$
$$= A_1'[Z_1'Z_2]_{\perp Z_3}A_2 \ .$$

(12) Let A_1 be any $N_1 \times p_1$ matrix and A_2 be any $N_2 \times p_2$ matrix. Then from (I.58) we obtain:

$$(I.62)\quad [Z_1A_1 + Z_2A_2]_{\parallel Z_3} = \left\{[Z_1 \vdots Z_2]\begin{bmatrix}A_1\\A_2\end{bmatrix}\right\}_{\parallel Z_3} = [Z_1 \vdots Z_2]_{\parallel Z_3}\begin{bmatrix}A_1\\A_2\end{bmatrix}$$
$$= [Z_1]_{\parallel Z_3}A_1 + [Z_2]_{\parallel Z_3}A_2 \ .$$

Similarly from (I.59) we obtain:

$$(I.63)\quad [Z_1A_1 + Z_2A_2]_{\perp Z_3} = [Z_1]_{\perp Z_3}A_1 + [Z_2]_{\perp Z_3}A_2 \ .$$

(13) If $Z_3'Z_4 = 0$ (i.e., Z_3 is orthogonal to Z_4),

$$(I.64) \quad [Z_1]_{\perp}[Z_3 \vdots Z_4] = ([Z_1]_{\perp}Z_3)_{\perp}Z_4 = ([Z_1]_{\perp}Z_4)_{\perp}Z_3 \ .$$

That (I.64) holds for $Z_3'Z_4 = 0$ can be seen by writing out each of the terms and observing that they are the same:

$$[Z_1]_{\perp}[Z_3 \vdots Z_4] = Z_1 - [Z_3^* \vdots Z_4^*]\{[Z_3^* \vdots Z_4^*]'[Z_3^* \vdots Z_4^*]\}^{-1}[Z_3^* \vdots Z_4^*]'Z_1 \ .$$

However, for $Z_3'Z_4 = 0$:

$$\{[Z_3^* \vdots Z_4^*]'[Z_3^* \vdots Z_4^*]\}^{-1} = \begin{bmatrix} Z_3^{*\prime}Z_3^* & 0 \\ 0 & Z_4^{*\prime}Z_4^* \end{bmatrix}^{-1} = \begin{bmatrix} [Z_3^{*\prime}Z_3^*]^{-1} & 0 \\ 0 & [Z_4^{*\prime}Z_4^*]^{-1} \end{bmatrix} ;$$

hence, $[Z_1]_{\perp}[Z_3 \vdots Z_4]$ becomes:

$$Z_1 - [Z_3^* \vdots Z_4^*]\begin{bmatrix} [Z_3^{*\prime}Z_3^*]^{-1} & 0 \\ 0 & [Z_4^{*\prime}Z_4^*]^{-1} \end{bmatrix}\begin{bmatrix} Z_3^{*\prime}Z_1 \\ Z_4^{*\prime}Z_1 \end{bmatrix} =$$

$$Z_1 - Z_3^*[Z_3^{*\prime}Z_3^*]^{-1}Z_3^{*\prime}Z_1 - Z_4^*[Z_4^{*\prime}Z_4^*]^{-1}Z_4^{*\prime}Z_1 \ .$$

On the other hand $([Z_1]_{\perp}Z_3)_{\perp}Z_4 = [Z_1 - Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}Z_1]_{\perp}Z_4$

$$= [Z_1 - Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}Z_1] - Z_4^*(Z_4^{*\prime}Z_4^*)^{-1}Z_4^{*\prime}Z_1$$

$$+ Z_4^*(Z_4^{*\prime}Z_4^*)^{-1}Z_4^{*\prime}Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}Z_1$$

$$= Z_1 - Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}Z_1 - Z_4^*(Z_4^{*\prime}Z_4^*)^{-1}Z_4^{*\prime}Z_1 \quad \text{since} \quad Z_4^{*\prime}Z_3^* = 0.$$

Similarly $([Z_1]_{\perp}Z_4)_{\perp}Z_3$ becomes the same.

(14) (I.64) does not in general hold for $Z_3'Z_4 \neq 0$, however. Since $[Z_3 \vdots [Z_4]_{\perp}Z_3]$ spans the same space as $[Z_3 \vdots Z_4]$ and since $Z_3'[Z_4]_{\perp}Z_3 = 0$ [see (I.45)] applying (I.64) we have:

(I.65) $[z_1]_{\perp}[z_3:z_4] = [z_1]_{\perp(z_3:[z_4]_{\perp z_3})} = ([z_1]_{\perp z_3})_{\perp([z_4]_{\perp z_3})}$.

(15) $([z_1'z_2]_{\parallel z_3})_{\parallel z_4}$ is defined as $([z_1]_{\parallel z_3})'_{\parallel z_4}([z_2]_{\parallel z_3})_{\parallel z_4}$.

From (I.54) we obtain:

(I.66) $([z_1'z_2]_{\parallel z_3})_{\parallel z_4} = ([z_1]_{\parallel z_3})'_{\parallel z_4}([z_2]_{\parallel z_3})_{\parallel z_4}$

$= [z_1]'_{\parallel z_3}([z_2]_{\parallel z_3})_{\parallel z_4} = ([z_1]_{\parallel z_3})'_{\parallel z_4}[z_2]_{\parallel z_3}$.

(16) However:

(I.67) $([z_1'z_2]_{\parallel z_3})_{\parallel z_4} \neq (z_1'[z_2]_{\parallel z_3})_{\parallel z_4}$ nor $([z_1]'_{\parallel z_3}[z_2])_{\parallel z_4}$,

since

$([z_1'z_2]_{\parallel z_3})_{\parallel z_4} = z_1'z_3^*(z_3^{*\prime}z_3^*)^{-1}z_3^{*\prime}z_4^*(z_4^{*\prime}z_4^*)^{-1}z_4^{*\prime}z_3^*(z_3^{*\prime}z_3^*)^{-1}z_3^{*\prime}z_2$

and $(z_1'[z_2]_{\parallel z_3})_{\parallel z_4} = z_1'z_4^*(z_4^{*\prime}z_4^*)^{-1}z_4^{*\prime}z_3^*(z_3^{*\prime}z_3^*)^{-1}z_3^{*\prime}z_2$.

Thus, we can perform transformations of the form (I.54) on the outermost $\parallel$ operator, only, e.g., as in (I.66).

Additional examples showing permissible transformations based on the outermost operator only follow:

(I.68) $[z_1']'_{\parallel z_3}[z_2]_{\parallel z_4} = ([z_1]_{\parallel z_3})'_{\parallel z_4}[z_2]_{\parallel z_4} = ([z_1]'_{\parallel z_3}z_2)_{\parallel z_4}$

$\neq ([z_1'z_2]_{\parallel z_3})_{\parallel z_4}$.

(I.69) $[z_1]'_{\parallel z_3}[z_2]_{\parallel z_4} = [z_1]'_{\parallel z_3}([z_2]_{\parallel z_4})_{\parallel z_3} = (z_1'[z_2]_{\parallel z_4})_{\parallel z_3}$

$\neq ([z_1'z_2]_{\parallel z_4})_{\parallel z_3}$.

(17) Similarly (I.66) through (I.69) hold true if the $\parallel$ operator is replaced by the $\perp$ operator. In particular; the following may

be easily verified by writing out the matrices involved in that same manner as for (I.66) through (I.69):

$$(I.70)\quad ([Z_1'Z_2]_{\perp Z_3})_{\perp Z_4} = ([Z_1]_{\perp Z_3})'_{\perp Z_4}([Z_2]_{\perp Z_3})_{\perp Z_4} \quad .$$

$$(I.71)\quad ([Z_1'Z_2]_{\perp Z_3})_{\perp Z_4} = ([Z_1]_{\perp Z_3})'_{\perp Z_4}([Z_2]_{\perp Z_3})_{\perp Z_4}$$

$$= [Z_1]'_{\perp Z_3}([Z_2]_{\perp Z_3})_{\perp Z_4} = ([Z_1]_{\perp Z_3})'_{\perp Z_4}[Z_2]_{\perp Z_3} \quad .$$

$$(I.72)\quad ([Z_1'Z_2]_{\perp Z_3})_{\perp Z_4} \neq (Z_1'[Z_2]_{\perp Z_3})_{\perp Z_4} \text{ nor } ([Z_1]'_{\perp Z_3}[Z_2])_{\perp Z_4} \quad .$$

$$(I.73)\quad [Z_1]'_{\perp Z_3}[Z_2]_{\perp Z_4} = ([Z_1]_{\perp Z_3})'_{\perp Z_4}[Z_2]_{\perp Z_4} = ([Z_1]'_{\perp Z_3}Z_2)_{\perp Z_4}$$

$$\neq ([Z_1'Z_2]_{\perp Z_3})_{\perp Z_4} \quad .$$

$$(I.74)\quad [Z_1]'_{\perp Z_3}[Z_2]_{\perp Z_4} = [Z_1]'_{\perp Z_3}([Z_2]_{\perp Z_4})_{\perp Z_3} = (Z_1'[Z_2]_{\perp Z_4})_{\perp Z_3}$$

$$\neq ([Z_1'Z_2]_{\perp Z_4})_{\perp Z_3} \quad .$$

(18) The direct relationship $[Z_1'Z_2] = [Z_1'Z_2]_{\parallel Z_3} + [Z_1'Z_2]_{\perp Z_3}$ holds for the outermost $\parallel$ and $\perp$ operators, only. For example, the following may be easily verified by writing out the matrices involved in the same manner as for (I.66) through (I.69):

$$(I.75)\quad ([Z_1'Z_2]_{\parallel Z_3})_{\parallel Z_4} + ([Z_1'Z_2]_{\parallel Z_3})_{\perp Z_4} = [Z_1'Z_2]_{\parallel Z_3}$$

$$(I.76)\quad ([Z_1'Z_2]_{\perp Z_3})_{\parallel Z_4} + ([Z_1'Z_2]_{\perp Z_3})_{\perp Z_4} = [Z_1'Z_2]_{\perp Z_3} \quad ;$$

however:

$$(I.77)\quad ([Z_1'Z_2]_{\parallel Z_3})_{\parallel Z_4} + ([Z_1'Z_2]_{\perp Z_3})_{\parallel Z_4} \neq [Z_1'Z_2]_{\parallel Z_4} \quad .$$

$$(I.78)\quad ([Z_1'Z_2]_{\parallel Z_3})_{\perp Z_4} + ([Z_1'Z_2]_{\perp Z_3})_{\perp Z_4} \neq [Z_1'Z_2]_{\perp Z_4} \quad .$$

2. Computation of matrices of the form $[Z_1'Z_2]_{\perp Z_3}$ and $[Z_1'Z_2]_{\parallel Z_3}$ by direct orthogonalization[1]

The procedure given in this section is very general in that it may be used to calculate matrices of the form $[Z_1'Z_2]_{\perp Z_3}$ in which:

(1) Z_1, Z_2, and Z_3 contain jointly dependent or predetermined variables, or both.

(2) Variables in any of the matrices may also occur in the other two as well. (If z_1 is contained in both Z_1 and Z_3, the row of $[Z_1'Z_2]_{\perp Z_3}$ corresponding to z_1 will be zero at the completion of the orthogonalization. Similarily if z_2 is contained in both Z_2 and Z_3, the column of $[Z_1'Z_2]_{\perp Z_3}$ corresponding to z_2 will be zero at the completion of the orthogonalization.)

(3) Z_1, Z_2, and Z_3 may have less than full column rank.

In this paper, Z_1 and Z_2 will most commonly be Y, the matrix of jointly dependent variables in a system of equations; Y_A, the matrix of jointly dependent variables in a subsystem of the equations; or $_+Y_\mu$, the matrix of jointly dependent variables in a single equation. Z_3 will most commonly be X_I, a matrix of instrumental variables; X_μ, the matrix of predetermined variables

---

[1]The orthogonalization method outlined here is very well known among mathematicians and statisticians; however, oddly enough the writer has never seen reference to its use in the field of econometrics for which it would seem to have considerable application.

in a single equation; or X, the matrix of predetermined variables in the system. Thus, the orthogonalization procedure outlined in this section will be most commonly used to calculate matrices of the form $[Y'Y]_{\perp X_I}$, $[Y_A'Y_A]_{\perp X_I}$, $[{}_+Y_\mu' \; {}_+Y_\mu]_{\perp X_I}$, and $[{}_+Y_\mu' \; {}_+Y_\mu]_{\perp X_\mu}$.

Matrices of the form $[Z_1'Z_2]_{\parallel Z_3}$ are calculated as $[Z_1'Z_2] - [Z_1'Z_2]_{\perp Z_3}$.

Rather than use the more common formula $[Z_1'Z_2]_{\perp Z_3} = Z_1'Z_2 - Z_1'Z_3(Z_3'Z_3)^{-1}Z_3'Z_2$, we will calculate $[Z_1'Z_2]_{\perp Z_3}$ by direct orthogonalization thereby eliminating the requirement that $(Z_3'Z_3)^{-1}$ exists, i.e., thereby permitting calculation of $[Z_1'Z_2]_{\perp Z_3}$ when Z_3 has less than full column rank. Even if Z_3 has full column rank, calculation of $[Z_1'Z_2]_{\perp Z_3}$ by direct orthogonalization is advantageous from the fact that

(1) fewer computer locations may be conveniently used to compute $[Z_1'Z_2]_{\perp Z_3}$,

(2) fewer arithmetic operations are required thereby saving computer time,

(3) $[Z_1'Z_2]_{\perp Z_3}$ may be computed to a higher degree of accuracy, and

(4) rk Z_3 is calculated as a byproduct of the computational procedure.

A computational procedure for calculating $[Z_1'Z_2]_{\perp Z_3}$ by direct orthogonalization follows:[1]

(1) Let Z be a $T \times N$ matrix containing all of the variables which occur in either Z_1 or Z_2. If desired Z could be

---

[1] A verification that the computational procedures produces the correct matrix follows the presentation of the computational procedure.

defined as $Z = [Z_1 \vdots Z_2]$; however, there is no need to repeat variables common to both Z_1 and Z_2.[1] If $Z_1 = Z_2$, then $Z = Z_1 = Z_2$. Z may contain variables in addition to those in Z_1 and Z_2 if desired.

Calculate the moment matrix (sums of squares and cross products matrix) of $[Z_3 \vdots Z]$, i.e., calculate:

$$(I.79) \qquad [Z_3 \vdots Z]'[Z_3 \vdots Z] = \begin{bmatrix} Z_3'Z_3 & Z_3'Z \\ N_3{}^{X}N_3 & N_3{}^{X}N \\ Z'Z_3 & Z'Z \\ N^{X}N_3 & N^{X}N \end{bmatrix} .$$

(2) Do elementary row operations on the rows of the matrix until the first N_3 columns are reduced to zeros below the diagonal. (It doesn't matter whether the diagonal elements are set to 1 or not.) This is equivalent to starting a forward solution of the Doolittle inversion procedure but stopping after the $N_3{}^{th}$ row. The above matrix (I.79) will have become

$$(I.80) \qquad \begin{bmatrix} A_{11} & A_{12} \\ N_3{}^{X}N_3 & N_3{}^{X}N \\ 0 & [Z'Z]_{\perp Z_3} \\ N^{X}N_3 & N^{X}N \end{bmatrix}$$

where A_{11} contains zeros below the diagonal and the results

---

[1] Repeating variables in the Z matrix causes no computational difficulty.

of the elementary row operations on and above the diagonal. A_{12} merely contains the results of the elementary row operations.

(3) $[Z_1'Z_2]_{\perp Z_3}$ is a submatrix occupying the same elements of $[Z'Z]_{\perp Z_3}$ as $Z_1'Z_2$ occupied of $Z'Z$.[1]

To increase accuracy, it is advisable to rearrange the first N_3 rows and columns at each step so that the largest diagonal element from among the remaining diagonal elements of the $Z_3'Z_3$ matrix is used as the pivot at each step. (This will not affect a row or a column of $[Z'Z]_{\perp Z_3}$; therefore, there is no requirement that track be kept of which rows and columns are switched. On the other hand, the information as to which diagonal elements have served as pivots can be used to derive a minimum subset of predetermined variables spanning the space of the columns of Z_3.) If the largest diagonal element becomes smaller than a preset or precalculated value, $\epsilon > 0$, the procedure is stopped, since all of the values to be reduced to zero will already be within ϵ of zero and $[Z'Z]_{\perp Z_3}$ will already be the moment matrix of the part of Z orthogonal to Z_3. The number of columns of Z_3 already operated on before the largest remaining diagonal element became less than ϵ is the rank of Z_3. The predetermined variables corresponding to the diagonal elements used

---

[1] It is noted further on that only the triangular part of $[Z'Z]$ need be formed and operated on; hence, $[Z_1'Z_2]_{\perp Z_3}$ is extracted from a triangular matrix representing the symmetric matrix $[Z'Z]_{\perp Z_3}$.

as pivots constitute a basis spanning the same space as the columns of Z_3. Each of the remaining predetermined variables may be expressed as a linear combination of this set of rk Z_3 variables.[1]

Since (1) no use will be made of the matrices A_{11} and

---

[1]The procedure outlined above also provides the starting point of a procedure for getting a set of least squares coefficients for one or more equations even if the matrix of independent variables has less than full column rank. Let $Z = Z_3 \Pi' + V$ be a set of equations for which a set of least squares coefficients is desired. Assume that rk $Z_3 = N_3^*$. Then at the completion of the orthogonalization procedure outlined above [i.e., (I.80)], N_3^* diagonal elements will have been selected and the remaining diagonal elements will have become less than ϵ. The calculation of a set of least squares coefficients is completed by:

(1) Setting the last $N_3 - N_3^*$ rows of A_{12} to zero. (These elements will already be approximately zero, but they will not, in general, be exactly zero due to rounding error.)

(2) Dividing each of the elements on or above the diagonals of the first N_3^* rows of $[A_{11} \vdots A_{12}]$ by the diagonal elements for the row. (The diagonal elements of the first N_3^* rows will then be 1.)

(3) Performing a back solution in the usual Doolittle manner, i.e., by reducing all elements above the diagonal elements of the first N_3^* columns to 0.

(4) Rearranging the first N_3 rows into their original order (in terms of the Z_3 matrix). A set of least squares coefficients ($\hat{\Pi}'$) is then given by the $N_3 \times N$ matrix in the position originally occupied by $Z_3'Z$ in (I.79). Estimated values of Z (i.e., $[Z]_{\| Z_3}$) and residuals for Z (i.e., $[Z]_{\perp Z_2}$) calculated through use of the $\hat{\Pi}'$ matrix calculated in this manner will be the same as the estimated values of Z and residuals calculated through use of any of the many possible sets of least squares coefficients. (Even though the least squares coefficients are not unique, the estimated values of the dependent variable and the residuals are unique--the same estimated values being obtained from any set of least squares coefficients.)

A_{12}, (2) the initial matrix is symmetric, and (3) the $[Z'Z]_{\perp Z_3}$ matrix is symmetric, all elements on one side of the diagonal need not be formed or operated on; that is, only a triangular matrix need be formed and all operations may be performed on this triangular matrix thereby saving computer memory.

Verification That The Computational Procedure Produces $[Z'Z]_{\perp Z_3}$[1]

That the matrix labeled $[Z'Z]_{\perp Z_3}$ is indeed the matrix $Z'Z - Z'Z_3(Z_3'Z_3)^{-1}Z_3'Z$ if Z_3 has full column rank [hence, $(Z_3'Z_3)^{-1}$ exists] can be readily demonstrated as follows:

Performing the above elementary row operations is equivalent to premultiplying (I.79) by a nonsingular matrix $\begin{bmatrix} E_{11} & 0 \\ E_{21} & I \end{bmatrix}$ such that:

$$\begin{bmatrix} E_{11} & 0 \\ E_{21} & I \end{bmatrix}\begin{bmatrix} Z_3'Z_3 & Z_3'Z \\ Z'Z_3 & Z'Z \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} .$$

Thus:

$$E_{21}Z_3'Z_3 + Z'Z_3 = 0 \quad \text{or} \quad E_{21} = -Z'Z_3(Z_3'Z_3)^{-1}$$

and:

$$E_{21}Z_3'Z + Z'Z = A_{22} .$$

Substituting for E_{21} into the last equation we get:

$$-Z'Z_3(Z_3'Z_3)^{-1}Z_3'Z + Z'Z = A_{22} \quad \text{or}$$

$$A_{22} = Z'Z - Z'Z_3(Z_3'Z_3)^{-1}Z_3'Z = [Z'Z]_{\perp Z_3} . \tag{I.81}$$

---

[1]The proof for Z_3 of full column rank was suggested by Professor Robert L. Gustafson.

The same set of manipulations may also be used to show that the matrix labeled $[Z'Z]_{\perp Z_3}$ is indeed that matrix even in the case of a Z_3 having rank $N_3^* < N_3$ (i.e., in the case of a Z_3 having less than full column rank). If rk $Z_3 = N_3^*$, row operations are performed on the columns corresponding to N_3^* of the variables in Z_3 before the diagonal elements corresponding to the remaining variables become less than ϵ (for a suitable choice of ϵ). The orthogonalization stops at this point. This is equivalent to performing row operations on the following submatrix of (I.79) (letting Z_3^* be a submatrix of Z_3 containing the variables corresponding to the N_3^* diagonal elements used as pivots):

$$\begin{bmatrix} Z_3^{*\prime}Z_3^* & Z_3^{*\prime}Z \\ N_3^{*\times}N_3^* & N_3^{*\times}N \\ \\ Z'Z_3^* & Z'Z \\ N^{\times}N_3^* & N^{\times}N \end{bmatrix} . \tag{I.82}$$

The same derivation may now be performed on (I.82) as was performed with (I.80)--the only difference in the intermediate matrices obtained is that Z_3^* will occur in place of Z_3 wherever Z_3 presently occurs. Thus (I.81) becomes:

$$A_{22} = Z'Z - Z'Z_3^*(Z_3^{*\prime}Z_3^*)^{-1}Z_3^{*\prime}Z . \tag{I.83}$$

But this is $[Z'Z]_{\perp Z_3}$ [see (I.57)]. Thus, the desired matrix is obtained even in the case of Z_3 having less than full column rank.

CHAPTER II

COEFFICIENT ESTIMATION

A. Basic Double k-class Model and Summary of Methods

The basic single equation procedure presented in this paper is the double k-class model--the model developed in this chapter.[1] Variance-covariance formulas for the double k-class model are given in Chapter III and a method for directly imposing restrictions on direct least squares and two-stage least squares coefficients are given in Chapter IV.

In this chapter, the double k-class computational formula for the coefficients of an equation are first given, some matrices are defined and a derivation of the double k-class formula is presented. Specific members of the double k-class model such as direct least squares (DLS), two-stage least squares (2SLS), and limited information single equation maximum likelihood (LIML) are summarized and then presented in more detail. Variations of the double k-class model including instrumental variables techniques are presented. Finally a discussion of selection of instruments--especially the maximum number of instruments which will be effective--is presented.

The μ^{th} equation in a system of the equations was defined in (I.8). If we drop the subscript μ from many of the matrices,

---

[1] The double k-class model is given in Nagar [1962] and Theil [1961], pp. 354.

we may write the equation as:

$$(II.1) \qquad \underset{T\times 1}{y} = \underset{T\times m}{Y}\ \underset{m\times 1}{\gamma} + \underset{T\times \ell}{X_{\mu}}\ \underset{\ell\times 1}{\beta} + \underset{T\times 1}{u} \ .$$

The computational formula for double k-class estimated coefficients can be written as:[1]

---

[1]Of the double k-class estimators discussed, only in the case of LIML does it make no substantive difference in the estimated coefficients which jointly dependent variable is chosen as the normalizing variable. For the remaining procedures, a change in the selected normalizing variable will change the resulting coefficients by more than just a trivial division of all coefficients by the negative of the coefficient of the variable chosen as the normalizing variable. Regarding this effect, Fisher [1965], p. 604 states:

> "It can be argued that limited-information maximum likelihood has the desirable property of treating all included endogenous variables in an equation symmetrically; indeed, Chow has shown that it is a natural generalization of ordinary least squares in the absence of a theoretically given normalization rule" [footnote reference deleted].
>
> "On the other hand, such an argument seems rather weak, since normalization rules are in fact generally present in practice, each equation of the model being naturally associated with that particular endogenous variable which is determined by the decision-makers whose behavior is represented by the equation. The normalization rules are in a real sense part of the specification of the model, and the model is not completely specified unless every endogenous variable appears (at least implicitly) in exactly one equation in normalized form. For example, it is not enough to have price equating supply and demand, equations should also be present which explain pure quotations by sellers and buyers and which describe the equilibrating process. (For most purposes, of course, such additional equations can remain in the back of the model builder's mind, although the rules for choosing instrumental variables given below may sometimes require that they be made explicit.)
>
> "Thus, symmetry may be positively undesirable in a well-specified model where one feels relatively certain as to appropriate normalization, although it may be desirable if one wishes to remain agnostic as to appropriate normalization."

$$(II.2)\quad \hat{\delta}_{k_1,k_2} = \begin{bmatrix} \hat{\gamma} \\ \hat{\beta} \end{bmatrix}_{k_1,k_2} = \begin{bmatrix} Y'Y-k_1[Y'Y]_{\perp X_I} & Y'X_\mu \\ X_\mu'Y & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y'y-k_2[Y'y]_{\perp X_I} \\ X_\mu'y \end{bmatrix}$$

where k_1 and k_2 are scalars which determine the particular double k-class member.

X_I is the matrix of instrumental variables used to adjust the jointly dependent variables in the equation. X_I includes all of the predetermined variables in the equation plus additional instruments--the additional instruments usually being (but not restricted to being) all or some of the additional predetermined variables in the system.[1]

$[Y'Y]_{\perp X_I}$ is the moment matrix of the part of Y orthogonal

---

[1]The basic double k-class method is usually given with the entire matrix of predetermined variables in the system, X, being used as the matrix of instruments, X_I; however, in practice some of the predetermined variables in the system are often omitted from the X matrix, predetermined variables are sometimes linearly combined (e.g., by use of principal components), etc. Rather than use the X matrix in our notation and then continually point out variations, it seems more fruitful to merely designate the matrix of variables used to adjust the jointly dependent variables as a matrix of instruments.

Since the particular instruments used to adjust the jointly dependent variables have a considerable effect on the coefficients obtained it is imperative that the particular instruments used be listed when reporting results. We will return to the problem of selecting instruments in section II.G.

to X_I.[1] Since $[Y'y]_{\perp X_I} = Y'_{\perp X_I} y_{\perp X_I} = Y'_{\perp X_I} y$ [see (I.56)], $[Y'y]_{\perp X_I}$ may be regarded as either (1) the vector of sums of cross-products of the part of Y orthogonal to X_I with the part of y orthogonal to X_I or (2) the vector of sums of cross-products of the part of Y orthogonal to X_I with y. Thus, computationally, X_I may be regarded as the matrix of instruments used to adjust Y or the matrix of instruments used to adjust $[y \vdots Y]$. In light of the equivalence of $[Y'y]_{\perp X_I}$ and $[Y]'_{\perp X_I} y$, (II.2) may also be written as:

$$\text{(II.3)} \qquad \hat{\delta}_{k_1,k_2} = \begin{bmatrix} Y'Y - k_1[Y'Y]_{\perp X_I} & Y'X_\mu \\ X'_\mu Y & X'_\mu X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y' - k_2 Y'_{\perp X_I} \\ X'_\mu \end{bmatrix} y \quad ;$$

however, for actual computations (II.2) should be used since $[Y'y]_{\perp X_I}$ may be computed by direct orthogonalization in the manner indicated in section I.D.2.

A tentative proof that $\hat{\delta}_{k_1,k_2}$ is a consistent estimator of δ given the statistical assumptions of section I.C.3 and the assumption $\text{plim}(k_1 - 1) = \text{plim}(k_2 - 1) = 0$ is given in appendix C.[2]

---

[1] The orthogonalization notation used here and the concept of orthogonalization are discussed in section I.D. A method to compute $[Y'Y]_{\perp X_I}$ and $[Y'y]_{\perp X_I}$ by direct orthogonalization is given in section I.D.2. During the calculation of $[Y'Y]_{\perp X_I}$ and $[Y'y]_{\perp X_I}$ it is recommended that rk X_μ be calculated and, if LIML coefficients are to be calculated, that the $[{}_+Y'{}_+Y]_{\perp X_\mu}$ matrix be saved. A method for calculating rk X_μ and $[{}_+Y'{}_+Y]_{\perp X_\mu}$ as an intermediate step in the calculation of $[Y'Y]_{\perp X_I}$ and $[Y'y]_{\perp X_I}$ is given in appendix A.

[2] Members of the double k-class family for which $\text{plim}(k - 1) = \text{plim}(k_2 - 1) = 0$ include 2SLS, LIML, Nagar's unbiased to $O(T^{-1})$ in probability estimator (UBK), and Nagar's minimum second moment estimator (MSM). $\text{plim}(k_1 - 1) = \text{plim}(k_2 - 1) = -1$ for DLS; hence, DLS is not shown to be consistent.

(II.2) is not the most general form of the double k-class formula, since it may be desired that different sets of instrumental variables be used in theadjustment of the different jointly dependent variables. Let $Y = [y_1 \ \cdots \ y_m]$ and let X_I^1 denote the set of instruments used to adjust y_1, X_I^2 denote the set of instruments used to adjust y_2, ... , X_I^m denote the set of instruments used to adjust y_m, and X_I^y denote the set of instruments used to adjust the normalizing jointly dependent variable, y. (X_I^y may be null, i.e., it may be desired that y not be adjusted.) Then, the following matrix could be used in (II.2) in place of the $[Y'Y]_{\perp X_I}$ matrix:

$$(II.4)\quad \begin{bmatrix} [y_1'y_1]_{\perp X_I^1} & [y_1]'_{\perp X_I^1}[y_2]_{\perp X_I^2} & \cdots & [y_m]'_{\perp X_I^1}[y_m]_{\perp X_I^m} \\ [y_2]'_{\perp X_I^2}[y_1]_{\perp X_I^1} & [y_2'y_2]_{\perp X_I^2} & \cdots & [y_m]'_{\perp X_I^2}[y_m]_{\perp X_I^m} \\ \vdots & \vdots & & \vdots \\ [y_m]'_{\perp X_I^m}[y_1]_{\perp X_I^1} & [y_m]'_{\perp X_I^m}[y_2]_{\perp X_I^2} & \cdots & [y_m'y_m]_{\perp X_I^m} \end{bmatrix}$$

and the following vector could be used in place of the $[Y'y]_{\perp X_I}$ vector:

$$(II.5)\quad \begin{bmatrix} [y_1]'_{\perp X_I^1}[y]_{\perp X_I^y} \\ [y_2]'_{\perp X_I^2}[y]_{\perp X_I^y} \\ \vdots \\ [y_m]'_{\perp X_I^m}[y]_{\perp X_I^y} \end{bmatrix}$$

A method of computing (II.4) and (II.5) by direct orthogonalization is given in appendix B.

The suggestion that separate instruments be used for the adjustment of each jointly dependent variable for 2SLS is made by Franklin Fisher.[1] He points out that such a procedure introduces certain problems of inconsistency depending on the assumptions of the model[2]; however, he then argues effectively as to why the inconsistency introduced is likely to be low.

We will not consider properties of the double k-class estimator if separate instruments are used in the adjustment of each jointly dependent variable. We merely note that if it is desired that a separate set of instruments be used in the adjustment of each jointly dependent variable, a computational method for doing this is given in appendix B. In section II.G we will return to the problem of selecting instruments and note that due to the use of direct orthogonalization, it may be feasible to merely note the separate instruments which the researcher would prefer to use in the adjustment of each jointly dependent variable and then use all of these instruments in the X_I matrix; that is, adjust all of the jointly dependent variables in the equation by the same set of instruments. If this approach is taken, computational form (II.2) is used.

If 2SLS is the basic computational method used, it is often suggested that asymptotic efficiency will be increased

---

[1] Fisher [1965], pp. 602-603, 625-633. Fisher points out (p. 603) that "if different predetermined variables are used in replacing each included endogenous variable (as suggested below), it is not clear how limited-information maximum likelihood carried over to such cases."

[2] Fisher [1965], p. 631.

in the estimation of δ if each explanatory jointly dependent variable is adjusted only by the predetermined variables in the reduced form equation in which this jointly dependent variable occurs rather than adjusting all explanatory jointly dependent variables by the entire set of predetermined variables in the system.[1] (Assuming that all predetermined variables do not occur with non-zero coefficients in each of the reduced form equation being estimated) this leads, of course, to the use of (II.4) in place of $[Y'Y]_{\perp X_I}$ and (II.5) in place of $[Y'y]_{\perp X_I}$ (except that y is usually not adjusted).

Derivation of (II.2)[2]

If the coefficients of (II.1) are estimated by DLS, due to the fact that some of the explanatory variables are contemporaneously correlated with the disturbance of the equation, the DLS coefficients do not possess even the property of consistency.[3]

To take account of the occurrence of explanatory variables which are contemporaneously correlated with the disturbance, let us first rewrite the equation in an adjusted form and then apply

---

[1] Predetermined variables will tend to occur with readily recognized zero coefficients in reduced form equation in systems which are recursive with respect to the coefficients. 2SLS (modified by II.4 and II.5) is still more appropriate than recursive system estimation methods if it is not assumed that the off-diagonal elements of the disturbance variance-covariance matrix are zero.

[2] The derivation in this section parallels the derivation of the single k-class method by Chow [1964], pp. 546-548.

[3] Except possibly in special recursive models. See Fisher [1965], pp. 592, 593.

DLS to the adjusted equation. As a first step let us divide the variation of each variable in Y and y into two parts--that part of the variation which is in the space spanned by the instruments and that part which is orthogonal to the instruments, i.e., we will divide Y and y into:

$$Y = Y_{\|X_I} + Y_{\perp X_I} \tag{II.6}$$

$$y = y_{\|X_I} + y_{\perp X_I} \quad . \tag{II.7}$$

Since the instruments are assumed to be asymptotically uncorrelated with the disturbances of all equations, the columns of $Y_{\|X_I}$ and $y_{\|X_I}$ which lie in the space spanned by X_I are also asymptotically uncorrelated with the disturbances of all equations, and in particular asymptotically uncorrelated with the disturbance of the μ^{th} equation, u. To adjust Y and y for the part asymptotically correlated with the disturbance, let us subtract a constant, g_1, times $Y_{\perp X_I}$ from Y and (recognizing the special role played by the normalizing variable) subtract another constant, g_2, times $y_{\perp X_I}$ from y. If $g_1 Y_{\perp X_I}$ and $g_2 y_{\perp X_I}$ are subtracted from Y and y respectively, equation (II.1) becomes:[1]

---

[1]It might be argued that it is unnecessary to adjust y, the normalizing jointly dependent variable. If y is not adjusted, one obtains Theil's h-class which contain DLS and 2SLS as particular cases. Some readers may consider the adjustment of y more justified if this adjustment is regarded as the result of a two step process as follows: First, $g_1 Y_{\perp X_I}$ is subtracted from Y giving us:

$$y \equiv [Y - g_1 Y_{\perp X_I}]\gamma + X_\mu \beta + [u + g_1 Y_{\perp X_I}\gamma]$$

and then $g_2 y_{\perp X_I}$ is subtracted from the disturbance to (possibly) make the resulting disturbance more homogeneous or (possibly) to reduce the asymptotic correlation between the disturbance and the explanatory variables, $[Y - g_1 Y_{\perp X_I} \vdots X_\mu]$, in the equation. If

$$(II.8)\quad [y - g_2 y_{\perp X_I}] = [Y - g_1 Y_{\perp X_I}]\gamma + X_\mu \beta + [u + g_1 Y_{\perp X_I}\gamma - g_2 y_{\perp X_I}]$$

or

$$(II.9)\quad [y - g_2 y_{\perp X_I}] = [Y - g_1 Y_{\perp X_I} \vdots X_\mu]\delta + [u + g_1 Y_{\perp X_I}\gamma - g_2 y_{\perp X_I}] \quad .$$

The double k-class estimator of δ in (II.9) may be written as:

$$(II.10)\quad \hat{\delta}_{k_1,k_2} =$$

$$\{[Y - g_1 Y_{\perp X_I} \vdots X_\mu]'[Y - g_1 Y_{\perp X_I} \vdots X_\mu]\}^{-1}[Y - g_1 Y_{\perp X_I} \vdots X_\mu]'[y - g_2 y_{\perp X_I}]$$

or

$$(II.11)\quad \hat{\delta}_{k_1,k_2} =$$

$$\begin{bmatrix} [Y-g_1 Y_{\perp X_I}]'[Y-g_1 Y_{\perp X_I}] & [Y-g_1 Y_{\perp X_I}]'X_\mu \\ X_\mu'[Y-g_1 Y_{\perp X_I}] & X_\mu' X_\mu \end{bmatrix}^{-1} \begin{bmatrix} [Y-g_1 Y_{\perp X_I}]'[y-g_2 y_{\perp X_I}] \\ X_\mu'[y-g_2 y_{\perp X_I}] \end{bmatrix} .$$

However, $[Y - g_1 Y_{\perp X_I}]'X_\mu = Y'X_\mu - g_1 Y'_{\perp X_I} X_\mu = Y'X_\mu - g_1 0 = Y'X_\mu$ and similarly $X_\mu'[y - g_2 Y_{\perp X}] = X_\mu' y$.[1] Also, since $Y'_{\perp X_I} Y = Y'Y_{\perp X_I} = [Y'Y]_{\perp X_I}$ and $Y'_{\perp X_I} y = Y'y_{\perp X_I} = [Y'y]_{\perp X_I}$ [see (I.56)], $[Y - g_1 Y_{\perp X_I}]'[Y - g_1 y_{\perp X_I}] = Y'Y - g_1 Y'Y_{\perp X_I} - g_1 Y'_{\perp X_I} Y + g_1^2 [Y'Y]_{\perp X_I} = Y - (2g_1 - g_1^2)[Y'Y]_{\perp X_I}$, and (II.11) may be rewritten as:

---

this approach is used, we obtain: $y = [Y-g_1 Y_{\perp X_I}]\gamma + X_\mu \beta + g_2 y_{\perp X_I} + [u + g_1 Y_{\perp X_I}\gamma - g_2 y_{\perp X_I}]$ which may be rewritten as (II.8).

If it is felt that the same adjustment should be made to all jointly dependent variables, then $g_1 = g_2$ and the single k-class estimators are obtained, of which LIML is an example.

[1] $Y'_{\perp X_I} X_\mu = 0$ since X_μ is in the space spanned by X_I and $Y_{\perp X_I}$ is orthogonal to X_I. See (I.38) and (I.53).

$$(II.12)\quad \hat{\delta}_{k_1,k_2} =$$

$$\begin{bmatrix} Y'Y-(2g_1-g_1^2)[Y'Y]_{\perp X_I} & Y'X_\mu \\ X_\mu'Y & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y'y-(g_1+g_2-g_1g_2)[Y'y]_{\perp X_I} \\ X_\mu'y \end{bmatrix} .$$

The basic estimating formula (II.2) is obtained from (II.12) by letting $k_1 = 2g_1 - g_1^2$ and $k_2 = g_1 + g_2 - g_1g_2$.

Summary of Methods.

If y is left unadjusted (i.e., g_2 is set to zero) and if g_1 is set to $1 - h$, Theil's h-class model is obtained as a particular case of the above double k-class model, since k_1 becomes $2g_1 - g_1^2 = 2(1 - h) - (1 - h)^2 = 1 - h^2$ and k_2 becomes $g_1 + g_2 + g_1g_2 = 1 - h$. Theil's h-class formula may be written:[1]

$$(II.13)\quad \hat{\delta}_h = \begin{bmatrix} Y'Y-(1-h^2)[Y'Y]_{\perp X_I} & Y'X_\mu \\ X_\mu'Y & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y'y-(1-h)[Y'y]_{\perp X_I} \\ X_\mu'y \end{bmatrix} .$$

As with the formula for any double k-class member, y may be substituted for $y_{\perp X_I}$ in the right hand side vector, since $[Y'y]_{\perp X_I} = Y'_{\perp X_I} y$.[2]

If the same basic adjustment is made to all jointly dependent variables (i.e., if g_1 and g_2 are restricted to the

---

[1] The h-class model may be found in Theil [1961], pp. 353-354 and Nagar [1962], p. 171.

[2] $[Y'y]_{\perp X_I} = Y'_{\perp X_I} y_{\perp X_I} = Y'_{\perp X_I}[y - y_{\perp X_I}] = Y'_{\perp X_I} y - Y'_{\perp X_I} y_{\perp X_I} = Y'_{\perp X_I} y - 0 = Y'_{\perp X_I} y$.

same scalar value g), the single k-class model with $k = 2g - g^2$ is obtained. The single k-class formula may be written:[1]

$$(II.14) \quad \hat{\delta}_k = \begin{bmatrix} Y'Y-k[Y'Y]_{\perp X_I} & Y'X_\mu \\ X_\mu'Y & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y'y-k[Y'y]_{\perp X_I} \\ X_\mu'y \end{bmatrix}$$

If both g_1 and g_2 are set to zero, the DLS estimation procedure with $k_1 = k_2 = 0$ is obtained. DLS may be regarded as either a k-class or an h-class member.

If g_1 is set to 1, even though g_2 is taken as any value we have $k_1 = 2g_1 - g_1^2 = 1$ and $k_2 = g_1 + g_2 - g_1g_2 = 1 + g_2 - g_2 = 1$; therefore, setting g_1 to 1 automatically gives $k_1 = k_2 = 1$--the 2SLS estimator. Thus, 2SLS may be regarded as the single k-class member with $k = 1$ or the h-class member with $h = 0$.

The LIML estimator is the particular single k-class member in which k equals the smallest eigenvalue (characteristic root) of the matrix $[Y'Y]_{\perp X_I}^{-1}[Y'Y]_{\perp X_\mu}$.[2] This eigenvalue will be greater than 1 except for the particular case $\text{rk } X_I = n$,[3] in which case the eigenvalue will be 1. An eigenvalue greater than 1 implies a g which must be expressed as a complex number containing an imaginary part. The eigenvalue itself will, of course, be real.

---

[1]The k-class model may be found in Theil [1961], pp. 231-237.

[2]The LIML estimator is discussed in greater detail in section II.C.1.

[3]$n = m + \ell$ is the number of "explanatory" variables in the equation.

Two other particular single k-class members which will be considered further on are members suggested by Nagar which we will refer to as unbiased to $0(T^{-1})$ in probability k (UBK) and minimum second moment (MSM).

B. Methods Which are Both h-class and Single k-class

1. Direct least squares (DLS)

As indicated earlier, direct least squares (DLS) coefficients may be obtained by setting g_1 and g_2 to 0 which implies a k of 0 and an h of 1. (II.14) becomes:

$$\text{(II.15)} \qquad \hat{\delta}_{DLS} = \begin{bmatrix} Y'Y & Y'X_\mu \\ X_\mu'Y & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y'y \\ X_\mu'y \end{bmatrix} .$$

In estimating by DLS, all of the jointly dependent variables except the normalizing variable are treated the same as the predetermined variables. If more than one jointly dependent variable occurs in the equation, the DLS estimated coefficients are not even consistent, since jointly dependent variables which are not even asymptotically uncorrelated with the disturbance are used as independent variables. Even in this case, however, DLS coefficients have some desirable properties such as small dispersion of coefficients about their expected values and finite sample coefficient variance-covariance matrices.[1]

---

[1] See Fisher [1965], pp. 591-592, 604-605.

2. Two-stage least squares (2SLS)[1]

As indicated earlier, if g_1 equals 1, g_2 may be anything without changing the resulting coefficients which we call the two-stage least squares (2SLS) coefficients. If g_2 is considered to be 0, 2SLS may be considered the h-class member with $h = 0$. If g_2 is considered to be 1, 2SLS may be considered the k-class member with $k = 1$. The 2SLS estimating formula becomes:[2]

$$\text{(II.16)} \qquad \hat{\delta}_{2SLS} = \begin{bmatrix} Y'Y - [Y'Y]_{\perp X_I} & Y'X_\mu \\ X_\mu'Y & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y'y - [Y'y]_{\perp X_I} \\ X_\mu'y \end{bmatrix} .$$

Since $[Y'Y]_{\| X_I} = Y'Y - [Y'Y]_{\perp X_I}$ and $[Y'y]_{\| X} = Y'y - [Y'y]_{\perp X_I}$, (II.16) may also be written as:

$$\text{(II.17)} \qquad \hat{\delta}_{2SLS} = \begin{bmatrix} [Y'Y]_{\| X_I} & Y'X_\mu \\ X_\mu'Y & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} [Y'y]_{\| X_I} \\ X_\mu'y \end{bmatrix} .$$

Derivation of 2SLS as a Two-stage Process[3]

As a first stage, let us calculate the predicted value of each explanatory jointly dependent variable (each variable in Y)

---

[1] Basic referenceson 2SLS include Theil [1961], pp. 228-240, 336-344, and Basmann [1957]. 2SLS is referred to as the generalized classical linear (GCL) estimator by Basmann. (Basmann derived 2SLS at approximately the same time as but independently of Theil.) More recently, Basmann extended his GCL estimator to a partial system or full system estimator.

[2] Theil and Basmann used the X matrix as the matrix of instruments, X_I.

[3] Theil [1961], pp. 228-230 derives 2SLS as a two-stage procedure.

by DLS, using the variables in X_I as explanatory variables in the DLS calculations. The resulting matrix of predicted values of variables in Y (often denoted $\hat{Y}$) is exactly the matrix $Y_{\|X_I}$ as noted in section I.D.1. Since X_I is assumed asymptotically uncorrelated with u, $Y_{\|X_I}$ will be asymptotically uncorrelated with u, also.

As a second stage, let us substitute $Y_{\|X}$ for Y into (II.1) and estimate the vector of coefficients, δ, by DLS, i.e., let us apply DLS using y as the dependent variable and the variables in the matrix $[Y_{\|X_I} \vdots X_\mu]$ as explanatory variables. We get:

$$\text{(II.18)} \qquad \hat{\delta}_{\text{2nd stage}} = \begin{bmatrix} Y'_{\|X_I} Y_{\|X_I} & Y'_{\|X_I} X_\mu \\ X'_\mu Y_{\|X_I} & X'_\mu X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y'_{\|X_I} y \\ X'_\mu y \end{bmatrix} .$$

However, $Y'_{\|X_I} Y_{\|X_I} = [Y'Y]_{\|X_I}$ (by definition of $[Y'Y]_{\|X_I}$), $Y'_{\|X_I} y = [Y'y]_{\|X_I}$ [see (I.54)], and since all variables in X_μ are also contained in X_I, $Y'_{\|X_I} X_\mu = [Y'X_\mu]_{\|X_I} = Y'[X_\mu]_{\|X_I} = Y'X_\mu$ [see (I.54) and (I.40)]; hence, (II.18) is equivalent to (II.16) and (II.17). Although $\hat{\delta}_{\text{2SLS}}$ calculated in two steps (i.e., by actually calculating $Y_{\|X_I}$) is algebraically the same as the computational formula for $\hat{\delta}_{\text{2SLS}}$ given in (II.16), the computational formula given in (II.16) should result in less rounding error since the computations are more direct.

Derivation of 2SLS as an Instrumental Variables Estimator Technique

In section II.D.2, 2SLS is derived as an instrumental variables estimator technique.

Derivation of 2SLS as an Application of Generalized Least Squares

In section IV.D, 2SLS is derived as an application of Aitken's generalized least squares.

Derivation of 2SLS as the Least Variance Difference

Consider linearly combining the jointly dependent variables in the equation into a single jointly dependent variable, y*, by postmultiplying these variables by a vector of coefficients, i.e., consider the calculation of $y^* = [y \vdots Y]\begin{bmatrix} 1 \\ -\hat{\gamma}^* \end{bmatrix}$. Let the residual sum of squares from regressing y^* on the predetermined variables in the equation be denoted $\hat{u}'\hat{u}$ (i.e., $\hat{u}'\hat{u} = [y^{*\prime}y^*]_{\perp X_\mu}$) and the residual sum of squares from regressing y^* on all of the instruments be denoted $\tilde{u}'\tilde{u}$ (i.e., $\tilde{u}'\tilde{u} = [y^{*\prime}y^*]_{\perp X_I}$). Then $\hat{u}'\hat{u} - \tilde{u}'\tilde{u}$ will be minimized if $\hat{\gamma}_{2SLS}$ is used as $\hat{\gamma}^*$.[1] Also, if $\hat{\gamma}_{2SLS}$ is used as $\hat{\gamma}^*$, the DLS coefficients obtained by regressing y^* on the predetermined variables in the equation will be $\hat{\beta}_{2SLS}$.[2]

The above least variance difference (LVD) property is intuitively desirable since it causes the jointly dependent variables to be linearly combined such that the instruments which are specified

---

[1] Basmann [1960a], pp. 100-102.

[2] Ibid.

a priori as being outside the equation add as little as possible to the explanation of the combined dependent variable (y*). Such an intuitively desirable property can, however, be easily over-emphasized. The LVR (least variance ratio) property of LIML (limited information single equation maximum likelihood) estimates would seem to be as appealing.

LVR estimates may be derived in the same manner as LVD estimates, except that instead of selecting $\hat{\gamma}^*$ to minimize $\hat{u}'\hat{u} - \tilde{u}'\tilde{u}$, $\hat{\gamma}^*$ is selected to minimize $\hat{u}'\hat{u}/\tilde{u}'\tilde{u}$, which is equivalent to minimizing $(\hat{u}'\hat{u} - \tilde{u}'\tilde{u})/\tilde{u}'\tilde{u}$. If $\hat{\gamma}_{LIML}$ is used as $\hat{\gamma}^*$, then $\hat{u}'\hat{u}/\tilde{u}'\tilde{u}$ will be a minimum.[1] Also, $\hat{\beta}_{LIML}$ will be the DLS coefficients obtained by regressing y* (calculated by using $\hat{\gamma}_{LIML}$ as $\hat{\gamma}^*$) on the predetermined variables in the equation.[2]

---

[1] $\hat{u}'\hat{u}/\tilde{u}'\tilde{u} = [\hat{u}'\hat{u}/\tilde{u}'\tilde{u}] - 1 + 1 = [(\hat{u}'\hat{u} - \tilde{u}'\tilde{u})/\tilde{u}'\tilde{u}] + 1$. Since $\hat{u}'\hat{u}/\tilde{u}'\tilde{u}$ and $[(\hat{u}'\hat{u} - \tilde{u}'\tilde{u})/\tilde{u}'\tilde{u}] + 1$ differ by an additive constant, they are minimized by the same values of y*. Also, $k_{LIML} = \hat{u}'\hat{u}/\tilde{u}'\tilde{u}$. See Koopmans and Hood [1953], pp. 166-169.

[2] Ibid.

C. Additional Single k-class Methods

1. Limited information single equation maximum likelihood (LIML)[1]

If k_{LIML} is calculated as the smallest eigenvalue (characteristic root) of the matrix $[{}_+Y'{}_+Y]^{-1}_{\perp X_I}[{}_+Y'{}_+Y]_{\perp X_\mu}$ then the limited information single equation maximum likelihood (LIML) coefficients may be calculated by the usual single k-class formula.[2] (As noted earlier, ${}_+Y = [y \vdots Y]$; hence, $[{}_+Y'{}_+Y]_{\perp X_I}$ and $[{}_+Y'{}_+Y]_{\perp X_\mu}$ are the moment matrices of the parts of the jointly dependent variables in the equation orthogonal to the instruments and to the predetermined variables in the equation, respectively. Computational formulas for these matrices are given in sections I.D.2 and appendix B.)[3]

If the matrix X is used as the matrix of instruments, X_I, and it is assumed that the matrix of disturbances of the system has the multivariate normal distribution, then the LIML estimates are maximum likelihood estimates given the limited amount of information used (the jointly dependent variables in the equation, the predetermined variables in the system, and which predetermined

---

[1] Basic references for the LIML estimator are Anderson and Rubin [1949], Koopmans and Hood [1953], pp. 162-170, and Chernoff and Divinsky [1953], pp. 240-246.

[2] See Theil [1961], p. 231. The computational equivalence of k_{LIML} as defined above and the more commonly expressed formulas for the calculation of k_{LIML} is noted further on in this section. That the matrix $[{}_+Y'{}_+Y]^{-1}_{\perp X_I}[{}_+Y'{}_+Y]_{\perp X_\mu}$ is not a symmetric matrix must be taken into account in extracting k_{LIML}.

[3] $[{}_+Y'{}_+Y]_{\perp X_\mu}$ may be calculated as an intermediate step of the calculation of $[{}_+Y'{}_+Y]_{\perp X}$ in the manner noted in appendix A.

variables have zero coefficients in the equation).[1] If the matrix X is not used as the matrix of instruments, then the resulting coefficients are not, strictly speaking, the usual limited information maximum likelihood coefficients, since the predetermined variables in a matrix of instruments have been substituted for the predetermined variables in the system.

LIML estimation utilizes the same information as 2SLS, and the LIML coefficients have the same asymptotic coefficient variance-covariance matrix as the 2SLS coefficients.[2] As noted at the end of section II.B.2, LIML coefficients may be derived as the coefficients with the least variance ratio (LVR).

If $\text{rk } X_I = n$ then $k_{LIML} = 1 = k_{2SLS}$; therefore, the coefficients for LIML and 2SLS coincide.[3] If $\text{rk } X_I > n$, then $k_{LIML} > 1$.[4] If $\text{rk } X_I < n$, a singular matrix is encountered during

---

[1] Koopmans and Hood [1953], pp. 166-170.

[2] Theil [1961], p. 232.

[3] Theil [1961], p. 232. As noted in section II.B.2, $k_{LIML} = \text{LVR} = \hat{u}'\hat{u}/\tilde{u}'\tilde{u}$ where $\hat{u}'\hat{u}$ and $\tilde{u}'\tilde{u}$ are as defined in section II.B.2. If $\text{rk } X_I = n$, then there are effectively only $m = n - \ell$ instruments in addition to the predetermined variables in the equation. The $m + 1$ jointly dependent variables may, therefore, be combined into a single jointly dependent variable in such a way that $\tilde{u}'\tilde{u} = \hat{u}'\hat{u}$. See section II.D for additional detail showing the equivalence of 2SLS and LIML for the case $\text{rk } X_I = n$.

[4] If $\text{rk } X_I > n$, then $\hat{u}'\hat{u} \neq \tilde{u}'\tilde{u}$ since there are effectively more than m instruments in addition to the predetermined variables in the equation; hence, $k_{LIML} = \hat{u}'\hat{u}/\tilde{u}'\tilde{u} > 1$. See Koopmans and Hood [1953], pp. 171-175.

estimation (a unique solution does not exist).

Alternative LIML Formulas

The LIML formula given above is not the most common formula for LIML. To see the relationship to more commonly quoted LIML formulas, we will first note some relationships between eigenvalues (characteristic roots) and eigenvectors.

Let A and B be $n \times n$ symmetric positive definite matrices.[1] Then the determinantal equation[2]

$$\det(A - c_i B) = 0 \tag{II.19}$$

has n solutions, $c_1 \ldots c_n$ of which some of the c_i (the eigenvalues) may be duplicates (i.e., there may be only m distinct roots with $m \leq n$). Since a determinantal equation is not changed by multiplying both sides by a constant, the determinantal equation

$$\det(B^{-1}) \cdot \det(A - c_i B) = \det(B^{-1}A - c_i I) = 0 \tag{II.20}$$

has the same eigenvalue solutions (the same c_i) as $\det(A - c_i B) = 0$.

(II.19) may also be converted to another problem--the calculation of the eigenvalues of the equation

$$(A - c_i B)d_i = 0 \tag{II.21}$$

---

[1] The matrix B in this section is any $n \times n$ positive definite symmetric matrix--not the matrix of coefficients of predetermined variables as in other sections of this paper.

[2] det denotes determinant.

where associated with each eigenvalue, c_i, is an $n \times 1$ eigenvector, d_i. Premultiplying (II.21) by d_i' we have:

$$(II.22) \qquad d_i'(A - c_i B)d_i = d_i'Ad_i - c_i d_i'Bd_i = 0$$

or

$$(II.23) \qquad d_i'Ad_i = c_i d_i'Bd_i$$

or

$$(II.24) \qquad c_i = \frac{d_i'Ad_i}{d_i'Bd_i} \quad .$$

That is, each eigenvalue, c_i, must meet relationship (II.24) with its associated eigenvector, d_i.

Similarly from either (II.20) or (II.21) we can derive that

$$(II.25) \qquad (B^{-1}A - c_i I)d_i = 0 \quad ; \text{ thus,}$$

$$(II.26) \qquad d_i'(B^{-1}A - c_i I)d_i = 0 \quad \text{and}$$

$$(II.27) \qquad c_i = \frac{d_i'(B^{-1}A)d_i}{d_i'd_i} \quad .$$

For LIML estimation $A = [{}_+Y' {}_+Y]_{\perp X_\mu}$, $B = [{}_+Y' {}_+Y]_{\perp X_I}$, the minimum c_i from any of the above formulations becomes k_{LIML}, and the corresponding d_i becomes ${}_+\hat{\gamma}_{LIML}$. $\hat{\beta}_{LIML}$ may be calculated as $\hat{\beta}_{LIML} = -[Y_\mu'X_\mu]^{-1}X_\mu' {}_+Y_\mu {}_+\hat{\gamma}_{LIML}$.[1] In the formula which is given in this paper, the smallest eigenvalue of $[{}_+Y' {}_+Y]_{\perp X_I}^{-1}[{}_+Y' {}_+Y]_{\perp X_\mu}$ becomes k_{LIML} which is substituted into

---

[1] Koopmans and Hood [1953], pp. 170-173.

the general k-class formula to calculate $\hat{\delta}_{LIML}$.

Another equivalent formulation is to use $[{}_+Y'{}_+Y]_{\perp X_\mu} - [{}_+Y'{}_+Y]_{\perp X_I}$ as A and ${}_+Y'{}_+Y$ as B. (This is the formulation of Anderson and Rubin [1949] and Chernoff and Divinsky [1953].) The minimum c_i becomes $k_{LIML} - 1$ but the associated eigenvector d_i is still ${}_+\hat{\gamma}_{LIML}$.

The eigenvalues of $A^{-1}B$ are the reciprocals of the eigenvalues of $B^{-1}A$; hence, instead of extracting the smallest eigenvalue of $[{}_+Y'{}_+Y]^{-1}_{\perp X_I}[{}_+Y'{}_+Y]_{\perp X_\mu}$, the largest eigenvalue of $[{}_+Y'{}_+Y]^{-1}_{\perp X_\mu}[{}_+Y'{}_+Y]_{\perp X_I}$ may be extracted and then k_{LIML} calculated as 1 divided by this eigenvalue.

Neither $A^{-1}B$ nor $B^{-1}A$ is symmetric; hence, a non-symmetric eigenvalue computer subroutine is required to extract the desired eigenvalues. Computational procedures for extracting eigenvalues of a matrix of the special form $A^{-1}B$ with A positive definite and B positive semi-definite are available or a computational scheme for more general non-symmetric matrices may be used.

Extraction of an eigenvalue as k_{LIML} and substitution of k_{LIML} into the usual k-class formula in order to calculate the LIML coefficients makes it unnecessary to calculate the corresponding eigenvector while the root is calculated thereby saving computer time. Use of the eigenvector corresponding to k_{LIML} as ${}_+\hat{\gamma}_{LIML}$ and then calculating $\hat{\beta}_{LIML}$ as $-[X'X]^{-1}X'{}_+Y{}_+\hat{\gamma}_{LIML}$ also requires special programming;

thereby, again giving incentive to calculate LIML coefficients through use of the k-class formula. (In addition as is noted in Chapter III, calculation of the estimated coefficient variance-covariance matrix does not require special programming if the general k-class coefficient variance-covariance formula is used.)

The two smallest eigenvalues provide information. The smallest is used as k_{LIML} and the first two smallest eigenvalues may be used in a test of identifiability of a structural equation.[1] The closeness of the second eigenvalue to the smallest eignevalue gives an indication of the "explosiveness" of the resulting LIML coefficients as noted in Klein and Nakamura [1962], pp. 294-295.

---

[1]Koopmans and Hood [1953], pp. 103-184.

2. Nagar's unbiased to $O(T^{-1})$ in probability k (UBK)

If k is calculated as $k = 1 + (\Lambda - n - 1)/T$ and $\hat{\delta}_k$ is calculated using the usual k-class formula, the coefficients obtained will be unbiased estimates of the true parameters, δ, to $O(T^{-1})$ in probability provided another statistical assumption is added to those made at the start of this paper (section I.C.3).[1] Rather than merely assuming that the predetermined variables are contemporaneously independent of the disturbance of the equation, u, the stronger assumption that the predetermined variables are non-stochastic is made, thereby not permitting lagged jointly dependent variables to occur as predetermined variables in the equation.[2]

$\Lambda - n - 1 = K^{**} - m - 1$ will be greater than or equal to zero for an over-identified equation and less than zero for a just-identified equation. Thus, $k_{UBK} \geq 1$ for an over-identified equation and $k_{UBK} < 1$ for a just-identified equation.

If X is used as X_I and rk X is less than Λ, it seems

---

[1] $O(T^{-1})$ in probability is defined in section I.B. In that section it is written more compactly as $O_p(T^{-1})$.

[2] The UBK method is given in Nagar [1959] and Theil [1961], pp. 349-350. Since Nagar's derivation of UBK is based on the reduced form equations which contain the "explanatory" jointly dependent variables in the equation, it is usual to use the X matrix as the matrix of instruments, X_I. Even though lagged jointly dependent variables occurred in the X matrix for the computational illustration in Nagar's article (Klein's model I), Nagar used the entire X matrix as the matrix of instruments, X_I.

appropriate to calculate k_{UBK} as $k_{UBK} = 1 + (rk\ X - n - 1)/T$, since if only the maximum number of lineraly independent predetermined variables were retained and the problem recalculated, the new Λ would be $rk\ X$.

When contemplating UBK estimation, one should remember that UBK coefficients are only asymptotically unbiased. Further, one should be reminded that even if the assumptions inherent in the derivation of the UBK method hold, the degree of bias of an estimator is only one characteristic in evaluating it. Another important characteristic is dispersion, especially dispersion about the true value. The UBK estimator is noted in this paper mainly because UBK coefficients are quite easy to calculate.

3. Nagar's minimum second moment k (MSM)

Nagar derived the moment matrix of $\hat{\delta}_k$ around δ, the true parameter vector, to $O(T^{-2})$ in probability where $\hat{\delta}_k$ is the estimated coefficient vector for any k-class member in which (1) k is non-stochastic[1] and (2) $k - 1 = O(T^{-1})$ in probability.[2] (DLS does not meet the latter requirement.) As for UBK, the stronger assumption that the predetermined variables are non-stochastic is also made. Nagar derives the value of k which minimizes the determinant of this matrix as:

$$k = 1 + \frac{1}{T}\left[\Lambda - 2(m + \ell) - 3 - \frac{tr(C_2Q)}{tr(C_1Q)}\right] \tag{II.28}$$

where tr denotes trace and C_1, C_2, and Q are matrices defined by Nagar.[3] C_1, C_2, and Q must be estimated, the estimate of Q suggested by Nagar being $\begin{bmatrix} [Y'Y]_{\perp X} & Y'X_\mu \\ X_\mu'Y & X_\mu'X_\mu \end{bmatrix}^{-1}$, and the estimates of C_1 and C_2 being defined in his article.[4] (Further on in

---

[1] k non-stochastic means in this case that k is independent of any stochastic variable (in particular, k is independent of u).

[2] The latter restriction implies that $\hat{\delta}_k$ is asymptotically unbiased and has the same asymptotic variance-covariance matrix as δ_{2SLS}. (Nagar [1959], p. 578.)

[3] The MSM method is given in Nagar [1959] and Theil [1961], pp. 349-353. Since Nagar's derivation of MSM is based on the reduced form equations which contain the "explanatory" jointly dependent variables in the equation, it is usual to use the X matrix as the matrix of instruments, X_I. Even though lagged jointly dependent variables occurred in the X matrix for the computational illustration in Nagar's article (Klein's model I), Nagar used the entire X matrix as the matrix of instruments, X_I.

[4] Nagar [1959] gives his suggested estimates for Q, C, and C_1 on p. 589. Although Nagar is not explicit on the point, it is apparent that he is suggesting that C_2 be calculated as $\hat{C}_2 = \hat{C} - \hat{C}_1$.

this paper the symbol Q will be used to denote quite a different matrix.)

The above method of estimating k (and thereby $\hat{\delta}$) will be referred to as the MSM (for minimum second moment) method. k will be greater than 1 only for Λ very large. Nagar notes that the sample size will usually restrict the number of predetermined variables which can be regarded as being in the system thereby causing this to happen rarely. Since $tr(C_2Q)/tr(C_1Q)$ can become large in certain problems, it is possible to obtain a k which is less than zero.

If X is used as X_I and X has less than full column rank it seems appropriate to substitute $rk\ X$ for Λ in the MSM formula for calculating k_{MSM}, since if only the maximum number of linearly independent variables were retained and the problem recalculated, the new Λ would be $rk\ X$.

Even if it is assumed that the assumptions inherent in the derivation of MSM hold, it must be realized that the derivation is made in terms of certain population matrices. The method attempts to minimize by the choice of k the moment matrix of $\hat{\delta}_k$ about δ, by first estimating these population matrices and then substituting them into the MSM formulas. The dispersion of these estimates about the actual population matrices may be considerable. As a result, the actual dispersion of $\hat{\delta}_{MSM}$ about δ could very well be much larger (instead of smaller) than the dispersion for 2SLS.

D. Methods Requiring $\text{rk } X_I = K = n$

In this section we will consider two methods--indirect least squares (ILS) and the instrumental variables estimator (IV)[1] which are applicable only if the matrix of instruments (X_I) actually used in the computation has full column rank (K) with the number of instruments being equal to the number of coefficients being estimated.[2,3] ILS and IV are often referred to as just-identified equation methods, because if the matrix X is used as X_I and X has full column rank, then based on a commonly used counting rule for identification of the coefficients of an equation in a system of equations, the equation is just-identified.[4]

---

[1] We have been treating all of the double k-class estimators as instrumental variables methods to the extent that jointly dependent variables are adjusted by a matrix of instruments in the computational procedure. In this section a computational method called the IV method will be discussed and its relationship to 2SLS and LIML noted. (The IV method was in use before the LIML and 2SLS methods were devised.)

[2] The matrix of instruments actually used in the computation may, of course, be selected or formed from a larger set of instruments.

[3] If X_μ has full column rank then $\text{rk } X^{**}_{\perp X_\mu} = m$ implies that $\text{rk } X_I = n$. ($n = m + \ell$ where m is the number of explanatory jointly dependent variables in the equation and ℓ is the number of predetermined variables in the equation. X** is the $T^{\times}(K - \ell)$ matrix composed of the instruments which are not predetermined variables in the equation being estimated.) Although $\text{rk } X_I = n$ implies $\text{rk } X_\mu = \ell$ and $\text{rk } X^{**} = m$, $\text{rk } X_\mu = \ell$ and $\text{rk } X^{**} = m$ do not together imply $\text{rk } X_I = n$. ($\text{rk } X_\mu = \ell$ and $\text{rk } X^{**}_{\perp X_\mu} = m$ together imply $\text{rk } X_I = n$.)

[4] The counting rule is a necessary but not sufficient condition for identifiability of the coefficients of the equation being estimated. See Koopmans and Hood [1953], pp. 135-142.

If (1) the matrix of predetermined variables in the system (X) is used as the matrix X_I in the estimation of 2SLS and LIML, (2) X is used as the matrix of instruments in the IV method, and (3) rk X = Λ = n, then the same estimated coefficients are obtained from ILS, IV, 2SLS, and LIML; hence, 2SLS might as well be used to estimate the coefficients in this case.[1] Thus, it is unnecessary to develop a special ILS computer program. If the matrix of instruments used in the IV method contains the predetermined variables in the equation, then the IV coefficients will also be the same as the corresponding 2SLS and LIML coefficients. Even if not all of the predetermined variables in the equation are used as instruments, the IV problem may be reformulated as a special 2SLS problem and readily solved;[2] hence, it is unnecessary that a special IV computer program be developed, either.

---

[1] That 2SLS and ILS coefficients coincide is shown in section II.D.1. Koopmans and Hood [1953], pp. 173-174 show that in the case of rk X_I = K = n (with X used as X_I), LIML and ILS coefficients coincide and $k_{LIML} = 1$. (Since k_{2SLS} is always 1, this shows that LIML and 2SLS coefficients coincide.)

[2] This is shown in section II.D.2.

1. Indirect least squares (ILS)

In the indirect least squares (ILS) method, reduced form coefficients are estimated directly by DLS (these reduced form coefficient estimates are consistent since each equation contains only 1 jointly dependent variable) and then structural coefficient estimates are obtained algebraically from the reduced form coefficient estimates.[1] Since the structural coefficients are the same continuous function of the reduced form coefficients as the estimated structural coefficients are of the estimated reduced form coefficients, and since the estimated reduced form coefficients are consistent, the ILS coefficients are consistent.[2]

Derivation of the ILS Estimating Equations

We will assume that X has full column rank so that unique unrestricted least squares estimates for reduced form equations containing all predetermined variables in the system exist.

To estimate the coefficients of the μ^{th} equation by ILS, we need not estimate the coefficients of all of the reduced form equations--only those with jointly dependent variables contained in the μ^{th} equation. If we divide the $G \times \Lambda$ matrix of reduced

---

[1] The ILS method may be found in Koopmans and Hood [1953], pp. 135-141.

[2] This follows from a theorem by Slutsky. See Goldberger [1954], pp. 118-119, 128, 326-327.

form coefficients, Π into $\begin{bmatrix}\Pi_1\\ \Pi_2\end{bmatrix}$ where Π_1 is the $(m_\mu + 1)\times\Lambda$ matrix of coefficients of the equations with the same jointly dependent variables as those in equation μ, then we can estimate Π_1 by DLS as:[1]

$$\hat{\Pi}_1' = [X'X]^{-1}X'_+Y \quad . \tag{II.29}$$

From (I.19) we have:

$$\Pi = -\Gamma^{-1}B \quad \text{or} \quad \Gamma\Pi = -B \quad \text{or} \quad \Gamma\Pi + B = 0$$

or in terms of the coefficients of the μ^{th} equation above, we have:[2]

$$[{}_+\gamma' \vdots 0']\Pi + [\beta' \vdots 0'] = 0' \quad \text{or} \quad {}_+\gamma'\Pi_1 + [\beta' \vdots 0'] = 0'$$

$$\text{or} \quad \Pi_1' {}_+\gamma + \begin{bmatrix}\beta\\ 0\end{bmatrix} = 0 \quad .$$

Substituting the estimate $\hat{\Pi}_1' = [X'X]^{-1}X'_+Y$ for Π_1', we get the ILS estimating equations:

$$[X'X]^{-1}X'_+Y\hat{\gamma}_{ILS} + \begin{bmatrix}\hat{\beta}_{ILS}\\ 0\end{bmatrix} = 0 \quad . \tag{II.30}$$

Substituting $[y \vdots Y]$ for ${}_+Y$ and $\begin{bmatrix}-1\\ \hat{\gamma}\end{bmatrix}$ for ${}_+\hat{\gamma}$ and rewriting, we have:

$$[X'X]^{-1}X'Y\hat{\gamma}_{ILS} + \begin{bmatrix}\hat{\beta}_{ILS}\\ 0\end{bmatrix} = [X'X]^{-1}X'y \quad . \tag{II.31}$$

---

[1] ${}_+Y = [y \vdots Y]$ as before.

[2] ${}_+\gamma = \begin{bmatrix}-1\\ \gamma\end{bmatrix}$ as before.

This is a system of Λ non-homogeneous equations in n $(= m + \ell)$ unknowns. For (II.31) to have a unique solution, Λ must equal n. If $\Lambda > n$, no solution exists.[1] For a solution to exist, $\Lambda - n$ of the equations given in (II.31) must be thrown away (i.e., $\Lambda - n$ of the rows of $[X'X]^{-1}$ must be deleted), thereby ignoring information and making the actual estimates obtained depend on the particular rows of $[X'X]^{-1}$ deleted. Equivalently, $\Lambda - n$ of the predetermined variables in the system can be ignored thereby forcing the number of predetermined variables used in the estimation to equal n. This gives rise to a different set of ILS coefficients for each different set of predetermined variables ignored.

Further Derivation Assuming $\Lambda = n$

Let us assume that for the μ^{th} equation, $n = \Lambda$. Given our assumption that X has full column rank this implies that $Z_\mu'X$ is square. Let us further assume that $Z_\mu'X$ is non-singular. A set of simultaneous equations may be premultiplied by a non-singular matrix without changing the solution. Premultiplying (II.31) by $Z_\mu'X$ we get:

---

[1] Assuming X has full column rank, an equation in which $\Lambda = n$ is just-identified by the counting rule for identification and an equation in which $\Lambda > n$ is over-identified by the counting rule for identification. The case of just-identification is also referred to as the case of "minimum requisite information" and the case of over-identification is also referred to as the case of "extra information." $\Lambda \geq n$ is a necessary but not sufficient condition for the population coefficients to be identifiable. See Koopmans and Hood [1953], pp. 135-142.

$$(II.32) \qquad Z_\mu'X[X'X]^{-1}X'Y\hat{\gamma}_{ILS} + Z_\mu'X\begin{bmatrix}\hat{\beta}\\0\end{bmatrix} = Z_\mu'X[X'X]^{-1}X'y \quad .$$

Since $Z_\mu'X[X'X]^{-1}X'Y = [Z_\mu'Y]_{\parallel X} = \begin{bmatrix}[Y'Y]_{\parallel X}\\ [X_\mu'Y]_{\parallel X}\end{bmatrix} = \begin{bmatrix}[Y'Y]_{\parallel X}\\ X_\mu'Y\end{bmatrix}$

and $Z_\mu'X[X'X]^{-1}X'y = [Z_\mu'y]_{\parallel X} = \begin{bmatrix}[Y'y]_{\parallel X}\\ [X_\mu'y]_{\parallel X}\end{bmatrix} = \begin{bmatrix}[Y'y]_{\parallel X}\\ X_\mu'y\end{bmatrix}$.

(II.32) may be rewritten as:[1]

$$(II.33) \qquad \begin{bmatrix}[Y'Y]_{\parallel X}\\ X_\mu'Y\end{bmatrix}\hat{\gamma}_{ILS} + \begin{bmatrix}Y'X_\mu\\ X_\mu'X_\mu\end{bmatrix}\hat{\beta}_{ILS} = \begin{bmatrix}[Y'y]_{\parallel X}\\ X_\mu'y\end{bmatrix}$$

or

$$(II.34) \qquad \begin{bmatrix}[Y'Y]_{\parallel X} & Y'X_\mu\\ X_\mu'Y & X_\mu'X_\mu\end{bmatrix}\begin{bmatrix}\hat{\gamma}\\ \hat{\beta}\end{bmatrix}_{ILS} = \begin{bmatrix}[Y'y]_{\parallel X}\\ X_\mu'y\end{bmatrix} \quad .$$

Therefore,

$$(II.35) \qquad \hat{\delta}_{ILS} = \begin{bmatrix}[Y'Y]_{\parallel X} & Y'X_\mu\\ X_\mu'Y & X_\mu'X_\mu\end{bmatrix}^{-1}\begin{bmatrix}[Y'y]_{\parallel X}\\ X_\mu'y\end{bmatrix} \quad .$$

However, (II.35) is the 2SLS computational formula (II.17) in which X is used as the matrix of instruments, X_I. Thus, ILS coefficients may be computed through use of the 2SLS formula.

---

[1] $[X_\mu'Y]_{\parallel X} = X_\mu'Y$ and $[X_\mu'y]_{\parallel X} = X_\mu'y$ since the variables in X_μ are contained in X. See (I.42).

2. The instrumental variables estimator (IV)

In our general double k-class methods, we have been adjusting the jointly dependent variables in the equation by a matrix of instrumental variables. In this section, we will consider the calculation of coefficients by a computational method (fairly widely used before the k-class methods such as LIML and 2SLS were devised) called the instrumental variables (IV) estimating method.[1] In this method, the same number of instruments as there are variables in the equation are used. We will show that on the one hand, the IV coefficients may be calculated by the 2SLS computational method and on the other hand, 2SLS coefficients may be derived as a particular case of the IV method. The choice of instruments for 2SLS and IV should apparently be based on the same criteria except that only n instruments can be chosen for the IV method.[2]

Let a single equation from a system of equations be written as

$$y = Z_{\mu}\delta + u \tag{II.36}$$

where all of the matrices and vectors have the same dimension and meaning as before [see (I.10) and (II.1)].

Let X_{IV} be a matrix of n instrumental variables (the same number of instrumental variables as there are columns in Z_{μ})

---

[1] Goldberger [1964] contains a detailed treatment of the instrumental variables method.

[2] Choice of instruments is discussed in section II.G.

and assume that $X'_{IV}Z$ is nonsingular (hence that X_{IV} and Z_μ have full column rank and that correlation exists between the variables in X_{IV} and the variables in Z_μ).

Premultiplying (II.36) by X'_{IV} we get:

$$X'_{IV}y = X'_{IV}Z_\mu\delta + X'_{IV}u \quad . \tag{II.37}$$

If we let the estimating equations for δ_{IV} be:

$$\underset{n\times n}{X'_{IV}Z_\mu}\ \underset{n\times 1}{\hat{\delta}_{IV}} = \underset{n\times 1}{X'_{IV}y} \quad , \tag{II.38}$$

we get:

$$\hat{\delta}_{IV} = [X'_{IV}Z_\mu]^{-1}X'_{IV}y \quad . \tag{II.39}$$

If the variables in X_{IV} are contemporaneously independent of u (thus, $\text{plim}(1/T)X'_{IV}u = 0$) and correlated with Z so that $\text{plim}(1/T)X'_{IV}Z_\mu = \Omega_{X_{IV}Z_\mu}$ exists and is non-singular, then $\hat{\delta}_{IV}$ is a consistent estimate of δ.[1]

Calculation of IV Problems on a 2SLS Computer Routine

It is not necessary to develop a computer program to calculate IV estimates, since $\hat{\delta}_{IV}$ may be calculated on a double k-class computer routine as a 2SLS problem. Let us premultiply equation (II.38) by the non-singular square matrix $Z'_\mu X_{IV}[X'_{IV}X_{IV}]^{-1}$

---

[1] A proof of the consistency of $\hat{\delta}_{IV}$ is given in Goldberger [1964], p. 285.

(premultiplication by a non-singular matrix will not change the solution) so that the estimating equations become:

$$(II.40) \qquad Z_\mu' X_{IV} [X_{IV}' X_{IV}]^{-1} X_{IV}' Z_\mu \hat{\delta}_{IV} = Z_\mu' X_{IV} [X_{IV}' X_{IV}]^{-1} X_{IV}' y$$

or

$$(II.41) \qquad [Z_\mu' Z_\mu]_{\parallel X_{IV}} \hat{\delta}_{IV} = [Z_\mu' y]_{\parallel X_{IV}} \quad .$$

Then:

$$(II.42) \qquad \hat{\delta}_{IV} = [Z_\mu' Z_\mu]^{-1}_{\parallel X_{IV}} [Z_\mu' y]_{\parallel X_{IV}} \quad .$$

Comparison of (II.42) with (II.17) or (II.35) shows that (II.42) may be computed on a 2SLS routine by merely treating Z_μ as Y, (i.e., by treating all of the variables in the equation as if they were jointly dependent variables) and by using X_{IV} as the matrix of the instruments, X_I. (Only the upper left hand submatrix will remain in the matrix to be inverted and only the upper subvector will remain in the right hand side vector.) The variables in X_μ are treated as jointly dependent in the computation only--not in the interpretation of results.

The calculation of $\hat{\delta}_{IV}$ as a special 2SLS problem also provides a convenient way of calculating the estimated coefficient variance-covariance matrix as is noted in the section on coefficient variance-covariance estimation.

If some of the predetermined variables in the equation are used as their own instruments (or are even linear combination of variables in X_{IV}), these predetermined variables need not be treated as jointly dependent for computational purposes. Let

$Z_\mu = [Z_\mu^* \vdots X_\mu^*]$ where variables are rearranged in the equation to permit listing predetermined variables which serve as instruments last. These variables comprise the X_μ^* matrix. (II.42) may be rewritten as:

$$\text{(II.43)} \qquad \hat{\delta}_{IV} = \begin{bmatrix} [Z_\mu^{*\prime} Z_\mu^*]_{\|X_{IV}} & [Z_\mu^{*\prime} X_\mu^*]_{\|X_{IV}} \\ [X_\mu^{*\prime} Z_\mu^*]_{\|X_{IV}} & [X_\mu^{*\prime} X_\mu^*]_{\|X_{IV}} \end{bmatrix}^{-1} \begin{bmatrix} [Z_\mu^{*\prime} y]_{\|X_{IV}} \\ [X_\mu^{*\prime} y]_{\|X_{IV}} \end{bmatrix} .$$

Since X_μ^* is in the space spanned by X_{IV},

$[Z_\mu^{*\prime} X_\mu^*]_{\|X_{IV}} = Z_\mu^{*\prime} X_\mu^*$, $[X_\mu^{*\prime} X_\mu^*]_{\|X_{IV}} = X_\mu^{*\prime} X_\mu^*$, and $[X_\mu^{*\prime} y]_{\|X_{IV}} = X_\mu^{*\prime} y$

[see (I.39)]; therefore, (II.43) may be rewritten as:

$$\text{(II.44)} \qquad \hat{\delta}_{IV} = \begin{bmatrix} [Z_\mu^{*\prime} Z_\mu^*]_{\|X_{IV}} & Z_\mu^{*\prime} X_\mu^* \\ X_\mu^{*\prime} Z_\mu^* & X_\mu^{*\prime} X_\mu^* \end{bmatrix}^{-1} \begin{bmatrix} [Z_\mu^{*\prime} y]_{\|X_{IV}} \\ X_\mu^{*\prime} y \end{bmatrix} .$$

Comparison of (II.44) with (II.17) shows that (II.44) is the computational formula for 2SLS in which the variables in X_μ^* are treated as predetermined variables in the equation, the variables in Z_μ^* are treated as explanatory jointly dependent variables in the equation (for computational purposes--not for interpretation of results), and the variables in X_{IV} are treated as predetermined variables in the system.

If (as is the usual practice) all of the predetermined variables in the equation are used as instruments, then $Z_\mu^* = Y$, $X_\mu^* = X_\mu$, and (II.44) becomes:

$$\text{(II.45)} \qquad \hat{\delta}_{IV} = \begin{bmatrix} [Y'Y]_{\|X_{IV}} & Y'X_{\mu} \\ X_{\mu}'Y & X_{\mu}'X_{\mu} \end{bmatrix}^{-1} \begin{bmatrix} [X_{\mu}'y]_{\|X_{IV}} \\ X_{\mu}'y \end{bmatrix},$$

the usual 2SLS computational formula in which X_{IV} is used as the matrix of instruments, X_I.

The above computational methods do not require that predetermined variables in the equation be selected as instruments; however, most criteria for selecting instruments make the predetermined variables in the equation prime candidates as instruments. As noted above, not selecting a predetermined variable in an equation as an instrument has the same computational effect (assuming that the variable is not in the space spanned by the instruments) as if the variable had been reclassified as a jointly dependent variable.

If $\Lambda = n$, then (as in double k-class estimation) the instruments are usually taken to be the predetermined variables in the system so that $X_{IV} = X_I = X$ and (II.45) becomes the usual 2SLS computational procedure. If $\Lambda > n$, unique IV estimates do not exist if the instruments are <u>selected</u> from among the Λ predetermined variables in the system, since the number of predetermined variables in the system is greater than the number of instruments to be selected. On the other hand a unique set of m instruments can be <u>calculated</u> from the Λ predetermined variables (this set of instruments consists of the variables in $Y_{\|X}$) such that if $X_{IV} = [Y_{\|X} \vdots X_{\mu}]$, 2SLS estimates (with $X_I = X$) are obtained.

Derivation of 2SLS as an IV Method

We have been treating IV as a special case of 2SLS, at least computationally. It is interesting to note that even if in 2SLS estimation the matrix of instruments, X_I, is of rank greater than n_μ (hence, there are more instruments in the X_I matrix than variables in the equation), 2SLS may be considered to be a particular case of IV.[1] To demonstrate this, we use the variables in the matrix $[Y_{\blacksquare X_I} \vdots X_\mu]$ as instruments.[2] (II.39) becomes:

$$\text{(II.46)} \qquad \hat{\delta}_{IV} = \{[Y_{\blacksquare X_I} \vdots X_\mu]'[Y \vdots X_\mu]\}^{-1}[Y_{\blacksquare X_I} \vdots X_\mu]'y$$

$$= \begin{bmatrix} Y'_{\blacksquare X_I} Y & Y'_{\blacksquare X_I} X_\mu \\ X_\mu' Y & X_\mu' X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y'_{\blacksquare X_I} y \\ X_\mu' y \end{bmatrix} .$$

Since $Y'_{\blacksquare X_I} Y = [Y'Y]_{\blacksquare X_I}$, $Y'_{\blacksquare X_I} y = [Y'y]_{\blacksquare X_I}$, and $Y'_{\blacksquare X_I} X_\mu = [Y'X_\mu]_{\blacksquare X_I}$ $= Y'X_\mu$ [see (I.54) and (I.40)], the above may be rewritten as:

$$\text{(II.47)} \qquad \hat{\delta}_{IV} = \begin{bmatrix} [Y'Y]_{\blacksquare X_I} & Y'X_\mu \\ X_\mu' Y & X_\mu' X_\mu \end{bmatrix}^{-1} \begin{bmatrix} [Y'y]_{\blacksquare X_I} \\ X_\mu' y \end{bmatrix} = \hat{\delta}_{2SLS} .$$

The variables in $Y_{\blacksquare X_I}$ serve as valid instruments since they are linear combinations of variables in X_I.

---

[1] Goldberger [1964], p. 332.

[2] $Y_{\blacksquare X_I}$ is the part of Y in the space spanned by X_I.

E. No Predetermined Variables in an Equation

All of the formulas for estimating coefficients given earlier remain valid even if no predetermined variables occur in the equation to be estimated. In the LIML computations, $[{}_{+}Y'{}_{+}Y]_{\perp X_{\mu}}$ (the moment matrix of the part of ${}_{+}Y$ orthogonal to the columns of X_{μ}) becomes ${}_{+}Y'{}_{+}Y$.[1]

---

[1] ${}_{+}Y = [y \vdots Y]$.

F. Only One Jointly Dependent Variable in an Equation

If only one jointly dependent variable occurs in an equation, all of the double k-class methods become the same as DLS, since equation (II.2) becomes simply:

$$\hat{\delta}_{k_1,k_2} = [X_\mu' X_\mu]^{-1} X_\mu' y \quad . \tag{II.48}$$

G. Selection of Instruments

In this section we will discuss the selection of instruments for double k-class methods, limiting our discussion for the sake of simplicity to 2SLS for most of the section. Instruments selected for double k-class methods should have the same basic characteristics as instruments selected for the IV estimator except that it seems less desirable to form linear combinations of possible instruments for the double k-class members, since the number of instruments to be used is not restricted to a given number--the number of explanatory variables in the equation--as is the case with the IV method. We will not make explicit further reference to the IV method, since it is most fruitful to regard (and calculate) IV coefficients as the special case of 2SLS in which there are the same number of instruments as explanatory variables in the equation.[1]

We will assume that the predetermined variables in the equation are among the instruments in the X_I matrix. For consistency in double k-class estimation we will assume that the matrix of instruments is contemporaneously independent of the disturbance of the equation being estimated (see appendix C). This suggests that the predetermined variables in the system are prime candidates for choices as instruments; however, cases may arise where it is desirable that not all of the predetermined variables in the system be used as instruments or that additional variables which are not predetermined variables in the system be used as instruments. If

---

[1] See section II.D.2.

all of the assumptions made at the start of this paper strictly held, one would be led toward the use of all of the predetermined variables in the system as the set of instruments; however, since these assumptions are likely to hold only imperfectly in practice, it may be desirable that given instruments be eliminated from the X_I matrix. For example if the disturbances of an equation are "slightly" serially correlated, one would surely question the use of lagged jointly dependent variables outside the equation but in the system as instruments.[1]

---

[1] See Fisher [1965] and Goldberger [1964] for expositions on the choice of instruments.

It has been usual (where possible) to use all of the predetermined variables in the system as instruments, the justification for using all of them being based on the fact that the unrestricted reduced form equations corresponding to the jointly dependent variables contain all of the predetermined variables in the equation. Examination of many systems (especially those in which the matrix of coefficients of the structural equations has a recursive structure) often discloses that certain of the predetermined variables occur with zero coefficients in the reduced form equation corresponding to a given jointly dependent variable. Often a separate set of instruments is then used in the adjustment of each jointly dependent variable. (If a separate set of instruments is used for each jointly dependent variable, the formulas given for the double k-class estimators can be modified in the manner noted in section II.A and appendix B.) Alternatively only those predetermined variables which have zero coefficients in the reduced form equations corresponding to all of the jointly dependent variables may be eliminated.

Fisher takes a causal approach to the selection of instruments which leads him to the examination of the structural equations rather than the reduced form equations as the basis for choice of instruments, the key structural equation for each explanatory jointly dependent variable being the one in which that variable occurs as the normalizing variable. (Fisher takes the approach that in a fully specified system, each jointly dependent variable will occur as the normalizing variable in one structural equation.) This also leads him to the possibility of using a separate set of instruments to adjust each explanatory jointly dependent variable.

Examination of alternative assumptions made in the block recursive systems which Fisher examines leads him to question (in some cases) the use of lagged jointly dependent variables and to suggest the use of lagged exogenous variables as instruments.

A model may also be examined for partitioning into subsystems. After partitioning, the instruments in the estimation of each equation in a subsystem are based on the predetermined variables in

The minimum number of instruments which can be selected for 2SLS is n (at least m instruments must be selected in addition to the ℓ predetermined variables in the equation); since otherwise the $[Y_{\perp X_I} : X_\mu]$ matrix has rank less than n and therefore, the $\begin{bmatrix} [Y'Y]_{\perp X_I} & Y'X_\mu \\ X_\mu'Y & X_\mu'X_\mu \end{bmatrix}$ matrix is singular, i.e., unique 2SLS coefficients do not exist.[1]

There is absolutely no maximum number of instruments which may be chosen if direct orthogonalization is used as suggested in this paper; however, if $\text{rk } X_I = T$, all double k-class coefficients become the same as the DLS coefficients.[2] This comes about because if $\text{rk } X_I = T$, then $Y_{\perp X_I} = 0$ since any variables in the observation matrix falls in the space spanned by X_I (i.e., all of the variables may be expressed as an exact linear combination of the variables

---

that subsystem. One basis for partitioning is the degree of correlation of the disturbances (estimated in some manner or assumed a priori). Also, Hannan [1967] gives a method for subdividing a system of equations of a special form into non-intersecting "maximal" subsystems.

[1]This is also the minimum number of instruments for LIML, since if the number of instruments is less than or equal to n, the LVR (least variance ratio) for LIML is (see sections II.B.2 and II.C.1) $k_{LIML} = \hat{u}'\hat{u} / \tilde{u}'\tilde{u} = 1$ (since the jointly dependent variables may be combined such that $\hat{u}'\hat{u}$ is no larger than $\tilde{u}'\tilde{u}$); hence, LIML becomes the same as 2SLS for which (in this case) unique coefficients do not exist. (Depending on the computational formula for k_{LIML}, arbitrarily large numbers may be encountered during the computation of k_{LIML} if the number of instruments is less than n.)

[2]In this case, the particular instruments chosen will have no effect on the coefficients obtained (provided $\text{rk } X_I = T$) since the DLS solution is obtained in any case.

in X_I); hence, the part of any jointly dependent variable orthogonal to X_I is zero. Thus, $[Y'Y]_{\perp X_I} = Y'_{\perp X_I} Y_{\perp X_I} = 0'0 = 0$, $[Y'y]_{\perp X_I} = X'_{\perp X_I} y_{\perp X_I} = 0'0 = 0$ and the general double k-class formula (II.2) becomes:

$$(II.49) \quad \hat{\delta}_{k_1,k_2} = \begin{bmatrix} Y'Y - k_1 \cdot 0 & Y'X_\mu \\ X'_\mu Y & X'_\mu X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y'y - k_2 \cdot 0 \\ X'_\mu y \end{bmatrix}$$

$$= \begin{bmatrix} Y'Y & Y'X_\mu \\ X'_\mu Y & X'_\mu X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y'y \\ X'_\mu y \end{bmatrix} = \hat{\delta}_{DLS} .$$

Thus, when $\text{rk } X = T$, all double k-class methods give estimated coefficients which coincide with the DLS coefficients.

The fact that all double k-class coefficients coincide with DLS coefficients when $\text{rk } X_I = T$ does not destroy the consistency of given double k-class methods. All that is indicated is that there are insufficient observations to distinguish the double k-class members from each other based on the estimated coefficients obtained.[2]

---

[1] In this formula and the remainder of this section we will assume that only zero restrictions are imposed on the coefficients and that the matrix $[Y \vdots X_\mu]$ has full column rank (this implies that $n \leq T$).

[2] This was pointed out to the writer by Professor Robert L. Gustafson. Consistency is an asymptotic property and a small number of observations in a given sample certainly does not affect an asymptotic property. If the number of instruments used in the estimation is fixed (e.g., the number of predetermined variables in the system is fixed for a given model; hence, if X is used as X_I then the number of variables in X_I is fixed) then (if the double k-class formulas are followed; that is, a switch is not made to the DLS formula) as T increases, at some point there will be sufficient observations that $\text{rk } X_I < T$ and the coefficients of the double k-class members will not coincide with DLS coefficients.

Historical Perspective

The formulas presently quoted in the econometric literature for the calculation of $[Y'Y]_{\|X_I}$ and $[Y'y]_{\|X_I}$ are $[Y'Y]_{\|X_I} = Y'X_I(X_I'X_I)^{-1}X_I'Y$ and $[Y'y]_{\|X_I} = Y'X_I(X_I'X_I)^{-1}X_I'y$. As a result of focusing on these formulas, the problem occurring when $\text{rk } X_I = T$ has been expressed as one of how to select or construct a set of predetermined variables, say X_a, which captures the maximum effect of the predetermined variables in the system but whose sums of squares and cross-products matrix, $X_a'X_a$, is nonsingular.[1]

Initially the solution to the problem was based on the omission of sufficient variables from the X_I matrix that the inverse of the resulting $X_a'X_a$ matrix could be calculated accurately. The variables retained in the X_a matrix (in addition to the predetermined variables in the equation) were selected such that the resulting X_a matrix captured as much effect of the original X_I matrix as possible.

Kloek and Mennes took a different approach to the restriction of the space of predetermined variables. They suggested linearly combining the predetermined variables in the system into a set of fewer predetermined variables by calculating principal components of predetermined variables. In their

---

[1]As an example, Kloek and Mennes [1960], p. 46 state: "The table also shows, however, that Λ may easily grow in excess of the number of observations T on which the estimation is based. This is a serious problem, for it implies that the matrix of sums of squares and cross products of the predetermined variables ($X'X$) is singular; and the inverse of this matrix is needed for the estimation of the reduced-form disturbances, which are auxillary in the estimation of the parameters of the structural equations."

famous article they elaborated four methods of calculating principal components.[1] The most elaborate of these methods is the method which Kloek and Mennes refer to as method 2. In method 2 the moment matrix of $X^{**}_{\perp X_\mu}$ (the part of the predetermined variables in the system with zero coefficients in the equation orthogonal to the predetermined variables with non-zero coefficients in the equation) is first calculated and then a predesignated number of principal components of the $X^{**}_{\perp X_\mu}$ matrix are calculated. These principal components plus X_μ are then used in place of the X_I matrix in double k-class estimation.

Regarding Restriction of The Space of Instrumental Variables

The calculation of $[Y'Y]_{\perp X_I}$ and $[Y'y]_{\perp X_I}$ by direct orthogonalization (hence the calculation of $[Y'Y]_{\parallel X_I}$ as $Y'Y - [Y'Y]_{\perp X_I}$ and $[Y'y]_{\parallel X_I}$ as $Y'y - [Y'y]_{\perp X_I}$) changes the focus of the problem from one of eliminating sufficient multicollinearity that an inverse (or a set of reduced form coefficients) can be accurately calculated, since $[Y'Y]_{\parallel X_I}$ and $[Y'y]_{\parallel X_I}$ are already unique and automatically calculated by the computational method given in this paper in even the most extreme cases of multicollinearity among the instruments. The problem now becomes one of whether the subspace of the instruments should be restricted so that the solution obtained will not coincide with the DLS solution.

There is often a good basis for not using all of the predetermined variables in the system as variables in the X_I matrix

---

[1] Kloek and Mennes [1960].

provided this is done because of characteristics of the data (e.g., some degree of serial correlation in the disturbances of an equation may imply that certain lagged jointly dependent variables would not serve as desirable instruments) or of the model (e.g., the reduced form equations corresponding to the explanatory jointly dependent variables do not contain a set of the predetermined variables). There does not, however, seem to be a good basis for restricting the space of instruments merely to cause the resulting coefficients to differ from the DLS coefficients.[1,2]

On the other hand, since there will surely be researchers who will desire that the space of instruments be restricted so that the resulting coefficients do not coincide with DLS coefficients (and since some effective arguments for restricting the space may be forthcoming in the future) a few additional remarks regarding how the space might be restricted will be made.

---

[1] As noted before, an estimator is not made inconsistent just because the resulting coefficients do not differ from DLS coefficients. Also, by now it should be obvious that asymptotic properties are a poor guide as to how to procede when the problem is the result of a small number of observations.

[2] Even though DLS does not use any information from the remainder of the system, there may still be considerable advantage to carefully specifying the remainder of the system even if it is known ahead of time that given the number of observations available, rk X_I will equal T. If a complete system is specified, FIML (full information maximum likelihood) may be applied even if rk X_I = T. (rk X_I = T presents no difficulty in FIML estimation, since the matrix of coefficients of jointly dependent variables are recognized in a special fashion, but the jointly dependent variables themselves are not adjusted by a set of instruments. The confusion which exists regarding this point is clarified in footnote 1 of page 203.) The FIML solution will not, in general, coincide with the DLS solution even in the case rk X_I = T.

The change in focus of the problem away from the prevention of a large degree of multicollinearity considerably reduces the desirability of using principal components to restrict the space spanned by the instruments. If it is decided that the space spanned by the instruments is to be restricted, then it is much more straightforward to do this by merely eliminating certain of the initial instruments considered for the analysis. In that way the coefficients obtained can be related to the particular set of instruments used, whereas it is very difficult to evaluate the effect of using as instruments in the computation a set of principal components of a larger set of instruments. The best guide as to which instruments to retain would seem to be to rank (or group) the instruments according to their desirability as instruments in the estimation of the equation.[1]

One of the advantages claimed for the use of principal components in restricting the space of the instruments is that the use of principal components is less arbitrary than the selection of instruments to retain--one merely decides on the principal component method to use (which might as well be method 2) and how many principal components to retain; (the number to retain is quite arbitrary); however, there are certainly easier and more informative methods (in terms of evaluating the results) to accomplish even this advantage.

As an example (in addition to the methods of Fisher [1965]) the orthogonalization procedure described in appendix A can be

---

[1]See Fisher [1965] for some procedures.

modified so that c instruments[1] (in addition to the ℓ predetermined variables in the equation) can be selected from among a prespecified set of instruments so that no instrument selected is an exact linear combination of instruments previously selected. The modification consists of merely (1) incorporating the effect of the variables in X_μ, and then (2) stopping the orthogonalization procedure after c diagonal elements have been selected as pivots from among those corresponding to instruments not in the equation. After c pivots have been used, the part of the matrix corresponding to ${}_+Y'{}_+Y$ will have become $[{}_+Y'{}_+Y]_{\perp}[X_\mu : X_1 \ldots X_c]$ where X_1 through X_c are the instruments corresponding to the c pivots selected. Using this matrix in place of $[{}_+Y'{}_+Y]_{\perp}X_I$ will result in the same coefficients as if X_1 through X_c were initially listed as the only instruments in addition to the variables in X_μ.[2]

Assuming that the space of instruments is to be restricted,

---

[1] c could be the number of principal components which would have been selected if principal components had been used. c must be greater than or equal to m since otherwise (as noted earlier) the matrix inverted in computing 2SLS or LIML estimated coefficients is singular.

[2] This method is somewhat similar to principal components method 2 in that by first incorporating the variables of X_μ, the first pivot selected from among the part of the matrix corresponding to the variables in X** is selected as the largest diagonal element of the $[X^{**\prime}X^{**}]_{\perp}X_\mu$ matrix (the moment matrix from which principal components are calculated if method 2 is used). If the variable corresponding to the first pivot selected from the $[X^{**\prime}X^{**}]_{\perp}X_\mu$ matrix is denoted as X_1, then the second pivot is selected as the largest one in the $[X^{**\prime}X^{**}]_{\perp}[X_\mu : X_1]$ matrix, and so on until c pivots (and hence c variables) have been selected in the orthogonalization procedure.

both the predesignation of the variables to be treated as instruments and the use of an automatic method to select a set of variables have the distinct advantage over the use of principal components in that it is clear as to the exact instruments used in the computations; hence, it is possible to more fully evaluate the results obtained. Ease of computation is a second advantage.

Predesignation of the variables to use has an advantage over the automatic selection computational method in that more judgement may be used in the selection of instruments. The two methods may, of course, be combined by (1) not listing predetermined variables in the system which are clearly not desired, (2) selecting the pivots corresponding to all of predetermined variables clearly desired, and (3) letting the automatic procedure select the remaining variables (up to a given prespecified number) through choice of the largest pivot at each step.

The methods for selecting instruments given in Fisher [1965] tend to lead to a separate set of instruments to adjust each jointly dependent variable in the equation. Whereas the use of this approach is certainly feasible, it would seem that a more straightforward approach would be to select a set of instruments corresponding to each explanatory jointly dependent variable as Fisher suggests and then use all of the instruments selected for any jointly dependent variable in the X_I matrix, thereby again adjusting all of the explanatory jointly dependent

variables by the same set of instruments.[1] In addition to the computations being easier to perform, interpretation of results will be simpler.[2,3]

---

[1] Such a procedure is feasible if direct orthogonalization is used to prevent multicollinearity from making the computations unreliable.

[2] On the other hand if it is known that a predetermined variable has no relationship to a jointly dependent variable (i.e., the coefficient corresponding to the predetermined variable is zero in the reduced form equation containing the jointly dependent variable) it is possible that asymptotic efficiency will be increased by not adjusting all explanatory jointly dependent variables by the same set of instruments.

[3] Again the writer would like to emphasize that in the latter part of this section he is attempting to give methods which will give results which may be more readily interpreted and more easily computed than the methods commonly advocated in the case where it has already been determined that the space of the instruments is to be limited. The writer currently prefers estimates which coincide with DLS estimates rather than resorting to a restriction of the space of instruments merely to prevent the estimates obtained from coinciding with DLS estimates.

CHAPTER III

DISTURBANCE VARIANCE AND COEFFICIENT VARIANCE-COVARIANCE ESTIMATION

In this chapter we will first discuss estimation of the disturbance variance for all double k-class members satisfying $\text{plim}(k_1 - 1) = \text{plim}(k_2 - 1) = 0$ and then discuss coefficient variance-covariance estimation for the double k-class estimators satisfying $\text{plim}\sqrt{T}(k_1 - 1) = \text{plim}\sqrt{T}(k_2 - 1) = 0$ (which still includes 2SLS, LIML, UBK, and MSM). Finally, estimation of coefficient "t-ratios" is discussed.

A. Disturbance Variance Estimation

For any double k-class member for which $\text{plim}(k_1 - 1) = \text{plim}(k_2 - 1) = 0$, a consistent estimate of σ^2, the disturbance variance of the μ^{th} equation, is given by:[1]

$$\hat{\sigma}^2_{k_1,k_2} = \hat{u}'_{k_1,k_2}\hat{u}_{k_1,k_2}/(T - n) \tag{III.1}$$

where $\hat{u}'_{k_1,k_2}\hat{u}_{k_1,k_2}$ is the usual residual sum of squares, the residual being given by $\hat{u}_{k_1,k_2} = y - Z_\mu \hat{\delta}_{k_1,k_2}$.[2]

---

[1] To the writer's knowledge no formula (except the one given in Nagar [1962]) exists in the literature for the estimated disturbance variance of a double k-class estimator. (III.1) is consistent with the usual formula for single k-class estimators. A tentative proof of the consistency of $\hat{\sigma}^2_{k_1,k_2}$ calculated by (III.1) is given in appendix C.

[2] $\hat{u}'\hat{u}$ for any linear estimator may be calculated more simply, accurately, and with less computer time if calculated as $\hat{\delta}'[Z'_\mu Z_\mu]\hat{\delta}$ rather than by calculating $\hat{u}$ and then calculating the sum of squares of $\hat{u}$. That the two are mathematically equivalent is easily verified. $\hat{u}'\hat{u} = [y - Z_\mu\hat{\delta}]'[y - Z_\mu\hat{\delta}] = [Z_\mu\hat{\delta}]'[Z_\mu\hat{\delta}] = \hat{\delta}'[Z'_\mu Z_\mu]\hat{\delta}$. For notational simplicity, however, we will continue to write the residual sum of squares as $\hat{u}'\hat{u}$ rather than its computational form $\hat{\delta}'[Z'_\mu Z_\mu]\hat{\delta}$.

All of the specific double k-class members previously discussed meet the above plim requirement except DLS. This includes 2SLS, LIML, UBK, and MSM. Although the above does not provide a consistent estimate of σ^2 for DLS when multiple jointly dependent variables occur in the equation, it does agree with the usual formula used for calculating σ^2_{DLS} (which is almost surely biased downward when multiple jointly dependent variables occur in the equation).

The consistency of the estimator will not be changed if any denominator, d, with $\text{plim}(d/T) = 1$, is used in place of $T - n$. Thus, T may also be used as a denominator instead of $T - n$.

Regarding the Appropriate "Degrees of Freedom" for the Denominator

When estimating σ^2 by DLS, it is usual to use the following formula:

$$\hat{\sigma}^2_{DLS} = \hat{u}'_{DLS}\hat{u}_{DLS}/(T - n) \quad . \tag{III.2}$$

If only one jointly dependent variable occurs in the equation, (III.2) provides a consistent estimate of σ^2. If multiple jointly dependent variables occur in the equation, it is usual to continue to use (III.2) although this estimate is almost surely biased downward.[1] It would seem desirable to develop an estimate which takes account of the occurance of multiple jointly dependent

---

[1] Monte Carlo results of Cragg support this. See Cragg [1966] and Cragg [1967].

variables in the equation; however, the writer is not aware of any results of work in this area. The use of $\text{rk } X_I$ in an adjustment for "degrees of freedom" (e.g., a degrees of freedom of $T - \text{rk } X_I$) would not seem to be a fruitful approach for DLS, since X_I has no effect on the DLS coefficients.

On the other hand, when estimating by other double k-class members, the set of instruments affects the size of $\sigma^2_{k_1,k_2}$, since the instruments are actually used in the estimation procedure. Given a particular equation, the use of additional instruments will tend to reduce $\hat{u}'_{k_1,k_2}\hat{u}_{k_1,k_2}$; hence, the question immediately arises--should we reflect this by changing the "degrees of freedom" used as the denominator to something like $T - \Lambda$, $T - \text{rk } X_I$, or even $T - m - \text{rk } X_I$? All of these denominators will give consistent estimates of σ^2 under the assumptions $\text{plim}(k_1 - 1) = 0$ and $\text{plim}(k_2 - 1) = 0$.

By way of a partial answer, let us consider what happens to the $\hat{u}'_{k_1,k_2}\hat{u}_{k_1,k_2}$ of a given double k-class member calculated for a given equation, μ, and a fixed sample of size T if additional instruments are added to a given X_I matrix. If the additional instruments are not linear combinations of instruments already in the X_I matrix, the rank of the new X_I will increase and $\hat{u}'_{k_1,k_2}\hat{u}_{k_1,k_2}$ will tend to decrease. If sufficient predetermined variables are added that $\text{rk } X_I = T$ ($\text{rk } X_I$ cannot exceed T), $\hat{\delta}_{k_1,k_2}$ becomes equal to $\hat{\delta}_{DLS}$ and $\hat{u}'_{k_1,k_2}\hat{u}_{k_1,k_2}$ becomes equal to $\hat{u}'_{DLS}\hat{u}_{DLS}$. Thus, if the <u>maximum</u> number of predetermined variables which can affect the estimation are classified as being

in the system, $\hat{u}'_{k_1,k_2}\hat{u}_{k_1,k_2}$ will be decreased to $\hat{u}'_{DLS}\hat{u}_{DLS}$. The use of $T - rk\ X_I$ as the degrees of freedom is clearly inappropriate in this case since $T - rk\ X_I = 0$ and, therefore, $\hat{\sigma}_{k_1,k_2}$ would be arbitrarily large.

The above suggests that although some adjustment based on number of predetermined variables might be appropriate, the use of $T - rk\ X_I$ does not seem appropriate. The use of $T - rk\ X_I$ seems inappropriate also from the standpoint that it completely ignores the number of actual coefficients estimated--a factor which would seem to be of considerably more importance in any "degrees of freedom" adjustment than $rk\ X_I$. That the use of $T - n$ as the "degrees of freedom" adjustment gives satisfactory results in at least some cases for 2SLS and LIML is indicated by some of Cragg's Monte Carlo results.[1]

Some Work by Nagar

Nagar derived the bias to $0(T^{-1})$ in probability of $(1/T)\hat{u}'_k\hat{u}_k$ as:[2]

$$(III.3)\quad (1/T)\mathcal{E}(\hat{u}'_k\hat{u}_k) - \sigma^2 = -\sigma^2\{2 - [\Lambda - n - \varkappa - 1]\cdot tr(QC_1) - tr(QC) + (n/T)\}$$

where k is assumed nonstochastic with $k - 1 = 0(T^{-1})$ in

---

[1] Cragg's conclusions noted in section III.C pertain to estimated "t-ratios"; however, $\hat{\sigma}^2$ is used in this estimation.

[2] Nagar [1961]. The comparable formula for double k-class members is given in Nagar [1962].

probability; $\varkappa$ is related to k by the relation $k = 1 + (\varkappa/T)$, $\varkappa$ being assumed to be non-stochastic and independent of T; and the predetermined variables are assumed to be nonstochastic.[1] Q, C, and C_1 are matrices which must be estimated (see section II. C.3). $\mathcal{E}$ denotes expected value.

It might be suggested that σ^2 be estimated by the formula $(1/T)\hat{u}_k'\hat{u}_k$ and then (III.3) be used to adjust for bias; however, this does not seem to be a fruitful approach. First of all, the formula for bias contains σ^2 as a parameter. Thus, we are in the position of estimating the bias of a particular estimate of σ^2 with σ^2 itself being a parameter. It is possible to manipulate the above formula to eliminate σ^2 from the right hand side; however, the resulting formula is still not very helpful in adjusting for bias.

Especially noteworthy is the fact that unbiasedness is only one desirable characteristic (actually the above formula only gives an asymptotic estimate of bias to $O(T^{-1})$ in probability). Another important characteristic is dispersion, especially dispersion about the true value. The above formula is likely to give a large dispersion about any value (let alone the true value), since it contains traces of certain matrices which must be estimated, and the estimates of these traces (at least those suggested by Nagar) vary substantially in magnitude

---

[1] These are the same assumptions as were made in section II.C. 3 (MSM estimation). k nonstochastic means in this case that k is independent of any stochastic variable (in particular, k is independent of u). Although 2SLS, UBK, and MSM meet this requirement, LIML does not.

in response to only small changes in the data or model. Also, the particular estimates of these matrices may add a substantial error to our estimate of the bias.[1]

A Common Oversight in Estimating $\hat{u}'_{2SLS}\hat{u}_{2SLS}$

As noted in section II.B.2, 2SLS may be estimated as a two stage process in which $Y_{\|X_I}$ is calculated in the first stage, and in the second stage, $Y_{\|X_I}$ is substituted for Y in the calculation of $\hat{\delta}_{2SLS}$. Often, the error sum of squares from the second step is then used as $\hat{u}'_{2SLS}\hat{u}_{2SLS}$ in the calculation of $\hat{\sigma}_{2SLS}$ and , therefore, the calculation of coefficient standard errors is based on $Y_{\|X_I}$ instead of the original Y. This is generally regarded as a less desirable estimate than the estimator we have given, namely that based on $\hat{u}_{2SLS} = y - Y\hat{\gamma} - X_\mu\hat{\beta}$, or some equivalent formula. Thus, after calculating $\hat{\delta}_{2SLS}$, $\hat{u}'\hat{u}$ should be calculated by using the original Y in place of $Y_{\|X_I}$ or by the formula in the footnote on page 114.[2]

---

[1] Some estimates for these matrices are suggested in Nagar [1959].

[2] The $\hat{u}'_{2SLS}\hat{u}_{2SLS}$ obtained through recalculating may be either larger or smaller than the $\hat{u}'\hat{u}$ obtained by DLS at the second stage.

B. Coefficient Variance-Covariance Estimation[1]

1. Double k-class

For any double k-class member for which $\operatorname{plim} \sqrt{T}(k_1 - 1) = \operatorname{plim} \sqrt{T}(k_2 - 1) = 0$, a consistent estimate of $\operatorname{Var}(\hat{\delta}_{k_1,k_2})$ is

---

[1]Coefficient variance-covariance estimation for the double k-class estimators which we have considered other than DLS is complicated by the fact that the small sample sampling variance may be infinite in some cases. Fisher [1965], p. 605 states: "The principal point that has emerged on small sample properties of limited-information estimators is that the sampling variances involved are infinite, at least in some cases. Such a conclusion is borne out both from the analytic work that has been accomplished to date and by the results of the Monte Carlo experiments that have been performed." (Fisher's limited-information estimators include all of the specific k-class estimators treated in this paper except DLS. Fisher gives a number of analytical and Monte Carlo references.)

This does not imply that coefficient variance-covariance estimation is futile in finite samples. Basmann [1961], p. 621 states: "It is appropriate to mention that even though the *exact* distribution function $F(x)$ of some estimator fails to possess moments of lower order (say) a variance, it is still possible in many cases to *approximate* $F(x)$ by a distribution function $G(x)$ that does possess (say) a variance and even possesses moments of still higher order. Thus A.L. Nagar has made an important contribution to econometric statistics by working out formulas for the bias and moment matrices of *approximate* distributions for Theil's k-Class estimators," [reference to Nagar [1959] deleted]. "The reader will easily satisfy himself that Nagar's approximations do not depend on the *exact* distributions possessing a finite variance. Examine the *exact* frequency function exhibited in Figure 1 below. This frequency function does not possess a finite variance. Consider the approximation obtained by truncating the *exact* frequency function at the points $\nu_1 = -3$, $\nu_1 = 3$. The approximate frequency function obtained in this way possesses finite moments of all orders. The approximate distribution will be an excellent one, indeed."

given by:[1]

$$\text{(III.4)} \qquad \hat{\text{Var}}(\hat{\delta}_{k_1,k_2}) = \hat{\sigma}^2_{k_1,k_2} \begin{bmatrix} Y'Y - k_1[Y'Y]_{\perp X_I} & Y'X_\mu \\ X'_\mu Y & X'_\mu X_\mu \end{bmatrix}^{-1}$$

where $\hat{\sigma}^2_{k_1,k_2}$ is calculated by (III.1) or any other formula with plim $\hat{\sigma}^2_{k_1,k_2} = \sigma^2$ (e.g., T could be used in the denominator in place of T - n if desired).

(III.4) does not provide a consistent estimate of $\text{Var}(\hat{\delta}_{DLS})$ when multiple jointly dependent variables occur in the equation even though it does agree with the usual formula for calculating $\hat{\text{Var}}(\hat{\delta}_{DLS})$

---

[1] To the writer's knowledge no formula (except the one given in Nagar [1962]) exists in the literature for the estimated coefficient variance-covariance matrix of the double k-class estimator. If $k_1 = k_2$, (III.4) becomes the usual estimated coefficient variance-covariance formula for the single k-class estimator.

Christ [1966], p. 445 states that the asymptotic coefficient variance-covariance matrix of the double k-class estimator is the same as the asymptotic coefficient variance-covariance matrix of the 2SLS estimator provided plim $\sqrt{T}(k_1 - 1) =$ plim $\sqrt{T}(k_2 - 1) = 0$, but does not give a formula for $\hat{\text{Var}}(\hat{\delta}_{k_1,k_2})$. Let the formula given in (III.4) be denoted A. In appendix C it is tentatively shown [under the slightly less restrictive assumption plim$(k_1 - 1) =$ plim$(k_2 - 1) = 0$] that $(1/T)$plim T A equals the asymptotic coefficient variance covariance matrix of the 2SLS estimator; hence, under the assumption plim $\sqrt{T}(k_1 - 1) =$ plim $\sqrt{T}(k_2 - 1) = 0$, A [i.e., the formula given in (III.4)] is a consistent estimate of $\text{Var}(\hat{\delta}_{k_1,k_2})$. LIML, 2SLS, UBK, and MSM all meet this plim requirement.

(since $k_1 = 0$ for DLS). Cragg's Monte Carlo results indicate that this estimate of $\hat{\delta}_{DLS}$ has a substantial downward bias.[1]

Conversion of an instrumental variables problem to a 2SLS problem in the manner indicated in section II.D.2 also provides a convenient method of estimating $\text{Var}(\hat{\delta}_{IV})$, since (III.4) will give the same result as the usual IV formula.[2]

---

[1] Cragg's Monte Carlo results are considered in section III.C in the discussion of t-ratios.

[2] See Goldberger [1964], pp. 286, 332 for the usual formula--$\hat{\sigma}^2[X_{IV}'Z_\mu]^{-1}[X_{IV}'X_{IV}][Z_\mu'X_{IV}]^{-1}$--which Goldberger notes is consistent. That it is equivalent to the 2SLS formula for the converted IV problem may be noted from: $\hat{\sigma}^2[X_{IV}'Z_\mu]^{-1}[X_{IV}'X_{IV}][Z_\mu'X_{IV}]^{-1} = \hat{\sigma}^2[Z_\mu'X_{IV}(X_{IV}'X_{IV})^{-1}X_{IV}'Z_\mu]^{-1} = \hat{\sigma}^2[Z_\mu'Z_\mu]^{-1}_{\|X_{IV}}$ [see (I.36)], which is $\hat{\sigma}^2$ times the matrix inverted in the calculation of 2SLS if the IV problem is converted to a 2SLS problem (see II.39). If (as is usual) the variables in X_μ are also contained in X_{IV}, $\hat{\sigma}^2[Z_\mu'Z_\mu]^{-1}_{\|X_{IV}}$ becomes the more familiar

$$\hat{\sigma}^2\begin{bmatrix}[Y'Y]_{\|X_{IV}} & Y'X_\mu \\ X_\mu'Y & X_\mu'X_\mu\end{bmatrix}^{-1}$$ [see (II.45)].

2. Alternative estimate for LIML

An asymptotic coefficient variance-covariance matrix for LIML has been derived by Rubin as:[1]

$$\text{(III.5)} \quad \hat{\text{Var}}(\hat{\delta}_{\text{LIML}}) = \hat{\sigma}^2_{\text{LIML}} \begin{bmatrix} Y'Y - [Y'Y]_{\perp X_I} + \frac{(k-1)}{\hat{\gamma}' {}_+f} {}_+f\,{}_+f' & Y'X_\mu \\ X'_\mu Y & X'_\mu X_\mu \end{bmatrix}^{-1}$$

where k is the smallest eigenvalue of $[Y'Y]^{-1}_{\perp X_I}[Y'Y]_{\perp X_\mu}$ as before, ${}_+f = [{}_+Y' {}_+Y]_{\perp X_I} \hat{\gamma}_{\text{LIML}}$, and f is the same as ${}_+f$ except that the element corresponding to the normalizing variable is deleted, i.e., $f = [{}_+Y'Y]_{\perp X_I} \hat{\gamma}_{\text{LIML}}$.

(III.5) is given for completeness. (III.4) would seem to be as desirable and has the further advantages:

(1) $\text{Var}(\hat{\delta}_{\text{LIML}})$ estimated by formula (III.4) provides an estimate which can be compared more readily to the $\text{Var}(\hat{\delta}_k)$ of other k-class members.

(2) (III.4) may be obtained as a by-product of the calculation of $\hat{\delta}_{\text{LIML}}$ whereas (III.5) requires special programming.

It should be noted that $\hat{\text{Var}}(\hat{\delta}_{\text{LIML}})$ by formula (III.4) $\neq \hat{\text{Var}}(\hat{\delta}_{\text{LIML}})$ by formula (III.5).

---

[1] Personal conference. The estimate given in Chernoff and Divinsky [1953], p. 245 is the same as (III.5) as may be seen by writing down the formulas for inverting the matrix given in (III.5) in parts and noting that the result is exactly the same as the formula given by Chernoff and Divinsky.

3. Nagar's unbiased to $O(T^{-2})$ in probability estimates

Nagar derived the moment matrix of $\hat{\delta}_k - \delta$ and $\hat{\delta}_{k_1,k_2} - \delta$ to $O(T^{-2})$ in probability. The estimated coefficient variance-covariance matrix could be based on the formulas derived by Nagar. (Most of the matrices in his formulas are based on population parameters, but estimates of these matrices of the type which he suggests in the calculation of $\hat{\delta}_{MSM}$ could be used.) Those interested in following this approach are referred to Nagar's articles.[1]

In evaluating whether to follow the approach of estimating a number of population matrices and substituting them into the formulas one should be reminded that:

(1) Assumptions additional to those which we have specified are imposed in the derivation of Nagar's formulas.

(2) Although the resulting formulas are of a higher order of unbiasedness (assuming that the actual population parameters are available) than the usual formulas, they are still asymptotic.

(3) Unbiasedness is only one desirable property. Dispersion, especially about the true value, is a property which is also very important. Nagar's derivations are in terms of certain population matrices and traces of other population matrices. The estimation of these matrices and traces and their substitution into Nagar's formulas are likely to add greatly to the dispersion of the estimated coefficient variance-covariance matrix.

---

[1] Nagar [1959], and Nagar [1962].

C. Coefficient Standard Errors and t-ratios

The square roots of the diagonal elements of the estimated coefficient variance-covariance matrix (i.e., the square roots of the estimated coefficient variances) are often used as approximate coefficient standard errors and the ratios of the coefficients to the square roots of the estimated coefficient variances are often used as approximate coefficient t-ratios; however, very little information is available on how well these computed values serve as approximate standard errors and approximate t-ratios.

Cragg [1966] and [1967] reported the results of a Monte-Carlo experiment involving DLS, 2SLS, UBK, LIML, 3SLS, and FIML. The coefficient matrix for the basic model was as follows:[1]

$$\begin{bmatrix} -1 & \gamma_{12} & \gamma_{13} & \beta_{11} & \beta_{12} & 0 & 0 & \beta_{15} & 0 & 0 \\ \gamma_{21} & -1 & 0 & \beta_{21} & 0 & \beta_{23} & 0 & \beta_{25} & 0 & \beta_{27} \\ 0 & \gamma_{32} & -1 & \beta_{31} & 0 & \beta_{33} & \beta_{34} & 0 & \beta_{36} & 0 \end{bmatrix}$$

---

[1] Cragg [1967], p. 94. The large number of abandoned samples casts considerable doubt on the meaningfullness of the results; however, the obviously very large amount of rounding error which was encountered surely had a considerably smaller effect on the calculation of the single equation estimates than for the 3SLS and FIML estimates. From Cragg's description of the FIML results, it would appear that they should be totally ignored due to excessive rounding error in their computation. In addition to the large number of abandoned samples, a number of FIML estimates were retained even though their computed coefficient variances were negative. This indicates either convergence to a saddle point instead of a local maximum for many problems (convergences of FIML is discussed in chapter V) or that a high degree of rounding error was encountered (or both). (The writer is not suggesting that additional samples should have been eliminated based on the FIML results, but that the entire set of FIML results should have been ignored.)

The results reported in Cragg [1967] consisted of 26 experiments, each containing 50 samples, the experiments being summarized in the following table:[1]

EXPERIMENTS CONDUCTED

Experiment	Special Features*	Abandoned samples
1	None	1
2	Disturbance set 2	0
3	Disturbance set 3	1
4	Exogenous variable set 2	0
5	Exogenous variable set 3	1
6	Structure 2	0
7	Structure 3	8
8	Structure 4	1
9	Structure 5	1
10	Structure 6	0
11	Structure 7	0
12	Structure 8	0
13	Values of Σ 25% those of Table I	3
14	Values of Σ 4 times those of Table I	3
15	Values of Σ 9 times those of Table I	4
16	Values of Σ 16 times those of Table I	6
17	Values of Σ 25 times those of Table I	10
18	35 observations	0
19	50 observations	0
20	70 observations	0
21	Multicollinearity 1	1
22	Multicollinearity 2	0
23	Multicollinearity 3	2
24	Multicollinearity 4	6
25	Multicollinearity 5	8
26	Multicollinearity 6	7

*Unless otherwise noted twenty observations were used with no specially introduced multicollinearity in exogenous variable data 1, structural disturbance set 1, and structure 1.

[1] Cragg [1967], p. 94.

A formula equivalent to (II.4) was used as the formula for estimating the coefficient variance-covariance matrix; that is, a degrees of freedom adjustment of T - 5 was used for all samples. Approximate t-ratios were calculated as the coefficients divided by the square roots of the diagonal elements of the coefficient variance-covariance matrices.

In reporting the results of the 26 experiments Cragg states:

> "The adequacy of the standard errors was investigated by examining the ratios of the deviations of the coefficients from the true values to their standard errors, which we call for simplicity the t ratios. It is sometimes supposed that t ratios are distributed as Student's t. In investigating this supposition there were two difficulties: (1) should the standard errors be adjusted for 'lost degrees of freedom' and (2) what is the appropriate number of degrees of freedom for the t distribution. After examining some of the data it appeared that the standard errors of a coefficient in a particular structural equation should be adjusted for the number of coefficients to be estimated in that equation." [reference to footnote deleted] "The most appropriate number of degrees of freedom appeared to be the number of observations minus the number of coefficients to be estimated in the equation in which the coefficient fell. The hypothesis investigated was that not more than five per cent of the t ratios would fall outside the ninety-five percent confidence intervals for roughly five percent of the coefficients in most of the experiments. The number of the consistent t ratios falling outside the interval was significantly higher than five per cent for only one or two coefficients in most experiments and quite often there was none."[1]

Also, Cragg states:

> "The DLS standard errors were not apt to be reliable for making inferences about the true values of the structural coefficients. Much more frequently than for the consistent methods, the number of DLS t ratios falling outside the ninety-five per cent confidence intervals was significantly greater than five per cent of the total number of estimates of a coefficient."[2]

---

[1] Cragg [1967], p. 101.

[2] Cragg [1967], p. 102.

As one of his conclusions, Cragg states:

> "Usually use of the standard errors of the consistent methods would lead to reliable inferences, but this was not always the case. The standard errors of DLS were not useful for making inference about the true values of the coefficients."[1]

Cragg reported some additional experiments in which the model noted above was modified to examine the effect of (1) errors in the exogenous variables, (2) stochastic coefficients, and (3) heteroskedastic and autocorrelated disturbances. Results similar to those noted above are reported for these additional experiments.[2]

---

[1] Cragg [1967], p. 109.

[2] See Cragg [1966].

CHAPTER IV

GENERALIZED LEAST SQUARES

A. Unrestricted Generalized Least Squares (GLS)

The generalized least squares model (also called the Aitken model) may be expressed as:[1]

$$\underset{T\times 1}{y} = \underset{T\times n}{X}\ \underset{n\times 1}{\delta} + \underset{T\times 1}{u} \tag{IV.1}$$

where the same statistical assumptions are made as were made in Chapter I except that

(1) $\mathcal{E}uu' = \dot{\Sigma}$, where $\dot{\Sigma}$ is a $\dot{T}\times\dot{T}$ positive definite matrix known except for a multiplicative constant.

If (IV.1) represents a single stochastic equation from a system of stochastic equations, then $\dot{T} = T$ and $\mathcal{E}uu' = \dot{\Sigma}$ is a loosening of the assumption made earlier in this paper that $\mathcal{E}uu' = \sigma^2 I$. (The dot is used above the Σ to insure that the $\dot{\Sigma}$ matrix is not confused with the Σ matrix which is the $M\times M$ disturbance variance-covariance matrix for a system of M equations.)

(2) X is a matrix of variables assumed fixed rather than merely contemporaneously uncorrelated with u.

The X matrix is not the same as the X matrix of preceding chapters. If (IV.1) represents a single equation from a system of equation then $\dot{T} = T$ and the X matrix is the same as the X_μ or Z_μ matrix. In part II of this paper we will, at times,

---

[1] Aitken [1934-35], pp. 42-43.

rewrite an entire set of M stochastic equations in the form (IV.1). In this case $\dot{T} = M \cdot T$, y and u will become $MT \times 1$ vectors, X will become an $MT \times (\sum_{\mu=1}^{M} n_\mu)$ matrix constructed from the X_μ matrices and matrices of zeros, and $\dot{\Sigma}$ will become an $MT \times MT$ matrix. (The ZA [Zellner-Aitken] and the 3SLS [three-stage least squares] models will be derived as modifications of the GLS model.)

(3) In the GLS model, X is assumed to have full column rank. This assumption will be relaxed in the next section when we consider the RGLS (restricted generalized least squares) model.

The GLS estimator is:

$$\hat{\delta}_{GLS} = [X'\dot{\Sigma}^{-1}X]^{-1}[X'\dot{\Sigma}^{-1}y] \tag{IV.2}$$

and the variance of $\hat{\delta}_{GLS}$ is given by:

$$\text{Var}(\hat{\delta}_{GLS}) = [X'\dot{\Sigma}^{-1}X]^{-1} \ . \tag{IV.3}$$

Under the above statistical assumptions, $\hat{\delta}_{GLS}$ is the minimum variance linear unbiased estimator. Quite a few applications of the GLS model will be made in the remainder of this paper; however, some of the assumptions will not be met for these applications. Even in these applications, although $\hat{\delta}_{GLS}$ will no longer be best linear unbiased, $\hat{\delta}_{GLS}$ may still have desirable properties. Even though some of the assumptions are not met, we will still refer to estimates using (IV.2) and (IV.3) as GLS estimates.

B. Restricted Generalized Least Squares (RGLS)

The restricted generalized least squares (RGLS) model is the same as the GLS model except that the following restrictions are imposed on the coefficients:

$$\underset{N_R \times n}{R}\ \underset{n \times 1}{\delta} = \underset{N_R \times 1}{r} \tag{IV.4}$$

where R is an $N_R \times n$ matrix of known elements and r is an $N_R \times 1$ vector of known elements.[1]

In the GLS model, X is assumed to have full column rank. This assumption is relaxed in the RGLS model. A corresponding necessary (but not sufficient) condition for the RGLS model is that $\text{rk } X + \text{rk } R \geq n$.[2] R need not have full row rank; that is, $\text{rk } R$ may be less than N_R (i.e., redundent restrictions are permitted in R).

The RGLS estimator is given by:[3]

$$\hat{\delta}_{RGLS} = Q\{[Q'(X'\dot{\Sigma}^{-1}X)Q]^{-1}Q'[(X'\dot{\Sigma}^{-1}y) - (X'\dot{\Sigma}^{-1}X)q]\} + q \tag{IV.5}$$

or, equivalently by:

---

[1] N_R may be greater than, equal to, or less than n. If $N_R > n$, then (1) some of the restrictions are redundant, (2) some of the restrictions are inconsistent, and/or (3) δ is restricted to a fixed set of coefficients. The computational procedure given in this chapter detects but allows redundancy. Inconsistency is also detected.

[2] $\text{rk } R$ of the coefficients in δ may be solved for in terms of the remaining $n - \text{rk } R$ coefficients. If $\text{rk } X$ is not greater than or equal to $n - \text{rk } R$, then the remaining $n - \text{rk } R$ coefficients (and, therefore, the $\text{rk } R$ coefficients) will not be unique.

[3] To the writer's knowledge, these forms of the RGLS estimator have not appeared in the literature. A more common formula will be presented later (IV.23) and some advantages of (IV.5) or (IV.6) and (IV.7) over (IV.23) there discussed. The proofs of (IV.5), (IV.6) and (IV.7) are given after Q and q are defined (via their computation).

$$\text{(IV.6)} \qquad \hat{\delta}^{(1)}_{RGLS} = [Q'(X'\dot{\Sigma}^{-1}X)Q]^{-1}Q'[(X'\dot{\Sigma}^{-1}y) - (X'\dot{\Sigma}^{-1}X)q]$$

and

$$\text{(IV.7)} \qquad \hat{\delta}^{(2)}_{RGLS} = Q_2\hat{\delta}^{(1)}_{RGLS} + q_2 \ .$$

where:

Q_2 is a rk R$^{\times}$(n - rk R) matrix derived from R by reducing R to essentially a row echelon form by a series of row operations, then forming Q_2 as the negative of the resulting row echelon matrix (possibly rearranged slightly).

Q is an n$^{\times}$(n - rk R) matrix formed as Q_2 augmented by an (n - rk R)$^{\times}$(n - rk R) identity matrix (and the rows possibly rearranged).

q_2 is a rk R$^{\times}$1 vector derived from r by performing the row operations on the augmented matrix $[R \vdots r]$ instead of on R alone.

q is an n$^{\times}$1 vector formed as q_2 augmented by an (n - rk R)$^{\times}$1 vector of zeros (and the elements possibly rearranged to conform with the rearrangement of the rows of Q).

$\hat{\delta}^{(1)}_{RGLS}$ is composed of the elements of $\hat{\delta}_{RGLS}$ corresponding to the (n-rk R)$^{\times}$1 vector of zeros added to q_2 in forming q.

$\hat{\delta}^{(2)}_{RGLS}$ is composed of the remaining rk R elements of $\hat{\delta}_{RGLS}$.

The use of (IV.6) and (IV.7) is equivalent to (1) solving (IV.4) for rk R of the coefficients in terms of the remaining n - rk R coefficients; (2) substituting this solution for the rk R coefficients into (IV.1), and rewriting (i.e., redefining variables) so that in effect an unrestricted GLS model with n - rk R coefficients to be

estimated is obtained; (3) estimating these $n - rk\ R$ coefficients by the GLS formula (IV.2); and (4) substituting these estimates back into the solution of step (1) to obtain the estimates of the $rk\ R$ coefficients which were there "solved out."

The use of formula (IV.5) is equivalent to the use of the two separate formulas (IV.6) and (IV.7). Computational procedures for forming Q, q, Q_2 and q_2 are given in the next section. The precise difinition of these matrices is given by their computational procedure.

The variance-covariance matrix for $\hat{\delta}_{RGLS}$ is:[1]

$$\text{(IV.8)} \qquad \text{Var}(\hat{\delta}_{RGLS}) = Q[Q'(X'\dot{\Sigma}^{-1}X)Q]^{-1}Q' \quad .$$

---

[1] A derivation of (IV.8) is given after the derivation of $\hat{\delta}_{RGLS}$.

1. Computation of Q and q

First of all, the augmented matrix $[R \vdots r]$ is formed. Any row operation which is performed on R will be performed on r as well. The matrix Q and the vector q may be formed as follows:[1]

1st Series of Steps -- Reduction of R to Row Echelon Form

(1a) Let abs $R_{i_1 j_1}$ be the largest element in absolute value of R. If abs $R_{i_1 j_1}$ is less than ε_1 go to step m. (ε_1 is explained in step m.) Otherwise switch rows 1 and i_1 of the augmented matrix and columns 1 and j_1 of this matrix so that the largest element occurs in column 1 of row 1. Record the order of the new rows and columns in terms of the order of the original rows and columns of R.

(1b) Perform row operations on the resulting augmented matrix to reduce the first column to column 1 of an $N_R \times N_R$ identity matrix. Denote the resulting augmented matrix as $[I^{(1)} \vdots R^{(1)} \vdots r^{(1)}]$, where $I^{(1)}$ is the first column of an $N_R \times N_R$ identity matrix.

---

[1] As indicated in Part III (Programming Considerations), to reduce rounding error, all variables should be normalized so that all elements of the Z'Z matrix will be of comparable magnitude and unaffected by the multiplication of any variable by a positive constant such as a power of 10 (i.e., unaffected by shifting the decimal point of the variable). Normalization such that the variables inherent in the Z'Z matrix all have length 1 or their deviations from means have length 1 is suggested. Thus, a step which should precede the first step outlined in this section is to normalize the columns of R and r to take account of the normalization of variables. The scaling of variables up or down by a user will then have no effect on the normalized R matrix.

If the elements in the rows of R and r differ greatly in magnitude, the R matrix and r vector should also be normalized rowwise to reduce rounding error. (Multiplication of any row of $[R \vdots r]$ by a constant does not change the restriction.)

(2a) Let abs $R_{i_2 j_2}$ be the largest element in absolute value in rows 2 through N_R of $R^{(1)}$. If abs $R_{i_2 j_2}$ is less than ϵ_2, go to step m. Otherwise switch rows 2 and i_2 of the augmented matrix and columns 2 and j_2 of this matrix so that this element occurs in column 2 of row 2. Record the order of the new rows and columns in terms of the order of the original rows and columns of R.

(2b) Perform row operations on the resulting augmented matrix to reduce the second column to column 2 of a $N_R \times N_R$ identity matrix. Denote the resulting augmented matrix as $[I^{(2)} : R^{(2)} : r^{(2)}]$, where $I^{(2)}$ is the first two columns of a $N_R \times N_R$ identity matrix. In general, at the k^{th} step:

(ka) Let abs $R_{i_k j_k}$ be the largest element in absolute value of $R^{(k-1)}$. If abs $R_{i_k j_k}$ is less than ϵ_k, go to step m. Otherwise, switch rows k and i_k of the augmented matrix and columns k and j_k of this matrix so that this element occurs in column k of row k. Record the order of the new rows and columns in terms of the order of the original rows and columns of R.

(kb) Perform row operations on the resulting augmented matrix to reduce the k^{th} column to column k of an $N_R \times N_R$ identity matrix. Denote the augmented matrix as $[I^{(k)} : R^{(k)} : r^{(k)}]$, where $I^{(k)}$ is the first k columns of a $N_R \times N_R$ identity matrix.

(m) The procedure is continued until either (1) all N_R rows have been treated (i.e., $I^{(k)}$ is an $N_R \times N_R$ identity matrix) in which case R has full row rank, i.e., rk R = N_R; or (2) at the m^{th} step, $R_{i_m j_m} < \epsilon_m$ in which case rk R = m - 1. Let us assume

the latter which we will call step m. Thus, at step m we have:[1]

$$(IV.9)\quad [I^{(m-1)} \vdots R^{(m-1)} \vdots r^{(m-1)}] =$$

$$\left[\begin{array}{c|c|c} \underset{\text{rk R}\times\text{rk R}}{I} & \underset{\text{rk R}\times(n-\text{rk R})}{A_1} & \underset{\text{rk R}\times 1}{b_1} \\ \hline \underset{(N_R-\text{rk R})\times\text{rk R}}{0} & \underset{(N_R-\text{rk R})\times(n-\text{rk R})}{A_2} & \underset{(N_R-\text{rk R})\times 1}{b_2} \end{array}\right].$$

(If rk R = N_R, the matrix $[0 \vdots A_2 \vdots b_2]$ will not occur.)

If no rounding error occurred, and if the remaining rows were exact linear combinations of the preceding m - 1 rows, then abs $R_{i_m j_m}$ would be zero. We must, however, allow for the possibility of rounding error; hence, we can detect R having less than full row rank only if we consider an ϵ_k which is greater than zero at each step. A preset $\epsilon_1 = \epsilon_2 = \cdots = \epsilon_{N_R}$ can be assumed or calculated before the procedure is started, or an ϵ_k can be calculated at each step to reflect the number of operations performed.

---

[1] During the calculation of $[I^{(m-1)} \vdots R^{(m-1)}]$, columns were rearranged; therefore, the column corresponding to a given coefficient in δ will have been moved. Suppose that the coefficients of δ are now rearranged so that each coefficient will be in the same order as its corresponding column of $[I^{(m-1)} \vdots R^{(m-1)}]$. Let us designate δ in its rearranged order as δ^*. Let us also rearrange the columns of R so that they are in the same order as their corresponding coefficients in δ^* (same order as the columns of $[I^{(m-1)} \vdots R^{(m-1)}]$) and let us delete from R and r the rows (if any) corresponding to $[0 \vdots A_2 \vdots b_2]$. Let us designate the new matrix obtained as R* and the new vector as r*. Then the original set of restrictions could be rewritten as

$$R^*\delta^* = r^* \text{ or } R_1^*\delta^{(1)*} + R_2^*\delta^{(2)*} = r^*$$

where R_1^* is a rk R×rk R matrix. By our method of calculation, $[I \vdots A_1 \vdots b_1] = (R_1^*)^{-1}[R^* \vdots r^*] = (R_1^*)^{-1}[R_1^* \vdots R_2^* \vdots r^*]$, i.e., $A_1 = R_1^{*-1}R_2^*$ and $b_1 = (R_1^*)^{-1}r^*$.

rk $R < N_R$ may occur in either of two cases:

(1) The remaining $N_R - m + 1$ restrictions are implied by (are linear combinations of) the preceding $m - 1$ restrictions treated; therefore, they can be ignored. In this case, b_2 will be approximately a vector of zeros. (If no rounding error occurred, b_2 would be exactly a vector of zeros; however, due to rounding error we should compare the absolute value of the elements of b_2 with some positive constant.)

(2) At least one of the remaining $N_R - m + 1$ restrictions is inconsistent with the preceding $m - 1$ restrictions treated; hence, action should be taken by the user to remove the inconsistency. Inconsistency is detected by comparing the absolute value of the elements of b_2 with a small positive constant. $abs(b_2)_i > \epsilon_{(b_2)_i}$ implies that row i of $[0 \vdots A_2 \vdots b_2]$ is inconsistent with the rows of $[I \vdots A_1 \vdots b_1]$. Since we recorded the row number of the original matrix R corresponding to each row of these augmented matrices, the set of equations (in terms of the row numbers of R) with which this equation is inconsistent can be noted so that the set of restrictions can be corrected by the user. Even after finding one i for which $abs(b_2)_i > \epsilon_{(b_2)_i}$, the remaining $(b_2)_i$ should be checked so that all inconsistencies are noted for correction.

2nd Series of Steps--Formation of Q *and* q *from* A_1 *and* b_1

Form a matrix Q^* and a vector q^* as:

$$(IV.10) \qquad \underset{n\times(n-rk\ R)}{Q^*} = \begin{bmatrix} -A_1 \\ rk\ R \times (n-rk\ R) \\ \\ I \\ (n-rk\ R)\times(n-rk\ R) \end{bmatrix}; \qquad \underset{n\times 1}{q^*} = \begin{bmatrix} b_1 \\ rk\ R \times 1 \\ \\ 0 \\ (n-rk\ R)\times 1 \end{bmatrix}.$$

In forming the matrix $[I^{(m-1)} : R^{(m-1)}]$ we rearrange the columns of R noting their revised order in terms of their original order. If the rows of Q^* and q^* are considered to be in the revised order (the same order as the columns of $[I^{(m-1)} : R^{(m-1)}]$) then the matrices Q and q can be formed from Q^* and q^* by rearranging the rows of Q^* and q^* so that they are in the same order as the original columns of R. Thus,

$Q = Q^*$ with the rows arranged to the original order of the columns of R.

$q = q^*$ with the rows arranged to the original order of the columns of R.

The columns of Q need not be rearranged, since they correspond to the rows of R, and the order of the rows of R is of no consequence.

Computation of Q_2, q_2

The above procedure also gives a method of separating out $n-rk\ R$ coefficients which may be estimated directly and then the remaining $rk\ R$ coefficient estimates solved for from these estimates.

Let Q_2 be formed from Q and q_2 be formed from q by deleting the $(n - rk\ R)\times(n - rk\ R)$ identity submatrix of Q and the

corresponding rows of q.[1] Let $\hat{\delta}^{(1)}$ be the coefficients corresponding to the rows of the identity submatrix which were deleted from the Q matrix and let $\hat{\delta}^{(2)}$ be the coefficients corresponding to the rows of Q_2, then $\hat{\delta}^{(1)}_{RGLS}$ may be calculated directly by (IV.6) and $\hat{\delta}^{(2)}_{RGLS}$ may be calculated by (IV.7) from Q_2, q_2, and $\hat{\delta}^{(1)}_{RGLS}$.

In the calculation of Q and q, it is not actually necessary to search for the largest element in the remaining submatrix at any step. The rows could be taken in turn and the first non-zero element encountered used for $R_{i_k j_k}$; however, the extra searching and non-sequential selection of rows will reduce rounding error for many problems.

Proof of (IV.5), (IV.6), and (IV.7)

The method of deriving Q, Q_2, q, and q_2 forms most of the proof of the formulas. In the first computational method, we reduced the $[R \vdots r]$ matrix to the reduced augmented matrix $\begin{bmatrix} I & A_1 & b_1 \\ 0 & A_2 & b_2 \end{bmatrix}$, where A_2 and b_2 are within rounding error of zero (assuming no inconsistency in the original set of restrictions). Only row operations were used in reducing the $[R \vdots r]$ matrix; therefore, the above reduced augmented matrix incorporates the full set of restrictions. In fact, if A_2 and b_2 are within rounding error of zero, the full set of restrictions are contained in the $[I \vdots A_1 \vdots b_1]$ matrix which contains $rk\ R$ rows. Thus, the restrictions may be expressed as:

$$[I \vdots A_1]\delta^* = b_1 \tag{IV.11}$$

---

[1] Q_2 may be formed directly from $-A_1$ and q_2 may be formed directly from b_1 be rearranging the rows of $-A_1$ and b_1, respectively.

where the * denotes that the coefficients in the δ vector have been rearranged into the same order as the columns of $[I \vdots A_1]$. For notational convenience, let us respecify the order of variables in the original problem so that they are in the same order as the columns of $[I \vdots A_1]$. Thus, under the renumbering, $\delta^* = \delta$ and $b_1 = q_2$. (IV.11) may be rewritten as:

$$(IV.12) \qquad [I \vdots A_1]\delta = [I \vdots A_1]\begin{bmatrix}\delta^{(2)}\\ \delta^{(1)}\end{bmatrix} = \delta^{(2)} + A_1\delta^{(1)} = q_2$$

or

$$(IV.13) \qquad \delta^{(2)} = -A_1\delta^{(1)} + q_2 = Q_2\delta^{(1)} + q_2 \ ,$$

since $Q_2 = -A_1$ under our revised numbering of variables. Our basic model is now:

$$(IV.14) \qquad y = [X_1 \vdots X_2]\begin{bmatrix}\delta^{(1)}\\ \\ \delta^{(2)}\end{bmatrix} + u$$

subject to

$$(IV.15) \qquad \delta^{(2)} = Q_2\delta^{(1)} + q_2$$

or, substituting for $\delta^{(2)}$ into (IV.14):

$$(IV.16) \qquad y = [X_1 \vdots X_2]\begin{bmatrix}\delta^{(1)}\\ \\ Q_2\delta^{(1)}+q_2\end{bmatrix} + u$$

or

$$(IV.17) \qquad y - X_2q_2 = [X_1 \vdots X_2]\begin{bmatrix}I\\ \\ Q_2\end{bmatrix}\delta^{(1)} + u$$

or

(IV.18) $$y - Xq = XQ\delta^{(1)} + u,$$

since $q = \begin{bmatrix} 0 \\ q_2 \end{bmatrix}$ and $Q = \begin{bmatrix} I \\ Q_2 \end{bmatrix}$ under our revised numbering.

Applying the GLS formula to (IV.18) [letting $y - Xq$ and XQ be the y and X respectively of (IV.2)] we get:

(IV.19)
$$\begin{aligned} \hat{\delta}^{(1)} &= [Q'X'\dot{\Sigma}^{-1}XQ]^{-1}[Q'X'\dot{\Sigma}^{-1}(y - Xq)] \\ &= [Q'X'\dot{\Sigma}^{-1}XQ]^{-1}[Q'X'\dot{\Sigma}^{-1}y - Q'X'\dot{\Sigma}^{-1}Xq] \end{aligned}$$

which is exactly formula (IV.6).

If we replace $\delta^{(1)}$ in (IV.5) by $\hat{\delta}^{(1)}$, we have $\hat{\delta}^{(2)}$ expressed as in formula (IV.7). Further,

(IV.20) $$\hat{\delta} = \begin{bmatrix} \hat{\delta}^{(1)} \\ \hat{\delta}^{(2)} \end{bmatrix} = \begin{bmatrix} I\hat{\delta}^{(1)} \\ Q_2\hat{\delta}^{(1)} \end{bmatrix} + \begin{bmatrix} 0 \\ q_2 \end{bmatrix} = Q\hat{\delta}^{(1)} + q \ ;$$

hence, substituting for $\hat{\delta}^{(1)}$ we get:

$$\hat{\delta} = Q[Q'X' \dot{\Sigma}^{-1}XQ]^{-1}[Q'X' \dot{\Sigma}^{-1}y - Q'X' \dot{\Sigma}^{-1}Xq] + q \ ,$$

which is exactly (IV.5).

To derive the variance formula (IV.8) we note that (by IV.20):[1]

(IV.21) $$Var(\hat{\delta}) = Q[Var(\hat{\delta}^{(1)})]Q' \ .$$

But by the GLS formula for variance (IV.3), substituting XQ from (IV.18) for the X in (IV.1):

(IV.22) $$Var(\hat{\delta}^{(1)}) = [Q'X'\dot{\Sigma}^{-1}XQ]^{-1} \ .$$

Substituting this into (IV.21) we get (IV.8).

---

[1] If $y = Ax + b$ with A a matrix of fixed elements and b a vector of fixed elements, then $Var(y) = A[Var(x)]A'$.

2. Relationship to another restriction formula

If the further assumptions that R has full row rank and X has full column rank are imposed, an alternative formula for $\hat{\delta}_{RGLS}$ is given by:

$$\text{(IV.23)} \quad \hat{\delta}_{RGLS} = \hat{\delta}_{GLS} - (X'\dot{\Sigma}^{-1}X)^{-1}R'[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}[R\hat{\delta}_{GLS} - r]$$

In order to show the relationship between this and our previous formula for $\hat{\delta}_{RGLS}$ we will derive (IV.23) from (IV.5) using the above additional assumptions ($\text{rk } R = N_R$ and $\text{rk } X = n$).[1]

Derivation of (IV.23)

Since $\text{rk } R = N_R$, we may reorder columns and variables if necessary and partition R as $[R_1 \vdots R_2]$ with R_1 square and nonsingular. Then application of row operations to reduce $R = [R_1 \vdots R_2]$ to $[I \vdots A_1]$ is equivalent to premultiplying R by a non-singular matrix C with C such that $CR_1 = I$. Then,

$$\text{(IV.24)} \quad RQ = IRQ = C^{-1}CRQ = C^{-1}[I \vdots A_1]\begin{bmatrix} -A_1 \\ \\ I \end{bmatrix} = C^{-1}\cdot[-A_1 + A_1] = C^{-1}\cdot 0$$

$$= 0 \quad .$$

---

[1] A method for deriving (IV.23) using Lagrangian multipliers is given in Stroud and Zellner [1962]. Another derivation of (IV.23) is given in Chipman and Rao [1964a].

The Ψ matrix and α vector of Chipman and Rao are the same as our R matrix and r vector respectively. Our Q matrix satisfies the requirements for their Φ matrix and our q vector satisfies the requirements for their $\tilde{G}\alpha$ vector. (Chipman and Rao do not present an actual computational scheme for Φ, $\tilde{G}$, or α.) Using these substitutions, the essence of (IV.5) is contained as an intermediate step of Chipman and Rao's derivation of (IV.23). Since the proof of (IV.5) given in this paper differs from Chipman and Rao's, the derivation of (IV.23) from (IV.5) differs from Chipman and Rao's also.

Thus, the columns of Q are orthogonal to the rows of R.[1]

Before going further we will find it useful to establish the following lemma:

(IV.25) *Lemma:* Let E and F be square symmetric idempotent matrices of order p with $EF = FE = 0$, $\text{rk } E = n$ and $\text{rk } F = p - n$; then, $E + F = I$.

Proof: Let $E + F = G$. Then G is symmetric idempotent and has rank $n + (p - n) = p$.[2] Thus, G has full row and column rank; hence, G^{-1} exists.

$$E + F = G = GI = G(GG^{-1}) = GG(G^{-1}) = GG^{-1} = I$$

(End of Proof of Lemma)

From (IV.5) we have:

$$\text{(IV.26)} \quad \hat{\delta}_{RGLS} = Q[Q'X'\dot{\Sigma}^{-1}X\,Q]^{-1}[Q'X'\dot{\Sigma}^{-1}y - Q'X'\dot{\Sigma}^{-1}Xq] + q$$

$$= Q[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}Q'(X'\dot{\Sigma}^{-1}X)[(X'\dot{\Sigma}^{-1}X)^{-1}X'\dot{\Sigma}^{-1}y - q] + q$$

$$= Q[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}Q'(X'\dot{\Sigma}^{-1}X)[\hat{\delta}_{GLS} - q] + q$$

Let $E = Q[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}Q'(X'\dot{\Sigma}^{-1}X)$ and

$$F = (X'\dot{\Sigma}^{-1}X)^{-1}R'[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}R\ .$$

Then, E and F are symmetric idempotent with $EF = FE = 0$, since:[3]

---

[1] By the definition of orthogonal complement given in Koopmans, Rubin and Leipnik [1950], p. 89, Q is the orthogonal complement of R; that is, Q has full column rank and satisfies $RQ = 0$.

[2] For symmetric idempotent matrices E and F: $(E + F)^2 = E + F$ if and only if $EF = FE = 0$. Also, $EF = FE = 0$ implies $\text{rk}(E + F) = \text{rk } E + \text{rk } F$. See Chipman and Rao [1964b].

[3] Each underlined expression is a matrix times its inverse, hence an identity matrix which may be suppressed.

(1) $EE = Q[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}\underline{Q'(X'\dot{\Sigma}^{-1}X)Q[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}}Q'(X'\dot{\Sigma}^{-1}X)$

$= Q[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}Q'(X'\dot{\Sigma}^{-1}X) = E$

(2) $FF = (X'\dot{\Sigma}^{-1}X)^{-1}R'[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}\underline{R(X'\dot{\Sigma}^{-1}X)^{-1}R'[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}}R$

$= (X'\dot{\Sigma}^{-1}X)^{-1}R'[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}R = F$

(3) $EF = Q[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}Q'\underline{(X'\dot{\Sigma}^{-1}X)(X'\dot{\Sigma}^{-1}X)^{-1}}R'[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}R$

$= Q[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}Q'R'[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}R = 0$, since $Q'R' = 0$.

(4) $FE = (X'\dot{\Sigma}^{-1}X)^{-1}R'[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}RQ[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}Q'[X'\dot{\Sigma}^{-1}X] = 0$,

since $RQ = 0$.

To apply the previous lemma, we need to derive the ranks of E and F. In the process of deriving these ranks we will point out two additional assumptions about rank that need to be made [see (4) and (8) below]. Those willing to assume that $\text{rk } E = n - \text{rk } R$ and $\text{rk } F = \text{rk } R$ may skip the derivation of the ranks of E and F in (1) through (10), below.

(1) $\dot{\Sigma}$ was assumed positive definite, hence of rank $\dot{T}$.

(2) X was assumed to have full column rank ($\text{rk } X = n$).

(3) Since $\dot{\Sigma}$ is positive definite, $\dot{\Sigma}^{-1}$ will be positive definite; therefore, $\dot{\Sigma}^{-1}$ can be expressed as $\dot{\Sigma}^{-1} = P'P$ where P is non-singular. Thus, $\text{rk}(X'\dot{\Sigma}X)^{-1} = \text{rk}(X'\dot{\Sigma}^{-1}X) = \text{rk}(X'P'PX) = \text{rk}(PX) = \text{rk } X = n$.[1]

(4) $\text{rk}(XQ) \leq \min(\text{rk } X, \text{rk } Q) = n - \text{rk } R$. We will assume that $\text{rk}(XQ) = n - \text{rk } R$.[2]

---

[1] For any matrix A, $\text{rk}(A'A) = \text{rk}(AA') = \text{rk } A = \text{rk } A'$. Also, if B_1 and B_2 are any non-singular matrices of order compatible with A (A may be rectangular) $\text{rk}(B_1A) = \text{rk}(AB_2) = \text{rk } A$.

[2] $\text{rk}(XQ) = \text{rk}(X_1 + X_2Q_2)$ where $\text{rk } X_1 = n - \text{rk } R$.

(5) $rk[Q'X'\dot{\Sigma}^{-1}XQ]^{-1} = rk[Q'X'\dot{\Sigma}^{-1}XQ] = rk[Q'X'P'PXQ]$

$= rk(PXQ) = rk(XQ) = n - rk\ R$.

(6) $rk(Q[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}Q') = rk[Q'X'\dot{\Sigma}^{-1}XQ]^{-1} = n - rk\ R$.

The first equality of (6) comes from, first, $rk(Q[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}Q')$ cannot exceed $n - rk\ R$, since the rank of the product of matrices cannot exceed the rank of any matrix in the product; second, $rk(Q[Q'X'\dot{\Sigma}^{-1}Q]^{-1}Q')$ cannot be less than $n - rk\ R$ since:

$$Q[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}Q' = \begin{bmatrix} I \\ Q_2 \end{bmatrix}[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}[I \vdots Q_2']$$

$$= \begin{bmatrix} [Q'X'\dot{\Sigma}^{-1}XQ]^{-1} & [Q'X'\dot{\Sigma}^{-1}XQ]^{-1}Q_2' \\ Q_2[Q'X'\dot{\Sigma}^{-1}XQ]^{-1} & Q_2[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}Q_2' \end{bmatrix}$$

and the rank of any matrix is greater than or equal to the rank of any submatrix in the matrix. The second equality comes from (5).

(7) $rk\ E = rk(Q[Q'X'\dot{\Sigma}^{-1}XQ]^{-1}Q') = n - rk\ R$.

The first equality comes from $(X'\dot{\Sigma}^{-1}X)$ being a non-singular matrix and the second equality comes from (6).

(8) (IV.23) assumes that $R(X'\dot{\Sigma}^{-1}X)^{-1}R'$ is non-singular; i.e., that $[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}$ exists. This implies that $rk[R(X'\dot{\Sigma}^{-1}X)^{-1}R'] = rk\ R$.

(9) $rk(R'[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}R) = rk\ R$.

To see this, let $[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1} = B_1$. Then $R'B_1R = [C^{-1}CR]'B_1[C^{-1}CR] = (CR)'(C^{-1})'B_1C^{-1}CR$ where C is the C matrix used in (V.24). Let $B_2 = (C^{-1})'B_1C^{-1}$. Then

$\text{rk } B_2 = \text{rk } B_1 = \text{rk } R$, since $(C^{-1})'$ and C^{-1} are non-singular.

$(CR)'B_2CR = \begin{bmatrix} I \\ A_1' \end{bmatrix} B_2 [I \vdots A_1] = \begin{bmatrix} B_2 & B_2A_1 \\ A_1'B_2 & A_1'B_2A_1 \end{bmatrix}$ and has the same rank as B_2 for the same reason as the explanation in (6).

(10) $\text{rk } F = \text{rk}(R'[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}R) = \text{rk } R$.

The first equality comes from $(X'\dot{\Sigma}^{-1}X)^{-1}$ being a non-singular matrix and the second equality comes from (9).

Since E and F are square idempotent matrices of order n with $EF = FE = 0$, $\text{rk } E = n - \text{rk } R$ and $\text{rk } F = \text{rk } R$, then by the above lemma (IV.25), $E + F = I$ or $E = I - F$.

Substituting $I - F$ for E into (IV.26) we have:

$$\hat{\delta}_{RGLS} = [I - (X'\dot{\Sigma}^{-1}X)^{-1}R'[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}R][\hat{\delta}_{GLS} - q] + q$$

$$= [\hat{\delta}_{GLS} - q] - (X'\dot{\Sigma}^{-1}X)^{-1}R'[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}[R\hat{\delta}_{GLS} - Rq] + q$$

$$= \hat{\delta}_{GLS} - (X'\dot{\Sigma}^{-1}X)^{-1}R'[R(X'\dot{\Sigma}^{-1}X)^{-1}R']^{-1}[R\hat{\delta}_{GLS} - r]$$

(since $r = R\hat{\delta}_{RGLS} = R[Q\hat{\delta}^{(1)} + q] = RQ\hat{\delta}^{(1)} + Rq = 0\ \hat{\delta}^{(1)} + Rq = Rq$) which is exactly equation (IV.23).

Advantages of the Q Formula over the R Formula

Formula (IV.5) (which we will call the Q formula) has the following advantages over formula (IV.23) (which we will call the R formula):

(1) X need not have full column rank if the Q formula is used but must have full column rank for the R formula. Thus, the Q formula may permit calculation of $\hat{\delta}_{RGLS}$ when a unique unrestricted $\hat{\delta}_{GLS}$ does not exist; however, to use the R formula

requires $\hat{\delta}_{GLS}$.

(2) For the Q formula, the restrictions imposed on the coefficients need not be linearly independent (R need not have full row rank); however, use of the R formula requires restrictions which are linearly independent.

(3) In calculating Q, an explicit check is made for inconsistent equations. In the R formula, both linearly dependent and inconsistent equations lead to the same result--singularity of the $R(X'\dot{\Sigma}^{-1}X)^{-1}R'$ matrix.

(4) The largest matrix inverted by the Q formula is of order n - rk R. The largest matrix inverted by the R formula is of order n.

(5) If an iterative procedure such as iteration on 3SLS (three-stage least squares) or ZA (Zellner-Aitken estimator) is being used, once the Q matrix and q vector are formed, they can be used for all iterations; therefore, the calculation of the restricted coefficients requires less time at each iteration for the Q formula than for the R formula.

(6) Use of the Q formula permits a unified treatment of restrictions on coefficients for DLS, ZA, IZA (iterative Zellner-Aitken), 3SLS, I3SLS (iterative three-stage least squares), SML (limited information subsystem maximum likelihood), LML (linearized maximum likelihood, and FIML (full information maximum likelihood), as

will become evident as each of these methods is discussed.

The Q formula has the disadvantage that the Q matrix and q vector must be calculated, but this is only a small task which can be accomplished rapidly on a computer.

C. Restrictions Imposed on Direct Least Squares Coefficients

DLS estimation is the particular case of GLS estimation in which $\dot{\Sigma}$ is assumed to be of the form $\sigma^2 I$ which is known except for a multiplication constant (σ^2) as required by the GLS model. Thus, if the model is written as:

$$y = X_\mu \delta + u \quad , \tag{IV.27}$$

the DLS estimator is

$$\hat{\delta}_{DLS} = [X_\mu' X_\mu]^{-1} X_\mu' y \tag{IV.28}$$

and the variance of $\hat{\delta}_{DLS}$ is given by

$$\text{Var}(\hat{\delta}_{DLS}) = \sigma^2 [X_\mu' X_\mu]^{-1} \quad . \tag{IV.29}$$

Assuming that X_μ contains only fixed variables, an unbiased estimate of σ^2 is[1]

$$\hat{\sigma}^2 = \hat{u}_{DLS}' \hat{u}_{DLS} / (T - n) \quad . \tag{IV.30}$$

If the restrictions

$$\underset{N_R \times n}{R} \quad \underset{n \times 1}{\delta} = \underset{N_R \times 1}{r} \tag{IV.31}$$

are imposed on the coefficients, the RDLS (restricted DLS) model is the particular case of the RGLS model in which $\dot{\Sigma}$ is assumed to be of the form $\sigma^2 I$. Thus, the restricted DLS solution may be obtained by

---

[1] Goldberger [1964], p. 268 states that the variables in X_μ may be stochastic provided they are distributed independently of u_μ.

substituting X_μ for X and $\sigma^2 I$ for $\dot{\Sigma}$ in RGLS formula (IV.5) which gives:[1,2]

$$(IV.32) \qquad \hat{\delta}_{RDLS} = Q\{[Q'(X_\mu'X_\mu)Q]^{-1}Q'[X_\mu'y - (X_\mu'X_\mu)q]\} - q$$

where Q and q are calculated from R and r in the manner given previously for the RGLS model. Q is an $n \times (n - \text{rk } R)$ matrix, and q is an $n \times 1$ vector. As with the GLS model, the use of (IV.32) obtains the same result as if (1) rk R coefficients are solved out in terms of the remaining n - rk R coefficients, (2) the n-rk R coefficients are estimated by DLS, and (3) the rk R coefficients which were solved out are calculated from the n - rk R coefficients which were estimated directly. Provided $[Q'X_\mu'X_\mu Q]^{-1}$ exists [this inverse will exist if $\text{rk}(X_\mu Q) = n - \text{rk } R$], the RDLS estimates are unique. The solution obtained is the solution which minimizes (by choice of $\hat{\delta}$) $(y - X_\mu\hat{\delta})'(y - X_\mu\hat{\delta})$ subject to $R\hat{\delta} = r$.

---

[1]The multiplicative constant σ^2 cancels out since

$$\begin{aligned}\hat{\delta}_{RDLS} &= Q\{[Q'(X_\mu'\sigma^2 I X_\mu)Q]^{-1}Q'[X_\mu'\sigma^2 I y - (X_\mu'\sigma^2 I X_\mu)q]\} - q \\ &= Q\{(1/\sigma^2)\sigma^2[Q'(X_\mu'X_\mu)Q]^{-1}Q'[X_\mu'y - (X_\mu'X_\mu)q]\} - q \\ &= Q\{[Q'(X_\mu'X_\mu)Q]^{-1}Q'[X_\mu'y - (X_\mu'X_\mu)q]\} - q.\end{aligned}$$

[2]Calculation of the Q matrix and q vector also gives a means of separating out rk R coefficients, $\hat{\delta}^{(1)}_{RDLS}$, which may be calculated from the remaining n-rk R "unrestricted" coefficients, $\hat{\delta}^{(2)}_{RDLS}$. Thus, the following pair of formulas are together equivalent to (IV.32):

$$\hat{\delta}^{(1)}_{RDLS} = [Q'(X_\mu'X_\mu)Q]^{-1}Q'[X_\mu'y - (X_\mu'X_\mu)q] \; ; \quad \hat{\delta}^{(2)}_{RDLS} = Q_2\,\hat{\delta}^{(1)}_{RDLS} + q_2$$

where Q_2 and q_2 are the subparts of Q and q noted earlier.

The variance-covariance matrix for $\hat{\delta}_{RDLS}$ is

$$\text{(IV.33)} \qquad \text{Var}(\hat{\delta}_{RDLS}) = \sigma^2 Q[Q'(X_\mu' X_\mu)Q]^{-1}Q'$$

with an unbiased estimate of σ^2 being

$$\text{(IV.34)} \qquad \hat{\sigma}^2_{RDLS} = \hat{u}'_{RDLS}\hat{u}_{RDLS}/(T - n + rk\ R)$$

provided X_μ is a matrix of "fixed" variables.[1]

If we relax our assumptions to those of the double k-class model, permitting, in particular, jointly dependent variables in the matrix X_μ, then the RDLS estimates will have the same properties as the DLS estimates noted in the double k-class section. Although the formulas within this section no longer provide unbiased or even consistent estimates, $\hat{\delta}_{RDLS}$ will still be the $\hat{\delta}$ which minimizes $\hat{u}'\hat{u}$ (subject to the restrictions, of course), and $\hat{\sigma}^2_{RDLS}$ will still be the estimate of σ^2 which would be obtained if the restrictions were solved out and then the usual DLS formulas applied.

---

[1] Or the variables in X_μ are distributed independently of u.

D. Restrictions Imposed on Two-stage Least Squares Coefficients

Derivation of 2SLS as a GLS Method

In the 2SLS model used earlier, we considered an equation of the form:

$$y = Z_\mu \delta + u \tag{IV.35}$$

with $\mathrm{Var}(u) = \sigma^2 I$.

If (IV.35) is premultiplied by X_I', we have[1]

$$X_I'y = X_I'Z_\mu \delta + X_I'u \tag{IV.36}$$

with $\mathrm{Var}(X_I'u) = X_I'[\mathrm{Var}(u)]X_I = X_I'[\sigma^2 I]X_I = \sigma^2 X_I'X_I$

provided that we assume that the X_I matrix contains "fixed" variables only. If GLS is applied to (IV.36) by using $X_I'y$ as the GLS y, $X_I'Z_\mu$ as the GLS X, and $X_I'u$ as the GLS u in the GLS computational formula (IV.2) the following GLS estimator is obtained:[2]

$$\begin{aligned}\hat{\delta} &= [Z_\mu'X_I(\sigma^2 X_I'X_I)^{-1}X_I'Z_\mu]^{-1}[Z_\mu'X_I(\sigma^2 X_I'X_I)^{-1}X_I'y] \\ &= [Z_\mu'X_I(X_I'X_I)^{-1}X_I'Z_\mu]^{-1}[Z_\mu'X_I(X_I'X_I)^{-1}X_I'y]\end{aligned} \tag{IV.37}$$

But, by (I.36) this is

$$[Z_\mu'Z_\mu]^{-1}_{\|X_I}[Z_\mu'y]_{\|X_I} = \begin{bmatrix}[Y'Y]_{\|X_I} & [Y'X_\mu]_{\|X_I} \\ [X_\mu'Y]_{\|X_I} & [X_\mu'X_\mu]_{\|X_I}\end{bmatrix}^{-1}\begin{bmatrix}[Y'y]_{\|X_I} \\ [X_\mu'y]_{\|X_I}\end{bmatrix},$$

and, since the variables in X_μ are contained in X_I, this is [see

---

[1] As before, we assume that the matrix of instruments, X_I, includes the variables in X_μ .

[2] This derivation of 2SLS is given in Zellner and Theil [1962], p. 56.

(I.40)]:

$$\begin{bmatrix} [Y'Y]_{\|X_I} & Y'X_\mu \\ X_\mu'Y & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} [Y'y]_{\|X_I} \\ X_\mu'y \end{bmatrix} = \hat{\delta}_{2SLS} \quad .$$

Thus, 2SLS may be derived by an application of GLS. Not all of the assumptions of the GLS model are met; hence, $\hat{\delta}_{2SLS}$ does not have all of the desirable properties that $\hat{\delta}_{GLS}$ has. Particularly affecting the properties of 2SLS is the fact that the m jointly dependent variables in the $X_I'Y$ matrix (a submatrix of $X_I'Z$) are asymptotically correlated with the disturbance $X_I'u$.

Calculation of 2SLS When Restrictions are Imposed on the Coefficients

Since 2SLS may be derived as a case of GLS, it is natural to question whether restrictions can be imposed on the coefficients and a restricted 2SLS estimator derived as an application of RGLS. If we denote:

(IV.38)
$$A = \begin{bmatrix} [Y'Y]_{\|X_I} & Y'X_\mu \\ X_\mu'Y & X_\mu'X_\mu \end{bmatrix}, \quad b = \begin{bmatrix} [Y'y]_{\|X_I} \\ X_\mu'y \end{bmatrix}$$

then the GLS-2SLS solution (IV.37) is:

(IV.39)
$$\hat{\delta}_{2SLS} = A^{-1}b \quad .$$

If we impose the restrictions

(IV.40)
$$R\delta = r$$

on the coefficients of (IV.35), then the RGLS solution corresponding to (IV.39) [we will denote this solution as the R2SLS solution] is:[1]

$$\text{(IV.41)} \qquad \hat{\delta}_{R2SLS} = Q\{[Q'AQ]^{-1}Q'[b - Aq]\} + q$$

where Q and q are calculated from R and r in the same way as in the usual RGLS model. Q is an $n^{X}(n\text{-rk } R)$ matrix and q is an $n^{X}1$ vector.

However, the coefficients obtained through use of the computational method given above may differ from the coefficients obtained if the restrictions, (IV.40), are used to reduce the number of coefficients to be estimated and then 2SLS is used to estimate the remaining coefficients. One reason that the coefficients may differ is that if the usual procedure of using the restrictions to reduce the number of coefficients to be estimated is followed, predetermined variables are often linearly combined with jointly dependent variables and the newly constructed variables (those that are linear combinations of jointly dependent and predetermined variables) are then labeled jointly dependent. The predetermined variables which are linearly combined with jointly dependent variables then no longer occur in the X_{μ} matrix. Since these variables are not longer in the X_{μ} matrix, unless they

---

[1] $$\text{Var}(\hat{\delta}_{R2SLS}) = \sigma^2[Q'AQ]^{-1}$$

with a consistent estimate of σ^2 being given by:

$$\hat{\sigma}^2_{R2SLS} = \hat{u}'_{R2SLS}\hat{u}_{R2SLS}/(T - n + \text{rk } R) \ .$$

If T were substituted for T - n + rk R in the denominator, the resulting $\hat{\sigma}^2_{R2SLS}$ would still be consistent.

are specifically entered into the X** matrix ($X_I = [X_\mu \vdots X^{**}]$), the space spanned by the X_I matrix formed after the coefficients are solved into the equation is likely to be a proper subspace of the space spanned by the X_I matrix which would have been formed before the coefficients were solved into the equation. Hence, $Y_{\blacksquare X}$ and $y_{\blacksquare X}$ are likely to change.

Another reason why the resulting coefficients may differ is that in using restrictions to reduce the number of coefficients, a set of predetermined variables may be linearly combined into a single predetermined variable; hence, the space spanned by X_I may again change.

To make the above remarks clearer, we will illustrate the effect of a restriction on two coefficients which effectively linearly combines a jointly dependent and predetermined variable if the usual procedure is applied. Suppose that the equation to be estimated is:

$$y_1 = \alpha_1 y_2 + \alpha_2 x_2 + \alpha_3 x_3 + u \tag{IV.42}$$

and the restriction

$$\alpha_1 - \alpha_2 = 0 \tag{IV.43}$$

is imposed on the coefficients in the equation.

Then, $y = y_1$, $Y = y_2$, and $X_\mu = [x_2, x_3]$. Suppose also that the R2SLS coefficients are calculated with the matrix of instruments being

$$X_I = [X_\mu \vdots X^{**}] = [x_2, x_3 \vdots x_4, x_5, x_6] \tag{IV.44}$$

and that (for simplicity) X_I has full column rank (i.e., that none of the variables in X_I can be expressed as a linear combination of the

remaining variables in X_I.) Let the coefficients obtained through application of the R2SLS formulas be denoted $[\hat{\alpha}_1]_{R2SLS}$, $[\hat{\alpha}_2]_{R2SLS}$, and $[\hat{\alpha}_3]_{R2SLS}$.

As an alternative means of estimating the coefficients, let us use the usual procedure of using the restrictions to reduce the number of coefficients to be estimated. We get:

$$y_1 = \alpha_1(y_2 + x_2) + \alpha_3 x_3 + u \quad . \tag{IV.45}$$

Thus, $y = y_1$, $Y = y_2 + x_2$, $X_\mu = x_3$, and if X^{**} is left unchanged, then

$$X_I = [X_\mu \vdots X^{**}] = [x_3 \vdots x_4, x_5, x_6] \quad . \tag{IV.46}$$

Application of the usual 2SLS formulas to (IV.45) will give a solution which we will designate the 2SLSR solution. $[\alpha_1]_{2SLSR}$ is not in general equal to $[\alpha_1]_{R2SLS}$ and $[\alpha_3]_{2SLSR}$ is not in general equal to $[\alpha_3]_{R2SLS}$, since the space of X_I has been restricted by omitting x_2.

The coefficients obtained in estimating (IV.42) subject to (IV.43) by the R2SLS formula (IV.41) would be the same as the coefficients obtained from estimating (IV.45) if:

(1) x_2 were added to X^{**} in estimating (IV.42) (i.e., the X_I given in (IV.46) were changed to $[x_3 \vdots x_4, x_5, x_6, x_2]$); therefore, the R2SLS coefficients would be obtained in both cases;[1] or if

---

[1] The R2SLS coefficients would be obtained because $[y_2 + x_2]_{\| X_I} = [y_2]_{\| X_I} + [x_2]_{\| X_I}$ [see (I.62)] which (if x_2 is contained in X_I) equals $[y_2]_{\| X_I} + x_2$ [see (I.38)].

(2) x_2 were listed as jointly dependent instead of predetermined and omitted from X_I (i.e., the X_I given in (IV.44) were changed to $[x_3 \vdots x_4, x_5, x_6]$); therefore, the 2SLSR coefficients would be obtained in both cases.

The preceding does not imply that either the $\hat{\delta}_{R2SLS}$ solution or the $\hat{\delta}_{2SLSR}$ solution is incorrect. It merely shows the importance of the particular instruments selected.

PART II

MULTIPLE EQUATIONS METHODS

CHAPTER V

FULL INFORMATION MAXIMUM LIKELIHOOD (FIML)

A. Properties of the Full Information Maximum Likelihood Estimator

The full information maximum likelihood (FIML) estimator which is considered in this section is maximum likelihood if in addition to the basic statistical assumptions of this paper (section I.C.3) we add the assumption that the matrix of disturbances, U, has the multivariate normal distribution. If the matrix of disturbances is not normally distributed then the FIML computational formulas given in this section give estimates which have been termed quasi-maximum likelihood estimates. Quasi-maximum likelihood estimates may still possess some desirable properties.[1]

The FIML estimator is a "full information" estimator in the sense that account is taken of all structural equations in the system (including identity equations) in deriving estimates of the population coefficients. In the single equation techniques of part I, consideration was given to the structure of only a single equation at a time. For some of these single equation techniques the predetermined variables in equations other than the equation being estimated were used, but no account was taken of the structure of the remaining equations. In the FIML method the coefficients of all stochastic equations are estimated simultaneously, a distinction being made between jointly dependent and predetermined variables in each equation, and explicit

---

[1] See Koopmans and Hood [1953], pp. 144-147.

account being taken of all structural coefficient restrictions and any identity equations which may complete the system.[1]

The FIML method may only be applied if the number of jointly dependent variables in the system equals the number of equations including identity equations.

Recording the particular variables which are said to occur in an equation is equivalent to restricting the coefficients of the equation corresponding to the remaining variables in the system to zero. The coefficient of one jointly dependent variable in the equation is also restricted to -1 to provide a normalization rule (see section I.C. 1).[2] Initially, these are the only types of restriction that will be permitted on the coefficients; however, in section V.F FIML estimation will be generalized to take account of arbitrary linear restrictions imposed on the coefficients.[3] In no place in this paper is consideration given to FIML estimation with restrictions imposed on the covariance matrix of the disturbances. In particular, not considered is the much simpler computational method called full-information diagonal (often abbreviated to FID) which is obtained by assuming that all off-diagonal elements of the disturbance variance-covariance matrix

---

[1]Actually, the coefficients of predetermined variables in the identity equations are not used explicitly in the computation of the structural coefficients; however, if they were not known, the equation would not be a true identity equation. Thus, an identity equation which contains no jointly dependent variable adds no information to the system and may therefore be deleted. Also, the specification of the model is not changed if all predetermined variables in an identity equation are multiplied by their respective known coefficients and combined into a single predetermined variable.

[2]As noted in section I.C.1, for FIML estimation it makes no substantive difference which jointly dependent variable is singled out as the normalizing jointly dependent variable.

[3]In section V.F the normalization rule is also generalized.

are zero. Thus, the FIML procedure developed in this paper is the full-information non-diagonal procedure (often denoted elsewhere as FIND).

Although considerable progress has been made in developing FIML procedures which permit estimation of equations in which coefficients enter in a non-linear fashion,[1] consideration in this paper will be restricted to estimation of a system in which coefficients enter each equation in a linear fashion.

In this chapter, we will assume that the system is identified. This means that each equation must be not only just-identified or over-identified in the single equation sense usually treated, but in a multiple equation sense as well.[2] Although some additional requirements must be met over and above those required for identification for single equation estimation, the single equation identification rules applied separately to each equation provide a good starting point.

A misconception fairly generally held is that there must be more observations in the sample than number of coefficients in the system to be estimated. This is certainly not true. Provided that there are sufficient observations that the estimated disturbance variance-covariance matrix is not singular (there must be more observations than number of equations with disturbances), only sufficient observations that DLS can be applied to each equation separately is all that is in

---

[1] Eisenpress has made considerable progress in this area. See Eisenpress and Greenstadt [1964].

[2] Identification is treated in the multiple equations sense in Koopmans, Rubin, and Leipnik [1950].

general required.[1]

(At this point the reader may want to review the notation developed in section I.C for expressing a system of equations including notation related to the reduced form equations.)

Another property which is certainly worth noting applies to the reduced form coefficients estimated from the FIML structional coefficients. Let, as before, the reduced form equations be expressed as $Y = X\Pi' + V$. Let $\hat{\Pi}$ be any estimate of Π which is not inconsistent with the restrictions imposed on the coefficients of the structural equations. Also, let $\hat{V} = Y - X\hat{\Pi}'$ be the matrix of reduced form residuals and $(1/T)\hat{V}'\hat{V}$ be the estimated variance-covariance matrix of the reduced form disturbances. Then if $\hat{\Pi}_{FIML}$ is calculated from the estimated FIML coefficients of the structural equations, i.e., $\hat{\Pi}_{FIML} = -\hat{\Gamma}^{-1}_{FIML}\hat{B}_{FIML}$, the resulting $\det(\frac{1}{T}\hat{V}'_{FIML}\hat{V}_{FIML})$ will be less than or equal to $\det(\frac{1}{T}\hat{V}'\hat{V})$ obtained from any set of reduced form coefficients which are not inconsistent with the restrictions imposed on the coefficients of the structural equations.[2] Thus, of the estimating

---

[1] That the estimated disturbance variance-covariance matrix, S, will be singular if the number of observations exceeds the number of equations is easily shown. As noted further on in (V.19), $S = \frac{1}{T}\hat{\alpha}_I(Z'Z)\hat{\alpha}'_I = (1/T)\hat{U}'\hat{U}$; hence, rk S = rk $\hat{U}$. If $T < M$, then rk $\hat{U}$ and therefore rk S will be less than M; i.e., S will be singular.

Since S is calculated in the same manner for limited information subsystem maximum likelihood (SML), the Zellner-Aitken estimator (ZA), and three-stage least squares (3SLS), the requirement that there not be fewer observations than stochastic equations applies to these methods as well.

[2] For this to be a meaningful statement we must assume that identity equations have been incorporated into the system by solving out one jointly dependent variable for each identity equation (thereby imposing less convenient restrictions as will be noted further on). Otherwise,

procedures which take the full information of the system into account, the FIML method gives the minimum estimated generalized variance of the disturbances of the reduced form equations.[1] For this reason, it is common to refer to FIML estimates as least generalized variance (LGV) estimates. The LGV property does not, of course, rely on an assumption of normality. It is a property similar to the least squares property for a single equation.

$$\det\left(\frac{1}{T}\hat{V}'\hat{V}\right) = \det\left(\frac{1}{T}\{[\hat{U} \vdots 0]\hat{\Gamma}^{-1}\}'[\hat{U} \vdots 0]\hat{\Gamma}^{-1}\right) = \det\left(\frac{1}{T}[\hat{\Gamma}^{-1}]'\begin{bmatrix}\hat{U}'\hat{U} & 0\\ 0 & 0\end{bmatrix}\hat{\Gamma}^{-1}\right) = 0$$

for any method meeting the restrictions of the structural equations. (Use of identity equations to solve out jointly dependent variables is necessary to explicitly specify properties, only. In the computational procedure which is presented, the identity equations are explicitly recognized by the computational procedure rather than used to eliminate jointly dependent variables from the system.)

[1] See Goldberger [1964], pp. 352-354.

B. Derivation of the Likelihood Function to be Maximized

Before continuing, let us designate what is meant by maximizing the likelihood function. Our equation system is a system of M equations containing disturbances and G - M identity equations which may be written as

$$Z\alpha' + [U \vdots 0] = 0 \tag{V.1}$$

or equivalently as

$$\alpha Z' + \begin{bmatrix} U' \\ 0' \end{bmatrix} = 0' \quad . \tag{V.2}$$

The matrix of coefficients, α, may be subdivided into the matrix of coefficients of jointly dependent variables, Γ, and the matrix of coefficients of predetermined variables, B. A further subdivision can be made on the basis of whether the coefficients are coefficients of stochastic equations or coefficients of identity equations (hence are known constants). Thus, α may be subdivided as:

$$\alpha = [\Gamma : B] = \begin{bmatrix} \alpha_I \\ \alpha_{II} \end{bmatrix} = \begin{bmatrix} \Gamma_I & B_I \\ \Gamma_{II} & B_{II} \end{bmatrix} \tag{V.3}$$

where $\alpha_I = [\Gamma_I \vdots B_I]$ represents the coefficients of the M stochastic equations and $\alpha_{II} = [\Gamma_{II} \vdots B_{II}]$ represents the coefficients of the G - M identity equations.

As a step in the derivation of the likelihood function, we will use the G - M identity equations to temporarily eliminate G - M jointly dependent variables from the system. (The eliminated variables will be reentered into the system at a later step.) Suppose we divide

our jointly dependent variables into two groups, $Y = [Y_1 \vdots Y_2]$, where Y_2 contains $G - M$ jointly dependent variables to be temporarily eliminated from the system and Y_1 contains the remaining M jointly dependent variables.

To reflect our subdivisions, we may rewrite (V.2) as:

$$\Gamma_{11}Y_1' + \Gamma_{12}Y_2' + B_I X' + U' = 0' \tag{V.4}$$

$$\Gamma_{21}Y_1' + \Gamma_{22}Y_2' + B_{II}X' = 0' \tag{V.5}$$

where the Γ matrix has been further subdivided to reflect the division of jointly dependent variables into those which are to be temporarily eliminated and those which will remain, i.e.,

$$\underset{G\times G}{\Gamma} = \begin{bmatrix} \underset{M\times G}{\Gamma_I} \\ \\ \underset{(G-M)\times G}{\Gamma_{II}} \end{bmatrix} = \begin{bmatrix} \underset{M\times M}{\Gamma_{11}} & \underset{M\times (G-M)}{\Gamma_{12}} \\ \\ \underset{(G-M)\times M}{\Gamma_{21}} & \underset{(G-M)\times (G-M)}{\Gamma_{22}} \end{bmatrix} . \tag{V.6}$$

We can assume that Γ_{22} is non-singular, since if it were singular we could merely select a different set of jointly dependent variables to be temporarily eliminated thereby rearranging the columns of Γ until a nonsingular Γ_{22} is obtained.[1]

Solving (V.5) for Y_2' we obtain:

$$\Gamma_{22}Y_2' = -\Gamma_{21}Y_1' - B_{II}X' \quad ; \tag{V.7}$$

hence,

$$Y_2' = -\Gamma_{22}^{-1}\Gamma_{21}Y_1' - \Gamma_{22}^{-1}B_{II}X' \quad . \tag{V.8}$$

---

[1] Γ was assumed to be nonsingular in section I.C.3, assumption (4).

Substituting (V.8) into (V.4), we obtain:

$$\Gamma_{11}Y_1' - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}Y_1' - \Gamma_{12}\Gamma_{22}^{-1}B_{II}X' + B_IX' + U' = 0' \tag{V.9}$$

or

$$\underset{M^{X}M}{\Gamma^*}\ \underset{M^{X}T}{Y_1'} - \underset{M^{X}\Lambda}{B^*}\ \underset{\Lambda^{X}T}{X'} + \underset{M^{X}T}{U'} = \underset{M^{X}T}{0'} \tag{V.10}$$

or

$$\underset{M^{X}(M+\Lambda)}{\alpha^*}\ \underset{(M+\Lambda)^{X}T}{Z_1'} + \underset{M^{X}T}{U'} = \underset{M^{X}T}{0'} \tag{V.11}$$

where $\Gamma^* = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}$ is the square $M^{X}M$ matrix of coefficients of the remaining jointly dependent variables.

$B^* = B_I - \Gamma_{12}\Gamma_{22}^{-1}B_{II}$ is the $M^{X}\Lambda$ matrix of coefficients of predetermined variables,

$\alpha^* = [\Gamma^* \vdots B^*]$ is the $M^{X}(M + \Lambda)$ matrix of coefficients of all the variables remaining in the system, and

$Z_1 = [Y_1 \vdots X]$.

If we assume that the $T^{X}M$ matrix of disturbances of the stochastic equations,

$$U = [U_1 \cdots U_M] = \begin{bmatrix} U_{[1]} \\ \vdots \\ U_{[T]} \end{bmatrix} = \begin{bmatrix} u_{11} & \cdots & u_{1M} \\ \vdots & & \vdots \\ u_{11} & \cdots & u_{TM} \end{bmatrix} ,$$

has a multivariate normal distribution with $\mathcal{E}U = 0$, $\mathcal{E}U_{[t]}'U_{[t]} = \Sigma$ for all t and $\mathcal{E}U_{[t]}'U_{[t]} = 0$ for $t \neq t'$, Σ being a positive definite matrix, we can write the density function for U as:

$$f_1(U,\Sigma) = (2\pi)^{-T/2}\det{}^{-T/2}\Sigma \exp\left(-\frac{1}{2}\sum_{t=1}^{T} U_{[t]}\Sigma^{-1}U_{[t]}'\right) . \tag{V.12}$$

We can convert this density function to the likelihood function by using (V.11) to transform the system from U to Z_1 and α^*, taking account of the Jacobian of the transformation. Also, the logarithm of the likelihood function is easier to work with in this case than the actual likelihood function. (Since the logarithmic transformation is a strictly increasing one, the logarithm of the likelihood will be maximized at the same point as the actual likelihood.) After these transformations are made, our "logarithmic likelihood function" may be written:[1]

$$(V.13)\quad f_2(Z_1,\alpha^*,\Sigma) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log[\det \Sigma] + T\log[\mathrm{abs}(\det \Gamma^*)] - \frac{1}{2}\sum_{t=1}^{T} Z_{1[t]}\alpha^{*\prime}\Sigma^{-1}\alpha^* Z'_{1[t]} \quad .$$

If (V.13) is maximized first with respect to $\hat{\Sigma}$ (thereby concentrating the function onto $\hat{\alpha}^*$ and Z_1), we get the relation:

$$(V.14)\qquad \hat{\Sigma} = \frac{1}{T}\hat{\alpha}^*(Z_1'Z_1)\hat{\alpha}^{*\prime} \quad .$$

Substituting $\hat{\Sigma}$ from (V.14) into (V.13) and dividing by $T/2$, the following function is obtained:[2]

$$(V.15)\quad f_3(\hat{\alpha}^*,Z_1) = c_1 + \log[\det{}^2\hat{\Gamma}^*] - \log[\det(\frac{1}{T}\hat{\alpha}^*(Z_1'Z_1)\hat{\alpha}^{*\prime})]$$

---

[1] See Koopmans and Hood [1953], pp. 143-160, 190-191.

[2] For a more detailed discussion of the "stepwise" maximization procedure see Koopmans and Hood [1953], pp. 160, 161, 191.

where c_1 is a constant.[1]

However, $\det^2(\Gamma^*) = \det^2(\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}) = \det^2\Gamma/\det^2\Gamma_{22}$ [2] and $\frac{1}{T}\alpha^*(Z_1'Z_1)\alpha^{*\prime} = \frac{1}{T}\alpha_I(Z'Z)\alpha_I'$;[3] hence, (V.15) may be rewritten as

$$(V.16) \qquad f_4(\hat{\alpha},Z) = c_1 - \log[\det^2\Gamma_{22}] + \log[\det^2\hat{\Gamma}] - \log[\det(\tfrac{1}{T}\hat{\alpha}_I[Z'Z]\hat{\alpha}_I')],$$

---

[1] $$c_1 = \frac{2}{T}(-\frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T} Z_{1[t]}\hat{\alpha}^{*\prime}\hat{\Sigma}^{-1}\hat{\alpha}^*Z_{1[t]}')$$

$$= -\log 2\pi - \frac{2}{T}(\frac{1}{2}\mathrm{tr}\{\frac{1}{T}Z_1\hat{\alpha}^{*\prime}[\hat{\alpha}^*(Z_1'Z_1)\hat{\alpha}^{*\prime}]^{-1}\hat{\alpha}^*Z_1'\})$$

$$= -\log 2\pi - \frac{1}{T}\mathrm{tr}\{\frac{1}{T}\hat{\alpha}^*Z_1'Z_1\hat{\alpha}^{*\prime}[\hat{\alpha}^*Z_1'Z_1\hat{\alpha}^{*\prime}]^{-1}\}$$

$$= -\log 2\pi - \frac{1}{T}\mathrm{tr}\{\frac{1}{T}I\} = -\log 2\pi - \frac{1}{T}\cdot 1 = -\log 2\pi - \frac{1}{T}$$

where tr denotes trace and we have used the relationship $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for any matrices A and B provided AB and BA are defined (i.e., provided the number of rows of A equals the number of columns of B and the number of rows of B equals the number of columns of A).

[2] Since $\det\begin{bmatrix} I & A \\ 0 & I \end{bmatrix} = \det I = 1$ for any matrix A and since $(\det B)\cdot(\det C) = \det(BC)$ for any square matrices B and C of the same order, we have

$$\det\Gamma = \det\begin{bmatrix} I & \Gamma_{12}\Gamma_{22}^{-1} \\ 0 & I \end{bmatrix}\cdot\det\begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}$$

$$= \det\left\{\begin{bmatrix} I & -\Gamma_{12}\Gamma_{22}^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}\right\} = \det\begin{bmatrix} \Gamma_{11}-\Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21} & 0 \\ 0 & \Gamma_{22} \end{bmatrix}$$

$= \det[\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}]\cdot\det\Gamma_{22}$; hence, $\det[\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}] = \det\Gamma/\det\Gamma_{22}$.

[3] $U = -Z_1\alpha^{*\prime} = -Z\alpha_I'$ by (V.11) and (V.4); therefore $\alpha^*Z_1'Z_1\alpha^{*\prime} = \alpha_I Z'Z\alpha_I'$.

or since Γ_{22} is a known matrix (coefficients of identity equations are assumed known), $\log[\det^2\Gamma_{22}]$ is merely a constant; therefore, (V.16) may be written as:

$$(V.17) \quad f_5(\hat{\alpha},Z) = c_2 + \log[\det^2\hat{\Gamma}] - \log[\det(\tfrac{1}{T}\hat{\alpha}_I[Z'Z]\hat{\alpha}_I')]$$

where c_2 is a constant.[1] Note that Γ is the matrix of coefficients of all G jointly dependent variables in the G equations (including identity equations) whereas α_I is the matrix of coefficients of all $(G + \Lambda)$ variables but only for the M stochastic equations.

To calculate FIML estimates we will select $\hat{\alpha}_I$[2] such that the concentrated logarithmic likelihood function (V.17) is a maximum given our particular sample Z and the restrictions which we have imposed on α (some elements of α_I are assumed to be zero, others are assumed to be -1 for normalization, and all of the elements of α_{II} are assumed known). Notice that it is not necessary to use the identity equations to eliminate jointly dependent variables in order to write down the concentrated logarithmic likelihood function. (We temporarily eliminated some jointly dependent variables in the derivation only.) The computational procedure which will be used to maximize $f_5(\alpha,Z)$ also will not require that the identity equations be used to eliminate jointly dependent variables.[3]

---

[1] $c_2 = c_1 - \log[\det^2\Gamma_{22}]$.

[2] Since $\hat{\alpha} = \begin{bmatrix}\hat{\alpha}_I \\ \hat{\alpha}_{II}\end{bmatrix} = \begin{bmatrix}\hat{\alpha}_I \\ \alpha_{II}\end{bmatrix}$ (α_{II} is assumed known), only $\hat{\alpha}_I$ need be estimated to complete the estimation of $\hat{\alpha}$.

[3] Rothenberg and Leenders [1964], pp. 72, 73 first showed that it is unnecessary to use identity equations to solve out jointly dependent variables to maximize the likelihood function.

The matrix $(1/T)\hat{\alpha}_I(Z'Z)\hat{\alpha}_I'$ is used repeatedly in the elaboration of the computational procedure which follows; hence, it will prove convenient to denote this matrix as S. If we use as an estimate of U the $T^{X}M$ matrix

$$\hat{U} = -Z\hat{\alpha}_I' \quad , \tag{V.18}$$

we can write S as:

$$\underset{M^{X}M}{S} = \hat{\Sigma} = \frac{1}{T}\hat{\alpha}_I(Z'Z)\hat{\alpha}_I' = \frac{1}{T}\hat{U}'\hat{U} \quad . \tag{V.19}$$

Let a_μ be the vector of estimated coefficients of the μ^{th} equation which are not restricted to either zero or -1, and let $_+a_\mu$ be defined as:

$$_+a_\mu = \begin{bmatrix} -1 \\ a_\mu \end{bmatrix} \quad , \tag{V.20}$$

i.e., $_+a_\mu$ is the vector of non-zero coefficients of the μ^{th} equation including the normalizing coefficient. Then S may also be defined as $S = [s_{\mu\mu'}]$ with:

$$s_{\mu\mu'} = \frac{1}{T}{}_+a_\mu'(_+Z_\mu'\ {}_+Z_{\mu'})\ {}_+a_{\mu'} = \frac{1}{T}\hat{u}_\mu'\hat{u}_{\mu'} \tag{V.21}$$

(since $\hat{u}_\mu = -{}_+Z_\mu\ {}_+a_\mu$).

Let us also group all of the unrestricted coefficients in the system into a single vector of coefficients and denote this vector as:

$$a = \begin{bmatrix} a_1 \\ \vdots \\ a_M \end{bmatrix} \quad , \tag{V.22}$$

i.e., $\mathbf{a}$ is the vector containing all of the unrestricted coefficients in $\hat{\alpha}$, all unrestricted coefficients of the first equation (first row of $\hat{\alpha}$) being listed first as the vector a_1, then the unrestricted coefficients of the second equation as the vector a_2, and finally the unrestricted coefficients of the M^{th} equation, as the vector a_M. (The coefficients of the $G - M$ identity equations are all known, i.e., restricted; hence, they are not included in the $\mathbf{a}$ vector.)

Since Z is fixed for any given sample, we choose the unrestricted coefficients of $\hat{\alpha}$ such that $f_5(\hat{\alpha},Z)$ is a maximum; however, for any given structure, the only elements of $\hat{\alpha}$ which are allowed to vary are the elements of the vector $\mathbf{a}$. Thus, for an assumed structure and a given sample, $f_5(\hat{\alpha},Z)$ may be considered a function of the vector $\mathbf{a}$ only, i.e.,

$$(IV.23) \qquad f(a) = f_5(\hat{\alpha},Z) = c_2 + \log(\det {}^2\hat{\Gamma}) - \log(\det S) \quad .$$

Another function which will be maximized when $f(a)$ is maximized is the function[1]

$$(IV.24) \qquad f^*(a) = f_5^*(\hat{\alpha},Z) = \det {}^2\hat{\Gamma}/\det S \quad .$$

We cannot readily maximize either $f(a)$ or $f^*(a)$ by setting partial derivatives equal to zero and solving for $\mathbf{a}$, since the partial derivatives are complicated nonlinear functions of the elements of $\mathbf{a}$. Some iterative procedure is required, and the iterative procedure proposed in this paper is outlined in the next section. In this procedure,

---

[1]The logarithm of $f^*(a)$ differs from $f(a)$ only by a constant, since $\log[\det {}^2\Gamma)/\det S] = \log(\det {}^2\Gamma) - \log(\det S)$.

a set of starting values for the vector **a** is assumed, and then the coefficients in **a** are progressively changed until f(a) reaches a maximum. The first and second partial derivatives of f(a) play a key role in the direction in a-space that **a** is changed at each step of the maximization procedure; however, it is convenient to base the amount of change in a given direction on the function f*(a).

C. Computational Procedure

1. A maximization procedure for functions non-linear in the parameters

Let us first consider maximizing a function $f(a)$ (with a being an n dimensional vector) and assume that (1) $f(a)$ and its derivatives are sufficiently "well behaved" to permit use of a Taylor expansion of sufficiently high order in approximating $f(a)$ about the starting estimates and subsequent points[1] and (2) $f(a)$ has only a single maximum with no additional local maxima in the region which is considered.[2] $f(a)$ may have saddle points in the n dimensional space of a without causing difficulty. As the procedure is outlined for the n dimensional case, it will be illustrated graphically for the 2 dimensional ($n = 2$) case. Further on, in applying the procedure being described, the n ($= \sum_{\mu=1}^{M} n_\mu$) element vector a defined in (V.22) will become our vector a, i.e., a will consist of the vector of all unrestricted coefficients in the system which are to be estimated; however, at this point, a is merely any parameter vector.

---

[1] The use of a Taylor expansion in deriving some properties of the maximization procedure outlined here follows Crockett and Chernoff [1955]. Crockett and Chernoff [1955], p. 34 state that: "For most arguments the use of third or fourth order expansions will suffice."

[2] At this point $f(a)$ may be any function which meets these assumptions. In sections V.C.2 and V.C.3 the procedure will be specialized to the maximization of the $f(a)$ given by (V.23). It is an unanswered question as to whether for the $f(a)$ given by (V.23) multiple local maxima may occur in a region which has a positive probability of being entered through choice of starting coefficients. For a number of problems, the writer has used a set of starting coefficients and calculated the coefficients which maximize the $f(a)$ of (V.23) and then assumed a set of starting coefficients in which all coefficients varied widely from the original set of starting coefficients. In all cases the same maximum was reached. This is encouraging but does not, of course, show that for many problems (or even these problems) multiple local maxima do not occur in regions of interest.

Figure V.25

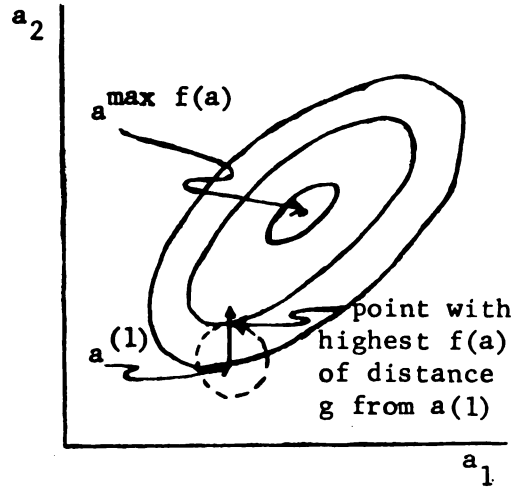
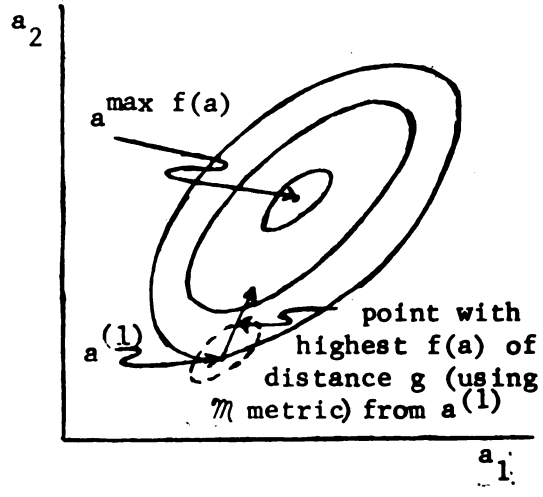


Figure V.26



In Figure V.26 the contour lines represent the same function as the contour lines for Figure V.25. (Each contour line is a locus of points having the same value of the function f(a).) The outermost contour line has the lowest value of f(a), and the innermost one the highest value. The maximum value of f(a) occurs at $a^{\max f(a)}$.

Suppose that we start at an initial point $a^{(1)}$ and consider all points which lie a distance of exactly g units from the point $a^{(1)}$. Next, suppose that the direction d is chosen such that at the distance g from $a^{(1)}$, $f(a^{(1)} + gd)$ is the maximum value which can be reached. (I.e., the locus of all points a distance of g from $a^{(1)}$ is the surface of a sphere of radius g centered at the point $a^{(1)}$. The arrow is drawn through the point on the surface of this sphere [circle in the case of Figure V.25] with the maximum f(a). The angle of the arrow indicates the direction d.)

If we continue in the direction d given by the arrow in Figure V.25, we will pass quite a way to the left of $a^{\max f(a)}$. Suppose that instead of considering the locus of points on the circle, we had considered the locus of points on an ellipse (ellipsoid in the case of an n-dimensional vector a) of approximately the same shape as the contour lines. In Figure V.26 an arrow has been drawn through the point with maximum f(a) on the surface of the ellipse. Notice that the arrow now points much more closely toward the maximum.

An ellipsoid may be traced out instead of a sphere by merely making our concept of distance more general. Let the distance g between any other point, $a^{(i)}$, and $a^{(1)}$ be defined as:

$$g = \sqrt{(a^{(i)} - a^{(1)})' \mathcal{M} (a^{(i)} - a^{(1)})} \tag{V.27}$$

where $\mathcal{M}$ is a positive definite matrix. Then all points at a distance g from $a^{(1)}$ will lie on the surface of an ellipsoid instead of a sphere. In Figure V.26 the matrix $\mathcal{M}$ is such that the resulting ellipsoid is approximately the same shape as the contour lines, while in Figure V.25 the identity matrix has been used as the matrix $\mathcal{M}$ and therefore a circle has been traced out. The matrix $\mathcal{M}$ is termed a metric. The Euclidean metric is represented by $\mathcal{M} = I$ as in Figure V.25.

Suppose that in selecting the direction to move, an arbitrarily small distance g is selected by letting g approach 0. Then, assuming a given metric $\mathcal{M}$, it can be shown that the direction in which f(a) increases the most rapidly from the point $a^{(1)}$ is given by

$d = \mathcal{M}^{-1}{}_{\iota}^{*(1)}$ where ${}_{\iota}^{*(1)}$ is the partial derivative of $f(a)$ evaluated at the point $a^{(1)}$, i.e.,

$$(V.28) \qquad {}_{\iota}^{*(1)} = \frac{\partial f(a)}{\partial a}\Big|_{a=a^{(1)}} = \begin{bmatrix} \frac{\partial f(a)}{\partial a_1} \\ \vdots \\ \frac{\partial f(a)}{\partial a_n} \end{bmatrix}_{a=a^{(1)}} .$$

Crockett and Chernoff show that for any positive definite matrix $\mathcal{M}$ and choice of h sufficiently small, $f(a^{(1)} + h \cdot \mathcal{M}^{-1}{}_{\iota}^{*(1)}) > f(a^{(1)})$ provided ${}_{\iota}^{*(1)} \neq 0$.[1]

Although any positive definite matrix $\mathcal{M}$ can be used for the metric and for h sufficiently small the direction will be sufficiently good that the function will increase (the function will stay the same if we are already at the maximum), some metrics will obviously be better choices than others, e.g., compare the two metrics used in Figures V.25 and V.26.

If $f(a)$ is expanded about the point $a^{(1)}$ in a Taylor expansion, the following is obtained:

$$(V.29) \qquad f(a) = f(a^{(1)}) + [a - a^{(1)}]'{}_{\iota}^{*(1)} + [a - a^{(1)}]'\mathcal{L}^{*(1)}[a - a^{(1)}] + \text{higher order terms}$$

where ${}_{\iota}^{*(1)}$ is the $n \times 1$ vector of first partial derivatives of $f(a)$

---

[1] Crockett and Chernoff [1955], p. 35. If $a^{(1)}$ is the maximum then ${}_{\iota}^{*(1)} = 0$ and $f(a^{(1)} + h\mathcal{M}^{-1}{}_{\iota}^{*(1)}) = f(a^{(1)} + h\mathcal{M}^{-1}0) = f(a^{(1)})$; i.e., no movement is made away from the maximum.

with respect to the elements of a evaluated at $a^{(1)}$ as before, and

$\mathcal{L}^{*(1)}$ is the $n^{X}n$ matrix of second partial derivatives of $f(a)$ with respect to the elements of a evaluated at the point $a^{(1)}$.

For any function $f(a)$ with continuous first and second partial derivatives, a local maximum occurs at any point satisfying $\frac{\partial f(a)}{\partial a} = 0$ if $\frac{\partial^2 f(a)}{\partial a^2}$ is a negative definite matrix at that point.

If the partial derivative of the Taylor expansion given in (V.29) is taken with respect to a, ignoring all higher order terms, we get:

$$\left(\frac{\partial f(a)}{\partial a}\right)' = {}^{*(1)}_{\iota} + \frac{1}{2}\cdot 2\cdot \mathcal{L}^{*(1)}[a - a^{(1)}] \quad . \tag{V.30}$$

Setting $\frac{\partial f(a)}{\partial a}$ to zero and solving for a, we get:

$$[a - a^{(1)}] = (\mathcal{L}^{*(1)})^{-1}{}^{*(1)}_{\iota} \tag{V.31}$$

or

$$a = a^{(1)} - (\mathcal{L}^{*(1)})^{-1}{}^{*(1)}_{\iota} \quad . \tag{V.32}$$

Also, taking the partial derivative of (V.30) with respect to a we get:

$$\frac{\partial^2 f(a)}{\partial a^2} = \mathcal{L}^{*(1)} \quad . \tag{V.33}$$

Since the second partial derivative must be negative definite for the point to be a local maximum, we get the additional condition $\mathcal{L}^{*(1)}$ must be negative definite or $-\mathcal{L}^{*(1)}$ must be positive definite

for the point to represent a local maximum.

To summarize, if it is assumed that consideration of only the first three terms of a Taylor expansion about $a^{(1)}$ will give a sufficiently close approximation to $f(a)$, then the following holds in a sufficiently small region of $a^{(1)}$:

$$(V.34) \qquad a^{\max f(a)} = a^{(1)} - (\mathscr{L}^{*(1)})^{-1} {}_{1}^{*(1)} \quad .$$

Multiplying the metric and the partial derivative by T will not change the result. Thus, (V.34) is equivalent to:[1]

$$(V.35) \qquad a^{\max f(a)} = a^{(1)} + (\mathscr{L}^{(1)})^{-1} {}_{1}^{(1)}$$

where $\mathscr{L}^{(1)} = -T\mathscr{L}^{*(1)}$ and ${}_{1}^{(1)} = T {}_{1}^{*(1)}$.

Since $a^{(1)}$ will usually be some distance from $a^{\max f(a)}$, the procedure indicated by formula (V.35) will be modified as follows:

(1) If $\mathscr{L}^{(1)}$ is not positive definite, it is adjusted to form another matrix, $|\mathscr{L}^{(1)}|$, which is positive definite, and $|\mathscr{L}^{(1)}|$ is used as the metric.[2]

(2) Instead of using a step size of 1 as (V.34) implies, a step size of $h^{(1)}$ is used, i.e., $a^{(1)} + h^{(1)} \cdot |\mathscr{L}^{(1)}|^{-1} {}_{1}^{(1)}$ is used.[3]

(3) A check will be made that, given the step size $h^{(1)}$, $f(a^{(1)} + h^{(1)} \cdot |\mathscr{L}^{(1)}|^{-1} {}_{1}^{(1)})$ is indeed greater than $f(a^{(1)})$.

---

[1] If the variables were not normalized in the manner indicated in Part III before computation begins, multiplication by T might be expected to increase rounding error; however, due to the normalization used, rounding error will not be increased.

[2] A method for forming the $|\mathscr{L}^{(1)}|$ matrix from the $\mathscr{L}^{(1)}$ matrix is given in section V.C.3.

[3] Determination of the step size $h^{(1)}$ is discussed in section V.C.4.

(4) A series of steps will be taken with the direction and step size recomputed at each step until $a^{\max} f(a)$ is reached.[1]

Instead of recalculating the metric each iteration, the same metric may be used for a number of iterations, only the vector of partial derivatives being recalculated each iteration. The writer has not compared total time required for convergence if the same metric is used for a number of iterations with the total time required for convergence if a new metric is calculated each iteration. Up to now, the writer has always recalculated the metric each iteration (and has obtained rapid convergence on all problems attempted).

---

[1] Determination of when convergence or maximimization is achieved is discussed in section V.C.5.

2. The vector of partial derivatives for FIML

In expressing the vector of partial derivatives and some alternative metrics for FIML, a number of intermediate matrices will be calculated from a given vector of coefficients, a, and the sample values of the variables, Z (more particularly from the Z'Z matrix), given the assumed structure. The Z'Z matrix and the assumed structure do not, of course, change between iterations. Only the vector of coefficients changes (given Z and the assumed structure, f(a) is a function of a, only). We will use a superscript on the a vector to indicate that a particular set of coefficients are used (e.g., $a^{(i-1)}$). No superscript will be used on the intermediate matrices calculated from the $a^{(i-1)}$ vector and Z'Z matrix, since it will be obvious from the formulas given which intermediate matrices change wherever a new vector of coefficients is used. The matrix $\hat{\Gamma}$ will be treated as an intermediate matrix of this form (i.e., $\hat{\Gamma}$ will not be superscripted), since it is formed from (1) the elements of the coefficient vector a, corresponding to jointly dependent variables in the system, (2) the coefficients of the identity equations corresponding to jointly dependent variables in the system, (3) zeros, and (4) -1's.

In computing the direction for the i^{th} iteration, $T\frac{\partial f(a)}{\partial a}$, evaluated at $a^{(i-1)}$ (i.e., with the coefficients obtained from the preceding iteration used as a) is the right hand side term by which the metric inverse is first multiplied. This term may be written as:[1]

---

[1]See Rothenberg and Leenders [1964], p. 61, 63-64 for a derivation of the vector of partial derivatives. Rothenberg and Leender's notation differs slightly, since their logarithmic likelihood function is 1/2 of the f(a) given above, i.e., their logarithmic likelihood function is

$$(V.36) \qquad {}_{1}^{(i-1)} = T\frac{\partial f(a)}{\partial a}\Big|_{a=a^{(i-1)}} = T\begin{bmatrix} \frac{\partial f(a)}{\partial a_1} \\ \vdots \\ \frac{\partial f(a)}{\partial a_M} \end{bmatrix}_{a=a^{(i-1)}} .$$

For FIML, the part of the right hand side corresponding to the unrestricted coefficients of the μ^{th} equation may be written as:

$$(V.37) \qquad {}_{1\mu}^{(i-1)} = T\frac{\partial f(a)}{\partial a_\mu}\Big|_{a=a^{(i-1)}} = T\begin{bmatrix} \hat{\gamma}\{\mu|\mu\} \\ 0 \end{bmatrix} + \sum_{\mu'=1}^{M} s^{\mu\mu'} Z_\mu' \hat{u}_{\mu'},$$

$$= T\begin{bmatrix} \hat{\gamma}\{\mu|\mu\} \\ 0 \end{bmatrix} + Z_\mu' \hat{U} \begin{bmatrix} s^{1\mu} \\ \vdots \\ s^{M\mu} \end{bmatrix}$$

where

$\hat{\gamma}\{\mu|\mu\}$ is the part of the μ^{th} column of $\hat{\Gamma}^{-1}$ corresponding to the unrestricted coefficients of the jointly dependent variables of the μ^{th} equation. ($\hat{\Gamma}$ is a $G \times G$ matrix; that is, the coefficients of jointly dependent variables in the identity equations are included in the $\hat{\Gamma}$ matrix. Only a part of $\hat{\Gamma}^{-1}$ is used--the part corresponding to the unrestricted coefficients of the stochastic equations. $\gamma\{\mu|\mu\}$ is an $m_\mu \times 1$ vector.[1]

$(1/2)f(a) = (1/2)[c_1 + \log(\det 2\hat{\Gamma}) - \log(\det S)]$

$= k' + (1/2)\cdot 2 \log[\text{abs}(\det \hat{\Gamma})] - 1/2 \log(\det S)$ where $k' = (1/2)c_2$.
Also the vector of partial derivatives and the metric have been multiplied by T in this paper as noted in the conversion from (V.34) to (V.35).

[1]The rather unique $\hat{\gamma}\{\mu|\mu\}$ notation should seem more justified when the $\hat{\gamma}\{\mu'|\mu\}$ and $\hat{\gamma}\{\mu|\mu'\}$ vectors are defined during the formation of the $\mathscr{L}$ metric (section V.C.3).

0 is an $\ell_\mu \times 1$ column vector of zeros corresponding to the predetermined variables in the μ^{th} equation.

$s^{\mu\mu'}$ is the element of the μ^{th} row and the μ'^{th} column of S^{-1}.

$$-Z'_{\mu}\hat{u}_{\mu'} = (Z'_{\mu} {}_{+}Z_{\mu'})_{+}a_{\mu'}^{(i-1)}$$

$$-Z'_{\mu}U = \left[(Z'_{\mu} {}_{+}Z_1)_{+}a_1^{(i-1)} \cdots (Z'_{\mu} {}_{+}Z_M)_{+}a_M^{(i-1)}\right] \quad {}_1$$

S is an $M \times M$ matrix corresponding to the M stochastic equations.

Notice that given the assumed structure and a set of sample values of variables, Z, the matrices and vectors given above are all calculated from a given vector of coefficients a or from intermediate matrices calculated through use of the vector a. A set of starting coefficients such as from 2SLS, LIML, DLS, 3SLS, or merely a set of "assumed" coefficients may serve as starting coefficients in calculating the first metric and right hand side.[2] The coefficients from the $i\text{-}1^{st}$ iteration are used in calculating the metric and right hand side for the i^{th} iteration.

${}_1{}_{+}a_{\mu}^{(i-1)} = \begin{bmatrix} -1 \\ a_{\mu}^{(i-1)} \end{bmatrix}$ and ${}_{+}Z_{\mu} = [y_{\mu} \vdots Z_{\mu}]$ as before.

[2] As noted in section V.C.1 the question of whether multiple local maxima occur in the region of interest is an unanswered question. To date the writer is not aware of a problem in which two sets of starting coefficients have led to different local maxima. (If rounding error were large any problem might appear to have many local maxima.)

3. Metrics for FIML

A rationale for suggesting the use of $-T\frac{\partial^2 f(a)}{\partial a^2}$ as the metric to use in determining direction was derived earlier. We will call this metric the $\mathcal{L}$ metric.[1] Often $a^{(i-1)}$, the vector of unrestricted coefficients to use in calculating the metric and the vector of partials for the i^{th} iteration will be sufficiently far away from $a^{\max f(a)}$ that $\mathcal{L}^{(i-1)}$ is not positive definite. For this reason, we will consider three other metrics--the $\mathcal{P}$ metric, the $\mathcal{R}$ metric, and the $|\mathcal{L}|$ metric. Of these metrics we will use only the $|\mathcal{L}|$ metric in our FIML iterations. The formulas for the $\mathcal{P}$ and $\mathcal{R}$ metrics are given so that we can draw correspondences to other methods involving one or more iterations such as the Zellner-Aitken estimator (ZA), iteration on the ZA estimator (IZA), 3SLS, and I3SLS. (All of these methods are discussed in subsequent chapters.) The $\mathcal{P}$ and $\mathcal{R}$ metrics are used during early iterations in some FIML computational schemes, but not the computational scheme given in this paper.

The metrics are most easily defined by dividing them into blocks which correspond to the division of the coefficient vector a according to the equations from which a came. Let $\mathcal{M}$ be an arbitrary metric subdivided as:

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_{11} & \cdots & \mathcal{M}_{1M} \\ \vdots & & \vdots \\ \mathcal{M}_{M1} & \cdots & \mathcal{M}_{MM} \end{bmatrix} . \tag{V.38}$$

---

[1] $\frac{\partial^2 f(a)}{\partial a^2}$ is the usual Newton metric.

where $\mathcal{M}_{\mu\mu'}$ is the block of the metric whose n_μ rows correspond to the unrestricted coefficients of the μ^{th} equation and whose $n_{\mu'}$ columns correspond to the unrestricted coefficients of the μ'^{th} equation.

In this section, we will omit from the metric the iteration number specifying the set of coefficients used as the vector a in evaluating the metric.

The $\mu\mu'^{th}$ block of the $\mathcal{P}$ metric is:[1]

$$\text{(V.39)} \qquad \mathcal{P}_{\mu\mu'} = s^{\mu\mu'} Z_\mu' Z_{\mu'} \quad .$$

The $\mathcal{P}$ metric is the $-T\frac{\partial^2 f(a)}{\partial a^2}$ matrix derived if the $\log(\det^{2}\hat{\Gamma})$ term of (V.23) is ignored in defining $f(a)$. [The $\log(\det^{2}\hat{\Gamma})$ term will not appear in $f(a)$ if the Jacobian of the transformation from U to α and Z is 1, or if it is ignored.]

The $\mu\mu'^{th}$ block of the R metric is:[2]

$$\text{(V.40)} \qquad \mathcal{R}_{\mu\mu'} = s^{\mu\mu'}[Z_\mu' Z_{\mu'}]_{\parallel X} = s^{\mu\mu'} \begin{bmatrix} [Y_\mu' Y_{\mu'}]_{\parallel X} & Y_\mu' X_{\mu'} \\ X_\mu' Y_{\mu'} & X_\mu' X_{\mu'} \end{bmatrix}$$

$$= s^{\mu\mu'} \begin{bmatrix} Y_\mu' Y_{\mu'} - [Y_\mu' Y_{\mu'}]_{\perp X} & Y_\mu' X_{\mu'} \\ X_\mu' Y_{\mu'} & X_\mu' X_{\mu'} \end{bmatrix} .$$

($[Y_\mu' X_{\mu'}]_{\parallel X} = Y_\mu' X_{\mu'}$, $[X_\mu' Y_{\mu'}]_{\parallel X} = X_\mu' Y_{\mu'}$, and $[X_\mu' X_{\mu'}]_{\parallel X} = X_\mu' X_{\mu'}$, since X_μ and $X_{\mu'}$ are submatrices of X [the matrix of predetermined variables in the system]-- see (I.54) and (I.40).)

---

[1] The $\mathcal{P}$ metric changes each iteration since S is calculated from the vector of coefficients and, therefore, $s^{\mu\mu'}$ changes each iteration.

[2] The $\mathcal{R}$ metric changes each iteration since S is calculated from the vector of coefficients and, therefore, $s^{\mu\mu'}$ changes each iteration.

If X has full column rank, then (V.40) may also be written as:

$$(V.41) \qquad \mathcal{R}_{\mu\mu'} = s^{\mu\mu'} Z_\mu' X (X'X)^{-1} X' Z_{\mu'} \quad ;$$

however, (V.40) is a preferable computational formula since $[Y'Y]_{\perp X}$ may be calculated by the orthogonalization procedure of section I.D.2, $[Y'Y]_{\shortparallel X}$ calculated as $Y'Y - [Y'Y]_{\perp X}$, and all of the matrices of the form $[Y_\mu' Y_{\mu'}]_{\shortparallel X}$ extracted as submatrices of $[Y'Y]_{\shortparallel X}$.

The $\mathcal{R}$ metric was derived by Rubin as a matrix whose inverse is asymptotically the same as $\mathcal{L}^{-1}$, $(\mathcal{L}^{\max f(a)})^{-1}$ being the asymptotic maximum likelihood estimator of the FIML coefficient variance-covariance matrix.[1]

The $\mu\mu'^{\text{th}}$ block of $\mathcal{L}$ for the FIML estimator is:[2]

$$(V.42) \qquad \mathcal{L}_{\mu\mu'} = T H_{\mu\mu'} + s^{\mu\mu'} Z_\mu' Z_{\mu'} - (1/T) Z_\mu' \hat{U} F_{\mu\mu'} \hat{U}' Z_{\mu'}$$

where:[3]

$$-Z_\mu' \hat{U} = -Z_\mu' [\hat{u}_1 \cdots \hat{u}_M] = [(Z_\mu' {}_+Z_1) {}_+a_1 \cdots (Z_\mu' {}_+Z_M) {}_+a_M]$$

---

[1] Its use is suggested in Rubin [1948] and Chernoff and Divinsky [1953]. Later yet, the $\mathcal{R}$ matrix became important as the matrix inverted in the calculation of 3SLS. By showing that $\hat{\text{Var}}(\hat{\delta}_{3SLS})$ asymptotically equals $\hat{\text{Var}}(\hat{\delta}_{FIML})$, Zellner and Theil [1962] and others have also shown that asymptotically $\mathcal{R}^{-1}$ equals $\mathcal{L}^{-1}$, since the usual asymptotic $\hat{\text{Var}}(\hat{\delta}_{3SLS})$ is $\mathcal{R}^{-1}$ and the usual asymptotic $\hat{\text{Var}}(\hat{\delta}_{FIML})$ is $\mathcal{L}^{-1}$.

[2] See Rothenberg and Leenders [1964], pp. 63-65 for a derivation of $\mathcal{L}_{\mu\mu'}$. Rothenberg and Leender's logarithmic likelihood function is 1/2 of the logarithmic likelihood being maximized here. Also, the vector of partial derivatives and the metric have been multiplied by T in this paper (see footnote 1 p. 179).

The $\mathcal{L}$ matrix changes each iteration since the vector of coefficients is used in the calculation of the intermediate matrices given below.

[3] ${}_+a_\mu = \begin{bmatrix} -1 \\ a_\mu \end{bmatrix}$ and ${}_+Z_\mu = [y \vdots Z_\mu]$ as before.

$$F_{\mu\mu'} = \begin{bmatrix} s^{1\mu'} \\ \vdots \\ s^{M\mu'} \end{bmatrix} [s^{1\mu} \cdots s^{M\mu}] + s^{\mu\mu'} S^{-1}$$

$$H_{\mu\mu'} = \begin{bmatrix} \hat{\gamma}\{\mu'|\mu\}\hat{\gamma}\{\mu|\mu'\}' & 0 \\ m_\mu \times m_{\mu'} & m_\mu \times \ell_{\mu'} \\ 0 & 0 \\ \ell_\mu \times m_{\mu'} & \ell_\mu \times \ell_{\mu'} \end{bmatrix} \quad \text{with:}$$

$\hat{\gamma}\{\mu'|\mu\}$ being a vector containing the m_μ elements of the μ^{th} column of $\hat{\Gamma}^{-1}$ corresponding to the m_μ unrestricted coefficients of the jointly dependent variables of equation μ. ($\hat{\Gamma}$ is a GXG matrix, since the coefficients of jointly dependent variables in the identity equations are included in the Γ matrix.)

$\hat{\gamma}\{\mu|\mu'\}$ being a vector containing the $m_{\mu'}$ elements of the μ^{th} column of $\hat{\Gamma}^{-1}$ corresponding to the $m_{\mu'}$ unrestricted coefficients of the jointly dependent variables of equation μ'.

The $\mathcal{P}$ and $\mathcal{R}$ metrics are always positive definite or positive semi-definite. (Under some circumstances the $\mathcal{P}$ and $\mathcal{R}$ metrics will be singular; therefore, positive semi-definite.) The $\mathcal{L}$ metric will generally not be positive definite except close to the maximum; that is, the $\mathcal{L}$ metric will generally not be positive definite when iteration is started.[1]

---

[1] A notable exception is Klein's model I. Although it is not positive definite when starting from the 2SLS estimates, only a few iterations are usually required to move the estimates into a region in which the $\mathcal{L}$ metric is positive definite. As a result of being so well conditioned, Klein's model I is somewhat misleading since many procedures which would have difficulty converging for many problems which may be encountered will converge easily for Klein's model I, the model on which they are usually tested.

Chernoff and Divinsky recommended use of the $\mathcal{P}$ metric initially, the $\mathcal{R}$ metric for a number of iterations and finally the $\mathcal{L}$ metric when the estimated coefficient vector, a, becomes close to the coefficients which maximize the likelihood. They also suggested some guides for switching from one metric to another.[1]

Eisenpress wrote the first large scale FIML computer routine available for general use and has calculated a wide variety of problems to it. Initially he programmed it using the $\mathcal{P}$, $\mathcal{R}$, and the $\mathcal{L}$ metrics as suggested by Chernoff and Divinsky. After considerable experimentation with speed of convergence (and whether convergence would occur at all for that matter), he along with John Greenstadt devised the $|\mathcal{L}|$ metric which he then used as the only metric.[2]

$\mathcal{L}$ can be expressed as $\mathcal{L} = E\lambda E'$ where λ is a diagonal matrix with the eigenvalues of $\mathcal{L}$ forming the diagonal elements and E is a matrix whose columns are eigenvectors of $\mathcal{L}$, the eigenvectors being in the same order as their corresponding eigenvalues on the diagonal of λ. (Any symmetric matrix may be decomposed in this fashion.) Let $|\mathcal{L}|$ be defined as:

$$|\mathcal{L}| = E|\lambda|E' \tag{V.43}$$

where $|\lambda|$ is the same as λ except that the absolute values of the eigenvalues replace the actual eigenvalues.

---

[1] Chernoff and Divinsky [1953].

[2] Personal conversation. Also, Eisenpress [date unknown].

As with any non-singular symmetric matrix, $\mathcal{L}^{-1}$ may be formed directly as:

$$(V.44) \qquad \mathcal{L}^{-1} = [E\lambda E']^{-1} = (E')^{-1}\lambda^{-1}E^{-1} = E\lambda^{-1}E'$$

(since $E^{-1} = E'$).

Thus, $|\mathcal{L}|^{-1}$ may also be formed directly as

$$(V.45) \qquad |\mathcal{L}|^{-1} = E|\lambda^{-1}|E'$$

where $|\lambda^{-1}|$ is formed by replacing each diagonal element of λ by 1 divided by its absolute value.[1]

Assuming that $\mathcal{L}$ is not singular, $|\mathcal{L}|$ will be positive definite by its method of calculation. Eisenpress has examined the effect of setting negative eigenvalues positive and recombining the matrices. The basic directions (the axis system of the metric being formed) are established by the eigenvectors with the eigenvalues determining the distance along each axis. The negative eigenvalues indicate a movement along an axis away from the maximum. Setting them positive results in movement along the axes (given by the eigenvectors) in the correct direction. Since the inverse of a metric corresponds to using the inverse of the eigenvalues, a large eigenvalue corresponds to a small movement along an eigenvector axis and a small eigenvalue corresponds to a large movement. Since negative eigenvalues are generally small in absolute value, setting them positive results in switching the direction from a large negative direction to a large positive

---

[1] Since all off-diagonal elements of $|\lambda^{-1}|$ are zero, $|\mathcal{L}|^{-1}$ can be formed more efficiently by forming the ij^{th} element of $|\mathcal{L}|^{-1}$ as $\sum_{k=1}^{n} \frac{1}{\text{abs } \lambda_k} e_{ik} e_{jk}$ where λ_k is the k^{th} eigenvalue of λ and $E = [e_{ij}]$.

direction. Eisenpress feels that this is correct based on the geometry of the situation. Convergence has been much more rapid when this is done than in some earlier experiments in which he substituted zero for the inverse of the negative eigenvalues (thereby moving along the eigenvectors only in the directions corresponding to positive eigenvalues).[1]

$\mathcal{L}^{-1}$ and $|\mathcal{L}|^{-1}$ coincide if all eigenvalues are positive and since direct inversion of $\mathcal{L}$ to form $\mathcal{L}^{-1}$ is faster than forming $|\mathcal{L}|^{-1}$ in the manner indicated above, the writer has modified the procedure as follows:

(1) Use the $|\mathcal{L}|^{-1}$ metric until 3 consecutive iterations have been performed in which all eigenvalues have been found to be positive.

(2) Use the $\mathcal{L}^{-1}$ metric by forming the metric through direct inversion of $\mathcal{L}$ for 5 iterations.

(3) Use $|\mathcal{L}|^{-1}$ for an iteration. If all eigenvalues are positive start with (2) again, otherwise start with (1) again.

(4) When convergence to $a^{\max f(a)}$ appears to have occurred, check as to whether all eigenvalues of $\mathcal{L}$ are positive. If any are negative, convergence has actually been to a saddle point. Switch to the $|\mathcal{L}|$ metric and start with (1) again so that movement will be past the saddle point toward the maximum. All eigenvalues of $\mathcal{L}$ will be positive at a local maximum.

It is not necessarily true that once a positive definite region for $\mathcal{L}$ is entered, $\mathcal{L}$ will continue to be positive definite. A non-positive definite region may again be encountered, e.g., one or more

---

[1]The writer hopes that he has not misrepresented Eisenpress and Greenstadt's developments in any of the above.

saddle points may be encountered as movement is made toward the maximum. No difficulty should be encountered in moving past a non-positive definite region if a reversion to the $|\mathcal{L}|$ metric is made.

4. Step size to use at each iteration

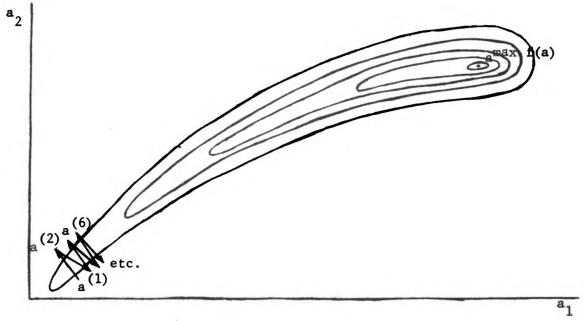
Earlier it was indicated that in the FIML convergence procedure, the coefficients for an iteration, $a^{(1)}$, are to be calculated from the coefficients of the previous iterations as:

$$a^{(i)} = a^{(i-1)} + h^{(i)} d^{(i)} \tag{V.46}$$

where $d^{(i)} = |\mathcal{H}^{(i-1)}|^{-1} \iota^{(i-1)}$ and $h^{(i)}$ is the step size for the iteration.

So far we have covered only the direction in which we will move from a set of coefficients in the space of unrestricted coefficients at any one iteration. The distance we move at any iteration can also be very important. Consider the situation given by Figure V.47.

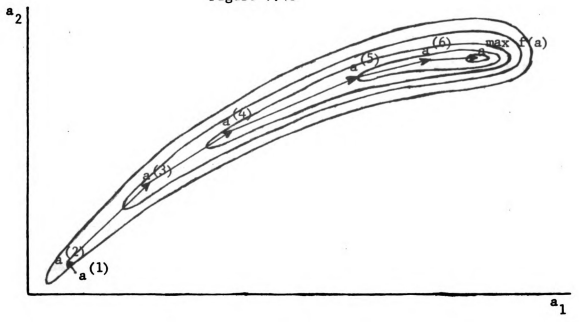
Figure V.47



Assume that $a^{\max f(a)}$ is the point of maximum likelihood and the contour lines represent points with equal value of the likelihood function. Assuming that direction d_1 is taken with step size h_1 for the first iteration, d_2 is taken with step size h_2 for the second iteration, etc., we may easily end up spending many iterations trying to move up the long narrow ridge (or we could even move down the ridge) due to using a series of step sizes which are too large. Figure V.47 illustrates this.

On the other hand, if for each step we could vary the step size such that we land somewhere near the top of the same ridge our situation could be as in Figure V.48.

Figure V.48



If too small a step is taken, the above may take many steps. Thus, our step size as well as our direction takes on considerable importance.

Following is a scheme which may be used to determine the step size, h, to be used for an iteration:

(1) For the usual iteration, a step size of $h_1^{(i)}$ is tried, i.e., $f^*(a^{(i-1)} + h_1^{(i)}d^{(i)})$ is calculated. (Determination of a starting value, $h_1^{(i)}$ is discussed further on.)

(a) If $f^*(a^{(i-1)} + h_1^{(i)}d^{(i)}) > f^*(a^{(i-1)})$, a trial step size twice as large, i.e., $2h_1^{(i)}$ is tried. If $f^*(a^{(i-1)} + 2h_1^{(i)}d^{(i)}) > f^*(a^{(i-1)} + h_1^{(i)}d^{(i)})$, a step size twice as large, i.e., $4h_1^{(i)}$ is tried. This process is continued until the j^{th} time a step size twice as large is tried, $f^*(a^{(i-1)} + 2^j h_1^{(i)}d^{(i)}) \leq f^*(a^{(i-1)} + 2^{j-1}h_1^{(i)}d^{(i)})$. At that time a quadratic approximation is used to calculate a step size which we will call h_2. If $f^*(a^{(i-1)} + h_2 d^{(i)}) > f^*(a^{(i-1)} + 2^{j-1}h_1^{(i)}d^{(i)})$, h_2 is used as the step size, h, for the iteration. Otherwise, $2^{(j-1)}h_1^{(i)}$ is used as the step size, $h^{(i)}$, for the iteration.

(b) On the other hand, if $f^*(a^{(i-1)} + h_1^{(i)}d^{(i)}) \leq f(a^{(i-1)})$, a trial step size half as large, i.e., $\frac{1}{2}h_1^{(i)}$ is tried. If $f^*(a^{(i-1)} + \frac{1}{2}h_1^{(i)}d^{(i)}) < f^*(a^{(i-1)})$, a step size half as large, i.e., $\frac{1}{4}h_1^{(i)}$ is tried. This process is continued until at the j^{th} time a step size half as large is tried either $f^*(a^{(i-1)} + (1/2)^j h_1^{(i)}d^{(i)}) \geq f^*(a^{(i-1)})$ or $(1/2)^j h_1^{(i)} < \varepsilon_h$.

(aa) If $f^*(a^{(i-1)} + (1/2)^j h_1^{(i)}d^{(i)}) > f^*(a^{(i-1)})$, a quadratic approximation is used to calculate a step size which we will call h_2. If $f^*(a^{(i-1)} + h_2 d^{(i)}) > f^*(a^{(i-1)} + (1/2)^j h_1^{(i)}d^{(i)})$, h_2 is used as the step size, $h^{(i)}$, for the iteration. Otherwise $(1/2)^j h_1^{(i)}$ is used as the step

size, $h^{(i)}$, for the iteration.

(bb) If $f^*(a^{(i-1)} + (1/2)^j h_1^{(i)} d^{(i)}) = f^*(a^{(i-1)})$, $(1/2)^j h_1^{(i)}$ is used as the step size, $h^{(i)}$, for the iteration.

(cc) If $(1/2)^j h_1^{(i)} < \epsilon_h$, a negative step, $-(1/2)^j h_1^{(i)}$ is tried.

(aaa) If $f^*(a^{(i-1)} - (1/2)^j h_1^{(i)} d^{(i)}) > f^*(a^{(i-1)})$, the trial step size is doubled [from $-(1/2)^j h_1^{(i)}$], redoubled, etc. in a negative direction in the same manner as was done in a positive direction in step (1a) above. When a step size such that $f^*(a^{(i-1)} - 2^k(1/2)^j h_1^{(i)} d^{(i)}) < f^*(a^{(i-1)} - 2^{(k-1)}(1/2)^j h_1^{(i)} d^{(i)})$ is reached, a quadratic approximation is used to calculate a step size--h_2.

If $f^*(a^{(i-1)} + h_2 d^{(i)}) > f^*(a^{(i-1)} - 2^{k-1}(1/2)^j h_1^{(i)} d^{(i)})$, h_2 is used as the step size, $h^{(i)}$, for the iteration. Otherwise, $-2^{(k-1)}(1/2)^j$ is used as the step size, $h^{(i)}$, for the iteration.

If $f^*(a^{(i-1)} - 2^k(1/2)^j h_1^{(i)} d^{(i)}) = f^*(a^{(i-1)} = 2^{k-1}(1/2)^j h_1^{(i)} d^{(i)})$ then $-2^k(1/2)^j h_1^{(i)}$ is used as the step size, $h^{(i)}$, for the iteration.

(bbb) If $f^*(a^{(i-1)} - (1/2)^j h_1^{(i)} d^{(i)}) \leq f^*(a^{(i-1)})$, $(1/2)^{j+1} h_1^{(i)}$ is used as the step size, $h^{(i)}$, for the iteration.

(2) If (a) the absolute value of all elements of ι are less than a preassigned epsilon, ϵ_ι, i.e., the partial convergence criterion (discussed farther on in the section on convergence criteria) has

been met; (b) $\mathcal{L}^{(i-1)}$ is positive definite; and (c) a positive step size was used the previous iteration, then an initial step size of 1 is tried.[1] If $f^*(a^{(i-1)} + d^{(i)}) \geq f^*(a^{(i-1)})$, the trial step size is doubled, redoubled, etc. and a quadratic approximation tried as for the usual iteration. If $f^*(a^{(i-1)} + d^{(i)}) < f^*(a^{(i-1)})$ a step size of 1 is used even though it leads to a slightly lower likelihood value. The reason for imposing a step size of at least 1 is that $a^{(i-1)}$ is apparently very close to the maximum. Given the small size of the elements of $\iota^{(i-1)}$ a small step size would lead to almost no movement. If the step size of 1 should be larger than optimal the next iteration can easily readjust the direction and size, since the $\mathcal{L}^{(i-1)}$ metric is very powerful close to the maximum. If the point is not close to the maximum, but the elements of ι are quite small due to the likelihood being almost flat, then again a large step is desirable so that a large movement will be made.

The Initial Step Size, $h_1^{(i)}$

The derivation of a maximization procedure based on the Taylor expansion given in section V.C.1 suggests a step size of 1 as an optimal step size; however, this argument is based on the Taylor expansion about $a^{(i-1)}$ being a sufficiently good approximation to the likelihood function. If $a^{(i-1)}$ is not very close to $a^{\max f(a)}$ then the Taylor expansion is unlikely to be a good approximation to the likelihood function. There

---

[1] ι_ι rather than ι^*_ι is compared to an epsilon due to the normalization of variables used in the computation. (Normalization of variables is discussed in chapter IX.)

are other arguments for a step size of 1 as being optimal but most of these also founder on some noble assumption. It is generally advantageous to allow the step size to vary even though the $\mathcal{L}$ metric is used and the current region is one in which the $\mathcal{L}$ metric is positive definite. Only when $a^{(i-1)}$ is virtually at the maximum does it seem somewhat desirable to limit the step size to 1 and even there the rules given previously for step size allow the step size to be greater than 1.

In applying the step size rules given previously, a step size of less than 1 is selected a far greater proportion of the steps than a step size of 1 in the FIML problems calculated by the writer. (An exception was Klein's model I where the average step size selected was .9.) As a result, the writer has currently set $h_1^{(i)}$ to .5 .

At first glance, it might seem that $h_1^{(i)}$ might be set to the step size used the preceding iteration or some proportion of this step size. This would be a very undesirable choice for $h_1^{(i)}$, however, as it turns out in practice that large step sizes tend to be followed by small ones, and vice-versa. If any "rule" is to be selected it should probably make $h_1^{(i)}$ inversely related to $h^{(i-1)}$, but the relationship is really to tenuous to be relied on.

It is quite possible that a variable $h_1^{(i)}$ which is better than any fixed h_1 could be calculated based on the eigenvalues of the metric for those iterations in which the eigenvalues and eigenvectors of the metric are calculated; however, the writer has not attempted to develop such a rule.

Minimum Step Size, Negative Step Size, and ϵ_h

Currently an ϵ_h of .001 is being used by the writer. As noted earlier, when the trial step value becomes less than ϵ_h a negative step of the same size is tried. If this gives a likelihood higher than the likelihood for the previous iteration, searching goes on in the negative direction for a higher likelihood value, yet. If the negative step does not give a higher likelihood value, half of the previous positive step is used as the likelihood value.

If a negative step is selected, the $|\mathcal{L}|$ metric is automatically used for at least the next 3 iterations. It is expected that a negative step will be selected only in rare pathological cases. It has been programmed into the procedure as a matter of interest to see whether such cases arise rather than in the expectation that it will provide a key element in the iteration scheme.

The selection of an ϵ_h of .001 is, of course, quite arbitrary. In selecting an ϵ_h, it is necessary to weigh the desirability of selecting a step size small enough that the likelihood is increased or at least not decreased against the extra time it takes to calculate the likelihood value at any given step size and the fact that if too small a step is taken, $a^{(i)}$ will almost coincide with $a^{(i-1)}$, (This follows the old army adage, "Do something, even if it is wrong!")

Quadratic Approximation

The quadratic approximation referred to previously consists of calculating the second degree polynomial which fits three equally spaced

points exactly.[1] If $a^{(i-1)}$ is the value of a at the start of the iteration, h^* is the step size with $f^*(a^{(i-1)} + h^*d^{(i)}) \geq f^*(a^{(i-1)})$, and $f^*(a^{(i-1)} + 2h^*d^{(i)}) \leq f^*(a^{(i-1)} + h^*d^{(i)})$, then the point $a^{(i-1)} + h^{**}d^{(i)}$ will be the maximum value of the quadratic function which goes through the three points $a^{(i-1)}$, $a^{(i-1)} + h^*d^{(i)}$, and $a^{(i-1)} + 2h^*d^{(i)}$ when h^{**} is calculated as:

(V.49) $h^{**} =$

$$h^*\left[1 + \frac{f^*(a^{(i-1)}) - f^*(a^{(i-1)} + 2h^*d^{(i)})}{2\{f^*(a^{(i-1)} + 2h^*d^{(i)}) + f^*(a^{(i-1)}) - 2f^*(a^{(i-1)} + h^*d^{(i)})\}}\right]$$

If in calculating (V.49) the denominator is zero, h^{**} is set to $2h^*$ if $f^*(a^{(i-1)} + 2h^*d^{(i)}) = f^*(a^{(i-1)})$ [and therefore $f^*(a^{(i-1)} + h^*d^{(i)})$ will also equal $f^*(a^{(i-1)})$] and h^{**} is set to h^* if $f^*(a^{(i-1)} + 2h^*d^{(i)}) \neq f^*(a^{(i-1)})$.

Formula (V.49) holds even if h^* is negative.

As noted earlier, if $f^*(a^{(i-1)} + h^*d^{(i)}) > f^*(a^{(i-1)} + h^{**}d^{(i)})$, h^* is used as the step size rather than h^{**}.

In problems calculated to date by the writer, the use of the quadratic approximation has not been a powerful procedure in the selection of a step size. For some iterations it has given a step with a higher likelihood value and for some iterations it has not. In fitting the quadratic, the $f^*(a)$ which we are using is only a montonic function of the likelihood function. It is very possible that better results would be obtained if some other monotonic function of the

---

[1] Koopmans, Rubin, and Leipnik [1950], p. 172 attribute the use of a quadratic approximation in the calculation of FIML problems to a suggestion by John Von Neumann.

likelihood function were used in deriving trial step sizes. Alternatively instead of the quadratic approximation, some other approximating function could be used.

In any event, until a better approximation is devised, the actual quadratic approximation calculation is trivial and the number of times it does lead to slightly better step sizes would appear to justify the additional time required to calculate the likelihood value at the new point, $a^{(i-1)} + h**d^{(i)}$.

"Local" methods could be derived for calculating an optimal step size based on, say, the eigenvalues of the metric, the ratios of elements of $d^{(i)}$, etc.; however, these methods require assumptions regarding the shape of the likelihood function. (An assumption often made in deriving a local method is that the likelihood function is approximately quadratic in the given direction, $d^{(i)}$, an assumption which does not appear justified.) As noted earlier, such local methods may be helpful for establishing an initial trial step, h_1, for an iteration, but they do not seem desirable as final determinants of the step size to be used for an iteration.

On the other hand, the suggested step size procedure outlined earlier is a "global" method in that no assumptions are required regarding the shape of the likelihood, except that it has only a single peak in the region under consideration and that there is no higher peak.

5. Convergence criteria

Following are some requirements which could be imposed in determining convergence:

(1) All eigenvalues of $\mathcal{L}^{(i-1)}$ must be positive for the iteration.

(2) Partial derivative convergence criterion:

$$\max_j(\text{abs } l^{(i)}_{[j]}) \leq \epsilon_{\ell} \quad , \tag{V.50}$$

i.e., the absolute value of all elements of the right hand side vector must be less than or equal to a preassigned constant.

(3) Coefficient convergence criterion:

$$\max_j \text{abs}(d^{(i)}_{[j]}/a_{[j]}) \leq \epsilon_a \tag{V.51}$$

where $d^{(i)}_{[j]}$ is the j^{th} element of the direction vector, $a_{[j]}$ is the j^{th} element of a and ϵ_a is a preassigned constant.

If a step size of 1 were imposed, $a^{(i)} - a^{(i-1)} = (a^{(i-1)} + d^{(i)}) - a^{(i-1)} = d^{(i)}$; therefore, (V.51) is equivalent to requiring that with a step size of 1, the absolute proportional change would be less than or equal to ϵ_a for all coefficients.

(4) Likelihood convergence criterion:

$$f^*(a^{(i-1)} + d^{(i)})/f^*(a^{(i-1)}) \leq \epsilon_{f^*} \quad , \tag{V.52}$$

i.e., if a step size of 1 were imposed, the ratio of the resulting likelihood to the likelihood for the preceding iteration is less than or equal to a preassigned constant.

If a user desired to iterate until each iteration produced a statistically insignificant change in the coefficients, then a

stopping criterion based on the relatives sizes of $f^*(a^{(i)})$ and $f^*(a^{(i-1)})$ might be considered. It should be recognized, however, that for many problems, coefficients differing considerably from those which maximize the likelihood function may not differ statistically from those which maximize the likelihood function, let alone those of the following iteration; hence, the coefficients derived through use of the likelihood convergence criterion may differ considerably from the maximum likelihood coefficients. Since it is the coefficients which maximize the likelihood function which are desired--not coefficients on a fairly flat surface away from the maximum--the likelihood convergence criterion does not appear to be a very fruitful criterion for convergence.

Up to now, in the FIML section of the AES STAT system (which is discussed in chapter IX), (1) and either "(2) and (3)" or (4) have been imposed for all problems. If none of the preassigned constants were supplied by the user, the problem iterated to the maximum number of iterations specified by the user. If ϵ_l but not ϵ_a was specified, ϵ_a was set to ϵ_l. If ϵ_a but not ϵ_l was specified, ϵ_l was set to $100 \cdot \epsilon_a$. An ϵ_l of .0000000001 and an ϵ_a of .000000000001 have worked well for a number of problems computed; however, both ϵ_l and ϵ_a are quite arbitrary. Close to the maximum, convergence usually becomes very rapid; so that a small ϵ_a or ϵ_l takes little additional computer time. On the other hand if rounding error becomes severe an extremely small ϵ_a or ϵ_l may be difficult to attain. The FIML section is being changed so that if the user specifies neither ϵ_l nor

ϵ_a, ϵ_a and ϵ_l will be automatically set to .0000000001 by the FIML section; hence, even if no convergence criterion is specified, iteration will no longer be merely to the maximum number of iterations specified.

D. Estimated Disturbance Variance-covariance Matrix

The maximum likelihood estimate of the disturbance variance-covariance matrix used in FIML estimation is the S matrix with the $\mu\mu'^{th}$ element of S calculated as:

$$(V.53) \qquad s_{\mu\mu'} = (1/T)\hat{u}_{\mu}'\hat{u}_{\mu'} = (1/T)\,{}_{+}a_{\mu}'({}_{+}Z_{\mu}'\,{}_{+}Z_{\mu'})\,{}_{+}a_{\mu'}$$

where ${}_{+}a_{\mu} = \begin{bmatrix} -1 \\ a_{\mu} \end{bmatrix}$ and a_{μ} is the part of $a^{\max} f(a)$ corresponding to the unrestricted coefficients of equation μ.[1]

In the single equation procedures previously discussed, the S matrix contained only a single element and so a "degrees of freedom" of $T-n_{\mu}$ was suggested as a possible denominator in calculating $s_{\mu\mu}$ for the μ^{th} equation, where n_{μ} is the number of "explanatory" jointly dependent variables plus the number of predetermined variables. A similar adjustment can be made in the calculation of the S matrix for FIML if an estimator more compatible with single equation techniques is desired. A more compatible estimator would be to use (V.53), but substitute $\sqrt{T - n_{\mu}} \cdot \sqrt{T - n_{\mu'}}$ for T in the formula.[2] The denominator for the μ^{th} diagonal element will then be $T - n_{\mu}$ as in the single equation procedures and the denominators for the off-diagonal elements will be such that the S matrix will still be positive definite.

---

[1] See (V.18) and (V.21).

[2] $\sqrt{T - n_{\mu}} \cdot \sqrt{T - n_{\mu'}}$ was suggested to the writer by Professor Arnold Zellner as an alternative to the use of T in the calculation of ZA estimates. Professor Zellner neither endorsed nor disparaged this adjustment for ZA. He merely listed it as an alternative.

It seems desirable that the maximum likelihood estimate of S (using the denominator of T) be used during iteration until convergence is complete so that $\hat{\alpha}_{FIML}$ will indeed be the maximum likelihood estimate of the coefficients. $\sqrt{T - n_\mu}\sqrt{T - n_\mu}$, would only be used in the denominator when printing out the estimated disturbance variance-covariance matrix corresponding to the FIML coefficients. The adjustment could also be used when printing out the estimated disturbance variance-covariance matrix corresponding to the coefficients from which the iterations were started or when printing out the estimated disturbance variance-covariance matrix corresponding to intermediate-stage coefficients.

$s_{\mu\mu}$, calculated by using T and $s_{\mu\mu}$, calculated by using $\sqrt{T - n_\mu}\sqrt{T - n_\mu}$, are asymptotically equivalent.

Often we are interested in the relative sizes of the estimated covariances (i.e., relative to the corresponding variances). In this case, the estimated disturbance variance-covariance matrix normalized so that 1's appear on the diagonal is useful. This matrix may be defined as:

(V.54) $$S_N = D_S S D_S$$

where $$D_S = \begin{bmatrix} \frac{1}{\sqrt{s_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{s_{MM}}} \end{bmatrix}$$

i.e., S_N is calculated by dividing each row and column of S by the square root of the corresponding diagonal element of S. S_N has the advantage of being independent of the scale of the normalizing variable--1's on the diagonal provide a convenient normalization. S_N is the same whether T or $\sqrt{T - n_\mu}\cdot\sqrt{T - n_{\mu'}}$ is used in the denominator in the calculation of the elements of S. The element in row μ and column μ' of S_N is:

(V.55) $$s_{\mu\mu'}/\sqrt{s_{\mu\mu}s_{\mu'\mu'}} \quad ,$$

the estimated simple correlation between the disturbance in equation μ and the disturbance in equation μ'.

E. Estimated Coefficient Variance-covariance Matrix

The maximum likelihood estimate of the coefficient variance-covariance matrix for FIML is:[1]

$$\text{(V.56)} \qquad \hat{\text{Var}}(a) = \frac{1}{T}\left[- \frac{\partial^2 f(a)}{\partial a^2} \Big|_{a=a^{\max} f(a)} \right]^{-1}$$

$$= \left[-T\frac{\partial^2 f}{\partial a^2} \Big|_{a=a^{\max} f(a)} \right]^{-1} = \left[\mathcal{L}^{\max f(a)} \right]^{-1} .$$

The estimated variances of the individual coefficients are given by the diagonal elements. The square roots of these diagonal elements are often used as asymptotic standard errors.

The elements of $\hat{\text{Var}}(a)$ could be adjusted in the same manner as the estimated disturbance variance-covariance matrix to provide an estimate more compatible with the usual single equation estimates. Let $\mathcal{L}^{-1}_{\mu\mu}$, be the $n_\mu \times n_\mu$, block of $[\mathcal{L}^{\max f(a)}]^{-1}$ with rows corresponding to unrestricted coefficients of equation μ and columns corresponding to unrestricted coefficients of equation μ'. Then, if $\mathcal{L}^{-1}_{\mu\mu}$, were multiplied by $\frac{T}{\sqrt{T - n_\mu}\sqrt{T - n_\mu}}$, the resulting estimated coefficient variance-covariance matrix would still be positive definite and asymptotically the same as the one given in (V.56), but more compatible with the estimated variance-covariance matrix given for the single

---

[1] Chernoff and Divinsky [1953], p. 259. This is the "information matrix" of Kendall and Stuart [1961], pp. 28, 54-55.

equation methods.[1] This adjustment should not be used during iteration to the FIML solution--only during printing out of the estimated coefficient variance-covariance matrix after a_{FIML} has been calculated.

Often we are interested in the relative size of the estimated covariance. In this case the estimated coefficient variance-covariance matrix normalized so that 1's appear on the diagonal is useful. This matrix may be defined as:

$$\hat{Var}_N(a) = D_C \, \hat{Var}(a) \, D_C \tag{V.57}$$

where $$D_C = \begin{bmatrix} \frac{1}{\sqrt{\hat{Var}(a_{(1)})}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\hat{Var}(a_{(2)})}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & \frac{1}{\sqrt{\hat{Var}(a_{(n)})}} \end{bmatrix},$$

i.e., $\hat{Var}_N(a)$ is calculated by dividing each row and column of $\hat{Var}(a)$ by the square root of the corresponding diagonal element of $\hat{Var}(a)$. $\hat{Var}_N(a)$ has the advantage of being independent of the scale of the variables; 1's on the diagonal provide a convenient normalization. $\hat{Var}_N(a)$ is the same whether T or $\sqrt{T - n_\mu}\sqrt{T - n_{\mu'}}$ is used in the

---

[1] Computationally the adjustment may be accomplished by (1) multiplying each row by $\sqrt{T/(T - n_\mu)}$ where n_μ is the number of unrestricted coefficients in the equation from which the coefficient corresponding to the row relates, and (2) multiplying each column by $\sqrt{T/(T - n_{\mu'})}$ where $n_{\mu'}$ is the number of unrestricted coefficients in the equation from which the coefficient corresponding to the column relates.

denominator in the calculation of $\hat{Var}(a)$. The element in row i and column j of $\hat{Var}_N(a)$ is the estimated simple correlation between a_i and a_j.

F. Arbitrary Linear Restrictions Imposed on the Coefficients

The iterative FIML procedure given in the previous section is designed to maximize $f(a)$ with respect to a; that is, adjust the elements of a until $f(a)$ is the highest possible, where $f(a)$ is defined in (V.23). In this section we will consider the problem:

$$\max_{a} f(a) \tag{V.58}$$

subject to:

$$\underset{N_R \times n}{R} \quad \underset{n \times 1}{a} = \underset{N_R \times 1}{r} \tag{V.59}$$

The vector a is the same vector as before--it contains all of the non-zero non-normalization elements of the rows of $\hat{\alpha}$ which correspond to stochastic equations in the system. In this section, N_R additional arbitrary linear restrictions are imposed on the $n = \sum_{\mu=1}^{M} n_\mu = \sum_{\mu=1}^{M} (m_\mu + \ell_\mu)$ coefficients of a. The equations need not be identifiable in the absence of the additional restrictions; however, they must be identifiable after the restrictions are imposed.

1. Illustration of linear restrictions on coefficients--Klein's model I

The placing of linear restrictions on coefficients is very straightforward and hardly needs illustrating. This example will, however, be used to demonstrate the effect of using identity equations to solve out jointly dependent variables in the system; since this is a technique that is commonly referred to, but apparently not very well understood.[1] In particular, we will note how the use of identity equations to solve out jointly dependent variables leads to the imposition of restrictions on coefficients, and how these restrictions may be expressed by the R matrix and r vector mention above.

Equations (I.5a) through (I.5h) present Klein's model I as an 8 equation model (3 stochastic equations and 5 identity equations) containing 8 jointly dependent variables (C, P, W, I, W_1, E, Y and K).[2] Often, the model is written as containing 3 stochastic equations and 3 identity equations by implicity carrying restrictions on certain coefficients rather than listing the last two identity equation. (The last two identity equations may be considered as having been solved into the equations as compared to our previous formulation of the model.) Following is an expression of Klein's model I as a 6 equation model:

(V.60a) Consumption: $C = \alpha_0^{[1]} + \alpha_1^{[1]} P + \alpha_2^{[1]} (W_1 + W_2) + \alpha_3^{[1]} P_{-1} + u_1$

---

[1] This example should convince most readers of the desirability of explicitly using the identity equations in the computational procedure instead of using the identity equation to eliminate jointly dependent variables before commencing computation.

[2] The definitions of the variables in Klein's model I follow (I.5h) in section I.C.1.

(V.60b) Investment: $I = \alpha_0^{[2]} + \alpha_1^{[2]}P + \alpha_2^{[2]}P_{-1} + \alpha_3^{[2]}K_{-1} + u_2$

(V.60c) Private Wage: $W_1 = \alpha_0^{[3]} + \alpha_1^{[3]}(Y + R - W_2) + \alpha_2^{[3]}(Y + R - W_2)_{-1} + \alpha_3^{[3]}t + u_3$

(V.60d) Product: $Y + R = C + I + G$

(V.60e) Income: $Y = P + W_1 + W_2$

(V.60f) Capital: $K = K_{-1} + I$

As the model is written above, it cannot be calculated by the FIML method, since it still contains 8 jointly dependent variables (C, P, $W_1 + W_2$, I, W_1, $Y + R - W_2$, Y, and K) but only 6 equations. It may be computed by FIML if we rewrite the first and third equations as:

(V.61a) $C = \alpha_0^{[1]} + \alpha_1^{[1]}P + \alpha_2^{[1]}W_1 + \alpha_4^{[1]}W_2 + \alpha_3^{[1]}P_{-1} + u_1$

(V.61b) $W_1 = \alpha_0^{[3]} + \alpha_1^{[3]}Y + \alpha_4^{[3]}(R - W_2) + \alpha_2^{[3]}(Y + R - W_2)_{-1} + \alpha_3^{[3]}t + u_3$

and impose the restrictions:

$$\alpha_2^{[1]} = \alpha_4^{[1]}$$
$$\alpha_1^{[3]} = \alpha_4^{[3]} \quad .$$

Thus, in eliminating the two identity equations, we have solved out two of the original eight jointly dependent variables, namely W (= $W_1 + W_2$) and E (= $Y + R - W_2$), and imposed restrictions on certain coefficients. At this point, we may write our problem as one of:

$$\max_a f(a)$$

subject to:

$$Ra = r$$

where:

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \qquad a_1 = \begin{bmatrix} \hat{\alpha}_1^{[1]} \\ \hat{\alpha}_2^{[1]} \\ \hat{\alpha}_0^{[1]} \\ \hat{\alpha}_4^{[1]} \\ \hat{\alpha}_3^{[1]} \end{bmatrix} \qquad a_2 = \begin{bmatrix} \hat{\alpha}_1^{[2]} \\ \hat{\alpha}_0^{[2]} \\ \hat{\alpha}_2^{[2]} \\ \hat{\alpha}_3^{[2]} \end{bmatrix} \qquad a_3 = \begin{bmatrix} \hat{\alpha}_1^{[3]} \\ \hat{\alpha}_0^{[3]} \\ \hat{\alpha}_4^{[3]} \\ \hat{\alpha}_2^{[3]} \\ \hat{\alpha}_3^{[3]} \end{bmatrix}$$

$$R = \begin{bmatrix} \alpha_1^{[1]} & \alpha_2^{[1]} & \alpha_0^{[1]} & \alpha_4^{[1]} & \alpha_3^{[1]} & \alpha_1^{[2]} & \alpha_0^{[2]} & \alpha_2^{[2]} & \alpha_3^{[2]} & \alpha_1^{[3]} & \alpha_0^{[3]} & \alpha_4^{[3]} & \alpha_2^{[3]} & \alpha_3^{[3]} \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \end{bmatrix}$$

$$r = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Now let us use the three remaining identity equations to solve out the jointly dependent variables Y, I, and C.[1] From (V.60a) we have:

(V.62e) $$Y = P + W_1 + W_2 \quad .$$

Rewriting (V.60f) we obtain:

(V.62f) $$I = K - K_{-1} \quad .$$

Substituting (V.62e) and (V.62f) into (V.60d) and rewriting it in terms of C we obtain:

(V.62d) $$C = P + W_1 + W_2 + R - K + K_{-1} - G \quad .$$

Substituting C, Y, and I as expressed by (V.62d), (V.62e), and (V.62f) we get the following expression of the model:

---

[1]These are the same variables solved out by Chernoff and Divinsky [1953].

$$(V.63a)\quad P + W_1 + W_2 + R - K + K_{-1} - G = \alpha_0^{[1]} + \alpha_1^{[1]}P + \alpha_2^{[1]}W_1 + \alpha_4^{[1]}W_2 + \alpha_3^{[1]}P_{-1} + u_1$$

$$(V.63b)\quad K - K_{-1} = \alpha_0^{[2]} + \alpha_1^{[2]}P + \alpha_2^{[2]}P_{-1} + \alpha_3^{[2]}K_{-1} + u_2$$

$$(V.63c)\quad W_1 = \alpha_0^{[3]} + \alpha_1^{[3]}(P + W_1 + W_2) + \alpha_4^{[3]}(R - W_2) + \alpha_2^{[3]}(Y + R - W_2)_{-1} + \alpha_3^{[3]}t + u_3$$

subject to:

$$\alpha_2^{[1]} = \alpha_4^{[1]}$$
$$\alpha_1^{[3]} = \alpha_4^{[3]} \quad .$$

Again, the model as expressed cannot be computed by FIML, since it has 5 jointly dependent variables ($P + W_1 + W_2 + R - K + K_{-1} - G$, P, W_1, $K - K_{-1}$, and $P + W_1 + W_2$) and only three equations. It may be re-written in the following manner to make it amenable for computation by FIML:

$$(V.64a)\quad K = -\alpha_0^{[1]} + (1 - \alpha_1^{[1]})P + (1 - \alpha_2^{[1]})W_1 + (1 - \alpha_4^{[1]})W_2 - \alpha_3^{[1]}P_{-1} + \alpha_5^{[1]}(G - R - K_{-1}) - u_1$$

$$(V.64b)\quad K = \alpha_0^{[2]} + \alpha_1^{[2]}P + \alpha_2^{[2]}P_{-1} + (\alpha_3^{[2]} + 1)K_{-1} + u_2$$

$$(V.64c)\quad W_1 = \alpha_0^{[3]} + \alpha_1^{[3]}P + \alpha_5^{[3]}W_1 + \alpha_4^{[3]}R + \alpha_2^{[3]}(Y + R - W_2)_{-1} + \alpha_3^{[3]} + u_3$$

subject to:

$$(1 - \alpha_2^{[1]}) = (1 - \alpha_4^{[1]})$$

$$\alpha_5^{[1]} = -1$$

$$\alpha_1^{[3]} = \alpha_5^{[3]} = \alpha_4^{[3]} \qquad \text{(2 restrictions)}$$

Thus, in terms of the **a** vector, R matrix and r vector, the model has become:

$$\mathbf{a} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix} \quad \text{where } \mathbf{a}_1 = \begin{bmatrix} \hat{\alpha}_1^{[1]*} \\ \hat{\alpha}_2^{[1]*} \\ \hat{\alpha}_0^{[1]*} \\ \hat{\alpha}_4^{[1]*} \\ \hat{\alpha}_3^{[1]*} \\ \hat{\alpha}_5^{[1]} \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} \hat{\alpha}_1^{[2]} \\ \hat{\alpha}_0^{[2]} \\ \hat{\alpha}_2^{[2]} \\ \hat{\alpha}_3^{[2]*} \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} \hat{\alpha}_1^{[3]} \\ \hat{\alpha}_5^{[3]} \\ \hat{\alpha}_0^{[3]} \\ \hat{\alpha}_4^{[3]} \\ \hat{\alpha}_2^{[3]} \\ \hat{\alpha}_3^{[3]} \end{bmatrix},$$

$$\hat{\alpha}_1^{[1]*} = 1 - \hat{\alpha}_1^{[1]}, \; \hat{\alpha}_2^{[1]*} = 1 - \hat{\alpha}_2^{[1]}, \; \hat{\alpha}_0^{[1]*} = -\hat{\alpha}_0^{[1]}, \; \hat{\alpha}_4^{[1]*} = 1 - \hat{\alpha}_4^{[1]},$$

$$\hat{\alpha}_3^{[1]*} = -\hat{\alpha}_3^{[1]}, \; \hat{\alpha}_3^{[2]*} = \hat{\alpha}_3^{[2]} + 1 ,$$

$$\mathbf{R} = \begin{array}{c} \begin{matrix} \alpha_1^{[1]*} & \alpha_2^{[1]*} & \alpha_0^{[1]*} & \alpha_4^{[1]*} & \alpha_3^{[1]*} & \alpha_5^{[1]} & \alpha_1^{[2]} & \alpha_0^{[2]} & \alpha_2^{[2]} & \alpha_3^{[2]*} & \alpha_1^{[3]} & \alpha_5^{[3]} \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

$$\begin{array}{c} \begin{matrix} \alpha_0^{[3]} & \alpha_4^{[3]} & \alpha_2^{[3]} & \alpha_3^{[3]} \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \end{array}, \quad \mathbf{r} = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} .$$

After converging to the FIML values for a, the desired coefficients $\hat{\alpha}_1^{[1]}$, etc. which have not been calculated directly are calculated as:

$$\hat{\alpha}_1^{[1]} = 1 - \hat{\alpha}_1^{[1]*}, \quad \hat{\alpha}_2^{[1]} = 1 - \hat{\alpha}_2^{[1]*}, \text{ etc.}^1$$

In our last formulation, we used an explicit normalization for each equation. For the first equation, it would probably have been slightly more convenient not to save out a normalizing coefficient, but instead use the R matrix and r vector to impose a normalization on the coefficients. Rewritten in this manner, the first equation becomes

$$\text{(V.65a)} \quad \alpha_0^{[1]} + (\alpha_1^{[1]} - 1)P + (\alpha_2^{[1]} - 1)W_1 + \alpha_6^{[1]}K + (\alpha_4^{[1]} - 1)W_2 + \alpha_3^{[1]}P_{-1} + \alpha_5^{[1]}(G - R - K_{-1}) + u_1 = 0$$

and the second and third equations are unchanged. The restrictions are:

---

[1] Since $\hat{\alpha}_1^{[1]}$ is a linear combination of only a constant with only a single coefficient $\hat{\alpha}_1^{[1]*}$, its variance and standard error are the same as the variance and standard error of $\hat{\alpha}_1^{[1]}$. Thus, asymptotic $\text{Var}(\hat{\alpha}_1^{[1]})$ = asymptotic $\text{Var}(\hat{\alpha}_1^{[1]*})$, asymptotic $\text{Var}(\hat{\alpha}_2^{[1]})$ = asymptotic $\text{Var}(\hat{\alpha}_2^{[1]*})$, etc. The ratios of coefficients to asymptotic standard errors (often called asymptotic t-ratios) will, of course, have to be recalculated, e.g.,

$$\frac{\hat{\alpha}_1^{[1]}}{\text{asymptotic } S_{\hat{\alpha}_1^{[1]}}} = \frac{\hat{\alpha}_1^{[1]}}{\text{asymptotic } S_{\hat{\alpha}_1^{[1]*}}} \neq \frac{\hat{\alpha}_1^{[1]*}}{\text{asymptotic } S_{\hat{\alpha}_1^{[1]}}}$$

where $S_{\hat{\alpha}_1^{[1]}}$ denotes the standard error of $\hat{\alpha}_1^{[1]}$.

$$(\alpha_1^{[1]} - 1) = (\alpha_2^{[1]} - 1)\ ,$$

$$\alpha_6^{[1]} = \alpha_5^{[1]} = 1 \qquad \text{(2 restrictions)},$$

$$\alpha_1^{[3]} = \alpha_5^{[3]} = \alpha_4^{[3]} \qquad \text{(2 restrictions)},$$

$$a_1 = \begin{bmatrix} \hat{\alpha}_1^{[1]**} \\ \hat{\alpha}_2^{[1]**} \\ \hat{\alpha}_6^{[1]} \\ \hat{\alpha}_0^{[1]} \\ \hat{\alpha}_4^{[1]**} \\ \hat{\alpha}_3^{[1]} \\ \hat{\alpha}_5^{[1]} \end{bmatrix} .$$

The R matrix and r vector are:

$$R = \begin{array}{c} \begin{matrix} \alpha_1^{[1]**} & \alpha_2^{[1]**} & \alpha_6^{[1]} & \alpha_0^{[1]} & \alpha_4^{[1]**} & \alpha_3^{[1]} & \alpha_5^{[1]} & \alpha_1^{[2]} & \alpha_0^{[2]} & \alpha_2^{[2]} & \alpha_3^{[2]*} \end{matrix} \\ \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$\begin{array}{c} \begin{matrix} \alpha_1^{[3]} & \alpha_5^{[3]} & \alpha_0^{[3]} & \alpha_4^{[3]} & \alpha_2^{[3]} & \alpha_3^{[3]} \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \end{bmatrix} \end{array} , \qquad r = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Summary of Klein Model I Illustration

The number of coefficients to be estimated in each of the formulations of Klein's model I may be summarized as follows:

Model Expressed by Equations	Number of Identity Equations	Number of Coefficients	Number of Restrictions	Number of "Unrestricted" Coefficients
(I.5a)-(I.5h)	5	12	0	12
(V.61a), (V.60b), (V.61c), (V.60d), (V.60e), (V.60f)	3	14	2	12
(V.64a)-(V.64c)	0	16	4	12
(V.65a)-(V.64b), (V.64c)	0	17	5	12

Thus, the dimensionality of the unrestricted coefficients space does not change as identity equations are used to solve out jointly dependent variables. Since the number of iterations required for convergence is related to the dimensionality of the coefficients space, all of the above expressions would require a comparable number of iterations. On the other hand, as we will note by the formulas of the next section, the formal listing of the identity equations rather than using them to solve out jointly dependent variables involves less cumbersome calculations each iteration thereby taking less total computer time. Explicit listing of identity equations also has the advantages: (1) it provides a more convenient way to formulate the problem (at least in Klein's model I) than solving out the identity equations, (2) all of the coefficients are derived directly, (3) estimated variances, and t-ratios are calculated directly, and (4) calculation of reduced form coefficients is simpler due to the direct calculation of the coefficients.

Restrictions on coefficients may arise from contexts other than using identity equations to eliminate jointly dependent variables. In these cases it will likely be more convenient to impose directly the restrictions onto the coefficients in the form $Ra = r$ than to attempt to create additional jointly dependent variables and identity equations to impose the restrictions.

The above illustration dealt with restrictions imposed on the coefficients of a single equation at a time; however, the computational formulas which follow will be equally applicable to restrictions which cut across equations.

2. Computational formulas

In chapter IV we noted a procedure for transforming an R matrix and r vector into a Q matrix and q vector with certain special properties relative to the R matrix and r vector.[1] The procedure gave a method of separating the vector of coefficients, a, into:

$$(V.66) \qquad a^*_{n\times 1} = \begin{bmatrix} a_{(1)} \\ (n-rk\ R)\times 1 \\ a_{(2)} \\ rk\ R\times 1 \end{bmatrix}$$

where the $n - rk\ R$ coefficients in the $a_{(1)}$ vector are estimated directly and the $rk\ R$ coefficients in the $a_{(2)}$ vector are calculated from the $a_{(1)}$ vector. a^* is the same as a except that the coefficients may have been rearranged.

The FIML computational formulas which take account of the additional restrictions are:[2]

$$(V.67) \qquad \underset{(n-rk\ R)\times 1}{a^{(i)}_{(1)}} = \underset{(n-rk\ R)\times 1}{a^{(i-1)}_{(1)}} + \underset{(1\times 1)}{h^{(i)}} \left| \underset{(n-rk\ R)\times(n-rk\ R)}{Q'\mathscr{L}^{(i-1)}Q} \right|^{-1} \underset{(n-rk\ R)\times 1}{Q'_1{}^{(i-1)}}$$

$$(V.68) \qquad \underset{rk\ R\times 1}{a^{(i)}_{(2)}} = \underset{rk\ R\times(n-rk\ R)}{Q_2} \underset{(n-rk\ R)\times 1}{a^{(i)}_{(1)}} + \underset{rk\ R\times 1}{q_2}$$

where:

$\mathscr{L}^{(i-1)}$ is T times the negative of the matrix of second partial derivatives of $f(a)$ with respect to a as before. (The additional restrictions are ignored in defining $\mathscr{L}^{(i-1)}$.)

---

[1] E.g., the Q matrix is the orthogonal complement of the R matrix.

[2] Derivations are given in the following sub-section.

$\iota^{(i-1)}$ is T times the vector of first partial derivatives of f(a) with respect to a as before. (The additional restrictions are ignored in defining ι.)

Q, Q_2, q, and q_2 are calculated from R and r as in section IV.B.1.

$a^{(i)}$ and $a^{(i-1)}$ denote the values of the coefficients for iteration i and (i-1), respectively.

The $|\ |$ lines around $|Q'\mathcal{L}^{(i-1)}Q|$ denote the positive definite matrix constructed from $(Q'\mathcal{L}^{(i-1)}Q)$ in the same manner that $|\mathcal{L}^{(i-1)}|$ is constructed from $\mathcal{L}^{(i-1)}$.

(V.67) and (V.68) may be combined into the single formula:

$$(V.69)\qquad \underset{n\times 1}{a^{(i)}} = \underset{n\times(n-rk\ R)}{Q}\left[\underset{(n-rk\ R)\times 1}{a^{(i-1)}_{(1)}} + h^{(i)}\,|Q'\mathcal{L}^{(i-1)}Q|^{-1}Q'\iota^{(i-1)}\right] + \underset{n\times 1}{q}$$

When convergence has been achieved, the estimated coefficient variance covariance matrix is given by:

$$(V.70)\qquad \text{asymptotic } \hat{Var}(a_{FIML}) = Q[Q'\mathcal{L}^{\max f(a)}_{n\times n}Q]^{-1}Q'$$

where $\max f(a)$ is the restricted maximum satisfying $Ra = r$.

The S matrix is calculated by formula (V.21) except that the restricted coefficients (V.69) obtained from the preceding iteration are used instead of the corresponding unrestricted coefficients in calculating the S matrix.

Any restriction may be imposed only on the coefficients of a single equation, or it may cut across equations. As in section IV.B, the restrictions need not be linearly independent, i.e., R need not have full row rank.

It is usually convenient and saving of computer time to list only the non-zero non-normalizing coefficients of the stochastic equations in the a vector; therefore, the above procedure is outlined with this in mind. It should be noted, however, that any coefficients of $\hat{\alpha}$ may be listed in the a vector and then restricted to -1, 0, or some other value by the R matrix and r vector. For some problems an implicit normalization of coefficients is more convenient than the explicit normalization which we have used--setting a coefficient to -1. Use of the R matrix and r vector allows for an implicit normalization--all that is required is that some normalization be imposed by the restrictions so that the coefficients of an equation cannot take on an infinite number of values due to lack of a normalization rule.[1]

Derivation of (V.66) through (V.70).

(V.68) is given by the method of calculating the Q matrix and q vector. (See section IV.B.1.) From (V.68), we derive that:

$$\text{(V.71)} \qquad \begin{bmatrix} a_{(1)}^{(i)} \\ a_{(2)}^{(i)} \end{bmatrix} = \begin{bmatrix} I \\ Q_2 \end{bmatrix} a_{(1)}^{(i)} + \begin{bmatrix} 0 \\ q_2 \end{bmatrix}$$

or

$$\text{(V.72)} \qquad a^{(i)} = Q a_{(1)}^{(i)} + q \quad .$$

In the following paragraphs, we omit the superscripts $^{(i)}$ for simplicity. Since the coefficients of a are a function of the

---

[1] An implicit normalization is used for the first equation of Klein's model I in the last version of the "restrictions" example of section V.F.1.

coefficients of $a_{(1)}$ ($a_{(1)}$ is a sub-vector of a), we can maximize $f(a)$ directly with respect to $a_{(1)}$. Let $a_{[i]}$ denote the i^{th} element of a, and $a_{(1)_{[j]}}$ denote the j^{th} element of $a_{(1)}$. Then from (V.72) we get:

$$\frac{\partial a_{[i]}}{\partial a_{(1)_{[j]}}} = Q_{ij} \tag{V.73}$$

where Q_{ij} is the ij^{th} element of Q.

Now consider the vector T times the partial derivatives, $T\frac{\partial f(a)}{\partial a_{(1)}}$. The j^{th} element is

$$T\frac{\partial f(a)}{\partial a_{(1)_{[j]}}} = T\sum_{t=1}^{n} \frac{\partial f(a)}{\partial a_i}\frac{\partial a_i}{\partial a_{(1)_{[j]}}} = \sum_{t=1}^{n} l_{[i]}Q_{ij}, \quad (j=1,\ldots,n\text{-rk } R),$$

where $l_{[i]}$ is the i^{th} element of l. Hence,

$$T\frac{\partial f(a)}{\partial a_{(1)}} = Q' l \ . \tag{V.74}$$

Next consider the matrix of $(-T)$ times the second-order partial derivatives, $-T\frac{\partial^2 f(a)}{\partial a_{(1)}^2}$. The element in row j, column k is (minus) the derivative of the j^{th} element of (V.73) with respect to $a_{(1)_{[k]}}$, i.e.,

$$-T\sum_{m=1}^{n}\sum_{i=1}^{n}\frac{\partial^2 f(a)}{\partial a_m \partial a_i}\frac{\partial a_i}{\partial a_{(1)_{[j]}}}\frac{\partial a_m}{\partial a_{(1)_{[k]}}} = \sum_{m=1}^{n}\sum_{i=1}^{n} \mathcal{L}_{im}Q_{ij}Q_{mk}$$

$$= \sum_{m=1}^{n}(Q_j'\mathcal{L}_m)Q_{mk} = Q_j'\sum_{m=1}^{n}\mathcal{L}_m Q_{mk} = Q_j'\mathcal{L}Q_k$$

where $\mathcal{L}_m$ is column m of $\mathcal{L}$ and Q_j, Q_k are columns j, k of Q. Thus,

$$-T\frac{\partial^2 f(a)}{\partial a_{(1)}^2} = Q'\mathcal{L}Q. \tag{V.75}$$

Taking account of the new matrix of second derivatives and the new matrix of partials we can now iterate to a maximum basing our iteration on the $n - rk\ R$ coefficients which we have separated out. Thus, (V.46) becomes:

$$(V.76)\quad a_{(1)}^{(i)} = a_{(1)}^{(i-1)} - h^{(i)} \left| -T\frac{\partial^2 f(a)}{\partial a_{(1)}^2}\right|_{a_{(1)}=a_{(1)}^{(i-1)}}^{-1} T\left[\frac{\partial f(a)}{\partial a_{(1)}}\right]_{a_{(1)}=a_{(1)}^{(i-1)}}$$

Substituting (V.74) and (V.75) into (V.76), we get (V.67).

(V.70) follows from (V.72) by a common variance relationship--if $y = Ax + b$ with A a matrix of fixed elements and b a vector of fixed elements, then $Var(y) = A[Var(x)]A'$.

G. Linearized Maximum Likelihood (LML)

The linearized maximum likelihood (LML) method is a complete system estimation method recently proposed by Rothenberg and Leenders.[1] LML estimates and asymptotic coefficient variances and covariances may be calculated as:

$$a_{LML} = a^{(1)} + [\mathcal{L}^{(1)}]^{-1} \cdot (1) \tag{V.77}$$

$$\text{asymptotic } \hat{\text{Var}}(a_{LML}) = [\mathcal{L}^{(1)}]^{-1} \tag{V.78}$$

where $\mathcal{L}^{(1)}$ and $\cdot^{(1)}$ are calculated from the starting estimates $a^{(1)}$.[2]

That is, the LML estimates are the estimates obtained at the end of the first iteration of the FIML iteration procedure, provided the following restrictions are imposed on the iteration:

(1) $\mathcal{L}^{(1)}$ is not converted to a positive definite matrix even though it is not positive definite. (Usually $\mathcal{L}$ will not be positive definite the first iteration.)

(2) A step size of 1 is automatically imposed.

The restriction formulas given for FIML may also be applied to LML as well without changing the basic LML properties.

Provided the starting estimates for LML ($a^{(1)}$) are consistent estimates of the underlying coefficients a with $a^{(1)} - a = 0(T^{-1/2})$ in probability and also provided that certain other regularity and existence conditions hold, then LML coefficients have the same asymptotic distribution as the corresponding FIML coefficients, i.e.,

---

[1] Rothenberg and Leenders [1964].

[2] A "degrees of freedom" adjustment may be made in asymptotic $\hat{\text{Var}}(a_{LML})$ in the same manner as for FIML.

$\mathcal{L}^{(1)}$ is asymptotically the same as $\mathcal{L}^{\max f(a)}$. This holds under a wide range of conditions including restrictions imposed on the disturbance variance-covariance matrix. Coefficients from 2SLS, LIML, or other methods meeting the consistency and probability requirements may be used as starting coefficients for the calculation of a_{LML}; however, the LML coefficients obtained if $a^{(1)} = a_{2SLS}$ will not, of course, equal the LML coefficients obtained if $a^{(1)} = a_{LIML}$. DLS estimates do not meet the consistency requirements; therefore, they cannot be used as starting estimates if the properties of LML coefficients are to be preserved.[1]

The LML coefficients may have a lower likelihood value than the starting estimates. This is because:

(1) $\mathcal{L}^{(1)}$ will in general not be positive definite; hence, our assurance that for a step size, h, sufficiently small, the calculation of $a^{(i)} = a^{(i-1)} + h^{(i)}(\mathcal{L}^{(i-1)})^{-1}{}_1{}^{(i-1)}$ leads to a higher likelihood value will not in general hold.

(2) Even if $\mathcal{L}^{(1)}$ is positive definite for a particular problem, the imposition of a step size of 1 may result in too large a step in the particular direction given by $(\mathcal{L}^{(1)})^{-1}{}_1{}^{(1)}$ with the result that a movement is made to a lower likelihood value.

Diagonal elements of $[\mathcal{L}^{(1)}]^{-1}$ may be negative, giving estimates of asymptotic coefficient variances which are negative.

---

[1] DLS estimates may be used as starting estimates for FIML, since iteration proceeds to the maximum of the likelihood, anyway. The FIML coefficients obtained will, of course, coincide with the FIML coefficients obtained from starting from 2SLS or LIML coefficients (assuming that multiple maxima do not occur in the regions of the coefficients).

CHAPTER VI

LIMITED INFORMATION SUBSYSTEM MAXIMUM LIKELIHOOD (SML)

A. Only Zero and Normalization Restrictions Imposed on Coefficients

Limited information subsystem maximum likelihood (SML) is a partial system maximum likelihood method which permits the simultaneous estimation of a subset of the equations in a system of equations.[1] In FIML estimation we distinguished between two kinds of equations--the

---

[1]The name limited information subsystem maximum likelihood comes from taking account of the structure of only a subsystem of the whole system in the estimation procedure. A basic reference for the SML method is Koopmans and Hood [1953]. Chernoff and Divinsky [1953] refer to the method as the LIS (limited information subsystem) method and give computational formulas. The computation of a particular SML problem is also given by Chernoff and Divinsky. We will use the abbreviation SML rather than LIS to emphasize the maximum likelihood character of SML which distinguishes SML from other "limited information subsystem" methods such as 3SLS (three-stage least squares) and I3SLS (iterative 3SLS).

Hannan [1967] recently derived an SML method (seemingly unaware of the prior existence of the more general SML method given by Koopmans and Hood [1953] which is discussed in this chapter) and showed the relationship between it and canonical correlation. As Chow and Ray-Chandhuri [1967] make clear, Hannan's method is only applicable as a very special case of the general SML procedure given by Koopmans and Hood. (The special case may be stated in our notation (pp. 226 *et seq.*), as follows. Let

$$\alpha_{I1} = (\Gamma_{I1} \quad B_{I1})$$

where α_{I1} is $M_1 \times (G + \Lambda)$, Γ_{I1} is $M_1 \times G$, and B_{I1} is $M_1 \times \Lambda$ (M_1 being the number of equations in the subsystem). Now imagine rearranging variables so that all G_1 dependent variables in the subsystem occur first among the dependent variables in the whole system, and all Λ_1 predetermined variables in the subsystem occur first among the predetermined variables in the whole system, so we can write

$$(\Gamma_{I1} \quad B_{I1}) = (\Gamma_{\Delta} \; 0 \; B_{*} \; 0)$$

where Γ_{Δ} is $M_1 \times G_1$ and B_{*} is $M_1 \times \Lambda_1$. Now consider the matrix

$$\alpha_{\Delta *} = (\Gamma_{\Delta} \quad B_{*})$$

which is $M_1 \times (G_1 + \Lambda_1)$. The special case that Hannan treats is that in which, *a priori*, there are exactly $M_1 - 1$ zeroes in each row of $\alpha_{\Delta *}$.)

M stochastic equations and the G - M identity equations. Thus, in FIML estimation we subdivided the system of equations into [see (V.3), (V.4), and (V.5)]:

$$\underset{M\times(G+\Lambda)}{\alpha_{I}}\ \underset{(G+\Lambda)\times T}{Z'} + \underset{M\times T}{U'} = \underset{M\times T}{0'} \tag{VI.1}$$

$$\underset{(G-M)\times(G+\Lambda)}{\alpha_{II}}\ \underset{(G+\Lambda)\times T}{Z'} \quad \underset{(G-M)\times T}{0'} \tag{VI.2}$$

In SML estimation we make a further division of the M stochastic equations into (1) the M_1 stochastic equations for which we specify the structure and (2) the remaining M_2 equations for which we estimate no coefficients and specify only any predetermined variables occurring in the M_2 equations which do not already occur in the M_1 equations. Thus, in deriving the SML computational formulas we will find it useful to subdivide the stochastic equations into two subsystems. The complete system may be written as:

$$\underset{M_1\times(G+\Lambda)}{\alpha_{I1}}\ \underset{(G+\Lambda)\times T}{Z'} + \underset{M_1\times T}{U_1'} = \underset{M_1\times T}{0'} \tag{VI.3}$$

$$\underset{M_2\times(G+\Lambda)}{\alpha_{I2}}\ \underset{(G+\Lambda)\times T}{Z'} + \underset{M_2\times T}{U_2'} = \underset{M_2\times T}{0'} \tag{VI.4}$$

$$\underset{(G-M)\times(G+\Lambda)}{\alpha_{II}}\ \underset{(G+\Lambda)\times T}{Z'} = \underset{(G-M)\times T}{0'} \tag{VI.5}$$

In SML estimation, only the structure of the first M_1 stochastic equations given by VI.3 (subsystem I1) is specified. Also, the predetermined variables in the entire system (including those which are in subsystem I2--the remaining M_2 stochastic equations--and subsystem II--the G - M identity equations) are specified (insofar as the researcher is able to specify additional predetermined variables in

subsystems I2 and II). In the derivation of the concentrated likelihood function which follows, we will see that the same likelihood function is obtained whether some or all of the identity equations are used to solve out jointly dependent variables in subsystem I1 or I2 or whether the identity equations are merely ignored (except that additional predetermined variables which occur in the identity equations are included in the set of predetermined variables recognized as being in the system in applying the computational procedure).

Let G_1 be the number of jointly dependent variables occurring in subsystem I1. Then (as will be shown further on), if G_1 equals M_1, and the rank of the matrix of predetermined variables in the entire system is less than $T - M_1 + 1$, the SML coefficients will coincide with the FIML coefficients obtained by applying the FIML computational procedure to the M_1 equations, only.[1] This holds whether M_1 is the entire system or whether M_1 is a subsystem of the entire system with additional predetermined variables in subsystems I2 and II occurring only with zero coefficients in the equations of subsystem I1. On the other hand, if subsystem I1 consists of only a single stochastic equation (i.e., the structure of only a single equation is specified) the resulting SML coefficients for the equation coincide with the usual

---

[1] rk $X < T - M_1 + 1$ is necessary for computation of SML coefficients as is shown in section VI.D; however, this condition is not a requirement for FIML computation, since the X matrix is not used in the adjustment of jointly dependent variables in FIML estimation.

LIML coefficients for the equation.[1]

Thus, (provided rk $X < T - M_1 + 1$) both FIML and LIML estimation may be considered to be particular cases of SML estimation; however, since the SML computational procedure is more cumbersome than the FIML computational procedure, it is more fruitful to calculate problems in which the number of jointly dependent variables in the system or subsystem to be estimated equals the number of equations by the FIML computational procedure rather than the SML computational procedure. Similarly if only a single stochastic equation is specified, it is more fruitful to use the much simpler LIML computational procedure than the SML computational procedure. The SML procedure may be applied, however, to the many cases in which the structure of more than one stochastic equation is specified but the entire system is not specified.

The SML method is sometimes referred to as the least generalized variance ratio (LGVR) method, since the coefficients obtained minimize the ratio of two generalized variances.[2] This is not as powerful a property as the LGV property of FIML, even though the properties do coincide if the number of jointly dependent variables equals the number of equations.

---

[1]Provided X_I (the matrix of instruments) in the LIML estimation is taken to be X^I (the matrix of predetermined variables for the system) or provided X_I is used in place of X in the SML calculations. A proof of the equivalence of LIML and SML in the case of only a single equation occurring in the subsystem being estimated is contained in Koopmans and Hood [1953], pp. 166-173.

[2]See Koopmans and Hood [1953], pp. 170-171.

1. Derivation of the likelihood function to be maximized[1]

The Likelihood To Be Maximized

Before indicating SML computational formulas, the function to be maximized by the SML procedure will be indicated. As a step in the derivation of the likelihood function, we will use the G - M identity equations to temporarily eliminate G - M jointly dependent variables from the system. (The eliminated variables will be reentered into the system at a later step.) Suppose that the jointly dependent variables are divided into two groups, $Y = [Y_1 \vdots Y_2]$, where Y_2 contains the G - M jointly dependent variables to be temporarily eliminated from the system and Y_1 contains the M jointly dependent variables which will remain.

To reflect our subdivisions we may rewrite (VI.3) through (VI.5) as:

$$\Gamma_{I1,1}Y_1' + \Gamma_{I1,2}Y_2' + B_{I1}X' + U_1' = 0' \qquad \text{(VI.6)}$$

$$\Gamma_{I2,1}Y_1' + \Gamma_{I2,2}Y_2' + B_{I2}X' + U_2' = 0' \qquad \text{(VI.7)}$$

---

[1]The derivation of the likelihood function to be maximized follows Koopmans and Hood [1953] and relies on that reference for some of the details of the derivation. The derivation given in this section differs primarily from the one given in Koopmans and Hood in that identity equations are explicitly treated in the derivation and it is shown that the same likelihood function is obtained as if the identity equations had been ignored, (except for using the predetermined variables from the identity equations in the computational method).

Professor Herman Rubin informed the writer that the same result is obtained whether some or all of the identity equations are used to eliminate jointly dependent variables from the stochastic equations whose structure is specified; however, the derivation of the likelihood function in a manner which shows that this is the case (given in this paper) was developed by the writer.

$$\Gamma_{21}Y_1' + \Gamma_{22}Y_2' + B_{II}X' = 0' \qquad \text{(VI.8)}$$

Up to this point the matrix of coefficients has been implicitly or explicitly subdivided in the following ways:

$$\underset{G\times(G+\Lambda)}{\alpha} = \underset{G\times G\ \ G\times\Lambda}{[\Gamma \vdots B]} = \begin{bmatrix} \underset{M_1\times(G+\Lambda)}{\alpha_{I1}} \\ \underset{M_2\times(G+\Lambda)}{\alpha_{I2}} \\ \underset{(G-M)\times(G+\Lambda)}{\alpha_{II}} \end{bmatrix} = \begin{bmatrix} \underset{M_1\times G}{\Gamma_{I1}} & \underset{M_1\times\Lambda}{B_{I1}} \\ \underset{M_2\times G}{\Gamma_{I2}} & \underset{M_2\times\Lambda}{B_{I2}} \\ \underset{(G-M)\times G}{\Gamma_{II}} & \underset{(G-M)\times\Lambda}{B_{II}} \end{bmatrix} \qquad \text{(VI.9)}$$

$$= \begin{bmatrix} \underset{M_1\times M}{\Gamma_{I1,1}} & \underset{M_1\times(G-M)}{\Gamma_{I1,2}} & \underset{M_1\times\Lambda}{B_{I1}} \\ \underset{M_2\times M}{\Gamma_{I2,1}} & \underset{M_2\times(G-M)}{\Gamma_{I2,2}} & \underset{M_2\times\Lambda}{B_{I2}} \\ \underset{(G-M)\times M}{\Gamma_{21}} & \underset{(G-M)\times(G-M)}{\Gamma_{22}} & \underset{(G-M)\times\Lambda}{B_{II}} \end{bmatrix}$$

We can use the G - M identity equations to temporarily eliminate the G - M jointly dependent variables in Y_2 in the same manner as for FIML [(V.7) through (V.11)] giving us:

$$\underset{M_1\times M}{\Gamma^*_{I1}}\ \underset{M\times T}{Y_1'} + \underset{M_1\times\Lambda}{B^*_{I1}}\ \underset{\Lambda\times T}{X'} + \underset{M_1\times T}{U_1'} = \underset{M_1\times T}{0'} \qquad \text{(VI.10)}$$

$$\underset{M_2\times M}{\Gamma^*_{I2}}\ \underset{M\times T}{Y_1'} + \underset{M_2\times\Lambda}{B^*_{I2}}\ \underset{\Lambda\times T}{X'} + \underset{M_2\times T}{U_2'} = \underset{M_2\times T}{0'} \qquad \text{(VI.11)}$$

or

$$\underset{M_1\times(M+\Lambda)}{\alpha^*_{I1}}\ \underset{(M+\Lambda)\times T}{Z_1'} + \underset{M_1\times T}{U_1'} = \underset{M_1\times T}{0'} \qquad \text{(VI.12)}$$

$$\underset{M_2 \times (M+\Lambda)}{\alpha^*_{I2}} \quad \underset{(M+\Lambda)\times T}{Z_1'} + \underset{M_2 \times T}{U_2'} = \underset{M_2 \times T}{0'} \tag{VI.13}$$

where

$$\Gamma^*_{I1} = \Gamma_{I1,1} - \Gamma_{I1,2}\Gamma_{22}^{-1}\Gamma_{21}$$

$$\Gamma^*_{I2} = \Gamma_{I2,1} - \Gamma_{I2,2}\Gamma_{22}^{-1}\Gamma_{21}$$

$$B^*_{I1} = B_{I1} - \Gamma_{I1,2}\Gamma_{22}^{-1}B_{II}$$

$$B^*_{I2} = B_{I2} - \Gamma_{I2,2}\Gamma_{22}^{-1}B_{II}$$

$$\alpha^* = [\Gamma^* \vdots B^*] = \begin{bmatrix} \alpha^*_{I1} \\ \\ \alpha^*_{I2} \end{bmatrix} = \begin{bmatrix} \Gamma^*_{I1} & B^*_{I1} \\ \\ \Gamma^*_{I2} & B^*_{I2} \end{bmatrix} .$$

In the derivation of the FIML likelihood function, we assumed that U has the multivariate normal density function given by (V.12) and derived the following intermediate likelihood function (V.13):

$$f_2(Z_1,\alpha^*,\Sigma) = \tag{VI.14}$$

$$-\frac{T}{2}\log 2\pi - \frac{T}{2}\log[\det \Sigma] + T \log[\text{abs}(\det \Gamma^*)] - \frac{1}{2}\sum_{t=1}^{T} Z_{1[t]}\alpha^{*\prime}\Sigma^{-1}\alpha^* Z'_{1[t]} .$$

Subdividing Σ to represent the subdivision of the stochastic equation into two groups we define:

$$\underset{M\times M}{\Sigma} = \begin{bmatrix} \underset{M_1 \times M_1}{\Sigma_{I1,I1}} & \underset{M_1 \times M_2}{\Sigma_{I1,I2}} \\ \\ \underset{M_2 \times M_1}{\Sigma_{I2,I1}} & \underset{M_2 \times M_2}{\Sigma_{I2,I2}} \end{bmatrix} \tag{VI.15}$$

where $\Sigma_{I1,I1}$ consists of the disturbance variance-covariance matrix

of the equations in subsystem I1 (whose structure has been specified), $\Sigma_{I1,I2}$ and $\Sigma_{I2,I1}$ $(= \Sigma'_{I1,I2})$ consist of disturbance covariances between the equations for subsystem I1 and subsystem I2, and $\Sigma_{I2,I2}$ consists of the disturbance variance-covariance matrix of the equations of subsystem I2.

Let us maximize $f_2(Z_1,\alpha^*,\Sigma)$ first with respect to α^*_{I2}, $\Sigma_{I1,I2}$, $\Sigma_{I2,I1}$, and $\Sigma_{I2,I2}$ taking no account of restrictions imposed on these matrices by the model--in particular taking no account of the structure of α^*_{I2}, e.g., ignoring the restrictions which would be imposed if account were taken that certain elements of α_{I2} are zero, other elements are -1, and, since $\alpha^*_{I2} = \alpha_{I2} - \Gamma_{I2,2}\Gamma_{22}^{-1}[\Gamma_{21} \vdots B_{II}]$, these restrictions imply restrictions on α^*_{I2}. The following logarithmic concentrated likelihood function is obtained:[1]

$$(VI.16)\quad g_2(Z_1,\hat{\alpha}^*_{I1},\hat{\Sigma}_{I1,I1}) = c_3 + \frac{T}{2}\log\left[\det\left(\frac{1}{T}\Gamma^{*\prime}_{I1}[Y'_1Y_1]_{\perp X}\hat{\Gamma}^*_{I1}\right)\right] - \frac{T}{2}\log[\det\hat{\Sigma}_{I1,I1}] - \frac{T}{2}\sum_{t=1}^{T} Z_{1[t]}\hat{\alpha}^{*\prime}_{I1}\hat{\Sigma}^{-1}_{I1,I1}\hat{\alpha}^{*\prime}_{I1}Z'_{1[t]}$$

where $[Y'_1Y_1]_{\perp X}$ is the moment matrix of the part of each jointly dependent variable orthogonal to the predetermined variables in the entire system (including predetermined variables in the identity equations and the part of the structure not specified). The $[Y'_1Y_1]_{\perp X}$ matrix is the same as the $[_+Y'_+Y]_{\perp X}$ matrix of the double k-class estimators except that instead of one row and column for each jointly dependent variable in equation μ, $[Y'_1Y_1]_{\perp X}$ contains one row and column for each jointly dependent variable not temporarily eliminated from the system, including the normalizing variables. (Also, we are

---

[1] Koopmans and Hood [1953], pp. 192-195.

using X as the matrix of instruments, X_I.) The usually quoted formula for calculating $[Y_1'Y_1]_{\perp X}$ is:[1]

$$(VI.17) \quad [Y_1'Y_1]_{\perp X} = Y_1'Y_1 - Y_1'X(X'X)^{-1}X'Y_1 = Y_1'(I - X(X'X)^{-1}X')Y_1 \ ;$$

however, there is considerable advantage to using direct orthogonalization to calculate $[Y_1'Y_1]_{\perp X}$ in the same manner as in calculating $[{}_+Y'{}_+Y]_{\perp X}$ rather than by using the above formula. (For one thing X need not have full column rank when direct orthogonalization is used.)

If $g_2(Z_1,\hat{\alpha}^*_{I1},\hat{\Sigma}_{I1,I1})$ is maximized with respect to $\hat{\Sigma}_{I1,I1}$ (thereby concentrating the function onto $\hat{\alpha}^*_{I1}$ and Z_1), the following relation is obtained:[2]

$$(VI.18) \quad \hat{\Sigma}_{I1,I1} = (1/T)\hat{\alpha}^*_{I1}[Z_1'Z_1]\hat{\alpha}^{*\prime}_{I1} \ .$$

Substituting for $\hat{\Sigma}_{I1,I1}$ into (VI.16) and dividing by T/2, we get the function:[3]

$$(VI.19) \quad g_3(\hat{\alpha}^*_{I1},Z_1) =$$

$$c_4 + \log[\det(\tfrac{1}{T}\hat{\Gamma}^*_{I1}[Y_1'Y_1]_{\perp X}\hat{\Gamma}^{*\prime}_{I1})] - \log[\det(\tfrac{1}{T}\hat{\alpha}^*_{I1}[Z_1'Z_1]\hat{\alpha}^{*\prime}_{I1})]$$

---

[1] Except, possibly, for a factor of T, $[Y_1'Y_1]_{\perp X}$ is the W matrix given in Koopmans and Hood [1953] and Chernoff and Divinsky [1953].

[2] Hood and Koopmans [1953], p. 166.

[3] Hood and Koopmans [1953], p. 195.

where c_4 is a constant.[1] However,[2]

$$\text{(VI.20)} \quad \det\left(\frac{1}{T}\Gamma^*_{I1}[Y_1'Y_1]_{\perp X}\Gamma^{*\prime}_{I1}\right) = \det\left(\frac{1}{T}\Gamma_{I1}[Y'Y]_{\perp X}\Gamma'_{I1}\right) = \det\left(\frac{1}{T}\Gamma_A[Y_A'Y_A]_{\perp X}\Gamma'_A\right)$$

where:

$[Y'Y]_{\perp X}$ is the $G^{X}G$ moment matrix of the part of the G jointly dependent variables in the system orthogonal to the Λ predetermined variables in the entire system. ($[Y'Y]_{\perp X}$ is the same as $[Y_1'Y_1]_{\perp X}$ but expanded to include all jointly dependent variables in the system.)

$[Y_A'Y_A]_{\perp X}$ is the $G_1{}^{X}G_1$ moment matrix of the part of the G_1 jointly dependent variables in subsystem I1 only (the M_1 stochastic equations for which the structure is specified) orthogonal to the Λ predetermined variables in the entire system.

Γ_A is the $G_1{}^{X}M_1$ matrix of coefficients of the G_1 jointly dependent variables which occur in the M_1 stochastic equations for which

---

[1] $$c_4 = \frac{2}{T}\left(c_3 - \frac{1}{2}\sum_{t=1}^{T} z_{1[t]}\hat{\alpha}^{*\prime}_{I1}\hat{\Sigma}^{-1}_{I1,I1}\hat{\alpha}^{*}_{I1}z'_{1[t]}\right)$$

$$= \frac{2}{T}\left(c_3 - \frac{1}{2}\mathrm{tr}\left\{\frac{1}{T}Z_1\hat{\alpha}^{*\prime}_{I1}[\hat{\alpha}^{*}_{I1}(Z_1'Z_1)\hat{\alpha}^{*\prime}_{I1}]^{-1}\hat{\alpha}^{*}_{I1}Z_1'\right\}\right)$$

$$= \frac{2}{T}\left(c_3 - \frac{1}{2}\mathrm{tr}\left\{\frac{1}{T}\hat{\alpha}^{*}_{I1}Z_1'Z_1\hat{\alpha}^{*\prime}_{I1}[\hat{\alpha}^{*}_{I1}Z_1'Z_1\hat{\alpha}^{*\prime}_{I1}]^{-1}\right\}\right)$$

$$= \frac{2}{T}\left(c_3 - \frac{1}{2}\mathrm{tr}\left\{\frac{1}{T}I\right\}\right) = \frac{2}{T}\left(c_3 - \frac{1}{2}\cdot 1\right)$$

$$= \frac{2}{T}c_3 - \frac{1}{T}$$

where tr denotes trace and we have used the relationship $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for any matrices A and B provided AB and BA are defined (i.e., provided the number of rows of A equals the number of columns of B and the number of rows of B equals the number of columns of A).

[2] The proofs of (VI.20) and (VI.21) follow (VI.21).

the structure is specified. If the variables in the Y matrix were rearranged to $Y = [Y_A \vdots Y_B]$ where Y_A contains the G_1 jointly dependent variables which are included in the M_1 equations for which the structure is specified and Y_B contains the remaining $G - G_1$ jointly dependent variables in the system, and if the columns of Γ_{I1} and Γ_A were rearranged into the same order, then

$$\underset{M_1 X G}{\Gamma_{I1}} = [\underset{M_1 X G_1}{\Gamma_A} \vdots \underset{M_1 X (G-G_1)}{0}] .$$

Also,

$$\text{(VI.21)} \quad \det(\tfrac{1}{T}\alpha^*_{I1}[Z_1'Z_1]\alpha^{*\prime}_{I1}) = \det(\tfrac{1}{T}\alpha_{I1}[Z'Z]\alpha'_{I1}) = \det(\tfrac{1}{T}\alpha_A[Z_A'Z_A]\alpha_A')$$

where:

Z_A is the $T X (G_1 + \Lambda_1)$ matrix of variables in the M_1 stochastic equations for which the structure is specified. (Z_A is the Z matrix with all variables which do not occur in the stochastic equations with structure specified deleted.)

α_A is the $M_1 X (G_1 + \Lambda_1)$ matrix of coefficients of the $G_1 + \Lambda_1$ variables which occur in the M_1 stochastic equations for which the structure is specified. If variables were rearranged as required:

$$\underset{M_1 X (G+\Lambda)}{\alpha_{I1}} = [\underset{M_1 X (G_1+\Lambda_1)}{\alpha_A} \vdots \underset{M_1 X [(G-G_1)+(\Lambda-\Lambda_1)]}{0}]$$

with $$\underset{M_1 X (G_1+\Lambda_1)}{\alpha_A} = [\underset{M_1 X G_1}{\Gamma_A} \vdots \underset{M_1+\Lambda_1}{B_A}] .$$

Showing (VI.20)

From (VI.8) we obtain:

$$Y_2' = -\Gamma_{22}^{-1}\Gamma_{21}Y_1' - \Gamma_{22}^{-1}B_{II}X' \quad ; \tag{VI.22}$$

hence,

$$Y_{\perp X} = [Y_1 \vdots Y_2]_{\perp X} = [Y_1 \vdots -Y_1\Gamma_{21}'(\Gamma_{22}^{-1})' - XB_{II}'(\Gamma_{22}^{-1})']_{\perp X} \quad . \tag{VI.23}$$

However, by (I.59), $[-Y_1\Gamma_{21}'(\Gamma_{22}^{-1})']_{\perp X} = -[Y_1]_{\perp X}\Gamma_{21}'(\Gamma_{22}^{-1})'$ and by (I.59) and (I.42), $[-XB_{II}'(\Gamma_{22}^{-1})']_{\perp X} = [X]_{\perp X}B_{II}'(\Gamma_{22}^{-1})' = -0B_{II}'(\Gamma_{22}^{-1})' = 0$.

Hence,

$$Y_{\perp X} = [[Y_1]_{\perp X} \vdots -[Y_1]_{\perp X}\Gamma_{21}'(\Gamma_{22}^{-1})'] \text{ , and } [Y'Y]_{\perp X} \text{ becomes:} \tag{VI.24}$$

$$[Y'Y]_{\perp X} = Y_{\perp X}'Y_{\perp X} \tag{VI.25}$$

$$= \begin{bmatrix} [Y_1]_{\perp X}'[Y_1]_{\perp X} & -[Y_1]_{\perp X}'[Y_1]_{\perp X}\Gamma_{21}'(\Gamma_{22}^{-1})' \\ -\Gamma_{22}^{-1}\Gamma_{21}[Y_1]_{\perp X}'[Y_1]_{\perp X} & \Gamma_{22}^{-1}\Gamma_{21}[Y_1]_{\perp X}'[Y_1]_{\perp X}\Gamma_{21}'(\Gamma_{22}^{-1})' \end{bmatrix}$$

$$= \begin{bmatrix} [Y_1'Y_1]_{\perp X} & -[Y_1'Y_1]_{\perp X}\Gamma_{21}'(\Gamma_{22}^{-1})' \\ -\Gamma_{22}^{-1}\Gamma_{21}[Y_1'Y_1]_{\perp X} & \Gamma_{22}^{-1}\Gamma_{21}[Y_1'Y_1]_{\perp X}\Gamma_{21}'(\Gamma_{22}^{-1})' \end{bmatrix} .$$

If Γ_{11} is correspondingly subdivided in the manner of (VI.9) we have:

(VI.26) $\det\{\Gamma_{I1}[Y'Y]_{\perp X}\Gamma'_{I1}\} =$

$$\det\left\{[\Gamma_{I1,1} \vdots \Gamma_{I1,2}]\begin{bmatrix}[Y'_1Y_1]_{\perp X} & -[Y'_1Y_1]_{\perp X}\Gamma'_{21}(\Gamma_{21}^{-1})' \\ -\Gamma_{22}^{-1}\Gamma_{21}[Y'_1Y_1]_{\perp X} & \Gamma_{22}^{-1}\Gamma_{21}[Y'_1Y_1]_{\perp X}\Gamma'_{21}(\Gamma_{21}^{-1})'\end{bmatrix}\begin{bmatrix}\Gamma'_{I1,1} \\ \Gamma'_{I1,2}\end{bmatrix}\right\}$$

$$= \det\{\Gamma_{I1,1}[Y'_1Y_1]_{\perp X}\Gamma'_{I1,1} - \Gamma_{I1,1}[Y'_1Y_1]_{\perp X}\Gamma'_{21}(\Gamma_{22}^{-1})'\Gamma'_{I1,2}$$

$$- \Gamma_{I1,2}\Gamma_{22}^{-1}\Gamma_{21}[Y'_1Y_1]_{\perp X}\Gamma'_{I1,1} + \Gamma_{I1,2}\Gamma_{22}^{-1}\Gamma_{21}[Y'_1Y_1]_{\perp X}\Gamma'_{21}(\Gamma_{22}^{-1})'\Gamma'_{I1,2}\}$$

$$= \det\{[\Gamma_{I1,1}-\Gamma_{I1,2}\Gamma_{22}^{-1}\Gamma_{21}][Y'_1Y_1]_{\perp X}\Gamma'_{I1,1}$$

$$- [\Gamma_{I1,1}-\Gamma_{I1,2}\Gamma_{22}^{-1}\Gamma_{21}][Y'_1Y_1]_{\perp X}\Gamma'_{21}(\Gamma_{22}^{-1})'\Gamma'_{I1,2}\}$$

$$= \det\{[\Gamma_{I1,1}-\Gamma_{I1,2}\Gamma_{22}^{-1}\Gamma_{21}][Y'_1Y_1]_{\perp X}[\Gamma_{I1,1}-\Gamma_{I1,2}\Gamma_{22}^{-1}\Gamma_{21}]'\}$$

$$= \det\{\Gamma^*_{I1}[Y'_1Y_1]_{\perp X}\Gamma^{*\prime}_{I1}\} .$$

The second equality of (VI.20) involves a mere rewriting to eliminate from $Y_{\perp X}$ variables for which no corresponding jointly dependent variables occurs in the set of stochastic equations whose structure is specified. If the variables are rearranged in the manner noted in the definition of Γ_A above, we have:

$$(VI.27)\quad \Gamma_{I1}[Y'Y]_{\perp X}\Gamma'_{I1} = [\Gamma_A \vdots 0]\begin{bmatrix}[Y'_AY_A]_{\perp X} & [Y'_AY_B]_{\perp X} \\ [Y'_BY_A]_{\perp X} & [Y'_BY_B]_{\perp X}\end{bmatrix}\begin{bmatrix}\Gamma'_A \\ 0'\end{bmatrix} = \Gamma_A[Y'_AY_A]_{\perp X}\Gamma'_A .$$

Showing (VI.21) Holds

By (VI.3) and (VI.12), $\alpha_{I1}Z' = -U'_1 = -\alpha^*_{I1}Z'_1$; hence, $\alpha^*_{I1}Z'_1Z_1\alpha^{*\prime}_{I1} = \alpha_{I1}Z'Z\alpha'_{I1}$.

The second equality of (VI.21) involves a mere rewriting to eliminate from Z variables which do not occur in the stochastic equations whose structure is specified. If the variables are rearranged in the manner noted in the definition of α_A we have:

$$(VI.28)\quad \alpha_{I1} Z'Z \alpha_{I1}' = [\alpha_A \vdots 0] \begin{bmatrix} Z_A'Z_A & Z_A'Z_B \\ Z_B'Z_A & Z_B'Z_B \end{bmatrix} \begin{bmatrix} \alpha_A' \\ 0' \end{bmatrix} = \alpha_A [Z_A'Z_A] \alpha_A' \quad .$$

Continuing the Derivation of The SML Likelihood Function

Substituting (VI.20) and (VI.21) into (VI.19) we have:

$$(VI.29)\quad g_4(\hat{\alpha}_A, Z) = c_4 + \log[\det(\tfrac{1}{T}\hat{\Gamma}_A [Y'Y]_{\perp X} \hat{\Gamma}_A')] - \log[\det(\tfrac{1}{T}\hat{\alpha}_A [Z_A'Z_A] \hat{\alpha}_A')]$$

which is the concentrated likelihood function which we will maximize in the calculation of SML estimates. Note that Γ_A and α_A are the coefficient matrices of the M_1 stochastic equations <u>before</u> using the identity equations to temporarily eliminate the G - M jointly dependent variables. Thus, the elimination of the G - M jointly dependent variables is convenient only for the derivation of the concentrated likelihood function and is unnecessary for the expression of the concentrated likelihood function or the computation of the SML coefficients.[1] The predetermined variables in the entire system (including those in the part of the system which was not specified and in the identity

---

[1] Thus, Chernoff and Divinsky [1953] would have obtained the same resulting coefficients and considerably simplified their computation of SML coefficients for Klein's model (b) (the same model as Klein's model I except that G and R are classified as jointly dependent instead of predetermined) if they had used the stochastic equations which were specified directly instead of using the identity equations to eliminate jointly dependent variables before commencing their iteration procedure.

equations) are used in the calculation of the $[Y'Y]_{\perp X}$ matrix.[1]

The matrices $\frac{1}{T}\hat{\alpha}_A[Z_A'Z_A]\hat{\alpha}_A'$ and $\frac{1}{T}\hat{\Gamma}_A[Y_A'Y_A]_{\perp X}\hat{\Gamma}_A'$ are used repeatedly in the elaboration of the SML computational procedure which follows; hence, it will prove convenient to denote these matrices as simply S and T, i.e.,

$$(VI.30)\quad S = \hat{\Sigma}_{I1,I1} = \frac{1}{T}\hat{\alpha}_A[Z_A'Z_A]\hat{\alpha}_A'$$

and

$$(VI.31)\quad T = \frac{1}{T}\hat{\alpha}_A[Z_A'Z_A]_{\perp X}\hat{\alpha}_A' = \frac{1}{T}\hat{\Gamma}_A[Y_A'Y_A]_{\perp X}\hat{\Gamma}_A' .$$

(The matrix T should not be confused with the number of observations, T.) Let $a = \begin{bmatrix} a_1 \\ \vdots \\ a_{M_1} \end{bmatrix}$ be the vector of unrestricted estimated coefficients of α_A (a is formed from $\hat{\alpha}_A$ in the same way that the a vector was formed from $\hat{\alpha}_I$ for FIML estimation [(V.20) and (V.22)]). Then S

---

[1]The preceding derivation showing that it matters not whether the identity equations are ignored in the computational procedure or used to eliminate jointly dependent variables from the subsystem being estimated assumes that the same X matrix is used in either case; that is, the X matrix is taken as the predetermined variables in the system before using the identity equations to eliminate jointly dependent variables. If instead of this approach the identity equations are used to eliminate jointly dependent variables from the subsystem being estimated and <u>then</u> the matrix of predetermined variables in the system is constructed as the predetermined variables in the newly modified subsystem being estimated plus the predetermined variables in the remaining stochastic equations, the X matrix constructed in this manner (we will call this matrix the "new" X matrix) will in general not coincide with the original X matrix, since some predetermined variables are likely to have been linearly combined with jointly dependent variables and the combined variables labeled jointly dependent. Since the space spanned by the new X matrix is likely to be smaller than the space spanned by the original X matrix, $[Y_A'Y_A]_{\perp X}$ will in general be changed and therefore, the coefficients obtained will, in general be different.

and T may also be defined as $S = [s_{\mu\mu'}]$ and $T = [t_{\mu\mu'}]$ with (see V.21):

(VI.31a) $$s_{\mu\mu'} = \frac{1}{T} + a_{\mu}'[+Z_{\mu}' + Z_{\mu'}]+a_{\mu'} = \frac{1}{T}\hat{u}_{\mu}'\hat{u}_{\mu'}$$

(VI.31b) $$t_{\mu\mu'} = \frac{1}{T} + \hat{\gamma}_{\mu}'[+Y_{\mu}' + Y_{\mu'}]+\hat{\gamma}_{\mu'} \quad .$$

We choose the unrestricted coefficients of $\hat{\alpha}_A$ such that $g_4(\hat{\alpha}_A, Z)$ is a maximum; however, since Z is fixed for any given sample, for any given structure the only elements of $\hat{\alpha}_A$ which are allowed to vary are the elements of the vector a. Thus, for an assumed structure and a given sample, $g_4(\hat{\alpha}_A, Z)$ may be considered a function of the vector a only, i.e.,

(VI.32) $$g(a) = g_4(\hat{\alpha}_A, Z) = c_4 + \log(\det T) - \log(\det S) \quad .$$

Another function which will be maximized when $g(a)$ is maximized is the function

(VI.33) $$g^*(a) = \det T / \det S \quad .$$

If the number of jointly dependent variables in the equation being estimated equals the number of equations being estimated, then Γ_A is square and $\det(\Gamma_A[Y_A'Y_A]_{\perp X}\Gamma_A') = \det \Gamma_A \cdot \det[Y_A'Y_A]_{\perp X} \cdot \det \Gamma_A' = \det^2\Gamma_A \cdot \det[Y_A'Y_A]_{\perp X}$. Hence, $\log[\det T] = \log[\det(\frac{1}{T}\hat{\Gamma}_A[Y_A'Y_A]_{\perp X}\hat{\Gamma}_A')] = \log[\det(\frac{1}{T}[Y_A'Y_A]_{\perp X})] + \log(\det^2\hat{\Gamma}_A)$. Substituting into (VI.32) we get $g(a) = c_5 + \log(\det^2\hat{\Gamma}_A) - \log[\det S]$ $(c_5 = c_4 + \log[\det(\frac{1}{T}[Y_A'Y_A]_{\perp X})])$,

the form of the logarithmic likelihood function for FIML [see (V.17)].[1] Thus, for this case, maximization of $g(a)$ using the SML procedure will result in the same a as maximization of the same equations using the FIML procedure.[2]

---

[1] This manipulation assumes that $\det T \neq 0$. If $\text{rk } X \geq T - G_1 + 1$ (where G_1 is the number of jointly dependent variables in the equations being estimated) then $\det T = 0$; hence, $\log[\det T] = \log[0]$ which is not defined. This case is discussed in more detail in section VI.D. $\text{rk } X \geq T - G_1 + 1$ causes no difficulty in FIML estimation.

Klein and Nakamura [1962], pp. 295-297 were evidently considering the calculation of FIML estimates as a particular case of SML in their comments of the effect of multicollinearity on FIML estimation relative to the effect of multicollinearity on other estimation procedures. This way of regarding FIML estimation may be undesirable both computationally and conceptionally. It may be undesirable computationally, since the SML computations are more severe and artifically impose problems when $\text{rk } X$ is large relative to the number of observations. It may be undesirable conceptionally since the FIML estimator has properties not possessed by the SML estimator (except as the two coincide), e.g., the FIML least generalized variance property (section V.A) is more powerful than the SML least variance ratio of two generalized variances property (section VI.A). Their conclusion that FIML is more sensitive to multicollinearity among the predetermined variables in the system than 2SLS and LIML does not follow from their arguments. They have given very good arguments as to why one would expect SML to be more sensitive to multicollinearity among the predetermined variables in the system than 2SLS or LIML.

[2] This is noted in Koopmans and Hood [1953], footnote 73, p. 165.

2. Computational formulas

The same computational procedure as was used for FIML is used for SML. Only the actual vector of partial derivatives and matrix of second partial derivatives differ from the FIML vector of partial derivatives and matrix of second partial derivatives. Also, g*(a) is used in place of f*(a) in determining the step size.

The vector of starting estimates, $a^{(1)}$, is arbitrary as in FIML. The starting estimates may be derived from single equation techniques such as DLS, 2SLS, and LIML, from multiple equation techniques such as 3SLS or I3SLS, or in some other manner. The coefficients for iteration i are calculated from the coefficients for iteration (i-1) by (V.48), i.e.,

$$a^{(i)} = a^{(i-1)} + h^{(i)} \cdot d^{(i)} \tag{VI.34}$$

where $d^{(i)} = |\mathcal{L}^{(i-1)}|^{-1} \iota^{(i-1)}$ and $h^{(i)}$ is the step size for the iteration. In what follows, we will omit the superscript giving the iteration number from the $a^{(i-1)}$, $d^{(i)}$, and $\iota^{(i)}$ vectors and the $\mathcal{L}^{(i)}$ matrix.

In SML estimation, the $\mu\mu'^{th}$ block of the $\mathcal{L}$ matrix is:

$$\mathcal{L}_{\mu\mu'} = -T\frac{\partial^2 g(a)}{\partial a_{\mu} a_{\mu'}} = s^{\mu\mu'} Z_{\mu}'Z_{\mu'} - (1/T) Z_{\mu}'\hat{U} F_{\mu\mu'} \hat{U}' Z_{\mu'}$$

$$- t^{\mu\mu'} [Z_{\mu}'Z_{\mu'}]_{\perp X} - (1/T)[Z_{\mu}'\hat{U}]_{\perp X} G_{\mu\mu'} [\hat{U}'Z_{\mu'}]_{\perp X} , \tag{VI.35}$$

and the μ^{th} block of the vector of the right hand side vector is:[1]

$$(VI.36)\quad \iota_\mu = T\frac{\partial g(a)}{\partial a_\mu} = \sum_{\mu'=1}^{M_1} s^{\mu\mu'} Z_\mu'\hat{u}_{\mu'} - \sum_{\mu'=1}^{M_1} t^{\mu\mu'}[Z_\mu'\hat{u}_{\mu'}]_{\perp X}$$

$$= Z_\mu'\hat{U}_1 \begin{bmatrix} s^{1\mu} \\ \vdots \\ s^{M_1\mu} \end{bmatrix} - [Z_\mu'\hat{U}_1]_{\perp X} \begin{bmatrix} t^{1\mu} \\ \vdots \\ t^{M_1\mu} \end{bmatrix} .$$

where:

$s^{\mu\mu'}$ is the element of the μ^{th} row and μ'^{th} column of S^{-1} [see (VI.30) and (VI.31a)].

$t^{\mu\mu'}$ is the element of the μ^{th} row and μ'^{th} column of T^{-1} [see (VI.31) and (VI.31b)].

---

[1] (VI.35) and (VI.36) may be derived by noting that $T\frac{\partial(\det S)'}{\partial a_\mu}$ is the same for FIML and SML-- $Z_\mu'U_1\begin{bmatrix} s^{1\mu} \\ \vdots \\ s^{M_1\mu} \end{bmatrix}$, the first term of (VI.36). The negative of the partial derivative of the first term of (VI.36) with respect to $a_{\mu'}$ gives $s^{\mu\mu'}Z_\mu'Z_{\mu'} - (1/T)Z_\mu'U_1F_{\mu\mu'}U_1'Z_{\mu'}$, the first two terms of (VI.35). To derive the second term of (VI.36) and the last two terms of (VI.35), note that $T = \frac{1}{T}\hat{\Gamma}[Y_A'Y_A]_{\perp X}\hat{\Gamma}' = \frac{1}{T}\hat{\alpha}_A([Y_A]_{\perp X} \vdots 0)'([Y_A]_{\perp X} \vdots 0)\hat{\alpha}_A' = \frac{1}{T}\hat{\alpha}_A[Z_A'Z_A]_{\perp X}\hat{\alpha}_A'$; hence, the same partial derivatives as for $\log(\det S)$ (with $S = \frac{1}{T}\hat{\alpha}_A[Z'Z]\hat{\alpha}_A'$) are obtained except that $[Z]_{\perp X}$ is substituted for Z in each of the terms.

Except for the lack of use of direct orthogonalization, essentially the same basic formulas are given in Chernoff and Divinsky [1963], pp. 261-263.

$$-Z_\mu'\hat{U}_1 = -Z_\mu'[\hat{u}_1 \cdots \hat{u}_{M_1}] = [(Z_\mu' + Z_1) + a_1 \cdots (Z_\mu' + Z_{M_1}) + a_{M_1}] \quad .$$

$$-[Z_\mu'\hat{U}_1]_{\perp X} = \{[Z_\mu' + Z_1]_{\perp X} + a_1 \cdots [Z_\mu' + Z_{M_1}]_{\perp X} + a_{M_1}\} \quad .$$

$$= \left\{\begin{bmatrix} [Y_\mu' + Y_1]_{\perp X} & 0 \\ & \\ 0 & 0 \end{bmatrix} + a_1 \cdots \begin{bmatrix} [Y_\mu' + Y_{M_1}]_{\perp X} & 0 \\ & \\ 0 & 0 \end{bmatrix} + a_{M_1}\right\}$$

$$= \left\{\begin{bmatrix} [Y_\mu' + Y_1]_{\perp X} + \hat{Y}_1 \\ m_\mu \times 1 \\ 0 \\ \ell_\mu \times 1 \end{bmatrix} \cdots \begin{bmatrix} [Y_\mu' + Y_{M_1}]_{\perp X} + \hat{Y}_{M_1} \\ m_\mu \times 1 \\ 0 \\ \ell_\mu \times 1 \end{bmatrix}\right\} \quad .$$

$$F_{\mu\mu} = \begin{bmatrix} s^{1\mu}' \\ \vdots \\ s^{M_1\mu}' \end{bmatrix} [s^{1\mu} \cdots s^{M_1\mu}] + s^{\mu\mu}' S^{-1} \quad .$$

$$G_{\mu\mu}' = \begin{bmatrix} t^{1\mu}' \\ \vdots \\ t^{M_1\mu}' \end{bmatrix} [t^{1\mu} \cdots t^{M_1}] + t^{\mu\mu}' T^{-1} \quad .$$

A "degrees of freedom" adjustment can be made in the estimated SML disturbance variance-covariance and the estimated SML coefficient variance-covariance matrices in the same manner as for FIML. These matrices can also be normalized in the manner suggested for FIML. (See section V.D. and V.E.) a_{SML} and $\mathscr{L}_{SML}^{-1}$ are, of course, used in place of a_{FIML} and $\mathscr{L}_{FIML}^{-1}$ in the calculation of these matrices.

B. Arbitrary Linear Restrictions Imposed on Coefficients

The iterative procedure given in the previous section is designed to maximize $g(a)$ with respect to the vector of coefficients, a. In this section the problem is to:

(VI.37) $$\max_{a} g(a)$$

subject to:

(VI.38) $$\underset{N_R \times n}{R} \quad \underset{n \times 1}{a} = \underset{N_R \times 1}{r}$$

where the R matrix and the r vector are as defined for restricting a in FIML estimation. The Q and Q_2 matrices and the q and q_2 vectors are calculated from the R matrix and the r vector in the same manner as for FIML, and the computation gives a means of separating out a vector of $n - rk\ R$ "unrestricted" coefficients, $a_{(1)}$, and a vector of $rk\ R$ coefficients, $a_{(2)}$, which may be calculated from the $a_{(1)}$ vector.

The computational formulas for FIML, (V.67) through (V.70), are applicable for SML as well, except that the $\mathscr{L}$ matrix and ι vector for SML are used in place of the $\mathscr{L}$ matrix and ι vector for FIML.

C. Using Instrumental Variables in SML Estimation

In deriving the logarithmic likelihood function and in the presentation of formulas for the SML estimator, the jointly dependent variables in the equations being estimated were adjusted by all of the jointly dependent variables in the system; i.e., the matrix X was used in the calculation of $[Y_A'Y_A]_{\perp X}$. A more general approach would be to adjust the jointly dependent variables in the equations being estimated by a matrix of instruments, X_I, where X_I contains all of the predetermined variables in the stochastic equations being estimated plus a set of additional instrumental variables.[1] Thus, the matrix $[Y_A'Y_A]_{\perp X_I}$ could be substituted for the matrix $[Y_A'Y_A]_{\perp X}$ in the above likelihood formulas and, therefore, in the SML coefficient formulas derived from the likelihood formulas. If such a substitution is made, the resulting SML coefficients will not be the same as the coefficients obtained through use of the $[Y_A'Y_A]_{\perp X}$ matrix and <u>cannot</u> be described as the coefficients which maximize the likelihood function given the structure of the equations estimated and <u>the predetermined variables in the system</u>.

---

[1] Selection of instrumental variables is discussed in section II.G.

D. SML Estimation when $\text{rk } X \geq T - M_1 + 1$

If $\text{rk } X \geq T - M_1 + 1$ (where M_1 is the number of equations in the system being estimated) then the T matrix will be singular; hence, the formulas given previously cannot be used to compute SML estimates.[1] In some cases, a matrix of instrumental variables, X_I, could be substituted for the X matrix as noted in section VI.C, with the number of instruments restricted so that $\text{rk } X_I < T - M_1 + 1$; however, the properties of the SML coefficients obtained from such a substitution have not been examined. There is especially a question as to whether or how much the statistical efficiency in the estimation of the SML coefficients is decreased if the X_I space is restricted.[2]

---

[1] $\det T = \det\{\frac{1}{T}\ddot{\Gamma}_A[Y_A'Y_A]_{\perp X}\ddot{\Gamma}_A'\} = \frac{1}{T}\det\{[\ddot{\Gamma}_A Y_A']_{\perp X}[Y_A\ddot{\Gamma}_A']_{\perp X}\}$ [the last equality comes from (I.80)]; hence the $T \times M_1$ matrix $[Y_A\ddot{\Gamma}_A']_{\perp X}$ must have full column rank if $\det T$ is not to be zero. But the space orthogonal to X has rank $T - \text{rk } X$; hence, if $[Y_A\ddot{\Gamma}_A']_{\perp X}$ is to have full column rank, $T - \text{rk } X$ must be greater than or equal to M_1, i.e., $\text{rk } X \leq T - M_1$. Thus, $\det T = 0$ if $\text{rk } X \geq T - M_1 + 1$.

[2] If the number of jointly dependent variables in the subsystem being estimated equals the number of equations being estimated, restriction of the X_I space will lead to the same estimates as would have been obtained if FIML had been applied only to that subsystem (see section VI.A); however, in this case it is more efficient (with respect to computer time and capacity requirements) to use the FIML computational method directly. ($\text{rk } X_I \geq T - M_1 + 1$ presents no difficulty in FIML estimation, since the joingly dependent variables are not adjusted in the FIML estimation procedure. Instead, the coefficients of the jointly dependent variables are taken explicitly into account.)

E. Iterative Limited Information Single Equation Maximum Likelihood (ILIML)

In the case of only a single jointly dependent variable per equation, Professor Lester Telser proposed an iterative DLS estimation procedure which estimates the coefficients of one equation at a time based on the predetermined variables in that equation plus the residuals from each of the remaining equations.[1] The new coefficients for an equation are used to estimate new residuals for the equation which are, in turn, used to estimate new coefficients for the remaining equations.

In chapter VII of this paper it is shown that Telser's iterative procedure leads to FIML estimates in the special case of a single jointly dependent variable per equation. In this section, we will demonstrate that the same procedure but using LIML instead of DLS leads to FIML estimates in the general case (multiple jointly dependent variables permitted in each equation).[2] In the case of an incomplete system, the method leads to SML estimates.

Derivation of The ILIML Method

As noted in (VI.33) the function $g^*(a) = \frac{\det T}{\det S} = \frac{\det\{\hat{\Gamma}_A[Y'_AY_A]_{\perp X}\hat{\Gamma}'_A\}}{\det\{\hat{\alpha}_A[Z'_AZ_A]\hat{\alpha}'_A\}}$ is maximized by the same coefficients that maximize the likelihood function. Suppose that we partition the matrix of coefficients such that the coefficients for one equation are distinguished

---

[1] Telser [1964], pp. 845-862.

[2] The method (ILIML) given in this section was proposed to the writer by Professor Herman Rubin in June, 1963. The key steps of the proof of the increase in the likelihood function at each iteration were indicated to the writer by Professor Rubin at that time.

from the coefficients for the remaining equations. Without loss of generality we can assume that we are distinguishing the first equation from the remaining equations whose structure is specified, since the equations can be rearranged. Thus, let α_A and Γ_A be subdivided as:

$$(VI.39)\quad \alpha_A = \begin{bmatrix} \alpha_1 \\ 1^{\times}(G_1+\Lambda_1) \\ \\ \alpha_2 \\ (M_1-1)^{\times}(G_1+\Lambda_1) \end{bmatrix} \quad \text{and} \quad \Gamma_A = \begin{bmatrix} \Gamma_1 \\ 1^{\times}G_1 \\ \\ \Gamma_2 \\ (M_1-1)^{\times}G_1 \end{bmatrix}$$

then $g^*(a)$ becomes:

$$(VI.40)\quad g^*(a) = \frac{\det\begin{bmatrix} \hat{\Gamma}_1[Y_A'Y_A]_{\perp X}\hat{\Gamma}_1' & \hat{\Gamma}_1[Y_A'Y_A]_{\perp X}\hat{\Gamma}_2' \\ \hat{\Gamma}_2[Y_A'Y_A]_{\perp X}\hat{\Gamma}_1' & \hat{\Gamma}_2[Y_A'Y_A]_{\perp X}\hat{\Gamma}_2' \end{bmatrix}}{\det\begin{bmatrix} \hat{\alpha}_1[Z_A'Z_A]\hat{\alpha}_1' & \hat{\alpha}_1[Z_A'Z_A]\hat{\alpha}_2' \\ \hat{\alpha}_2[Z_A'Z_A]\hat{\alpha}_1' & \hat{\alpha}_2[Z_A'Z_A]\hat{\alpha}_2' \end{bmatrix}}$$

or

$$(VI.41)\quad g^*(a) = \frac{\hat{\Gamma}_1[Y_A'Y_A]_{\perp X}\hat{\Gamma}_1' - \hat{\Gamma}_1[Y_A'Y_A]_{\perp X}\hat{\Gamma}_2'(\hat{\Gamma}_2[Y_A'Y_A]_{\perp X}\hat{\Gamma}_2')^{-1}\hat{\Gamma}_2[Y_A'Y_A]_{\perp X}\hat{\Gamma}_1'}{\hat{\alpha}_1Z_A'Z_A\hat{\alpha}_1' - \hat{\alpha}_1Z_A'Z_A\hat{\alpha}_2'(\hat{\alpha}_2Z_A'Z_A\hat{\alpha}_2')^{-1}\hat{\alpha}_2Z_A'Z_A\hat{\alpha}_1'} \cdot \frac{\det(\hat{\Gamma}_2[Y_A'Y_A]_{\perp X}\hat{\Gamma}_2')}{\det(\hat{\alpha}_2Z_A'Z_A\hat{\alpha}_2')}$$

(VI.41) is derived from (VI.40) by the determinental relationship shown in footnote 3 of page 167. The "det" operator may be omitted from the numerator and denominator of the first term of the product since each is a $1^{\times}1$ matrix.

The second term of (VI.41) consists only of matrices fixed for any sample ($[Y_A'Y_A]_{\perp X}$ and $Z_A'Z_A$) and coefficient matrices from equations other than equation 1. Thus for a given $\hat{\alpha}_2$, the second term may be considered to be a positive constant in the function to be

maximized at this point and we may consider the problem to be one of selecting $\hat{\alpha}_1$ to minimize:[1]

(VI.42) $g^{**}(a) =$

$$\frac{\hat{\alpha}_1 Z_A' Z_A \hat{\alpha}_1' - \hat{\alpha}_1 Z_A' Z_A \hat{\alpha}_2' (\hat{\alpha}_2 Z_A' Z_A \hat{\alpha}_2')^{-1} \hat{\alpha}_2 Z_A' Z_A \hat{\alpha}_1'}{\hat{\Gamma}_1 [Y_A' Y_A]_{\perp X} \hat{\Gamma}_1' - \hat{\Gamma}_1 [Y_A' Y_A]_{\perp X} \hat{\Gamma}_2' (\hat{\Gamma}_2 [Y_A' Y_A]_{\perp X} \hat{\Gamma}_2')^{-1} \hat{\Gamma}_2 [Y_A' Y_A]_{\perp X} \hat{\Gamma}_1'} .$$

The problem has been switched from maximizing a function to minimizing the reciprocal of the function ($\hat{\alpha}_2$ being temporarily held fixed) to make the correspondence with LIML easier and to make the derivation of some partial derivatives slightly simpler.

Let $\hat{U}_2$ be the $T^X(M_1 - 1)$ matrix of residuals of equation 2 through M_1 and assume that $\hat{U}_2$ has full column rank.[2] Thus,

(VI.43) $$\hat{U}_2 = -Z_A \hat{\alpha}_2' = -Y_A \hat{\Gamma}_2' - X_A \hat{B}_2' .$$

Then the numerator of (VI.42) may be written as:

(VI.44) $$\hat{\alpha}_1 [Z_A' Z_A - Z_A' \hat{U}_2 (\hat{U}_2' \hat{U}_2)^{-1} \hat{U}_2' Z_A] \hat{\alpha}_1' = \hat{\alpha}_1 [Z_A' Z_A]_{\perp \hat{U}_2} \hat{\alpha}_1' \quad \text{[see (I.37)]} .$$

That the denominator of (VI.42) may be written as:[3]

(VI.45) $$\hat{\Gamma}_1 [Y_A' Y_A]_{\perp [X : \hat{U}_2]} \hat{\Gamma}_1'$$

may be seen as follows:

---

[1] $g^*(a) = \frac{1}{g^{**}(a)} \cdot \frac{\det(\hat{\Gamma}_2 [Y_A' Y_A]_{\perp X} \hat{\Gamma}_2')}{\det(\hat{\alpha}_2 Z_A' Z_A \hat{\alpha}_2')}$

[2] If $\hat{\Sigma}_{I1,I1} = \frac{1}{T}[\hat{U}_1' \hat{U}_1]$ is assumed nonsingular; then $\hat{U}_1 = [\hat{u}_1 \vdots \hat{U}_2]$ and, therefore, $\hat{U}_2$ must already have full column rank.

[3] $[Y_A' Y_A]_{\perp [X : \hat{U}_2]}$ is the part of the moment matrix of Y_A orthogonal to both X and $\hat{U}_2$.

$$[\hat{U}_2]_{\perp X} = [-Y_A\hat{\Gamma}_2' - X_A\hat{B}_2']_{\perp X} = -[Y_A]_{\perp X}\hat{\Gamma}_2' - [X_A]_{\perp X}\hat{B}_2' \qquad \text{[by (I.63)]}$$

$$= -[Y_A]_{\perp X}\hat{\Gamma}_2' \text{ since } [X_A]_{\perp X} = 0 \qquad \text{[see (I.42)]}.$$

Thus,

$$[Y_A]_{\perp}[X\vdots\hat{U}_2] = [Y_A]_{\perp}(X\vdots[\hat{U}_2]_{\perp X}) = \{[Y_A]_{\perp}([\hat{U}_2]_{\perp X})\}_{\perp X} \qquad \text{[see (I.65)]}$$

$$= \{[Y_A]_{\perp}(-[Y_A]_{\perp X}\hat{\Gamma}_2')\}_{\perp X}$$

$$= \{Y_A - (-[Y_A]_{\perp X}\hat{\Gamma}_2')(\hat{\Gamma}_2[Y_A'Y_A]_{\perp X}\hat{\Gamma}_2')^{-1}(-\hat{\Gamma}_2[Y_A]_{\perp X}'Y_A\}_{\perp X} \qquad \text{[see (I.47)]}$$

$$= [Y_A]_{\perp X} - \{[Y_A]_{\perp X}\}_{\perp X}\hat{\Gamma}_2'(\hat{\Gamma}_2[Y_A'Y_A]_{\perp X}\hat{\Gamma}_2')^{-1}\hat{\Gamma}_2[Y_A]_{\perp X}'Y_A \qquad \text{[see (I.62)]}.$$

However, $\{[Y_A]_{\perp X}\}_{\perp X} = [Y_A]_{\perp X}$ and $[Y_A]_{\perp X}'Y_A = [Y_A'Y_A]_{\perp X}$ [see (I.56)]; hence,

$$\text{(VI.46)} \qquad [Y_A]_{\perp}[X\vdots\hat{U}_2] = [Y_A]_{\perp X} - [Y_A]_{\perp X}\hat{\Gamma}_2'(\hat{\Gamma}_2[Y_A'Y_A]_{\perp X}\hat{\Gamma}_2')^{-1}\hat{\Gamma}_2[Y_A'Y_A]_{\perp X}.$$

$[Y_A'Y_A]_{\perp}[X\vdots\hat{U}_2]$ may be written as $Y_A'[Y_A]_{\perp}[X\vdots\hat{U}_2]$. Using (VI.46), we have

$$\text{(VI.47)} \qquad [Y_A'Y_A]_{\perp}[X\vdots\hat{U}_2] = Y_A'[Y_A]_{\perp}[X\vdots\hat{U}_2]$$

$$= Y_A'[Y_A]_{\perp X} - Y_A'[Y_A]_{\perp X}\hat{\Gamma}_2'(\hat{\Gamma}_2[Y_A'Y_A]_{\perp X}\hat{\Gamma}_2')^{-1}\hat{\Gamma}_2[Y_A'Y_A]_{\perp X}$$

$$= [Y_A'Y_A]_{\perp X} - [Y_A'Y_A]_{\perp X}\hat{\Gamma}_2'(\hat{\Gamma}_2[Y_A'Y_A]_{\perp X}\hat{\Gamma}_2')^{-1}\hat{\Gamma}_2[Y_A'Y_A]_{\perp X}.$$

Hence, $\hat{\Gamma}_1[Y_A'Y_A]_{\perp}[X\vdots\hat{U}_2]\hat{\Gamma}_1'$ is the denominator of (VI.42) as claimed in (VI.45), and (VI.42) becomes:

$$\text{(VI.48)} \qquad g^{**}(a) = \frac{\hat{\alpha}_1[Z_A'Z_A]_{\perp\hat{U}_2}\hat{\alpha}_1'}{\hat{\Gamma}_1[Y_A'Y_A]_{\perp}[X\vdots\hat{U}_2]\hat{\Gamma}_1'}$$

But in the notation of part I of this paper (since rows and columns of $[Z_A'Z_A]_{\perp\hat{U}_2}$ and $[Y_A'Y_A]_{\perp}[X\vdots\hat{U}_2]$ corresponding to coefficients restricted to be zero in $\hat{\alpha}_1$ and $\hat{\Gamma}_1$ may be delected), (VI.48) may be rewritten

as:[1]

$$\text{(VI.49)} \qquad g^{**}(a) = \frac{[{}_{+}\hat{Y}_1' \vdots \hat{\beta}_1'][{}_{+}Z_1' \; {}_{+}Z_1]_{\perp\hat{U}_2}\begin{bmatrix} {}_{+}\hat{Y}_1 \\ \hat{\beta}_1 \end{bmatrix}}{{}_{+}\hat{Y}_1'[{}_{+}Y_1' \; {}_{+}Y_1]_{\perp[X\vdots\hat{U}_2]} \; {}_{+}\hat{Y}_1} \; .$$

We may minimize $g^{**}(a)$ first with respect to $\hat{\beta}_1$ thereby concentrating the function on ${}_{+}\hat{Y}_1$. First notice that the numerator of (VI.49) may be rewritten as:

$$\text{(VI.50)} \quad [{}_{+}\hat{Y}_1' \;\; \hat{\beta}_1'] \begin{bmatrix} {}_{+}Y_1' \; {}_{+}Y_1 & {}_{+}Y_1'X_1 \\ X_1' \; {}_{+}Y_1 & X_1'X_1 \end{bmatrix}_{\perp\hat{U}_2} \begin{bmatrix} {}_{+}\hat{Y}_1 \\ \hat{\beta}_1 \end{bmatrix} =$$

$$[{}_{+}\hat{Y}_1' \;\; \hat{\beta}_1'] \begin{bmatrix} [{}_{+}Y_1' \; {}_{+}Y_1]_{\perp\hat{U}_2} & [{}_{+}Y_1'X_1]_{\perp\hat{U}_2} \\ [X_1' \; {}_{+}Y_1]_{\perp\hat{U}_2} & [X_1'X_1]_{\perp\hat{U}_2} \end{bmatrix} \begin{bmatrix} {}_{+}\hat{Y}_1 \\ \hat{\beta}_1 \end{bmatrix}$$

$$= {}_{+}\hat{Y}_1'[{}_{+}Y_1' \; {}_{+}Y_1]_{\perp\hat{U}_2} \; {}_{+}\hat{Y}_1 + 2\hat{\beta}_1'[X_1' \; {}_{+}Y_1]_{\perp\hat{U}_2} \; {}_{+}\hat{Y}_1 + \hat{\beta}_1'[X_1'X_1]_{\perp\hat{U}_2}\hat{\beta}_1 \; .$$

Taking the partial derivative of $g^{**}(a)$ with respect to $\hat{\beta}_1$ and setting the partial derivative to zero we have:

$$\text{(VI.51)} \quad (1/d)\{2[X_1' \; {}_{+}Y_1]_{\perp\hat{U}_2} \; {}_{+}\hat{Y}_1 + 2[X_1'X_1]_{\perp\hat{U}_2}\hat{\beta}_1\} = 0$$

where d is the denominator of $g^{**}(a)$ in (VI.49). Solving for $\hat{\beta}_1$

---

[1] ${}_{+}\hat{Y}_\mu = \begin{bmatrix} -I \\ \hat{\gamma}_\mu \end{bmatrix}$ and ${}_{+}Z_\mu = [y_\mu \vdots Z_\mu]$.

we have:[1]

$$\text{(VI.52)} \qquad \hat{\beta}_1 = -[X_1'X_1]_{\perp\hat{U}_2}^{-1}[X_1' \; {}_+Y_1]_{\perp\hat{U}_2} \; {}_+\hat{Y}_1 \; .$$

Substituting for $\hat{\beta}_1$ into the numerator of g**(a) [i.e., substituting into (VI.50)], we have:

$$\begin{aligned}\text{(VI.53)} \quad & {}_+\hat{Y}_1'[{}_+Y_1' \; {}_+Y_1]_{\perp\hat{U}_2} \; {}_+\hat{Y}_1 - 2 \; {}_+\hat{Y}_1'[{}_+Y_1'X_1]_{\perp\hat{U}_2}[X_1'X_1]_{\perp\hat{U}_2}^{-1}[X_1' \; {}_+Y_1]_{\perp\hat{U}_2} \; {}_+\hat{Y}_1 \\ & + {}_+\hat{Y}_1'[{}_+Y_1'X_1]_{\perp\hat{U}_2}[X_1'X_1]_{\perp\hat{U}_2}^{-1}[X_1'X_1]_{\perp\hat{U}_2}[X_1'X_1]_{\perp\hat{U}_2}^{-1}[X_1' \; {}_+Y_1]_{\perp\hat{U}_2} \; {}_+\hat{Y}_1 = \\ & {}_+\hat{Y}_1'[{}_+Y_1' \; {}_+Y_1]_{\perp\hat{U}_2} \; {}_+\hat{Y}_1 - {}_+\hat{Y}_1'[{}_+Y_1'X_1]_{\perp\hat{U}_2}[X_1'X_1]_{\perp\hat{U}_2}^{-1}[X_1' \; {}_+Y_1]_{\perp\hat{U}_2} \; {}_+\hat{Y}_1 \\ & = {}_+\hat{Y}_1'[{}_+Y_1' \; {}_+Y_1]_{\perp[X_1 : \hat{U}_2]} \; {}_+\hat{Y}_1\end{aligned}$$

where $[{}_+Y_1' \; {}_+Y_1]_{\perp[X_1 : \hat{U}_2]}$ denotes the moment matrix of the part of ${}_+Y_1$ orthogonal to both X_1 and $\hat{U}_2$.[2]

---

[1] $\frac{\partial^2 g^{**}(a)}{\partial \hat{\beta}_1} = \frac{2}{d}[X_1'X_1]_{\perp\hat{U}_2}$, a positive definite matrix if $\det([X_1'X_1]_{\perp\hat{U}_2}) \neq 0$ and $d \neq 0$ (since d cannot be negative); hence, under these conditions the second order condition for the value of $\hat{\beta}_1$ given by (VI.52) to minimize (VI.49) is met.

[2] That the last equality in (VI.53) holds may be seen by writing out $[{}_+Y_1]_{\perp[X_1 : \hat{U}_2]}$ as follows:

$$[{}_+Y_1]_{\perp[X_1 : \hat{U}_2]} = [{}_+Y_1]_{\perp\{\hat{U}_2 : [X_1]_{\perp\hat{U}_2}\}} = \{[{}_+Y_1]_{\perp\hat{U}_2}\}_{\cdot\{[X_1]_{\perp\hat{U}_2}\}} \quad \text{[see (I.65)]}$$

$$= [{}_+Y_1]_{\perp\hat{U}_2} - [X_1]_{\perp\hat{U}_2}([X_1]_{\perp\hat{U}_2}'[X_1]_{\perp\hat{U}_2})^{-1}[X_1]_{\perp\hat{U}_2}'[{}_+Y_1]_{\perp\hat{U}_2} \; .$$

Hence, $[{}_+Y_1' \; {}_+Y_1]_{\perp[X_1 : \hat{U}_2]} = {}_+Y_1'[{}_+Y_1]_{\perp[X_1 : \hat{U}_2]}$ [see (I.56)]

$$= {}_+Y_1'[{}_+Y_1]_{\perp\hat{U}_2} - {}_+Y_1'[X_1]_{\perp\hat{U}_2}[X_1'X_1]_{\perp\hat{U}_2}^{-1}[X_1' \; {}_+Y_1]_{\perp\hat{U}_2}$$

$$= [{}_+Y_1' \; {}_+Y_1]_{\perp\hat{U}_2} - [{}_+Y_1'X_1]_{\perp\hat{U}_2}[X_1'X_1]_{\perp\hat{U}_2}^{-1}[X_1' \; {}_+Y_1]_{\perp\hat{U}_2} \; .$$

Thus, (VI.49) becomes:

$$\text{(VI.54)} \quad g^{**}(a)\Big|_{\min \hat{\beta}_1} = \frac{{}_{+}\hat{Y}_1'[{}_{+}Y_1' \; {}_{+}Y_1]\perp[X_1 \vdots \hat{U}_2] \; {}_{+}\hat{Y}_1}{{}_{+}\hat{Y}_1'[{}_{+}Y_1' \; {}_{+}Y_1]\perp[X \vdots \hat{U}_2] \; {}_{+}\hat{Y}_1} \; .$$

Comparison of (VI.54) with the alternative formulations of the LIML problem (section II.C.1) shows that the ${}_{+}\hat{Y}_1$ which minimizes $g^{**}(a)\Big|_{\min \hat{\beta}_1}$ is the eigenvector, d_i, corresponding to the smallest eigenvalue, c_i, of $A^{-1}B$ with $A = [{}_{+}Y_1' \; {}_{+}Y_1]\perp[X \vdots \hat{U}_2]$ and $B = [{}_{+}Y_1' \; {}_{+}Y_1]\perp[X \vdots \hat{U}_2]$. Further comparison with section II.C.1 shows that the eigenvalue is the value k_{LIML} in a LIML problem with ${}_{+}Y_1$ being the matrix of jointly dependent variables in the equation, $[X_1 \vdots \hat{U}_2]$ being the matrix of predetermined variables in the equation and $[X \vdots \hat{U}_2]$ being the matrix of instruments, X_I. The ${}_{+}\hat{Y}$ which minimizes $g^{**}(a)\Big|_{\min \hat{\beta}_1}$ may be calculated either as (1) the eigenvector, d_i, corresponding to the minimum eigenvalue, k_{LIML}, or (2) by substituting k_{LIML} into the usual k-class formula; thereby calculating $\hat{Y}_1$, $\hat{\beta}_1$, and $M_1 - 1$ additional coefficients corresponding to the variables in $\hat{U}_2$. (The $M_1 - 1$ coefficients corresponding to the variables in $\hat{U}_2$ are then ignored in further calculations.)

In summary, what we have shown is that given a set of coefficients for all equations except equation 1, the coefficients of equation 1 which will maximize the likelihood function is the LIML solution of equation 1 modified by including the residuals of equations 2 through M_1 as additional predetermined variables in equation 1.

If the coefficients of equations 2 through M_1 are assumed fixed, the coefficients of equation 1 ($\hat{Y}_1$ and $\hat{\beta}_1$, only--the coefficients corresponding to $\hat{U}_2$ are ignored) estimated in this manner

will increase the likelihood function over the value of the likelihood function obtained by the original LIML coefficients for equation 1 (assuming that the original LIML coefficients did not already maximize the likelihood function considering the coefficients of equations 2 through M_1 as fixed).

Now, let us define the calculation of new coefficients for equation 1 as step 1 of an iteration and calculate the residuals for that equation (ignoring the coefficients corresponding to the residuals for equations 2 through M_1 in the calculation of the residuals for equation 1). Let us define the calculation of new coefficients for equation 2 as step 2 of an iteration and calculate these coefficients as the LIML solution obtained if the new residuals from equation 1 and the residuals from equations 3 through M_1 are included as additional predetermined variables in equation 2. The new coefficients for equation 2 will increase the likelihood function over the value of the likelihood function at the end of step 1.

If the procedure is continued through step M_1 of the iteration, the new coefficients for equation M_1 will, in general, increase the likelihood function over the previous coefficients for equation M_1. Let us define a new iteration as starting with the calculation of new coefficients for equation 1, using the new residuals obtained from steps 2 through M_1 of equations 2 through M_1. Clearly the new coefficients obtained will give a likelihood value greater than or equal to the likelihood which was obtained at step 1 of the previous iteration, since the likelihood will have been increased (or at least not decreased) at each step of the previous iteration. The likelihood at any one step of an iteration will not be strictly higher than the

likelihood at the same step of the previous iteration only if the coefficients for all equations maximize the likelihood assuming that the coefficients of the remaining equations are held constant.

Summary of the ILIML Computational Method

(1) LIML may be used to estimate starting coefficients and residuals separately for each of the M_1 equations.[1]

(2) Step 1 of an iteration consists of estimating new LIML coefficients for equation 1 by using the residuals from equations 2 through M_1 as additional predetermined variables in the equation. A new vector of residuals is estimated for equation 1 from the new coefficients of equation 1, the coefficients corresponding to the residuals of equations 2 through M_1 being ignored in calculation of the vector of residuals for equation 1. The new vector of residuals replaces the old vector of residuals in steps 2 through M_1 (i.e., in the calculation of new coefficients for equations 2 through M_1).

Step 2 of an iteration consists of estimating new LIML coefficients for equation 2 by using the residuals from equation 1 and 3 through M_1 as additional predetermined variables in the equation. A new vector of residuals is estimated for equation 2 and this vector represents equation 2 in the calculation of new coefficients for the other equations until step 2 of the following iteration.

New coefficients and new residuals are estimated for each equation in turn until new coefficients have been calculated for the M^{th} equation.

---

[1] As with FIML, other starting estimates such as DLS or 2SLS estimates could be used.

(3) A new iteration is calculated in the same manner using the residuals from the preceding iteration in the calculation of new coefficients for each equation in turn.

(4) Iteration continues until all coefficients converge. The convergence criteria given for FIML in section V.C.5 are applicable for this computational method of calculating FIML and SML coefficients as well.

Increasing Computational Efficiency of ILIML

The iterative procedure may be made considerably more efficient computer time-wise and also more accurate rounding error-wise if the residuals are not explicitly calculated. (A large number of multiplications and additions are required to calculate the residuals and form the sums of cross-products of the residuals with the other variables in each equation.) All moment matrices needed for the calculations may be calculated directly from the vector of coefficients and the moment matrix of all variables in the M_1 equations through use of a number of alternative formulas with the optimal formulas to use depending on the amount of special programming which the user of the ILIML method is willing to do. Following are some relationships which may be worked into a procedure:

From (VI.54), we note that the k value for estimating the coefficients may be derived as the smallest eigenvalue of

$$[{}_+Y_\mu' \; {}_+Y_\mu]^{-1}[X:\hat{U}_{-\mu}][{}_+Y_\mu' \; {}_+Y_\mu][X_\mu:\hat{U}_{-\mu}]$$

where $\hat{U}_{-\mu}$ denotes the residuals for all equations except the μ^{th} equation. If desired, the eigenvector corresponding to the smallest root may be used as ${}_+\hat{\gamma}$ and $\hat{\beta}$ may be calculated by formula (VI.52).

$[{}_+Y'_\mu \ {}_+Y_\mu]_\perp[X \vdots \hat{U}_{-\mu}]$ and $[{}_+Y'_\mu \ {}_+Y_\mu]_\perp[X_\mu \vdots \hat{U}_{-\mu}]$ may be calculated by direct orthogonalization using the orthogonalization procedure given in appendix A.

Let Z be the matrix of all of the jointly dependent variables in the first M_1 equations and all of the predetermined variables in the system. Then $Z'\hat{u}_\mu = -[Z' \ {}_+Z_\mu]_+\hat{\delta}_\mu$ and $\hat{u}'_\mu\hat{u}_{\mu'} = {}_+\hat{\delta}'_\mu[{}_+Z'_\mu \ {}_+Z_{\mu'}]_+\hat{\delta}_{\mu'}$.

Remember that one variable in the $\hat{U}$ and $\hat{U}_{-\mu}$ matrices changes each time a new set of coefficients are calculated for an equation, and that the $\hat{U}_{-\mu}$ matrix for equation 2 differs from the $\hat{U}_{-\mu}$ matrix for equation 1 in that it includes residuals from equation 2 instead of residuals from equation 1.

Since any equation with $\text{rk } X = n_\mu$ does not affect the final FIML or SML coefficients of the remaining equations of a system of equations, some additional computational efficiency may be obtained by omitting all equations with $\text{rk } X = n_\mu$ from the iterative procedure until convergence has been obtained (maximum likelihood estimates have been obtained) for all equations with $\text{rk } X > n_\mu$.[1] The maximum likelihood coefficients for an equation with $\text{rk } X = n_\mu$ may then be directly calculated as a LIML problem which contains as the only "extra" predetermined variables in that equation the residuals (calculated from

---

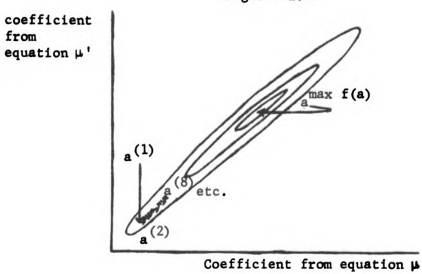
[1] An equation with $\text{rk } X = n_\mu$ is usually referred to as a just-identified equation and an equation with $\text{rk } X > n_\mu$ is usually referred to as an over-identified equation. (See section II.D.)

the converged coefficients) of the equations for which $rk\ X > n_\mu$.[1]

Comparison With FIML and SML Methods

At each step of each iteration in the ILIML method, the coefficients of only one equation are modified; therefore, it would seem likely that situations such as illustrated in Figure VI.55 could arise.

Figure VI.55



---

[1] A variation of this technique may also be used in the SML procedure. Convergence on an SML routine may be first obtained for those equations with $rk\ X > n_\mu$ and the system then enlarged to contain also the equations with $rk\ X = n_\mu$.

The two procedures may also be mixed, SML being used to obtain maximum likelihood coefficients for the equations for which $rk\ X > n_\mu$. The maximum likelihood coefficients for an equation with $rk\ X = n_\mu$ may then be calculated as a LIML problem which contains as the only "extra" predetermined variables in that equation residuals (calculated from the converged coefficients) of the equations for which $rk\ X > n_\mu$.

Since the coefficients from one equation are selected to maximize the likelihood function given the coefficients of the other equations, movement at a single step is only in the space of the coefficients of a single equation. Thus, (as in Figure VI.55) very little change in the coefficients may result at any one step if movement is up to a ridge which is not parallel to the space of coefficients of one of the equations. It is even quite conceivable that so little change would be made in the coefficients of any equation that it will be thought that the coefficients have converged while actually a considerable distance from the maximum of the likelihood function. On the other hand for the situation in Figure VI.55, movement to the maximum would be very rapid for the FIML and SML computational procedures given earlier, since all coefficients are simultaneously adjusted during any one step of the procedure.

The question of whether multiple local maxima occur in critical regions in the likelihood function for systems of equations estimated by FIML and SML is still an open question. It should be noted that the ILIML method is as likely as the FIML or SML method to move to a local maximum rather than some global maximum (assuming that it moves to a local maximum at all). However, added to this possibility is a considerably more serious possibility--that convergence may be to a point on a ridge. This is much more serious since (1) ridges which could cause difficulty are surely more likely to occur, and (2) convergence will be to a non-unique point even within the region. (E.g., if estimates result in movement to a point b on the ridge just above a stable point a on the ridge, movement will then be not toward a. Instead b will be considered a point of convergence or movement

farther up the ridge will occur.)

In general, the larger the number of equations and the larger the total number of coefficients in the entire system being estimated, the slower ILIML convergence is likely to be, since in these cases the coefficients of any equation are likely to span less of the total coefficients space. It is conceivable that convergence may be actually faster [computer time-wise] for the ILIML method than for the FIML and SML methods in very small models due to the simple computations performed; however, if such cases do occur, any computer time saved through use of the ILIML method would surely be trivial, since FIML and SML will converge rapidly in these cases, also. On the other hand for the large majority of problems to be calculated, the use of the FIML or SML method may save a very large amount of computer time over the use of the ILIML method.

In only 11 iterations a coefficient convergence criterion of .000 000 000 001 was satisfied when Klein's model I was calculated by the FIML procedure; however, when Klein's model I was calculated by the ILIML procedure, a coefficient convergence criterion of only .00001 was satisfied after 327 iterations. Convergence was occurring very slowly at this point with approximately 100 iterations being required per extra digit of convergence.[1] (The FIML and ILIML coefficients agreed up to the accuracy calculated by the ILIML procedure.)

---

[1]Klein's model I which contains only three stochastic equations is given in section I.C.1, and the FIML solution to Klein's model I is given in the reproduced computer output of section IX.K. The coefficient convergence criterion is defined in section V.C.5.

The primary advantages of the ILIML method over the FIML or SML method would appear to be:

(1) Ease of programming. (However, in general, the more advantage taken of programming relationships to reduce computer time in computation by the ILIML method, the less advantageous the ILIML method is from this standpoint.)

(2) With a given sized computer memory, it may be possible to calculate much larger problems by means of the ILIML method. (However, in general, the larger the problem being calculated, the less favorably the ILIML method may be expected to perform both with respect to speed of convergence and whether convergence becomes so slow that the problem appears converged when it is short of the maximum of the likelihood function.)

On balance, the ILIML method would not appear to warrant an investment in programming. Instead it seems desirable that FIML and SML be programmed by the formulas given previously.

CHAPTER VII

ZELLNER-AITKEN ESTIMATOR (ZA)

A. Only Zero and Normalization Restriction Imposed on Coefficients

The Zellner-Aitken estimator (ZA) is a multiple equations method which may be applied when each structural equation contains only a single jointly dependent variable.[1] If all of the equations in the system are of this form and ZA is applied to the complete system, then ZA is a complete system method.[2] ZA may also be applied to only part of the system, in which case ZA may then be regarded as a partial system method.

The ZA estimating equations can be derived as an application of GLS (in which an estimated disturbance variance-covariance matrix is substituted for the actual disturbance variance-covariance matrix) as follows:

Assuming that there are M equations in the system, the μ^{th} equation being of the form:

$$y_\mu = X_\mu \beta_\mu + u_\mu \quad , \tag{VII.1}$$

then the entire system can be written as:

$$\begin{aligned} y_1 &= X_1\beta_1 + \cdots + 0\beta_M + u_1 \\ &\vdots \\ y_M &= 0\beta_1 + \cdots + X_M\beta_M + u_M \end{aligned} \tag{VII.2}$$

---

[1] Zellner proposed the "efficient estimating procedure" which we are calling the Zellner-Aitken estimator in the article Zellner [1962].

[2] If the complete system contains only one jointly dependent variable per equation, then the structural equations and reduced form equations coincide. These reduced form equations will, of course, not be unrestricted but will incorporate all of the a priori restrictions of the structural equations.

or if we define the following matrices and vectors:

$$\dot{y} = \underset{MT\times 1}{\begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix}} \qquad \dot{X} = \underset{MT\times n}{\begin{bmatrix} X_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & X_M \end{bmatrix}} \qquad \beta = \underset{n\times 1}{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_M \end{bmatrix}} \qquad u = \underset{MT\times 1}{\begin{bmatrix} u_1 \\ \vdots \\ u_M \end{bmatrix}}$$

where $n = \sum_{\mu=1}^{M} n_\mu$ (n_μ is the number of explanatory variables in the μ^{th} equation), we can write the entire system as:

$$\dot{y} = \dot{X}\beta + u \quad . \tag{VII.3}$$

Initially we will make the statistical assumptions given at the start of this paper (section I.C.3). One of the assumptions implies that X_μ for each equation has full column rank which in turn implies that $\dot{X}$ has full column rank. This assumption will be relaxed in a later section of this chapter when restrictions on the coefficients are permitted.

Let $U = [u_1 \cdots u_M]$ be the $T\times M$ matrix of disturbances, with column μ the disturbances for equation μ and row t the disturbances for observation t, designated $U_{[t]}$. Then the assumptions of section I.C.3 that $\mathcal{E}U = 0$, $\mathcal{E}U'_{[t]}U_{[t]} = \underset{M\times M}{\Sigma}$ for all t, and $\mathcal{E}U'_{[t]}U_{[t']} = 0$ for all $t \neq t'$ imply that:

$$\mathcal{E}u = \underset{MT\times 1}{0} \tag{VII.4}$$

and

$$\mathcal{E}u_\mu u'_{\mu'} = \sigma_{\mu\mu'} I \tag{VII.5}$$

where I is $T\times T$ and $\sigma_{\mu\mu'}$ is the element in row μ and column μ' of Σ. (VII.5) is commonly expressed for the whole covariance matrix

of u by the "Kronecker product" ⊗ as:[1]

---

[1] If A is any $p \times q$ matrix and B is any $r \times s$ matrix, the Kronecker product $A \otimes B$ is defined as the $(p \cdot r) \times (q \cdot s)$ matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1q}B \\ \vdots & & \vdots \\ a_{p1}B & \cdots & a_{pq}B \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & \cdots & a_{11}b_{1s} & \cdots & a_{1q}b_{11} & \cdots & a_{1q}b_{1s} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{11}b_{r1} & \cdots & a_{11}b_{rs} & \cdots & a_{1q}b_{r1} & \cdots & a_{1q}b_{rs} \\ & & \vdots & & & & \vdots \\ a_{p1}b_{11} & \cdots & a_{p1}b_{1s} & \cdots & a_{pq}b_{11} & \cdots & a_{pq}b_{1s} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{p1}b_{r1} & \cdots & a_{p1}b_{rs} & \cdots & a_{pq}b_{r1} & \cdots & a_{pq}b_{rs} \end{bmatrix}$$

For A and B square, symmetric, and nonsingular matrices, $A \otimes B$ is also square, symmetric, and nonsingular and $[A \otimes B]^{-1} = A^{-1} \otimes B^{-1}$. This can be seen by (1) premultiplying $A \otimes B$ by $A^{-1} \otimes B^{-1}$ and observing that the identity matrix is obtained and (2) postmultiplying $A \otimes B$ by $A^{-1} \otimes B^{-1}$ and observing that the identity matrix is obtained; hence that $A^{-1} \otimes B^{-1}$ is the inverse of $A \otimes B$. Let us assume that A is $p \times p$ and B is $q \times q$. Premultiplying $A \otimes B$ by $A^{-1} \otimes B^{-1}$ we get:

$$[A^{-1} \otimes B^{-1}][A \otimes B] = \begin{bmatrix} a^{11}B^{-1} & \cdots & a^{1p}B^{-1} \\ \vdots & & \vdots \\ a^{p1}B^{-1} & \cdots & a^{pp}B^{-1} \end{bmatrix} \begin{bmatrix} a_{11}B & \cdots & a_{1p}B \\ \vdots & & \vdots \\ a_{p1}B & \cdots & a_{pp}B \end{bmatrix} .$$

The ij^{th} block in the product is $\sum_{k=1}^{p} [(a^{ik}B^{-1})(a_{kj}B)] = \sum_{k=1}^{p} a^{jk}a_{kj}(B^{-1}B)$ $= [\sum_{k=1}^{p} a^{ik}a_{kj}] \cdot I$, where I is $T \times T$. But $\sum_{k=1}^{p} a^{ik}a_{kj} = 1$ if $i = j$ and 0 if $i \neq j$; hence the product is:

$$\begin{bmatrix} 1 \cdot I & \cdots & 0 \cdot I \\ \vdots & & \vdots \\ 0 \cdot I & \cdots & 1 \cdot I \end{bmatrix} = \begin{bmatrix} I & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & I \end{bmatrix} = I_{pq \times pq} .$$

Similarily,

$$[A \otimes B][A^{-1} \otimes B^{-1}] = \begin{bmatrix} a_{11}B & \cdots & a_{1p}B \\ \vdots & & \vdots \\ a_{p1}B & \cdots & a_{pp}B \end{bmatrix} \begin{bmatrix} a^{11}B^{-1} & \cdots & a^{1p}B^{-1} \\ \vdots & & \vdots \\ a^{p1}B^{-1} & \cdots & a^{pp}B^{-1} \end{bmatrix} = \begin{bmatrix} 1 \cdot I & \cdots & 0 \cdot I \\ \vdots & & \vdots \\ 0 \cdot I & \cdots & 1 \cdot I \end{bmatrix}$$

$$= I_{pq \times pq} .$$

Hence, $[A \otimes B]^{-1} = [A^{-1} \otimes B^{-1}]$.

$$(VII.6)\quad \delta_{\mu\mu'} = \underset{M\times M}{\Sigma} \otimes \underset{T\times T}{I} = \underset{MT\times MT}{\begin{bmatrix} \sigma_{11}I & \cdots & \sigma_{1M}I \\ \vdots & & \vdots \\ \sigma_{M1}I & \cdots & \sigma_{MM}I \end{bmatrix}} = \underset{MT\times MT}{\begin{bmatrix} \sigma_{11} & \cdots & 0 & \cdots & \sigma_{1M} & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \sigma_{11} & \cdots & 0 & \cdots & \sigma_{1M} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \sigma_{M1} & \cdots & 0 & \cdots & \sigma_{MM} & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \sigma_{M1} & \cdots & 0 & \cdots & \sigma_{MM} \end{bmatrix}}$$

We will designate $\delta_{uu'}$ as $\underset{MT\times MT}{\overset{*}{\Sigma}}$.

If we treat $\dot{y}$ as the dependent variable, $\dot{X}$ as the matrix of predetermined variables, and u as the vector of disturbances and substitute into GLS formulas (V.12) and (V.13) we get:

$$(VII.7)\quad \hat{\beta}_{GLS} = [\underset{n\times MT}{\dot{X}'} \quad \underset{MT\times MT}{\overset{*}{\Sigma}{}^{-1}} \quad \underset{MT\times n}{\dot{X}}]^{-1}[\underset{n\times MT}{\dot{X}'} \quad \underset{MT\times MT}{\overset{*}{\Sigma}{}^{-1}} \quad \underset{MT\times 1}{\dot{y}}]$$

$$(VII.8)\quad \text{asymptotic } \hat{Var}(\hat{\beta}_{GLS}) = [\underset{n\times n}{\dot{X}' \overset{*}{\Sigma}{}^{-1} \dot{X}}]^{-1} .$$

If we had assumed that X, the matrix of predetermined variables, were fixed in repeated samples, then $\hat{Var}(\hat{\beta}_{GLS})$ would be a small sample estimate rather than an asymptotic estimate. Also, $\hat{\beta}_{GLS}$ would be best linear unbiased (assuming that we knew Σ).

Due to the particular form of $\overset{*}{\Sigma}$, $\dot{X}$, and $\dot{y}$, $\hat{\beta}_{GLS}$ and asymptotic $\hat{Var}(\hat{\beta}_{GLS})$ may be written out in more detail.

$$\overset{*}{\Sigma}{}^{-1} = [\Sigma \otimes I]^{-1} = \Sigma^{-1} \otimes I = \begin{bmatrix} \sigma^{11} & \cdots & \sigma^{1M} \\ \vdots & & \vdots \\ \sigma^{M1} & \cdots & \sigma^{MM} \end{bmatrix} \otimes I = \begin{bmatrix} \sigma^{11}I & \cdots & \sigma^{1M}I \\ \vdots & & \vdots \\ \sigma^{M1}I & \cdots & \sigma^{MM}I \end{bmatrix}$$

where I is a $T\times T$ identity matrix, and $\Sigma = [\sigma_{\mu\mu'}]$ and $\Sigma^{-1} = [\sigma^{\mu\mu'}]$ are $M\times M$ matrices. Thus:

(VII.9) $\hat{\beta}_{GLS} =$

$$\left\{\begin{bmatrix} X_1' & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & X_M' \end{bmatrix}\begin{bmatrix} \sigma^{11}I & \cdots & \sigma^{1M}I \\ \vdots & & \vdots \\ \sigma^{M1}I & \cdots & \sigma^{MM}I \end{bmatrix}\begin{bmatrix} X_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & X_M \end{bmatrix}\right\}^{-1} \times$$

$$\times \left\{\begin{bmatrix} X_1' & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & X_M' \end{bmatrix}\begin{bmatrix} \sigma^{11}I & \cdots & \sigma^{1M}I \\ \vdots & & \vdots \\ \sigma^{M1}I & \cdots & \sigma^{MM}I \end{bmatrix}\begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix}\right\}$$

$$= \begin{bmatrix} \sigma^{11}X_1'X_1 & \cdots & \sigma^{1M}X_1'X_M \\ \vdots & & \vdots \\ \sigma^{M1}X_M'X_1 & \cdots & \sigma^{MM}X_M'X_M \end{bmatrix}^{-1}\begin{bmatrix} \sum_{\mu=1}^{M}\sigma^{1\mu}X_1'y_\mu \\ \vdots \\ \sum_{\mu=1}^{M}\sigma^{M\mu}X_M'y_\mu \end{bmatrix}$$

and

$$\text{(VII.10)}\quad \text{asymptotic } \hat{\text{Var}}(\hat{\beta}_{GLS}) = \begin{bmatrix} \sigma^{11}X_1'X_1 & \cdots & \sigma^{1M}X_1'X_M \\ \vdots & & \vdots \\ \sigma^{M1}X_M'X_1 & \cdots & \sigma^{MM}X_M'X_M \end{bmatrix}^{-1} .$$

Since Σ is unknown, the following estimate of Σ is substituted into the formulas:[1]

---

[1]The S matrix will be singular if the number of equations exceeds the number of observations. Let $\underset{T\times M}{\hat{U}} = [\hat{u}_1 \cdots \hat{u}_M]$ where the $\hat{u}_\mu$ are DLS estimates. Then $\underset{M\times M}{S} = (1/T)\hat{U}'\hat{U}$; hence, rk S = rk $\hat{U}$. If $T < M$ then rk $\hat{U}$ and therefore rk S will be less than M, i.e., S will be singular. This is not in conflict with our derivation in section (VII.B) of DLS as the special case of ZA in which there is zero correlation between residuals across equations, since that derivation holds only for nonsingular S. Even though Σ is assumed nonsingular, a particular estimate of Σ may still be singular (and as shown above if $M > T$ then $\hat{\Sigma} = S$ is singular).

(VII.11) $$\hat{\Sigma} \overset{def}{=} S = [s_{\mu\mu'}]$$

with $s_{\mu\mu'} = (1/T)\hat{u}_{\mu}^{DLS\,\prime}\hat{u}_{\mu'}^{DLS} = (1/T)\,{}_{+}\hat{\beta}_{\mu}^{DLS\,\prime}[{}_{+}Z_{\mu}'\,{}_{+}Z_{\mu'}]\,{}_{+}\hat{\beta}_{\mu'}^{DLS}$, ${}_{+}\hat{\beta}_{\mu}^{DLS}$ being the vector of DLS coefficients (including the normalizing coefficient, -1) for equation μ and ${}_{+}Z_{\mu} = [y_{\mu} \vdots Z_{\mu}]$.[1]

Substituting S into (VII.9) and (VII.10) we get:

(VII.12) $$\hat{\beta}_{ZA} = [\varphi^{(1)}]^{-1}p^{(1)}$$

as the ZA estimate of the vector of coefficients of the M equations where:

---

[1] T is being used rather than some other divisor so that IZA (iterative ZA) coefficients can be derived as maximum likelihood coefficients further on. Two alternative divisors suggested by Professor Zellner follow:

(1) $$s^{*}_{\mu\mu'} = \frac{\hat{u}_{\mu}^{DLS\,\prime}\hat{u}_{\mu'}^{DLS}}{T - \ell_{\mu} - \ell_{\mu'} + k_{\mu\mu'}}$$

where $k_{\mu\mu'} = tr\{X_{\mu}[X_{\mu}'X_{\mu}]^{-1}X_{\mu}'X_{\mu'}[X_{\mu'}'X_{\mu'}]^{-1}X_{\mu'}'\}$

$$= tr\{[X_{\mu}'X_{\mu}]^{-1}[X_{\mu}'X_{\mu'}][X_{\mu'}'X_{\mu'}]^{-1}[X_{\mu'}'X_{\mu}]\} = \rho_{H}^{2}\cdot\min(\ell_{\mu},\ell_{\mu'})$$

where ρ_{H} is Hooper's trace correlation coefficient which measures the correlation between sets of variables X_{μ} and $X_{\mu'}$. As indicated in Stroud and Zellner [1962], S* estimated by this formula provides an unbiased estimate of Σ if X is assumed to be fixed in repeated samples. Although unbiased, S* would likely have more variance about Σ due to the evaluation of the trace terms.

In a discussion with the writer, Professor Zellner suggested that he did not favor the use of formula (1), since $S^{*} = [s^{*}_{\mu\mu'}]$ is not positive definite. Rather, he favored dividing by T or using the following estimate which is not unbiased, but provides a positive definite S matrix:

(2) $$s^{**}_{\mu\mu'} = \frac{\hat{u}_{\mu}^{DLS\,\prime}\hat{u}_{\mu'}^{DLS}}{\sqrt{T - \ell_{\mu}}\cdot\sqrt{T - \ell_{\mu'}}} \quad .$$

$$\varphi^{(1)} = \begin{bmatrix} s^{11}X_1'X_1 & \cdots & s^{1M}X_1'X_M \\ \vdots & & \vdots \\ s^{M1}X_M'X_1 & \cdots & s^{MM}X_M'X_M \end{bmatrix}$$

$$p^{(1)} = \begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu}X_1'y_\mu \\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu}X_M'y_\mu \end{bmatrix} = \begin{bmatrix} X_1'[y_1 \cdots y_M] \begin{bmatrix} s^{11} \\ \vdots \\ s^{M1} \end{bmatrix} \\ \vdots \\ X_M'[y_1 \cdots y_M] \begin{bmatrix} s^{M1} \\ \vdots \\ s^{MM} \end{bmatrix} \end{bmatrix}$$

where $S^{-1} = [s^{\mu\mu'}]$. Also,

(VII.13) $$\text{asymptotic } \hat{\text{Var}}(\hat{\beta}_{ZA}) = (\varphi^{(1)})^{-1} .$$

Even if X is assumed to be fixed for repeated samples, $\hat{\beta}_{ZA}$ is no longer best linear unbiased and $(\varphi^{(1)})^{-1}$ provides only as asymptotic estimate of $\text{Var}(\hat{\beta}_{ZA})$ due to the use of an estimate of Σ in the GLS formulas.[1]

A "degrees of freedom" adjustment can be made in the estimated ZA coefficient variance-covariance matrix [i.e., asymptotic $\hat{\text{Var}}(\hat{\delta}_{ZA})$] in the same manner as for FIML (section V.E). The estimated ZA coefficient variance-covariance matrix can also be normalized in the manner suggested for FIML.

If an estimate of the disturbance variance-covariance matrix is desired, it seems desirable to utilize the estimated ZA coefficients to calculate a new Σ instead of merely using the S matrix calculated from the DLS coefficients. Utilizing the ZA coefficients the estimated

---

[1] See Zellner [1962] for a proof of (VII.13).

disturbance variance-covariance matrix becomes:

$$\hat{\Sigma}_{ZA} = [s^{ZA}_{\mu\mu'}] \tag{VII.14}$$

with $s^{ZA}_{\mu\mu'} = (1/T)\hat{u}^{ZA'}_{\mu}\hat{u}^{ZA}_{\mu'} = (1/T)\,{}_{+}\hat{\beta}^{ZA'}_{\mu}[{}_{+}Z'_{\mu}\ {}_{+}Z_{\mu'}]\,{}_{+}\hat{\beta}^{ZA}_{\mu'}$, ${}_{+}\hat{\beta}^{ZA}_{\mu}$ being the vector of ZA coefficients (including the normalizing coefficient, -1) for equation μ and ${}_{+}Z_{\mu} = [y_{\mu} \vdots Z_{\mu}]$.

A "degrees of freedom" adjustment may be made in the $\hat{\Sigma}_{ZA}$ matrix and the $\hat{\Sigma}_{ZA}$ matrix normalized may be computed in the same manner as for FIML (section V.D.).

The ZA estimates and the DLS estimates coincide if either:

(1) The same predetermined variables occur in all equations,[1] or

(2) There is zero correlation between the residuals from DLS for all pairs of DLS equations, i.e., S is a diagonal matrix.[2]

---

[1] That ZA and DLS estimates coincide if the same predetermined variables occur in all equations can be seen as follows. Let $X = X_1 = X_2 = \cdots = X_M$. Then

$$\hat{\beta}_{ZA} = [S^{-1}\otimes(X'X)]^{-1}\begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu}X'y_{\mu} \\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu}X'y_{\mu} \end{bmatrix}$$

However, $[S^{-1}\otimes(X'X)]^{-1} = [S^{-1}]^{-1}\otimes(X'X)^{-1} = S\otimes(X'X)^{-1}$; hence, the vector of coefficients for the i^{th} equation becomes:

$$\hat{\beta}^{(i)}_{ZA} = [s_{i1}(X'X)^{-1} \cdots s_{iM}(X'X)^{-1}]\begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu}X'y_{\mu} \\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu}X'y_{\mu} \end{bmatrix}$$

$$= \sum_{\mu'=1}^{M}\sum_{\mu=1}^{M} s_{i\mu'}s^{\mu'\mu}[X'X]^{-1}X'y_{\mu} = \sum_{\mu=1}^{M}\{[X'X]^{-1}X'y_{\mu}\sum_{\mu'=1}^{M} s_{i\mu'}s^{\mu'\mu}\} \quad .$$

It is sometimes suggested that ZA be applied to the unrestricted reduced form equations to improve their efficiency over DLS applied to the unrestricted reduced form equations; however, there will be no improvement due to (1) above. On the other hand it is often the case that structural equation coefficient restrictions imply readily recognized restrictions on the coefficients of reduced form equations. If these restrictions are taken into account in estimating the reduced form equations, the ZA coefficients will not in general coincide with the DLS coefficients obtained from the reduced form equations, even if the restrictions are taken into account in the DLS estimates. In section VII.C a computational method for taking into account general linear restrictions on coefficients in ZA estimation will be presented. This computational method may prove helpful in direct estimation of reduced form coefficients, since structural equation restrictions may imply restrictions on the reduced form coefficients of a more general form than the special case that certain reduced form coefficients are zero. The computational method given in section VII.C is sufficiently general to take account of restrictions which cut across reduced form equations.[1]

However, $\sum_{\mu=1}^{M} s_{i\mu}, s^{\mu'\mu} = [s_{i1} \cdots s_{iM}] \begin{bmatrix} s^{1\mu} \\ \vdots \\ s^{M\mu} \end{bmatrix} = \begin{cases} 1 \text{ if } i = \mu \\ 0 \text{ if } i \neq \mu \end{cases}$;

hence, $\sum_{\mu=1}^{M} \{[X'X]^{-1}X'y_{\mu} \sum_{\mu'=1}^{M} s_{i\mu}, s^{\mu'\mu}\} = [X'X]^{-1}X'y_{\mu} = \hat{\beta}_{DLS}^{(i)}$. Goldberger [1964], pp. 248 and 263 also contains a proof of (1).

[2] A proof of (2) is given in the next section (section VII.B).

[1] It should not be forgotten that FIML may also be applied directly to reduced form equations and that FIML coefficients will not coincide with DLS coefficients in the same case that ZA coefficients do not coincide with DLS coefficients. Some relationships between FIML and ZA coefficients are discussed in section VII.D.1.

B. An Alternate Computational Procedure

Since (2) above (i.e., S a diagonal matrix) suggests an alternative method for calculating $\hat{\beta}_{ZA}$, we will verify (2). In this case:

$$S^{-1} = \begin{bmatrix} s_{11} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & s_{MM} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{s_{11}} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \frac{1}{s_{MM}} \end{bmatrix}$$ and the formula for ZA becomes

$$(\mathcal{O}^{(1)})^{-1} p^{(1)} = \begin{bmatrix} \frac{1}{s_{11}} X_1'X_1 & \cdots & 0\ X_1'X_M \\ \vdots & & \vdots \\ 0\ X_M'X_1 & \cdots & \frac{1}{s_{MM}} X_M'X_M \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{s_{11}} X_1'y_1 + \cdots + 0\ X_1'y_M \\ \vdots \\ 0\ X_M'y_1 + \cdots + \frac{1}{s_{MM}} X_M'y_M \end{bmatrix}$$

$$= \begin{bmatrix} s_{11}(X_1'X_1)^{-1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & s_{MM}(X_M'X_M)^{-1} \end{bmatrix} \begin{bmatrix} \frac{1}{s_{11}} X_1'y_1 \\ \vdots \\ \frac{1}{s_{MM}} X_M'y_M \end{bmatrix}$$

$$= \begin{bmatrix} (X_1'X_1)^{-1}X_1'y_1 \\ \vdots \\ (X_M'X_M)^{-1}X_M'y_M \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1^{DLS} \\ \vdots \\ \hat{\beta}_M^{DLS} \end{bmatrix} .$$

In the above verification the diagonal elements of S cancelled out; hence, if any diagonal matrix had been used in place of S the same result would have been obtained. In particular, an identity matrix could be used for S. This gives an alternative method for calculating ZA coefficients:

(1) Start the computation by using an identity matrix in place of the S matrix. (No initial coefficients are required, since they are used only in the calculation of the S matrix.)

Calculate $\varphi^{(0)}$ and $p^{(0)}$ using the same formulas as for $\varphi^{(1)}$ and $p^{(1)}$ except that the identity matrix is used for the S matrix. Then:

$$\hat{\beta}_{DLS} = (\varphi^{(0)})^{-1} p^{(0)} \quad .$$

(2) Calculate S from $\hat{\beta}_{DLS}$ in the same way as given before and use it to calculate a new $\varphi^{(1)}$ matrix and $p^{(1)}$ vector. Then:

$$\hat{\beta}_{ZA} = (\varphi^{(1)})^{-1} p^{(1)} \qquad \text{as before.}$$

Whereas the alternate computational method given above takes slightly more computer time at the 0^{th} iteration (for a large scale computer the additional time is hardly measurable), it is simpler to program--especially if provision is made to iterate on the ZA estimates in the manner indicated further on. Another advantage of starting in this manner will be noted in the discussion of the calculation of restricted ZA coefficients (section VII.C).

A disadvantage of starting with an identity matrix in place of the S matrix is that it imposes DLS estimates as the starting estimates for the ZA procedure, whereas it may be desired that other estimates be used as starting estimates for some or all of the equations; however, there seems little tendency to use estimates other than DLS estimates as starting estimates for ZA.

C. Arbitrary Linear Restrictions Imposed on Coefficients

As was noted earlier, the ZA formulas may be derived as an application of the GLS method. In like manner, restricted ZA (RZA) estimates in which linear restrictions are imposed on the coefficients being estimated may be derived as an application of the RGLS method.

If a set of N_R restrictions given by

$$\underset{N_R \times n}{R} \quad \underset{n \times 1}{\beta} = \underset{N_R \times 1}{r} \tag{VII.15}$$

is impose on the ZA model ($y = X\beta + u$ with assumptions as stated previously), the RZA formulas are given by:[1]

$$\underset{n\times 1}{\hat{\beta}^{RZA}} = \underset{n\times(n-rk\ R)}{Q} \quad \{\underset{(n-rk\ R)\times(n-rk\ R)}{[Q'\varphi^{(1)}Q]^{-1}} \quad Q'\underset{(n-rk\ R)\times 1}{[p^{(1)} - \varphi^{(1)}q]}\} + \underset{n\times 1}{q} \tag{VII.16}$$

$$\text{asymptotic } \hat{Var}(\hat{\beta}^{RZA}) = \underset{n\times n}{Q}[Q'\varphi^{(1)}Q]^{-1}Q' \tag{VII.17}$$

(VII.16) and (VII.17 are derived from substituting the ZA matrix $\varphi^{(1)}$ and the ZA vector $p^{(1)}$ into RGLS formulas (IV.5) and (IV.8). Q and q are calculated from R and r by the computational procedure of section IV.B.1.

---

[1] Calculation of the Q matrix and q vector also gives a means of separating out rk R coefficients, $\hat{\beta}^{RZA}_{(1)}$, which may be calculated from the remaining n-rk R "unrestricted" coefficients, $\hat{\beta}^{RZA}_{(2)}$. Thus, the following pair of formulas are together equivalent to (VII.16):

$$\underset{(n-rk\times R)\times 1}{\hat{\beta}^{RZA}_{(1)}} = \underset{(n-rk\ R)\times(n-rk\ R)}{[Q'\varphi^{(1)}Q]^{-1}} \quad \underset{(n-rk\ R)\times 1}{Q'[p^{(1)} - \varphi^{(1)}q]}$$

$$\underset{nk\ R\times 1}{\hat{\beta}^{RZA}_{(2)}} = \underset{rk\ R\times(n-rk\ R)}{Q_2} \quad \underset{(n-rk\ R)\times 1}{\hat{\beta}^{RZA}_{(1)}} + \underset{rk\ R\times 1}{q_2}$$

where Q_2 and q_2 are the subparts of Q and q noted in section IV.B.

In the calculation of the $\varphi^{(1)}$ matrix and $p^{(1)}$ vector, the S matrix corresponding to the restricted DLS rather than the unrestricted DLS estimates should be used.

Imposing Restrictions which Cut Across Equations on the Coefficients Used in Calculating the S Matrix

It seems clearly desirable that restrictions which do not cut across equations be imposed on the DLS coefficients used to estimate the S matrix. It would also seem desirable to impose restrictions which do cut across equations on the DLS coefficients as well. (They will then not be DLS coefficients for separate equations but coefficients closely related to DLS) A scheme for doing so follows.

Let $\hat{\beta}_{DLSME}$ be the set of coefficients which minimize $\hat{u}'\hat{u} = (\dot{y}-\dot{X}\hat{\beta})'(\dot{y}-\dot{X}\hat{\beta})$. Then the estimator

$$\underset{n\times 1}{\hat{\beta}_{DLSME}} = \underset{n\times 1}{\hat{\beta}_{DLS}} = \underset{(\sum_{\mu=1}^{M} n_\mu)\times 1}{\begin{bmatrix} \hat{\beta}_1^{DLS} \\ \vdots \\ \hat{\beta}_M^{DLS} \end{bmatrix}}$$

is the same as that obtained when each equation is estimated separately by DLS.

Let us now change the problem to one of:

(VII.18)
$$\min_{\hat{\beta}} (\dot{y}-\dot{X}\hat{\beta})'(\dot{y}-\dot{X}\hat{\beta})$$
subject to:
$$R\hat{\beta} = r$$

Then the resulting solution will be the restricted DLS solution if no restrictions cut across equations and a solution which we will call the

RDLSME (Restricted DLS Multiple Equations)solution if one or more restrictions cut across equations.

The RDLSME solution ignores the covariances between the disturbance of separate equations as does the restricted DLS solution; however, it does take account of all restrictions on the coefficients which cut across equations whereas the restricted DLS solution does not. RDLSME coefficients may be automatically calculated and then used to calculate the S matrix used to estimate the ZA coefficients by imposing restrictions on the alternate method given earlier for the calculation of the ZA estimates; that is:

(1) Start the computation by using an identity matrix in place of the S matrix. Calculate $\varphi^{(0)}$ and $p^{(0)}$ using the same formula as for $\varphi^{(1)}$ and $p^{(1)}$ except that the identity matrix is used for the S matrix:

$$\hat{\beta}_{RDLSME} = Q[Q'\varphi^{(0)}Q]^{-1}Q'[p^{(0)} - \varphi^{(0)}q] + q \quad .$$

(2) Calculate S from $\hat{\beta}_{RDLSME}$ and use it to calculate a new $\varphi^{(1)}$ matrix and $p^{(1)}$ vector. Then calculate $\hat{\beta}_{ZA}$ by formula (VII.16) as before.[1]

In addition to being simple to program, the use of the 0^{th} iteration method given above has the advantage that unique DLS coefficients need not exist provided the restrictions which cut across equations provide sufficient restrictions on the coefficients of all equations that

---

[1] If restrictions which cut across equations are ignored in the estimation of the S matrix, the resulting RZA coefficients will, of course, not coincide with the RZA coefficients obtained through taking account of the restrictions.

unique $\hat{\beta}_{RZA}$ estimates exist.[1]

Relationship to A Common RZA Formula

In applying all of the RZA formulas given so far, the R matrix need not have full row rank and the X_μ matrix for each equation need not have full column rank. If the additional requirements that (1) the R matrix has full row rank and (2) X_μ has full column rank for each equation are imposed, substitution of $\varphi^{(1)}$ and $p^{(1)}$ into RGLS formula (IV.23) gives the following alternative RZA formula:[2]

$$\text{(VII.20)} \quad \hat{\beta}_{RZA} = \hat{\beta}_{ZA} - (\varphi^{(1)})^{-1}R'[R(\varphi^{(1)})^{-1}R']^{-1}[R\hat{\beta}_{ZA} - r] \quad .$$

As with our other formulas, restricted DLS rather than unrestricted DLS estimates should be used to estimate S when calculating $\hat{\beta}_{ZA}$;

---

[1]As an example, suppose that a system of 4 equations contains only a single jointly dependent variable y_i in each equation and that each of the 4 equations contain the same 5 predetermined variables, $x_1 \cdots x_5$, plus 6 additional predetermined variables not contained in any of the other equations. Let $\hat{\beta}_j^{(i)}$ be the coefficient corresponding to x_j in equation i and assume that $\beta_j^{(1)} = \beta_j^{(2)} = \beta_j^{(3)} = \beta_j^{(4)}$ for $j = 1,\ldots,5$.

If we assume that the data consists of only 10 observations, unique DLS coefficients do not exist since each equation contains 11 variables but there are only 10 observations. Also, rk $X_i = 10$ (the maximum rank) implies a vector of residuals of all zeros for each equation; hence, $S = \frac{1}{T}\hat{U}'\hat{U} = 0$ where 0 is a 4×4 matrix. On the other hand the restrictions on coefficients given above imposes 15 independent restrictions which cut across equations; hence, it is probable that the coefficient space is sufficiently restricted that the DLSME and RZA coefficients are unique.

[2]This formula is given in Stroud and Zellner [1962], p. 10.

otherwise, information which is assumed to hold is ignored in the calculation of the S matrix. If both unrestricted $\hat{\beta}_{ZA}$ and $\hat{\beta}_{RZA}$ are desired, they should be calculated separately.

The advantages of the formulas which use the Q matrix and q vector over the formula which used the R matrix and r vector are given in section IV.B.2 for RGLS, but are applicable to RZA as well.

D. Iterative Zellner-Aitken Estimator (IZA)

1. Only zero and normalization restrictions imposed on coefficients

If the application of the ZA method to the DLS estimates results in an increase in statistical efficiency, the question naturally arises as to whether the ZA coefficients should then be used to estimate a new S matrix, new ZA estimates calculated, etc. If the process is continued, it may be hypothesized that the coefficients will converge; that is, that the proportional change in all coefficients in any one iteration will become less than a small preassigned constant, $\epsilon > 0$. If the coefficients converge, let us call the result the IZA (iterative ZA) coefficients. It is shown in this section that if the IZA procedure converges, the IZA and FIML coefficients will coincide; that is, if to the other statistical assumptions we add normality of the disturbances, then the converged IZA coefficient estimates are maximum likelihood estimates.

The FIML Procedure

First, let us examine the FIML estimating procedure in the case of only one jointly dependent variable per equation. Since the matrix of coefficients of the jointly dependent variables, Γ, takes on a particularly simple form--the identity matrix, the likelihood function in the form $f^*(a)$ [given in (V.24)] becomes $\det^2\Gamma/\det S = \det^2 I/\det S = 1/\det S$. Thus, maximization of the likelihood function implies the minimization of $\det S$.[1]

---

[1] This may also be derived by noting that FIML estimates minimize the estimated variance-covariance matrix of the restricted reduced form, and that in the ZA system, the structural equations may also be regarded as the reduced form equations in which account is taken of all coefficient restrictions in the model.

For any equation μ, ${}_{+}Y_{\mu}$ becomes y_{μ} (Y_{μ} is empty) and Z_{μ} becomes X_{μ}. Substituting into equations (V.37) and (V.42) we get:[1]

$$\text{(VII.21)} \quad \iota_{\mu}^{(i-1)} = T\frac{\partial f(a)}{\partial a_{\mu}}\bigg|_{a=a^{(i-1)}} = \sum_{\mu'=1}^{M} s^{\mu\mu'} X_{\mu}'\hat{u}_{\mu'} = X_{\mu}'\hat{U}\begin{bmatrix} s^{1\mu} \\ \vdots \\ s^{M\mu} \end{bmatrix}$$

and

$$\text{(VII.22)} \quad \mathcal{L}_{\mu\mu'} = -T\frac{\partial^2 f(a)}{\partial a_{\mu}\partial a_{\mu'}}\bigg|_{a=a^{(i-1)}} = s^{\mu\mu'}X_{\mu}'X_{\mu'} - (1/T)X_{\mu}'\hat{U}F_{\mu\mu'}\hat{U}'X_{\mu'},$$

where $$F_{\mu\mu'} = \begin{bmatrix} s^{1\mu'} \\ \vdots \\ s^{M\mu'} \end{bmatrix}\begin{bmatrix} s^{1\mu} & \cdots & s^{M\mu} \end{bmatrix} + s^{\mu\mu'}S^{-1} \quad .$$

At the i^{th} iteration, the new coefficient vector $a^{(i)}$ is formed from the old coefficient vector $a^{(i-1)}$ as:

$$\text{(VII.23)} \quad a^{(i)} = a^{(i-1)} + h^{(i)}d^{(i)}$$

where $h^{(i)}$ is the step size and the direction, $d^{(i)}$, is formed by $d^{(i)} = |\mathcal{L}^{(i-1)}|^{-1}\iota^{(i-1)}$ with:

$$\iota^{(i-1)} = \begin{bmatrix} \iota_1^{(i-1)} \\ 1 \\ \vdots \\ \iota_M^{(i-1)} \end{bmatrix} \quad \text{and} \quad \mathcal{L}^{(i-1)} = \begin{bmatrix} \mathcal{L}_{11}^{(i-1)} & \cdots & \mathcal{L}_{1M}^{(i-1)} \\ \vdots & & \vdots \\ \mathcal{L}_{M1}^{(i-1)} & \cdots & \mathcal{L}_{MM}^{(i-1)} \end{bmatrix} \quad .$$

---

[1] Since $\hat{\Gamma} = I$, $\det^{2\hat{\Gamma}} = \det^{2}I$ which is a constant; hence, terms derived from $\det^{2\hat{\Gamma}}$ drop out of $\frac{\partial f(a)}{\partial a}$ and $\frac{\partial^2 f(a)}{\partial a^2}$. Thus, $\hat{\gamma}\{\mu|\mu\}$ is deleted from ι_{μ} in (V.37) and H is deleted from $\mathcal{L}_{\mu\mu'}$ in (V.42).

The IZA Procedure

Now let us examine the IZA iterative procedure. At the i^{th} iteration:

$$\hat{\beta}^{(i)} = (\mathcal{O}^{(i-1)})^{-1} \begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu} X_1' y_\mu \\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu} X_M' y_\mu \end{bmatrix}$$

where $$\mathcal{O}^{(i-1)} = \begin{bmatrix} s^{11} X_1'X_1 & \cdots & s^{1M} X_1'X_M \\ \vdots & & \vdots \\ s^{M1} X_M'X_1 & \cdots & s^{MM} X_M'X_M \end{bmatrix},$$

$\hat{\beta}^{(i-1)}$ being used in calculating the S matrix which is in turn used to calculate $\mathcal{O}^{(i-1)}$ and $p^{(i-1)}$ for the current $\hat{\beta}^{(i)}$.

The increment added to $\hat{\beta}^{(i-1)}$ to form $\hat{\beta}^{(i)}$ is:

(VII.24) $$\hat{\beta}^{(i)} - \hat{\beta}^{(i-1)} =$$

$$(\mathcal{O}^{(i-1)})^{-1} \begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu} X_1' y_\mu \\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu} X_M' y_\mu \end{bmatrix} - (\mathcal{O}^{(i-1)})^{-1} \mathcal{O}^{(i-1)} \hat{\beta}^{(i-1)}$$

$$= (\mathcal{O}^{(i-1)})^{-1} \left\{ \begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu} X_1' y_\mu \\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu} X_M' y_\mu \end{bmatrix} - \begin{bmatrix} s^{11} X_1'X_1 & \cdots & s^{1M} X_1'X_M \\ \vdots & & \vdots \\ s^{M1} X_M'X_1 & \cdots & s^{MM} X_M'X_M \end{bmatrix} \begin{bmatrix} \hat{\beta}_1^{(i-1)} \\ \vdots \\ \hat{\beta}_M^{(i-1)} \end{bmatrix} \right\} .$$

$$= (\mathcal{P}^{(i-1)})^{-1}\left\{\begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu} X_1' y_\mu \\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu} X_M' y_\mu \end{bmatrix} - \begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu} X_1' X_\mu \hat{\beta}_\mu^{(i-1)} \\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu} X_M' X_\mu \hat{\beta}_\mu^{(i-1)} \end{bmatrix}\right\}$$

$$= (\mathcal{P}^{(i-1)})^{-1}\begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu} X_1'(y_\mu - X_\mu \hat{\beta}_\mu^{(i-1)}) \\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu} X_M'(y_\mu - X_\mu \hat{\beta}_\mu^{(i-1)}) \end{bmatrix} = (\mathcal{P}^{(i-1)})^{-1}\begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu} X_1' \hat{u}_\mu \\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu} X_M' \hat{u}_\mu \end{bmatrix}$$

$$= (\mathcal{P}^{(i-1)})^{-1}\, \iota^{(i-1)} .$$

i.e., $\hat{\beta}^{(i)} - \hat{\beta}^{(i-1)} = 1 \cdot (\mathcal{P}^{(i-1)})^{-1}\, \iota^{(i-1)}$. Thus, for the special case of one jointly dependent variable per equation the IZA procedure is exactly the same as the FIML procedure except that: (1) a step size of 1 is used for each iteration, and (2) the $\mathcal{P}$ metric is used instead of the $|\mathcal{H}|$ metric.[1]

If the formula for the $\mathcal{P}$ metric given in this chapter is compared to the formula for the $\mathcal{P}$ metric given in the FIML chapter (V.39), they will be seen to be the same, i.e., the $\mathcal{P}$ metric of the IZA procedure is the initial metric used by Chernoff and Divinsky in their FIML computational procedure. (The $\mathcal{P}$ metric was used as the initial metric due to its "safe" characteristics. For one thing it is positive definite thereby insuring that if a sufficiently small step size is chosen, there will of necessity be an increase in the likelihood function for $\iota^{(i)} \neq 0$.)

---

[1] The ι vector and $|\mathcal{H}|$ matrix are the right hand side vector [defined by (V.37)] and metric [defined by (V.42) and (V.43)] for FIML.

Since no jointly dependent variables other than the normalizing variable occur in any equation the $\mathcal{R}$ metric given by formula (V.40) also reduces to the $\mathcal{P}$ metric. (As noted in section V.C.3, the $\mathcal{R}$ metric is a metric which was developed to be asymptotically the same as the $\mathcal{L}$ metric.) Thus, the $\mathcal{P}$ metric given here is the same as the metrics used for the bulk of the iterations by Chernoff and Divinsky. Chernoff and Divinsky only shifted to the more powerful $\mathcal{L}$ metric close to the maximum. (In addition to the advantages of the more powerful metric, $\mathcal{L}^{-1}$ provides a maximum likelihood estimate of the estimated coefficient variance-covariance matrix rather than one which can be shown to be asymptotically the same, i.e., $\mathcal{R}^{-1}$, which in the case of one jointly dependent variable per equation is the same as $\mathcal{P}^{-1}$.)

Since the likelihood function for the ZA model is of a considerably simpler form than the likelihood function to be maximized in the general FIML case, the $\mathcal{P}$ metric (which in this case coincides with the $\mathcal{R}$ metric) should be adequate in most cases for convergence. On the other hand, the $|\mathcal{L}|$ metric which we developed earlier should still prove the more powerful in terms of number of iterations required for convergence. Also convergence could surely be speeded if a variable step size were computed in a fashion such as the one given in the FIML section. If a step size of 1 is imposed, it is conceivable [though unlikely because of the particularly simple form of f(a)] that the coefficients may diverge from the maximum or cycle in some fashion since an increase in the likelihood at each iteration is only guaranteed for a step sufficiently small.

Due to the more powerful metric used in the FIML procedure, total computing time may be expected to be less if the FIML procedure of section V.C. is used rather than the IZA procedure.[1] There would be even more advantage if the FIML formulas were modified by formulas (VII.21) and (VII.22) to take advantage of the simple form of Γ. Finally (if the disturbances are assumed to be normally distributed), the FIML procedure does provide a maximum likelihood estimate of the coefficient variance-covariance matrix rather than one which is only asymptotically the same as the maximum likelihood estimate.

On the other hand the IZA procedure is easier to program. Convergence would be speeded up considerably (at only a small cost in additional programming) if step size were varied by using a scheme such as that given in the FIML section. If the IZA procedure is used, $d^{(i)}$ can be calculated by (1) calculating the new coefficients, which would be obtained at the i^{th} iteration if the IZA procedure were used [i.e., $\hat{\beta}^{(i)} = (\mathscr{P}^{(i-1)})^{-1} p^{(i-1)}$], and (2) calculating $d^{(i)}$ as

$$d^{(i)} = \hat{\beta}^{(i)} - \hat{\beta}^{(i-1)} .$$

New coefficients for the iteration could then be calculated as $\hat{\beta}^{(i-1)} + h^{(i)} d^{(i)}$, the final step size $h^{(i)}$ being based on det S for trial values of $h^{(i)}$.[2]

---

[1] It is conceivable that there exist some exceedingly simple models in which the IZA procedure requires less computer time; however, these problems will surely be encountered rarely and even in these cases the amount of computer time saved by the IZA procedure will be hardly measurable. On the other hand, for the majority of problems encountered, the FIML procedure should result in considerably faster convergence computer time-wise than the IZA procedure. Also, the FIML procedure may be expected to converge for some problems for which the IZA procedure does not converge.

[2] Since only a single jointly dependent variable occurs in each equation, det S is minimized at the maximum of the likelihood.

A "degrees of freedom" adjustment may be made in the estimated disturbance variance-covariance and the estimated coefficient variance-covariance matrices which correspond to the IZA coefficients in the same manner as for FIML (sections V.D. and V.E.). If the IZA coefficients are to be maximum likelihood estimates, the adjustment to the disturbance variance-covariance matrix is made only after the coefficients have converged, i.e., a diviser of T is used during iteration.

2. Arbitrary linear restrictions imposed on coefficients

Restrictions may be imposed each iteration of the computation of the IZA coefficients in the same manner as for ZA. When convergence has been obtained, the resulting coefficients will satisfy the restrictions and will be the same as would be obtained if the restrictions were used in the FIML computational procedure.

E. Iterative Direct Least Squares (IDLS or Telser Method)

For the particular case of a system of equations in which each equation contains only one jointly dependent variable, Professor Lester Telser proposed a multiple equations computational procedure in which DLS is used as the primary computational procedure but in an iterative fashion.[1] In each step of the IDLS (iterative DLS) procedure the coefficients of only a single equation are calculated by DLS except that the residuals from all of the other stochastic equations in the system are included as extra explanatory variables; that is, the explanatory variables for the DLS calculation are taken to be the predetermined variables for the equation being estimated and the residuals from all of the other stochastic equations in the system. Only the coefficients of the predetermined variables of the equation are used. The coefficients corresponding to the residuals are ignored.

The IDLS procedure may be considered to be a special case of the ILIML procedure given in section VI.E, since DLS may be considered to be the particular case of LIML in which only one jointly dependent variable occurs in each equation. Thus, the computational procedure summarized on pages 256 and 257 for ILIML is the same as the IDLS computational procedure with LIML substituted for DLS as the basic computational method.

Derivation of the IDLS Method[2]

Let us add normality of the disturbances to the statistical assumptions previously made. Then, since the IDLS method is a particular

---

[1] Telser [1964].

[2] Telser [1964] uses a different approach in his derivation of the IDLS method.

case of the ILIML method, the derivation of the ILIML method is sufficient to show that at any one step in the procedure, the coefficients of an equation are selected to maximize the likelihood function, provided we consider the coefficients of the remaining equations in the system as fixed during that step. (This implies that if the coefficients of all other equations are maximum likelihood coefficients, then the single step will result in maximum likelihood coefficients being estimated for that equation as well.) Thus, it may be expected that, in general, if enough steps are taken (enough steps may be a very large number of steps in some cases) convergence will be to the maximum of the likelihood function. Particular cases similar to that posed following Figure VI.55 of section VI.E may arise where movement to a point short of the maximum may occur and from that point no further movement occurs or movement is so slow that it is thought that convergence has occurred.

Since the restriction to one jointly dependent variable per equation considerably simplifies the likelihood function, the remainder of this section will be devoted to giving a simpler derivation of the IDLS method than the one given for ILIML (however, the basic steps in the derivation are the same as for ILIML).

Let us separate out the first equation from the remainder of the system and select coefficients for that equation which maximize the likelihood function assuming that the coefficients of the other equations are fixed. Thus, we will divide the $M^{X}(M + \Lambda)$ matrix of coefficients α into two parts--α_1, the $1^{X}(M + \Lambda)$ matrix of coefficients of the first equation, and α_2, the $(M - 1)^{X}(M + \Lambda)$ matrix of coefficients of the remaining $M - 1$ equations:

$$\text{(VII.26)} \qquad \alpha = \begin{bmatrix} \underset{1\times(M+\Lambda)}{\alpha_1} \\ \\ \underset{(M-1)\times(M+\Lambda)}{\alpha_2} \end{bmatrix}$$

Let S be the estimated disturbance variance-covariance matrix corresponding to a set of estimated coefficients $\hat{\alpha}$ as before. Then, since $\hat{U} = -Z\hat{\alpha}'$,

$$\text{(VII.27)} \qquad S = (1/T)\hat{U}'\hat{U} = (1/T)\hat{\alpha}Z'Z\hat{\alpha}' \quad .$$

Selecting coefficients to maximize the likelihood function is equivalent to selecting coefficients to minimize det S (see section VII.D.1).

In the same manner as for ILIML (section VI.E), we may factor det S into:

$$\text{(VII.28)} \quad \det S = \frac{1}{T}\det(\hat{\alpha}Z'Z\hat{\alpha}') = \frac{1}{T}\alpha_1[Z'Z]_{\perp\hat{U}_2}\alpha_1' \cdot \det(\alpha_2[Z'Z]\alpha_2')$$

where:

$U_2 = -Z\alpha_2$ is the $T\times(M-1)$ matrix of residuals of equations 2 through M and $[Z'Z]_{\perp\hat{U}_2}$ is the moment matrix of the part of Z orthogonal to $\hat{U}_2$.

Let us subdivide Z as $Z = [y_1 \vdots X_1 \vdots Z_1^{**}]$ where y_1 is the $T\times 1$ vector of observed values of the jointly dependent variable of equation 1, X_1 is the $T\times n_1$ matrix of predetermined variables in equation 1, and Z_1^{**} is the $T\times(M+\Lambda-n_1-1)$ matrix of variables outside equation 1 but in the system. Then $\hat{\alpha}_1$ may be correspondingly subdivided as $\hat{\alpha}_1 = [-1 \vdots \hat{\beta}_1' \vdots 0']$ where $\hat{\beta}_1$ is the $n_1\times 1$ vector of coefficients of predetermined variables in the first equation, and 0 is the vector of coefficients of the variables outside equation 1. $\alpha_1[Z'Z]_{\perp\hat{U}_2}\alpha_1'$ may now be written as:

$$(VII.29)\quad [-1 \vdots \hat{\beta}_1' \vdots 0'] \begin{bmatrix} y_1'y_1 & y_1'X_1 & y_1'Z_1^{**} \\ X_1'y_1 & X_1'X_1 & X_1'Z_1^{**} \\ Z_1^{**\prime}y_1 & Z_1^{**\prime}X_1 & Z_1^{**\prime}Z_1^{**} \end{bmatrix}_{\perp \hat{U}_2} \begin{bmatrix} -1 \\ \hat{\beta}_1 \\ 0 \end{bmatrix} =$$

$$y_1'y_1 - 2[y_1'X_1]_{\perp \hat{U}_2}\hat{\beta}_1 + \hat{\beta}_1'[X_1'X_1]_{\perp \hat{U}_2}\hat{\beta}_1 .$$

Substituting (VII.29) into (VII.28), taking the partial derivative of det S with respect to $\hat{\beta}_1$ and setting the partial derivative to zero we have:

$$(VII.30)\quad \left(\frac{\partial \det S}{\partial \hat{\beta}_1}\right)' = \frac{1}{T}\det(\alpha_2[Z'Z]\alpha_2')(-2[X_1'y_1]_{\perp \hat{U}_2} + 2[X_1'X_1]_{\perp \hat{U}_2}\hat{\beta}_1) = 0$$

or solving for the minimizing value of $\hat{\beta}_1$ we have:[1]

$$(VII.31)\quad \hat{\beta}_1 = [X_1'X_1]^{-1}_{\perp \hat{U}_2}[X_1'y_1]_{\perp \hat{U}_2} .$$

But the solution given above is the least squares solution of the coefficients corresponding to X_1 which would be obtained if y_1 were used as the dependent variable and the variables in the matrix $[X_1 \vdots \hat{U}_2]$ were used as predetermined variables in the equation. This may be seen as follows:

Let $\hat{\Phi}_1$ be the $(M - 1)\times 1$ vector of least squares coefficients corresponding to the variables in $\hat{U}_2$. Then the least squares solution

---

[1] $\frac{\partial^2 \det S}{\partial \hat{\beta}_1^2} = \frac{2}{T}\det(\alpha_2[Z'Z]\alpha_2')[X_1'X_1]^{-1}_{\perp \hat{U}_2}$, a positive definite matrix provided $[X_1'X_1]^{-1}_{\perp \hat{U}_2}$ exists and $\det(\alpha_2[Z'Z]\alpha_2') \neq 0$; therefore, the second order condition for $\hat{\beta}_1$ to minimize det S is met.

of $[\hat{\beta}_1' \vdots \hat{\Phi}_1']$ is:[1]

$$(VII.32)\quad \begin{bmatrix}\hat{\beta}_1\\ \hat{\Phi}_1\end{bmatrix} = [[X_1 \vdots \hat{U}_2]'[X_1 \vdots \hat{U}_2]]^{-1}[X_1 \vdots \hat{U}_2]'y_1$$

$$= \begin{bmatrix} X_1'X_1 & X_1'\hat{U}_2 \\ \hat{U}_2'X_1 & \hat{U}_2'\hat{U}_2 \end{bmatrix}^{-1}\begin{bmatrix} X_1'y_1 \\ \hat{U}_2'y_1 \end{bmatrix}$$

$$= \begin{bmatrix} [X_1'X_1]_{\perp\hat{U}_2}^{-1} & -[X_1'X_1]_{\perp\hat{U}_2}^{-1} X_1'\hat{U}_2[\hat{U}_2'\hat{U}_2]^{-1} \\ -[\hat{U}_2'\hat{U}_2]_{\perp X_1}^{-1}\hat{U}_2'X_1[X_1'X_1]^{-1} & [\hat{U}_2'\hat{U}_2]_{\perp X_1}^{-1} \end{bmatrix}\begin{bmatrix} X_1'y_1 \\ \hat{U}_2'y_1 \end{bmatrix}$$

$$= \begin{bmatrix} [X_1'X_1]_{\perp\hat{U}_2}^{-1}X_1'y_1 - [X_1'X_1]_{\perp\hat{U}_2}^{-1}X_1'\hat{U}_2[\hat{U}_2'\hat{U}_2]^{-1}\hat{U}_2'y_1 \\ -[\hat{U}_2'\hat{U}_2]_{\perp X_1}^{-1}\hat{U}_2'X_1[X_1'X_1]^{-1}X_1'y_1 + [\hat{U}_2'\hat{U}_2]_{\perp X_1}^{-1}\hat{U}_2'y_1 \end{bmatrix}$$

$$= \begin{bmatrix} [X_1'X_1]_{\perp\hat{U}_2}^{-1}[X_1'y_1 - X_1'\hat{U}_2[\hat{U}_2'\hat{U}_2]^{-1}\hat{U}_2'y_1] \\ [\hat{U}_2'\hat{U}_2]_{\perp X_1}^{-1}[\hat{U}_2'y_1 - \hat{U}_2'X_1[X_1'X_1]^{-1}X_1'y_1] \end{bmatrix}$$

$$= \begin{bmatrix} [X_1'X_1]_{\perp\hat{U}_2}^{-1}[X_1'y_1]_{\perp\hat{U}_2} \\ [\hat{U}_2'\hat{U}_2]_{\perp X_1}^{-1}[\hat{U}_2'y_1]_{\perp X_1} \end{bmatrix} .$$

---

[1] The formula given here for the inverse of a partitioned matrix is derived in Faddeeva [1959], pp. 102-103 except that he writes $[X_1'X_1]_{\perp\hat{U}_2}$ in the form $X_1'X_1 - X_1'\hat{U}_2(\hat{U}_2'\hat{U}_2)^{-1}\hat{U}_2'X_1$. Also, to save an inversion, Faddeeva doesn't treat $X_1'X_1$ and $\hat{U}_2'\hat{U}_2$ uniformly but instead writes a different term for $[\hat{U}_2'\hat{U}_2]_{\perp X_1}$.

Thus, $\hat{\beta}_1 = [X_1'X_1]_{\perp\hat{U}_2}^{-1}[X_1'y_1]_{\perp\hat{U}_2}$ as claimed; however, [as derived in (VII.31)] this is the solution which maximizes the likelihood assuming that the coefficients of the remaining equations are fixed. (If the remaining coefficients are maximum likelihood coefficients, the solution given by (VII.32) gives maximum likelihood coefficients for this equation as well.)

New residuals may now be calculated for equation 1 using the new coefficient vector $\hat{\beta}_1$ but not including the coefficients in $\hat{\Phi}_1$, and new coefficients can be estimated for each equation in turn, using the new residuals calculated in previous steps as additional predetermined variables in the equation in the same manner as for ILIML. After estimating new coefficients for equation M, a new iteration is started by again estimating new coefficients for equation 1. Iteration continues until all coefficients converge.

Summary of the IDLS Computational Method

The summary of the ILIML computational method--pp. 256-257 of this paper--summarizes the IDLS computational method as well if DLS is used instead of LIML.

Increasing Efficiency of IDLS Computation

Except for the notes regarding eigenvalues and eigenvectors, the suggestions for increasing efficiency for the ILIML method are applicable to the IDLS method as well. In particular, the residuals need not be calculated, since:

(VII.33) $$Z'\hat{u}_\mu = -[Z'_+Z_\mu]_+\hat{\beta}_\mu$$

and

(VII.34) $$\hat{u}_\mu'\hat{u}_\mu = {}_+\hat{\beta}_\mu'[{}_+Z_\mu' {}_+Z_\mu]_+\hat{\beta}_\mu$$

where $_{+}Z_{\mu} = [y_{\mu} \vdots Z_{\mu}]$, $_{+}\hat{\beta}_{\mu} = \begin{bmatrix} -1 \\ \hat{\beta}_{\mu} \end{bmatrix}$, and Z is the matrix of variables in the system (or subsystem) being estimated.

In like manner to the ILIML method, any equation with $rk\ X = n_{\mu}$ does not affect the converged FIML, IZA, or IDLS coefficients; hence, some additional computational efficiency may be obtained by omitting all equations with $rk\ X = n_{\mu}$ from the iterative procedure until convergence has been obtained (maximum likelihood estimates have been obtained) for all equations with $rk\ X > n_{\mu}$.[1]

For the model we are now considering (only one jointly dependent variable per equation) $rk\ X = n_{\mu}$ will occur for the μ^{th} equation only if it is assumed that all of the predetermined variables occurring in any equation in the system being estimated occur also in the μ^{th} equation with non-zero coefficients. The maximum likelihood coefficients for an equation containing all of the predetermined variables in the system may then be directly calculated by including as the only "extra" predetermined variables in that equation the residuals (calculated from the converged coefficients) of all equations for which $rk\ X > n_{\mu}$.[2]

---

[1] An equation with $rk\ X = n_{\mu}$ is usually referred to as a just-identified equation and an equation with $rk\ X > n_{\mu}$ is usually referred to as an over-identified equation.

[2] This technique may also be used in the FIML and IZA procedures. Convergence may be first obtained (on a FIML or IZA routine) for those equations which do not contain all of the predetermined variables in the system and the system then enlarged to contain also the equations each of which contains all of the predetermined variables in the system.

The IDLS procedure may also be combined with the FIML or IZA procedure by using FIML or IZA to calculate the coefficients of those equations which do not contain all of the predetermined variables in the system. The coefficients of an equation which contains all of the predetermined variables in the system may then be directly calculated as a DLS problem which contains as extra predetermined variables the residuals (calculated from the converged coefficients) of the equations which do not contain all of the predetermined variables in the system.

Comparison with FIML and IZA Methods

The remarks comparing convergence of the ILIML method with convergence of the FIML and SML methods (pp. 259-262) are applicable for comparing the IDLS method with the FIML and IZA methods as well. (The IZA method [unlike the ILDS method] is similar to the FIML method in that adjustments are made to all coefficients in any step of the convergence procedure; hence, convergence will occur to at least a local maximum if convergence occurs.) Situations in which convergence will be short of the maximum of the likelihood function may occur for IDLS in the same manner as for ILIML.[1]

---

[1]Klein's model I was modified by reclassifying all explanatory jointly dependent variables in each of the 3 stochastic equations as predetermined. (The normalizing jointly dependent variable for each stochastic equation became the only jointly dependent variable in the equation, and the identity equations were deleted from the system.) The coefficients of the modified model were then estimated by FIML, IZA, and IDLS using a coefficient convergence criterion (see section V.C.5) of .000 000 001. FIML required 6 iterations to converge, IZA required 46 iterations to converge, and IDLS required 64 iterations to converge. (The estimates obtained coincided, of course, for the 3 methods.)

The Monte-Carlo experiment reported in Kmenta and Gilbert [1967] was calculated on the AES STAT system. In this experiment the FIML, IZA and IDLS estimates coincided for all samples for which all three methods were calculated. Even in the simple 2 equation models, the FIML computational procedure was much more powerful than the IZA and IDLS procedures, the FIML procedure requiring about 8 iterations for each problem and the IZA and IDLS procedures requiring about 23 iterations. (A coefficient convergence criterion of only .000 001 was used. Had a smaller convergence criterion been used, only a few additional iterations would have been required for the FIML procedure, since it is powerful close to the maximum of the likelihood. On the other hand the IDLS convergence procedure does not converge faster as the maximum is approached; hence, a number of additional iterations would have been required for IDLS.) The number of iterations required for convergence for IDLS was highly variable. In the 2 equation model, IDLS sometimes required more and sometimes fewer iterations than IZA, but never as few iterations as FIML. In a 4 equation model, there was a much greater advantage to using FIML than IZA or IDLS than for the 2 equation model. Also, the number of iterations for IDLS became much higher than for IZA. (The iteration results reported in this footnote are not reported in Kmenta and Gilbert [1967].)

CHAPTER VIII

THREE-STAGE LEAST SQUARES (3SLS)

A. Only Zero and Normalization Restrictions Imposed on Coefficients

Three-stage least squares (3SLS)[1] is usually thought of as a method for estimating the coefficients of a complete system of equations and its properties are usually compared with the properties of the FIML estimator. It should be noted, however, that the 3SLS estimation procedure may also be applied to a subsystem of equations--utilizing the structure of the subsystem being estimated plus additional instruments (usually the predetermined variables in the remainder of the system); hence, it is often more fruitful to compare the properties of the 3SLS estimator with the properties of the SML estimator. Also, as with SML estimation (and unlike FIML estimation) jointly dependent variables are adjusted by a matrix of variables contemporaneously independent of the disturbances of the equations being estimated. As a result, if the rank of the matrix of variables used in the adjustment of the jointly dependent variables equals the number of observations, the special adjustment has no effect. (This is shown in section VIII.C.)

Regarding identity equations, Zellner and Theil recommend that they simply be deleted from the three stage procedure. In a footnote, they go on to say: "It is sometimes recommended that such equations be eliminated by a substitution of variables. This is superfluous and makes the computations more complicated than necessary."[2] Predetermined variables in identity equations do serve as prime candidates as instruments for the X_I matrix which is used to adjust jointly dependent

---

[1] The basic article on 3SLS is Zellner and Theil [1962].

[2] Zellner and Theil [1962], p. 63.

variables in the two stage and three stage procedures.[1]

The 3SLS estimating equations can be derived as an application of GLS in which (1) an estimated disturbance variance-covariance matrix is substituted for the actual disturbance variance-covariance matrix, and (2) stochastic variables are included in the GLS X matrix. Following is a derivation.[2]

As in part I let the μ^{th} structural equation of the system or subsystem being estimated be

$$(VIII.1) \qquad \underset{T\times 1}{y_\mu} = \underset{T\times n_\mu}{Z_\mu}\ \underset{n_\mu\times 1}{\delta_\mu} + \underset{T\times 1}{u_\mu}$$

---

[1] The X_I matrix is defined further on in this section.

[2] Since 2SLS coefficients which are starting estimates for 3SLS may also be derived as an application of GLS, 3SLS might be said to be derivable as an application of the GLS procedure twice. This is the approach used in Zellner and Theil [1962].

The derivation in this paper follows Zellner and Theil's derivation except that:

(1) Instead of restricting the matrix of variables used to adjust the jointly dependent variables to the entire matrix of predetermined variables in the system (X), the jointly dependent variables are adjusted by a matrix of instrumental variables, X_I, with X_I containing all of the predetermined variables in the subsystem being estimated plus additional instruments. X may of course be used as the matrix X_I. A careful reading of the derivations given in Zellner and Theil [1962] will disclose that none of the properties which they claim for the 3SLS estimator will be affected by this substitution provided we make the same assumptions regarding the variables in X_I that they make regarding the variables in X--that the variables are fixed.

(2) Zellner and Theil assumed that the X matrix has full column rank. We will assume that the X_I matrix has full column rank for ease of deriving 3SLS as a GLS procedure, but present a computational procedure for which X_I may have less than full column rank.

where y_μ is the vector of observations of the normalizing jointly dependent variable in the equation, $Z_\mu = [Y_\mu \vdots X_\mu]$ is the matrix of explanatory variable in the equation (the $T \times m_\mu$ submatrix Y_μ is the matrix of explanatory jointly dependent variables in the equation and the $T \times \ell_\mu$ submatrix X_μ is the matrix of predetermined variables in the equation), $\delta_\mu = \begin{bmatrix} \gamma_\mu \\ \beta_\mu \end{bmatrix}$ is the vector of population coefficients of the explanatory variables of the μ^{th} equation (γ_μ is the $m_\mu \times 1$ subvector of the population coefficients of the explanatory jointly dependent variables and β_μ is the $\ell_\mu \times 1$ subvector of population coefficients of the predetermined variables), and u_μ is the vector of disturbances of the μ^{th} equation.

Let X_I (the subscript I denotes instruments) be a $T \times K$ matrix of instrumental variables containing the predetermined variables in the system or subsystem being estimated plus possibly additional instrumental variables. The discussion of selection of instruments in section II.G for the X_I matrix of the double k-class estimators is applicable to instrumental variables used in the X_I matrix for 3SLS as well. The predetermined variables in the remainder of the system (if 3SLS is applied to a subsystem of the entire system) and the predetermined variables in identity equations should certainly be considered as candidates for inclusion as instruments in the X_I matrix.[1] When reporting results of 3SLS estimation, the particular instruments included in the X_I matrix should be reported along with 3SLS coefficients obtained, since the particular instruments included in the X_I matrix affects the 3SLS coefficients obtained.

---

[1] It is usual to use the X matrix (the matrix of predetermined variables in the system) as X_I (the matrix of instrumental variables).

In what follows we will assume that X_I consists of "fixed" variables, only.[1]

Initially we will assume that X_I has full column rank for ease of deriving the computational formulas; however, (as is noted further on in the derivation) an X_I of less than full column rank presents no difficulty if the formulas which are presented in this paper are used. X_I must have rank at least equal to the maximum number of explanatory variables in any equation in the system or subsystem being estimated. (Even this lesser restriction will be relaxed in section VIII.D when we consider general linear restrictions on coefficients.)[2]

---

[1] Assuming that X_I contains "fixed" variables follows Zellner and Theil [1962]. This is a restrictive assumption since it excludes lagged jointly dependent variables from occurring as predetermined variables in the subsystem being estimated. The assumption that X_I contains "fixed" variables was apparently made by Zellner and Theil for convenience in deriving the 3SLS estimator and deriving properties regarding this estimator. It is common to use 3SLS even if equations contain lagged jointly dependent variables. In his derivation of the 3SLS estimator Goldberger [1964], p. 347 states without proof that: "For convenience we assume that all predetermined variables are exogenous variables distributed independently of the disturbances; the results however, carry over to the general case." (The assumptions regarding predetermined variables in section I.C.3 of this paper follow the assumptions of Goldberger's "general case". In particular, lagged jointly dependent variables are permitted as predetermined variables in Goldberger's general case.)

The discussion in Fisher [1965] regarding the use of lagged jointly dependent variables as instruments would appear applicable for instruments used in the X_I matrix in addition to the predetermined variables in the system or subsystem being estimated.

[2] The assumption made further on in the derivation [following (VIII.4)] that the matrices $X_I'Z_\mu$ have full column rank implies that rk $X \geq n_\mu$ for $\mu = 1,\ldots,M$. When we consider general linear restrictions on coefficients, the matrices $X_I'Z_\mu$ need not have full column rank (provided the coefficients space is sufficiently restricted that unique 3SLS coefficients exist); hence, the requirement rk $X_I \geq n_\mu$ may be relaxed somewhat in that section.

As in our derivation of 2SLS as a GLS method (section IV.D), let us premultiply each equation in the system or subsystem being estimated by the same matrix--the transpose of the X_I matrix defined above. The μ^{th} equation becomes:

$$\underset{K\times 1}{X_I'y_\mu} = \underset{K\times n_\mu}{X_I'Z_\mu}\ \underset{n_\mu\times 1}{\delta_\mu} + \underset{K\times 1}{X_I'u_\mu} \quad . \tag{VIII.2}$$

The entire system can be written as:

$$\begin{aligned} X_I'y_1 &= X_I'Z_1\delta_1 + \cdots + 0\,\delta_M + X_I'u_1 \\ &\vdots \\ X_I'y_M &= 0\,\delta_1 + \cdots + X_I'Z_M\delta_M + X_I'u_M \end{aligned} \tag{VIII.3}$$

or if we define the following matrices and vectors:

$$\dot{y} = \underset{MK\times 1}{\begin{bmatrix} X_I'y_1 \\ \vdots \\ X_I'y_M \end{bmatrix}}, \quad \dot{X} = \underset{MK\times n}{\begin{bmatrix} X_I'Z_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & X_I'Z_M \end{bmatrix}}, \quad \delta = \underset{n\times 1}{\begin{bmatrix} \delta_1 \\ \vdots \\ \delta_M \end{bmatrix}}, \quad \dot{u} = \underset{MK\times 1}{\begin{bmatrix} X_I'u_1 \\ \vdots \\ X_I'u_M \end{bmatrix}},$$

where $n = \sum_{\mu=1}^{M} n_\mu$, we can write the entire system as:

$$\underset{MK\times 1}{\dot{y}} = \underset{MK\times n}{\dot{X}}\ \underset{n\times 1}{\delta} + \underset{MK\times 1}{\dot{u}} \quad . \tag{VIII.4}$$

Initially we will assume that the matrices $X_I'Z_\mu$ have full column rank which implies that the matrix $\dot{X}$ has full column rank; however, this assumption will be relaxed in section VIII.D when formulas for imposing general linear restrictions on coefficients are presented.

Let $X_I'u_\mu = \dot{u}_\mu$, $\mu = 1,\ldots,M$. Then since X_I is fixed, $\mathcal{E}\dot{u}_\mu = \mathcal{E}X_I'u_\mu = X_I'\mathcal{E}u_\mu = X_I'0 = 0$; hence, the entire vector $\mathcal{E}\dot{u}_\mu = 0$.

Let $U = [u_1 \cdots u_M]$ be the $T\times M$ matrix of disturbances, with column μ the disturbances for equation μ and row t the disturbances

for observation t, designated $U_{[t]}$. Then from the assumptions of section I.C.3 that $\mathcal{E}U = 0$, $\mathcal{E}U'_{[t]}U_{[t]} = \underset{M\times M}{\Sigma} = [\sigma_{\mu\mu'}]$ for all t, and $\mathcal{E}U'_{[t]}U_{[t']} = 0$ for $t' \neq t$ plus the assumption that X_I is fixed we obtain that:

$$\text{(VIII.5)} \qquad \mathcal{E}\dot{u}_{\mu}\dot{u}'_{\mu'} = \mathcal{E}X'_I u_{\mu} u'_{\mu'} X_I = X'_I[\mathcal{E}u_{\mu}u'_{\mu'}]X_I = X'_I[\sigma_{\mu\mu'}I]X_I$$

$$= \sigma_{\mu\mu'}X'_I X_I \quad \text{for} \quad \mu, \mu' = 1,\ldots,M \ .$$

The latter set of relations is expressed in terms of the entire covariance matrix of $\dot{u}$ by means of the Kronecker product as:[1]

$$\text{(VIII.6)} \qquad \underset{MK\times MK}{\mathcal{E}\dot{u}\dot{u}'} = \underset{M\times M}{\Sigma} \otimes \underset{K\times K}{[X'_I X_I]} \ .$$

We will designate $\mathcal{E}\dot{u}\dot{u}'$ as $\underset{MK\times MK}{\overset{*}{\Sigma}}$.

The matrix $\dot{X}$ does not consist of fixed variables only, since some of the variables contain submatrices of the form $X'_I Y_{\mu}$, with Y_{μ} jointly dependent, so that (even if the $\overset{*}{\Sigma}$ matrix were known) if we used $\dot{X}$ as the GLS X in deriving the 3SLS estimator, the resulting coefficients would not have all of the GLS properties; however, the GLS derivation is used primarily as a means of suggesting the 3SLS estimator as an estimator with potentially desirable properties. Properties of the 3SLS estimator are derived and proved in Zellner and Theil [1962] after the computational formulas are established.

If we treat $\dot{y}$ as the GLS y, $\dot{X}$ as the GLS X, and $\dot{u}$ as the GLS u, GLS formula (IV.2) becomes:

---

[1] See footnote 1 of page 265 for a definition of the Kronecker product and a proof that for A and B square, symmetric, and nonsingular, $[A\otimes B]$ is also nonsingular and $[A\otimes B]^{-1} = A^{-1}\otimes B^{-1}$.

(VIII.7) $$\hat{\delta}_{GLS} = [\dot{X}'\overset{*}{\Sigma}^{-1}\dot{X}]^{-1}[\dot{X}'\overset{*}{\Sigma}^{-1}\dot{y}] \quad ;$$

however, $\overset{*}{\Sigma}^{-1} = [\Sigma \otimes X_I'X_I]^{-1} = \Sigma^{-1} \otimes [X_I'X_I]^{-1}$

$$= \begin{bmatrix} \sigma^{11} & \cdots & \sigma^{1M} \\ \vdots & & \vdots \\ \sigma^{M1} & \cdots & \sigma^{MM} \end{bmatrix} \otimes [X_I'X_I]^{-1} = \begin{bmatrix} \sigma^{11}[X_I'X_I]^{-1} & \cdots & \sigma^{1M}[X_I'X_I]^{-1} \\ \vdots & & \vdots \\ \sigma^{M1}[X_I'X_I]^{-1} & \cdots & \sigma^{MM}[X_I'X_I]^{-1} \end{bmatrix} .$$

Thus, (VIII.7) may be rewritten as:

(VIII.8) $\hat{\delta}_{GLS} =$

$$\left\{\begin{bmatrix} Z_1'X_I & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & Z_M'X_I \end{bmatrix}\begin{bmatrix} \sigma^{11}[X_I'X_I]^{-1} & \cdots & \sigma^{1M}[X_I'X_I]^{-1} \\ \vdots & & \vdots \\ \sigma^{M1}[X_I'X_I]^{-1} & \cdots & \sigma^{MM}[X_I'X_I]^{-1} \end{bmatrix}\begin{bmatrix} X_I'Z_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & X_I'Z_M \end{bmatrix}\right\}^{-1} \times$$

$$\times \left\{\begin{bmatrix} Z_1'X_I & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & Z_M'X_I \end{bmatrix}\begin{bmatrix} \sigma^{11}[X_I'X_I]^{-1} & \cdots & \sigma^{1M}[X_I'X_I]^{-1} \\ \vdots & & \vdots \\ \sigma^{M1}[X_I'X_I]^{-1} & \cdots & \sigma^{MM}[X_I'X_I]^{-1} \end{bmatrix}\begin{bmatrix} X_I'y_1 \\ \vdots \\ X_I'y_M \end{bmatrix}\right\}$$

$$= \begin{bmatrix} \sigma^{11}Z_1'X_I(X_I'X_I)^{-1}X_I'Z_1 & \cdots & \sigma^{1M}Z_1'X_I(X_I'X_I)^{-1}X_I'Z_M \\ \vdots & & \vdots \\ \sigma^{M1}Z_M'X_I(X_I'X_I)^{-1}X_I'Z_1 & \cdots & \sigma^{MM}Z_M'X_I(X_I'X_I)^{-1}X_I'Z_M \end{bmatrix}^{-1}\begin{bmatrix} \sum_{\mu=1}^{M}\sigma^{\mu 1}Z_1'X_I(X_I'X_I)^{-1}X_I'y_\mu \\ \vdots \\ \sum_{\mu=1}^{M}\sigma^{\mu M}Z_M'X_I(X_I'X_I)^{-1}X_I'y_\mu \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^{11}[Z_1'Z_1]_{\parallel X_I} & \cdots & \sigma^{1M}[Z_1'Z_M]_{\parallel X_I} \\ \vdots & & \vdots \\ \sigma^{M1}[Z_M'Z_1]_{\parallel X_I} & \cdots & \sigma^{MM}[Z_M'Z_M]_{\parallel X_I} \end{bmatrix}^{-1}\begin{bmatrix} \sum_{\mu=1}^{M}\sigma^{\mu 1}[Z_1'y_\mu]_{\parallel X_I} \\ \vdots \\ \sum_{\mu=1}^{M}\sigma^{\mu M}[Z_M'y_\mu]_{\parallel X_I} \end{bmatrix} .$$

That $Z_\mu'X_I(X_I'X_I)^{-1}X_I'Z_{\mu'} = [Z_\mu'Z_{\mu'}]_{\parallel X_I}$ (the cross-product of the part of Z_μ in the space spanned by X_I with the part of $Z_{\mu'}$ in the space spanned by X_I) and that $Z_\mu'X_I(X_I'X_I)^{-1}X_I'y_{\mu'} = [Z_\mu'y_{\mu'}]_{\parallel X_I}$ with

$[z'_\mu y_{\mu'}]_{\|X_I}$ similarily defined follows from (I.36).

Since Σ is unknown, we will substitute the following estimate of Σ into the formula:[1]

$$\hat{\Sigma} \overset{\text{def}}{=} S = [s_{\mu\mu'}] \tag{VIII.9}$$

with $s_{\mu\mu'} = (1/T)\hat{u}_\mu^{2SLS'}\hat{u}_{\mu'}^{2SLS} = (1/T)_+\hat{\delta}_\mu^{2SLS'}[{}_+Z'_\mu\ {}_+Z_{\mu'}]_+\hat{\delta}_{\mu'}^{2SLS}$, ${}_+\hat{\delta}_\mu^{2SLS}$ being the vector of 2SLS coefficients (including the normalizing coefficient, -1) for equation μ and ${}_+Z_\mu = [y_\mu \vdots Z_\mu]$.

Substituting S into (VIII.8) for Σ we get the 3SLS estimator:

$$\hat{\delta}_{3SLS} = (\mathcal{R}^{(1)})^{-1}\mathcal{r}^{(1)} \tag{VIII.10}$$

where $S^{-1} = [s^{\mu\mu'}]$,

$$\mathcal{R}^{(1)} = \begin{bmatrix} s^{11}[Z'_1Z_1]_{\|X_I} & s^{1M}[Z'_1Z_M]_{\|X_I} \\ s^{M1}[Z'_MZ_1]_{\|X_I} & s^{MM}[Z'_MZ_M]_{\|X_I} \end{bmatrix}, \text{ and} \tag{VIII.11}$$

$$\mathcal{r}^{(1)} = \begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu}[Z'_1y_\mu]_{\|X_I} \\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu}[Z'_My_\mu]_{\|X_I} \end{bmatrix}.$$

---

[1]The S matrix will be singular if the number of equations exceeds the number of observations. Let $\underset{T\times M}{\hat{U}} = [\hat{u}_1 \cdots \hat{u}_M]$ where the $\hat{u}_\mu$ are 2SLS estimates. Then $\underset{M\times M}{S} = (1/T)\hat{U}'\hat{U}$; hence, rk S = rk $\hat{U}$. If T < M then rk $\hat{U}$ and therefore rk S will be less than M, i.e., S will be singular. This is not in conflict with our derivation in section VIII.B of 2SLS as the special case of 3SLS in which there is zero correlation between residuals across equations, since this derivation holds only for nonsingular S. Even though Σ is assumed nonsingular, a particular estimate of Σ may still be singular and (as shown above) if M ≥ T, $\hat{\Sigma}$ = S is singular.

Let Y_A be the matrix of jointly dependent variables in the M equations being estimated by 3SLS, let X_A be the matrix of predetermined variables in the M equations being estimated by 3SLS, and let $Z_A = [Y_A \vdots X_A]$. Then $\mathcal{R}^{(1)}$ and $r^{(1)}$ can be computer efficiently by forming the matrix

$$[Z_A'Z_A]_{\|X_I} = \begin{bmatrix} [Y_A'Y_A]_{\|X_I} & Y_A'X_A \\ X_AY_A & X_A'X_A \end{bmatrix} \tag{VIII.12}$$

(this matrix may be formed in triangular form since it is symmetric) and then extracting the submatrices $[Z_{\mu}'Z_{\mu'}]_{\|X_I}$ used in $\mathcal{R}^{(1)}$ and the subvectors $[Z_{\mu}'y_{\mu'}]_{\|X_I}$ used in $r^{(1)}$ from $[Z_A'Z_A]_{\|X_I}$. That $[Y_A'X_A]_{\|X_I} = Y_A'X_A$ and $[X_A'X_A]_{\|X_I} = X_A'X_A$ follows from our definition of X_I as containing the matrix X_A [see (I.40) and (I.41)]. $[Y_A'Y_A]_{\|X_I}$ is computed as $[Y_A'Y_A] - [Y_A'Y_A]_{\perp X_I}$ with $[Y_A'Y_A]_{\perp X_I}$ computed by direct orthogonalization in the manner given in section I.D.2. Note that $[Z_A'Z_A]_{\|X_I}$ is unique and easily computed even for an X_I having less than full column rank.

If $\dot{X}$ had consisted only of fixed variables and Σ were known so that Σ rather than its estimate S were used in the calculation of $\mathcal{R}^{(1)}$ and $r^{(1)}$, then by GLS formula (IV.3), $\text{Var}(\hat{\delta}_{3SLS}) = (\mathcal{R}^{(1)})^{-1}$. As shown in Zellner and Theil [1962], even though $\dot{X}$ contains non-fixed variables and S is used instead of Σ,

$$\text{asymptotic } \hat{\text{Var}}(\hat{\delta}_{3SLS}) = (\mathcal{R}^{(1)})^{-1} . \tag{VIII.13}$$

A "degrees of freedom" adjustment can be made in the estimated 3SLS coefficient variance-covariance matrix [i.e., asymptotic $\hat{\text{Var}}(\hat{\delta}_{3SLS})$] in the same manner as for FIML (section V.E.). The estimated 3SLS

coefficient variance-covariance matrix can also be normalized in the manner suggested for FIML.

If an estimate of the disturbance variance-covariance matrix Σ is desired, it seems desirable to utilize the estimated 3SLS coefficients to calculate a new $\hat{\Sigma}_{3SLS}$ instead of merely using the S matrix calculated from the 2SLS coefficients. Utilizing the 3SLS coefficients, the estimated disturbance variance-covariance matrix becomes:

$$\hat{\Sigma}_{3SLS} = [s^{3SLS}_{\mu\mu'}] \tag{VIII.14}$$

with $s^{3SLS}_{\mu\mu'} = (1/T)\hat{u}^{3SLS\prime}_{\mu}\hat{u}^{3SLS}_{\mu'} = (1/T)\hat{\delta}^{3SLS\prime}_{\mu}[{}_{+}Z'_{\mu}\ {}_{+}Z_{\mu'}]\hat{\delta}^{3SLS}_{\mu'}$, ${}_{+}\hat{\delta}^{3SLS}_{\mu}$ being the vector of 3SLS coefficients for equation μ (including the normalizing coefficient, -1) and ${}_{+}Z_{\mu} = [y_{\mu} \vdots Z_{\mu}]$.

A "degrees of freedom" adjustment may be made in the $\hat{\Sigma}_{3SLS}$ matrix, and the $\hat{\Sigma}_{3SLS}$ matrix normalized may be computed in the same manner as for FIML (section V.D).

B. An Alternate Computational Procedure

3SLS estimates will be the same as 2SLS estimates if the S matrix is a diagonal matrix, i.e., there is zero correlation between the 2SLS residuals of each pair of equations.[1] This is easily verified by writing out the 3SLS estimating equations. In this case,

$$S^{-1} = \begin{bmatrix} s_{11} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & s_{MM} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{s_{11}} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \frac{1}{s_{MM}} \end{bmatrix}$$

and the formula for 3SLS becomes

$$(R^{(1)})^{-1} r^{(1)} =$$

$$\begin{bmatrix} \frac{1}{s_{11}}[Z_1'Z_1]\|X_I & \cdots & 0\cdot[Z_1'Z_M]\|X_I \\ \vdots & & \vdots \\ 0\cdot[Z_M'Z_1]\|X_I & \cdots & \frac{1}{s_{MM}}[Z_M'Z_M]\|X_I \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{s_{11}}[Z_1'y_1]\|X_I + \cdots + 0\cdot[Z_1'y_M]\|X_I \\ \vdots \\ 0\cdot[Z_1'y_M]\|X_I + \cdots + \frac{1}{s_{MM}}[Z_M'y_M]\|X_I \end{bmatrix}$$

$$= \begin{bmatrix} s_{11}[Z_1'Z_1]^{-1}\|X_I & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & s_{MM}[Z_M'Z_M]^{-1}\|X_I \end{bmatrix} \begin{bmatrix} \frac{1}{s_{11}}[Z_1'y_1]\|X_I \\ \vdots \\ \frac{1}{s_{MM}}[Z_M'y_M]\|X_I \end{bmatrix}$$

$$= \begin{bmatrix} [Z_1'Z_1]^{-1}\|X_I[Z_1'y_1]\|X_I \\ \vdots \\ [Z_M'Z_M]^{-1}\|X_I[Z_M'y_M]\|X_I \end{bmatrix} = \begin{bmatrix} \hat{\delta}_1^{2SLS} \\ \vdots \\ \hat{\delta}_M^{2SLS} \end{bmatrix}$$

In the above verification the diagonal elements of S cancelled out; hence, if any diagonal matrix had been used in place of S, the

---

[1] Zellner and Theil [1962], p. 58, note this special case.

same result would have been obtained. In particular, an identity matrix could be used for S. This gives a basis for an alternate method for calculating 3SLS coefficients which is essentially the same as the alternate method for calculating ZA coefficients:

(1) Start the computation by using an identity matrix in place of the S matrix. (No initial coefficients are required, since they are used only in the calculation of the S matrix.)

Calculate $\mathcal{R}^{(0)}$ and $r^{(0)}$ using the same formulas as for $\mathcal{R}^{(1)}$ and $r^{(1)}$ except that an identity matrix is used for the initial S matrix. Then:

$$\hat{\delta}_{2SLS} = \begin{bmatrix} \hat{\delta}_1^{2SLS} \\ \vdots \\ \hat{\delta}_M^{2SLS} \end{bmatrix} = (\mathcal{R}^{(0)})^{-1} r^{(0)}$$

(2) Calculate S from $\hat{\delta}_{2SLS}$ in the way given earlier in this chapter and use S to calculate a new $\mathcal{R}^{(1)}$ matrix and $r^{(1)}$ vector. Then:

$\hat{\delta}_{3SLS} = (\mathcal{R}^{(1)})^{-1} r^{(1)}$ as before.

Whereas the alternate computational method given above takes slightly more computer time at the 0^{th} iteration (for a large scale computer, the additional time required is hardly measurable), it is simpler to program--especially if provision is made to iterate on the 3SLS estimates in the manner indicated farther on. Another advantage of starting in this manner will be noted in the discussion of the calculation of restricted 3SLS coefficients.

A disadvantage of starting with an identity matrix in place of the S matrix is that it imposes 2SLS estimates as the starting

estimates of the 3SLS procedure, whereas it may be desired that other estimates be used as starting estimates for some or all of the equations; however, there seems little tendency to use estimates other than 2SLS estimates as starting estimates for 3SLS.[1]

---

[1] LIML estimates or other similar estimates meeting certain consistency requirements (DLS estimates do not meet these requirements) could be substituted for the 2SLS estimates in the 3SLS procedure without changing the proof of the derivation of the asymptotic moment matrix of 3SLS in Zellner and Theil's article [1962]; however, it is assumed that 3SLS estimates are based on 2SLS estimates unless stated otherwise. Use of estimates other than 2SLS estimates would, of course, change the resulting 3SLS estimates obtained.

C. 3SLS Estimation when $\text{rk } X_I = T$

In section II.G it was noted that if $\text{rk } X_I = T$, the estimated coefficients for all double k-class estimators become the same as the DLS coefficients. In similar fashion for 3SLS estimation, if $\text{rk } X_I = T$, $\hat{\delta}_{3SLS} = \hat{\delta}_{ZA}$; that is, the 3SLS coefficients obtained will be the same coefficients as if the explanatory jointly dependent variables of each equation were misclassified as predetermined and ZA applied. The 3SLS coefficients obtained will not, of course, have the same properties as ZA coefficients.[1]

That $\hat{\delta}_{3SLS} = \hat{\delta}_{ZA}$ is easily seen. Let Z_A be the matrix of variables in the subsystem being estimated; then $\text{rk } X_I = T$ implies that $[Z_A]_{\| X_I} = Z_A$, since all variables are in the space spanned by X_I .

$\mathcal{R}^{(1)}$ becomes $\begin{bmatrix} s^{11}Z_1'Z_1 & \cdots & s^{1M}Z_1'Z_M \\ \vdots & & \vdots \\ s^{M1}Z_M'Z_1 & \cdots & s^{MM}Z_M'Z_M \end{bmatrix}$ and $r^{(1)}$ becomes $\begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu}Z_1'y_\mu \\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu}Z_M'y_\mu \end{bmatrix}$;

hence, $\hat{\delta}_{3SLS} = (\mathcal{R}^{(1)})^{-1}(r^{(1)}) = (\mathcal{P}^{(1)})^{-1}(p^{(1)}) = \hat{\delta}_{ZA}$ as can be verified by comparison with ZA formula (VII.12).

---

[1] It may be recalled that SML estimation is so affected by the rank of the X_I matrix that if $\text{rk } X_I \geq T - M + 1$, the SML computations cannot be performed (at least the SML formulas given in Chapter VI cannot be used without some modification). On the other hand FIML estimation is not affected by the rank of the X_I matrix, since jointly dependent variables are not adjusted by the X_I matrix. (Instead, the matrix of coefficients of jointly dependent variables are used directly in the estimation procedure.)

The fact that $\hat{\delta}_{3SLS}$ coincides with $\hat{\delta}_{ZA}$ computed with jointly dependent variables misclassified does not destroy the consistency of $\hat{\delta}_{3SLS}$.[1] All that is indicated is that there are insufficient observations to distinguish $\hat{\delta}_{3SLS}$ estimates from $\hat{\delta}_{ZA}$ estimates (unless the space of X_I is restricted in some fashion).

The discussion of whether the space spanned by X_I should be restricted (and methods for restricting the space of X_I) in the case of the double k-class methods (section II.G) is applicable to 3SLS as well, except that if the subsystem being estimated contains very many predetermined variables, the rank of the predetermined variables in the subsystem being estimated may already equal T.

Let the matrix of predetermined variables in the subsystem being estimated be denoted X_A. It seems undesirable to restrict the subspace of X_I in a manner such that X_A is not in the space spanned by X_I. If the space of X_I is restricted such that X_A is not in the space spanned by X_I, the 3SLS formulas given previously are not valid, since (see VIII.12) $[X_A'Y_A]_{|X_I} \neq X_A'Y_A$ and $[X_A'X_A]_{|X_I} \neq X_A'X_A$. To take account of these non-equalities, any predetermined variable outside the space spanned by X_I must be adjusted in the same manner as the jointly dependent variables in the computational formulas;

---

[1] Consistency is an asymptotic property and a small number of observations in a given sample certainly does not affect an asymptotic property. If the number of instruments used in the estimation is fixed (e.g., the number of predetermined variables in the system is fixed for a given model, so if X is used as X_I, X_I will be fixed) then (if the $\hat{\delta}_{3SLS}$ formula is followed; that is, a switch is not made to the $\hat{\delta}_{ZA}$ formula) as T increases, at some point there will be sufficient observations that rk X_I < T and $\hat{\delta}_{3SLS}$ will not coincide with $\hat{\delta}_{ZA}$.

however, this is the same as misclassifying these variables as jointly dependent in the original model, i.e., this adjustment has the same computational effect as a change in the model in response to the small number of observations.

D. Arbitrary Linear Restrictions Imposed on Coefficients

As noted earlier, the 3SLS formulas may be derived as an application of the GLS method. In like manner, we may derive restricted 3SLS (R3SLS) estimates (in which arbitrary linear restrictions are imposed on the coefficients) as an application of the RGLS method.

If the set of N_R restrictions given by:

$$\underset{N_R \times n}{R} \quad \underset{n \times 1}{\delta} = \underset{N_R \times 1}{r} \tag{VIII.15}$$

is imposed on the 3SLS model ($\dot{y} = \dot{X}\delta + \dot{u}$ with assumptions as stated previously), the R3SLS formulas are given by:[1]

$$\underset{n\times 1}{\hat{\delta}_{R3SLS}} = \underset{n\times(n-rk\ R)}{Q} \quad \{\underset{(n-rk\ R)\times(n-rk\ R)}{[Q'\mathcal{R}^{(1)}Q]^{-1}} \quad \underset{(n-rk\ R)\times 1}{Q'[r^{(1)} - \mathcal{R}^{(1)}q]}\} + \underset{n\times 1}{q} \tag{VIII.16}$$

$$\text{asymptotic } \hat{Var}(\hat{\delta}_{R3SLS}) = Q[Q'\mathcal{R}^{(1)}Q]^{-1}Q' \ . \tag{VIII.17}$$

(VIII.16) and (VIII.17) are derived from substituting the 3SLS matrix $\mathcal{R}^{(1)}$ and the 3SLS vector $r^{(1)}$ into RGLS formulas (IV.5) and (IV.8). Q and q are calculated from R and r by the computational

---

[1] Calculation of the Q matrix and q vector also gives a means of separating out rk R coefficients, $\hat{\delta}_{(1)}$, which may be calculated from the remaining n-rk R "unrestricted" coefficients, $\hat{\delta}_{(2)}$. Thus, the following pair of formulas are together equivalent to (VIII.16):

$$\underset{(n-rk\ R)\times 1}{\hat{\delta}^{R3SLS}_{(1)}} = \underset{(n-rk\ R)\times(n-rk\ R)}{[Q'\mathcal{R}^{(1)}Q]^{-1}} \quad \underset{(n-rk\ R)\times 1}{Q'[r^{(1)} - \mathcal{R}^{(1)}q]}$$

$$\underset{rk\ R\times 1}{\hat{\delta}^{R3SLS}_{(2)}} = \underset{rk\ R\times(n-rk\ R)}{Q_2} \quad \underset{(n-rk\ R)\times 1}{\hat{\delta}^{R3SLS}_{(1)}} + \underset{rk\ R\times 1}{q_2}$$

where Q_2 and q_2 are the subparts of Q and q noted in chapter IV.

procedure given in section IV.B.1.[1]

The R matrix need not have full row rank and the $R^{(1)}$ matrix may be singular (the Z_μ matrices may have less than full column rank).

Although the R matrix and r vector can contain restrictions imposed on the coefficients of only a single equation and/or restrictions which cut across equations, it should be noted that the same answer will not in general be obtained if restrictions which affect the coefficients of a single equation are solved into that equation as if these restrictions are listed in the R matrix and r vector and R3SLS applied. The reason that the resulting coefficients may be different are outlined for R2SLS in section IV.D, but are equally applicable to R3SLS. Restrictions which cut across equations cannot, of course, be solved into a single equation; hence, must be imposed through use of a procedure such as the R3SLS computational procedure.

In calculating the $R^{(1)}$ matrix and $r^{(1)}$ vector, it seems desirable that the S matrix corresponding to the restricted 2SLS estimates rather than the unrestricted 2SLS estimates be used.

---

[1] If R is assumed to have full row rank and the $R^{(1)}$ matrix is assumed to have full column rank then the formula

$$\hat{\delta}_{R3SLS} = \hat{\delta}_{3SLS} - (R^{(1)})^{-1}R'[R(R^{(1)})^{-1}R'][R\hat{\delta}_{3SLS} - r]$$

may be used. (This formula is implied by Zellner and Theil [1962], p. 78 in their remark that restrictions may be applied to 3SLS estimates.) The advantages of the Q, q method over the R, r method for GLS given in section IV.B.2 are applicable to 3SLS as well.

Imposing Restrictions Which Cut Across Equations on the Coefficients Used to Calculate the S Matrix

It seems clearly desirable that restrictions which do not cut across equations be imposed on the 2SLS coefficients used to calculate the S matrix. It would also seem desirable to impose restrictions which do cut across equations on the 2SLS coefficients as well. (They will then not be 2SLS coefficients for separate equations, but coefficients closely related to 2SLS.) This may be done in the same way as was indicated in the ZA procedure, i.e., by using the alternate 3SLS computational method noted earlier (section VIII.B), i.e.:

(1) Start the computations by using an identity matrix in place of the S matrix. Calculate $R^{(0)}$ and $r^{(0)}$ using the same formula as for $R^{(1)}$ and $r^{(1)}$ except that the identity matrix is used in place of the S matrix. Then:

$$\text{(VIII.18)} \quad \hat{\delta}_{\text{R2SLSME}} = Q[Q'R^{(0)}Q]^{-1}Q'[r^{(0)} - R^{(0)}q] + q \quad .$$

(2) Calculate S from $\hat{\delta}_{\text{R2SLSME}}$ and use it to calculate a new $R^{(1)}$ matrix and $r^{(1)}$ vector. Then calculate $\hat{\delta}_{\text{R3SLS}}$ by formula (VIII.16) as before.[1]

In addition to being simple to program, the use of the 0^{th} iteration method has the advantage that unique 2SLS estimates (which

---

[1] $\hat{\delta}_{\text{R3SLS}}$ calculated in this manner will not be the same as $\hat{\delta}_{\text{R3SLS}}$ calculated by ignoring the restrictions which cut across equations in the calculation of the S matrix.

take account of restrictions which do not cut across equations) need not exist provided there is sufficient identification that unique R3SLS estimates exist.[1]

---

[1]It may be recalled that R2SLS as defined in section IV.D can incorporate restrictions which do not cut across equations whereas R2SLSME as in (VIII.18) can incorporate also restrictions which cut across equations. Unique estimates of the latter may exist even when unique estimates of the former do not. An example of this similar to the example for ZA (footnote 1 of page 277) could be constructed.

E. Iterative Three-stage Least Squares (I3SLS)

1. Only zero and normalization restrictions imposed on coefficients

In their concluding remarks, Zellner and Theil state: "The three-stage least squares estimator $\hat{\delta}$ implies in general a new estimator of $[\sigma_{\mu\mu'}]$ which differs from $[s_{\mu\mu'}]$. One can then set up a new stage based on this estimator rather than $[s_{\mu\mu'}]$ and proceed iteratively. No report on this method can be made as yet, but we hope to come back to it in the future."[1]

Such a procedure has been referred to as multi-stage least squares; however, this does not seem very appropriate, since going on to subsequent iterations is not the same as going on to subsequent stages as from the DLS 1st stage, to the 2SLS 2nd stage, to the 3SLS 3rd stage. Instead we will refer to the above procedure as iterative three-stage least squares (I3SLS). Madansky examines the question and concludes that iteration is not worthwhile "in the sense that there is no improvement in the asymptotic variance of the estimator. On the other hand, the effect of such iteration on the finite sample variance (or even on the finite-sample bias) of the estimator is still an open question."[2]

ZA may be regarded as the particular case of 3SLS in which there is only one jointly dependent variable per equation. As we saw earlier, IZA estimates coincide with FIML estimates. The question then arises: "Will iteration on 3SLS estimates lead to SML estimates if applied to a partial system or to FIML estimates if applied to a complete system

---

[1] Zellner and Theil [1962], p. 78.

[2] Madansky [1964], p. 55.

assuming that the iterations on the 3SLS estimates converge (so that we may refer to them as I3SLS estimates)?" The answers to both questions are in general, no, provided that $\text{rk } X_I > n_\mu$ for at least one equation in the system.[1] Thus, occurrence of the "explanatory" jointly dependent variables in the case of 3SLS and I3SLS has a considerable effect on the properties of 3SLS and I3SLS relative to ZA and IZA.

First of all, let us show that I3SLS and SML estimates do not in general coincide. To do so, assume that the partial system to be estimated consists of only a single equation with $\text{rk } X_I > n_\mu$ for that equation. If a third stage is applied to the single equation the two-stage estimates are again obtained, for the 3-stage formula becomes:

$$\hat{\delta}_{3SLS} = [s^{11}[Z_1'Z_1]_\perp X_I]^{-1}[s^{11}[Z_1'y_1]_\perp X_I] = [Z_1'Z_1]_\perp^{-1} X_I [Z_1'y_1]_\perp X_I = \hat{\delta}_{2SLS} .$$

Continued iteration will give only the 2SLS estimates for each iteration, that is, the system clearly converges to the 2SLS estimates.

Hood and Koopmans show that the SML estimates for a system with a single equation are the LIML estimates.[2] Thus, in this case the I3SLS and SML estimates will coincide only if the 2SLS and LIML estimates coincide which will not in general occur for an equation for which $\text{rk } X_I > n_\mu$.

The question might still arise--is there some peculiarity about a complete system which would make I3SLS estimates the same as FIML estimates. We can easily answer this in the negative by extending the

---

[1] If $\text{rk } X_I > n_\mu$ for an equation, the equation is usually termed "over-identified" and if $\text{rk } X_I = n_\mu$ for an equation, the equation is usually termed "just-identified" (see section II.B). If $\text{rk } X_I = n_\mu$ for all equations in the system, all of the following estimators (plus other estimators) give the same estimated coefficients: FIML, SML, 3SLS, I3SLS, LIML, and 2SLS.

[2] Koopmans and Hood [1953], pp. 170-171.

above special case.

Zellner and Theil show that in the estimation of the coefficients of a system of equations by 3SLS, any equation for which $\text{rk } X_I = n_\mu$ can be omitted in the calculation of the 3SLS coefficients of the remaining equations.[1] Since their proof makes no use of where the original estimates are derived, it applies to any iteration; hence, the equations for which $\text{rk } X_I = n_\mu$ may be omitted from the system in the calculation of the I3SLS coefficients of the equations for which $\text{rk } X_I > n_\mu$, i.e., the iterations may be performed on the equations for which $\text{rk } X_I > n_\mu$, only.

Now, consider a complete system in which $\text{rk } X_I > n_\mu$ for only one equation in the system and $\text{rk } X_I = n_\mu$ for the remainder of the equations. Then the I3SLS estimates of that one equation will be the 2SLS estimates, which will not, in general, be the same as the FIML estimates of that equation. Thus, I3SLS estimates will not, in general, be the same as FIML estimates.

The I3SLS Procedure

To see how the computation of the I3SLS estimates differs at each iteration from the computation of SML and FIML estimates, let us first note the change in the coefficient vector from one iteration

---

[1] Zellner and Theil [1962], pp. 63-68. Actually their proof is not complete as they only show that the R matrix can be subdivided. They do not show the effect of multiplying the subdivided R matrix by a similarily subdivided r vector; however, some additional manipulation shows that their conclusion is correct.

to the next in the I3SLS procedure.[1] At the i^{th} iteration of the I3SLS procedure: $\hat{\delta}^{(i)} = (\mathcal{R}^{(i-1)})^{-1}\mathcal{r}^{(i-1)}$ where $\mathcal{R}^{(i-1)}$ and $\mathcal{r}^{(i-1)}$ are calculated in the same manner [(VIII.10) and (VIII.11)] as $\mathcal{R}^{(1)}$ and $\mathcal{r}^{(1)}$ except that $\hat{\delta}^{(i-1)}$ is used instead of $\hat{\delta}^{(1)}$.

The increment added to $\hat{\delta}^{(i-1)}$ to form $\hat{\delta}^{(i)}$ is:

---

[1]Many assume that the relationship between FIML and 3SLS has been at least largely explicated in Chow [1964], pp. 548-550; however, Chow's derivation of 3SLS as a minimization of a particular determinant is incorrect from a couple of standpoints:
(1) Chow computes his estimate of $\sigma_{\mu\mu'}$ as:

$$\sigma_{\mu\mu'} = \frac{1}{T}([y_\mu]_{\|X} - [Y_\mu]_{\|X}\bar{\gamma}_\mu - X_\mu\bar{\beta}_\mu)'([y_{\mu'}]_{\|X} - [Y_{\mu'}]_{\|X}\bar{\gamma}_{\mu'} - X_{\mu'}\bar{\beta}_{\mu'})$$

$$= \frac{1}{T}(y_\mu - [Y_\mu]_{\|X}\bar{\gamma}_\mu - Z_\mu\bar{\beta}_\mu)'(y_{\mu'} - [Y_{\mu'}]_{\|X}\bar{\gamma}_{\mu'} - Z_{\mu'}\bar{\beta}_{\mu'})$$

where $\bar{\gamma}_\mu$ apparently is $\hat{\gamma}_\mu^{2SLS}$ and $\bar{\beta}_\mu$ apparently is $\hat{\beta}_\mu^{2SLS}$. (Chow does not divide by T, but a division by T cancels out in the actual estimation, anyway. Also, in his notation Chow used β's where we use γ's and vice versa.) On the other hand, Zellner and Theil [1962] compute their estimate of $\sigma_{\mu\mu'}$ as:

$$\hat{\sigma}_{\mu\mu'} = \frac{1}{T}(y_\mu - Y_\mu\hat{\gamma}_\mu^{2SLS} - X_\mu\hat{\beta}_\mu^{2SLS})'(y_{\mu'} - Y\hat{\gamma}_{\mu'}^{2SLS} - X_{\mu'}\hat{\beta}_{\mu'}^{2SLS})$$

that is, through use of the actual values of the jointly dependent variables rather than their adjusted values $Y_{\|X}$.

(2) Even if we accepted Chow's estimate of Σ instead of the estimate proposed by Zellner and Theil we would still not get 3SLS, since application of Chow's formula (7.10) would give a new estimate for the vector of coefficients which would lead to a new estimate of Σ, and so on.

$$\text{(VIII.19)}\qquad \hat{\delta}^{(i)} - \hat{\delta}^{(i-1)} = (\mathcal{R}^{(i-1)})^{-1}r^{(i-1)} - (\mathcal{R}^{(i-1)})^{-1}\mathcal{R}^{(i-1)}\hat{\delta}^{(i-1)}$$

$$= (\mathcal{R}^{(i-1)})^{-1}\left\{\begin{bmatrix}\sum_{\mu=1}^{M} s^{1\mu}[z_1'y_\mu]_{\|X_I}\\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu}[z_M'y_\mu]_{\|X_I}\end{bmatrix} - \begin{bmatrix}s^{11}[z_1'z_1]_{\|X_I} & \cdots & s^{1M}[z_1'z_M]_{\|X_I}\\ \vdots & & \vdots \\ s^{M1}[z_M'z_1]_{\|X_I} & \cdots & s^{MM}[z_M'z_M]_{\|X_I}\end{bmatrix}\begin{bmatrix}\hat{\delta}_1^{(i-1)}\\ \vdots \\ \hat{\delta}_M^{(i-1)}\end{bmatrix}\right\}$$

$$= (\mathcal{R}^{(i-1)})^{-1}\left\{\begin{matrix}\sum_{\mu=1}^{M} s^{1\mu}([z_1'y_\mu]_{\|X_I} - [z_1'z_\mu]_{\|X_I}\hat{\delta}_\mu^{(i-1)})\\ \vdots \\ \sum_{\mu=1}^{M} s^{M\mu}([z_M'y_\mu]_{\|X_I} - [z_M'z_\mu]_{\|X_I}\hat{\delta}_\mu^{(i-1)})\end{matrix}\right\} = (\mathcal{R}^{(i-1)})^{-1}m^{(i-1)}$$

where we are temporarily denoting the right hand side vector as $m^{(i-1)}$. The j^{th} block of $m^{(i-1)}$ is:

$$\text{(VIII.20)}\qquad m_j^{(i-1)} = \sum_{\mu=1}^{M} s^{j\mu}([z_j'y_\mu]_{\|X_I} - [z_j'z_\mu]_{\|X_I}\hat{\delta}_\mu^{(i-1)})$$

$$= \sum_{\mu=1}^{M} s^{j\mu}[z_j'y_\mu - z_j'z_\mu\hat{\delta}_\mu^{(i-1)}]_{\|X_I} \qquad \text{[see (I.62)]}$$

$$= \sum_{\mu=1}^{M} s^{j\mu}[z_j'(y_\mu - z_\mu\hat{\delta}_\mu^{(i-1)})]_{\|X_I} = \sum_{\mu=1}^{M} s^{j\mu}[z_j'\hat{u}_\mu]_{\|X_I}\ .$$

Let us subdivide the j^{th} block according to whether the variables in Z_j are jointly dependent or predetermined. Then, since X_j is contained in X_I, $[X_\mu'\hat{u}_\mu]_{\|X_I} = X_\mu'\hat{u}_\mu$ [see (I.54) and (I.40)] and (VIII.20) becomes:

$$\text{(VIII.21)} \qquad m_j^{(i-1)} = \begin{bmatrix} \sum_{\mu=1}^{M} s^{j\mu}[Y_j'\hat{u}_\mu]_{\|}X_I \\ \\ \sum_{\mu=1}^{M} s^{j\mu}X_j'\hat{u}_\mu \end{bmatrix} .$$

Thus,

$$\text{(VIII.22)} \qquad \hat{\delta}^{(i)} - \hat{\delta}^{(i-1)} = (\mathcal{R}^{(i-1)})^{-1} m^{(i-1)}$$

with

$$\text{(VIII.23)} \qquad m^{(i-1)} = \begin{bmatrix} \begin{bmatrix} \sum_{\mu=1}^{M} s^{1\mu}[Y_1'\hat{u}_\mu]_{\|}X_I \\ \\ \sum_{\mu=1}^{M} s^{1\mu}X_1'\hat{u}_\mu \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \sum_{\mu=1}^{M} s^{M\mu}[Y_M'\hat{u}_\mu]_{\|}X_I \\ \\ \sum_{\mu=1}^{M} s^{M\mu}X_M'\hat{u}_\mu \end{bmatrix} \end{bmatrix}$$

As noted earlier, the $\mathcal{R}$ metric is a positive definite metric which has been used in the past as the metric for a large number of iterations in the FIML and SML computational methods. A necessary condition for I3SLS convergence is that $\mathcal{R}^{-1}m$ becomes zero. A sufficient condition for this is that m goes to zero. If m can be shown to be the partial derivative of some function $f^{**}(\hat{\delta})$ which is unimodal in the region of iteration and satisfies certain regularity conditions, then if we allow a step size other than 1, i.e., calculate $\hat{\delta}^{(i)} = \hat{\delta}^{(i-1)} + h^{(i)}(\mathcal{R}^{(i-1)})^{-1}m^{(i-1)}$, then for $h^{(i)}$ sufficiently

small, $f^{**}(\hat{\delta}^{(i)}) > f^{**}(\hat{\delta}^{(i-1)})$. The condition $m = 0$ will imply that the maximum has been reached in the region. In checking possibilities for a function of interest being maximized or minimized, m need not be the partial derivative of the function. It is only necessary that f^{**} be a strictly increasing or decreasing function of the function of interest.

The SML and FIML Procedures

The vector of SML which is comparable with the vector $m^{(i-1)}$ of I3SLS is [as given in (VI.36)]:

$$\text{(VIII.24)} \qquad {}_{1}{}^{(i-1)} = \begin{Bmatrix} \sum_{\mu=1}^{M} \{s^{1\mu} Z_1' \hat{u}_\mu - t^{1\mu} [Z_1' \hat{u}_\mu]_{\perp X_I}\} \\ \vdots \\ \sum_{\mu=1}^{M} \{s^{M\mu} Z_M' \hat{u}_\mu - t^{M\mu} [Z_M' \hat{u}_\mu]_{\perp X_I}\} \end{Bmatrix}$$

$$= \begin{Bmatrix} \begin{bmatrix} \sum_{\mu=1}^{M} \{t^{1\mu} [Y_1' \hat{u}_\mu]_{\| X_I} + (s^{1\mu} - t^{1\mu}) Y_1' \hat{u}_\mu\} \\ \sum_{\mu=1}^{M} s^{1\mu} X_1' \hat{u}_\mu \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \sum_{\mu=1}^{M} \{t^{M\mu} [Y_M' \hat{u}_\mu]_{\| X_I} + (s^{M\mu} - t^{M\mu}) Y_M' \hat{u}_\mu\} \\ \sum_{\mu=1}^{M} s^{M\mu} X_M' \hat{u}_\mu \end{bmatrix} \end{Bmatrix}$$

The bottom part of each block of ${}_{1}$ corresponds to the predetermined variables of the equation and is the same as for I3SLS. It is only the top part of each block (which corresponds to the jointly dependent variables of the equations) that differs. Further, if S^{-1}

were substituted for T^{-1} in the SML formula, even the top parts of corresponding blocks would be the same. ($T = (1/T)\hat{\alpha}_A[Z_A'Z_A]_{\perp X_I}\hat{\alpha}_A'$ where $[Z_A'Z_A]_{\perp X_I}$ is the moment matrix of the part of Z_A orthogonal to X_I. $S = (1/T)\hat{\alpha}_A[Z_A'Z_A]\hat{\alpha}_A'$.)

In the special case that the number of jointly dependent variables in the system equals the number of equations, the SML estimates become FIML estimates and the vector corresponding to m becomes (from V.37):

$$\iota = \left\{ \begin{bmatrix} T\hat{\gamma}^{\{1|1\}} + \sum_{\mu=1}^{M} s^{1\mu} Y_1'\hat{u}_\mu \\ \sum_{\mu=1}^{M} s^{1\mu} X_1'\hat{u}_\mu \end{bmatrix} \right\}$$
$$\vdots$$
$$\left\{ \begin{bmatrix} T\hat{\gamma}^{\{M|M\}} + \sum_{\mu=1}^{M} s^{M\mu} Y_M'\hat{u}_\mu \\ \sum_{\mu=1}^{M} s^{M\mu} X_M'\hat{u}_\mu \end{bmatrix} \right\}$$

where $\hat{\gamma}^{\{\mu|\mu\}}$ is formed from $\hat{\Gamma}^{-1}$ as indicated following (V.37). The number of observations, T, in the FIML expression should not be confused with the T matrix in the SML expression.

Notice that the blocks of the right hand side corresponding to predetermined variables coincide for I3SLS and FIML, but the blocks of the right hand side corresponding to jointly dependent variables differ for the two methods. For I3SLS, the jointly dependent variables are adjusted to variables which are asymptotically uncorrelated with the disturbances whereas for FIML, the jointly dependent variables are left unadjusted and the special nature of the jointly dependent variables is taken into account through the use of elements of the inverse

of the matrix of estimated coefficients of the jointly dependent variables. [This adjustment is derived as the appropriate adjustment through recognition of the Jacobian of the transformation of the likelihood from U to α and Z--see (V.12) and (V.13).]

Determining additional characteristics of the extreme point to which I3SLS converges would be very helpful in determining whether such iteration is worthwhile.[1,2] If it were determined that I3SLS estimation is worthwhile and we knew a definite function being maximized or minimized, the speed of convergence could be considerably increased. As a start toward increasing the speed of convergence, the step size could be made more optimal at each iteration. Possibly a more efficient metric could also be devised.

---

[1]Convergence was obtained for some problems and appeared to be slowly occurring in all other I3SLS problems computed on the AES STAT system.

[2]If we define I2SLS as the same procedure as ILIML (see section VI.E) except that 2SLS is used as the basic computational scheme rather than LIML, then I2SLS estimates apparently do not coincide with I3SLS estimates. Using a coefficient convergence criterion (see section V.C.5) of .000 000 000 1, 61 iterations were required to calculate I2SLS estimates for Klein's model I. The coefficients obtained were:

	C	P	W	CONSTANT	P_{-1}
Eq 1	-1	.2013	.7821	16.0510	.1282
	I	P	CONSTANT	P_{-1}	K_{-1}
Eq 2	-1	.1050	22.9420	.6253	-.1680
	W_1	E	CONSTANT	E_{-1}	t
Eq 3	-1	.3578	2.5265	.2130	.1815

Klein's model I is given in section I.C.a and the 2SLS, 3SLS, and I3SLS solutions to Klein's model I are given in the reproduced computer output of section IX.K.

2. Arbitrary linear restrictions imposed on coefficients

Restrictions may be imposed each iteration of the computation of the I3SLS coefficients in the same manner as for 3SLS.

PART III

ADDITIONAL PROGRAMMING CONSIDERATIONS

CHAPTER IX

ADDITIONAL PROGRAMMING CONSIDERATIONS

The previous 8 chapters contain basic formulas for computation of simultaneous stochastic equations methods. The way that these formulas are actually programmed on the computer has a very large effect on the actual coefficients obtained due to: (1) rounding error in performing the computations and (2) misspecification of the particular equation system to the computer. (I.e., the computer solves the equation system which is specified to it. This is very commonly not the equation system which the user thinks that he is specifying to it.)

In this chapter we will consider programming procedures which may considerably reduce rounding error as compared to the programming practices usually used. The user control cards used in the stochastic equations portion of the AES STAT package of computer programs[1] will be presented to illustrate a form of user control cards which results in considerably fewer errors and provides much more flexibility in the

---

[1]Except for a few subroutines and some recent modifications, the simultaneous stochastic linear equations part of the AES STAT system was programmed by the writer as a Department of Agricultural Economics research project. Although the programming was spaced out over a few years the actual programming required approximately one year of actual programming time from the writer. The remainder of the system (the DLS methods and methods other than the stochastic simultaneous equations portion) was developed as part of the writer's Agricultural Experiment Station programming responsibilities. Although the writer commenced programming the AES STAT system as the only programmer, his staff has been increased by the addition of approximately one-half of a programmer's time each year. Presently there are three full time programmers beside the writer expanding the system (primarily the analysis of variance and covariance and the least squares portions), writing user descriptions, consulting with members of the university in the use of the routines, and developing other routines not a part of this system (i.e., with a different form of parameters). Programmers who have worked on the AES STAT system at one time or another include Mr. Donald F. Kiel, Miss Mary E. Rafter, Mrs. Barbara Bray, Mrs. Marylyn A. Donaldson, Mr. Richard J. Martz, Mrs. Sara J. Paulson, Mr. Peter M. Schwinn, Mr. Frederick J. Ball, Mr. Tim Walters, and Mr. John Geweke.

range of problems which may be calculated than the user control cards used for most simultaneous stochastic equations computer programs.[1] To illustrate the control cards, the control cards required for calculating 2SLS, LIML, 3SLS, I3SLS, LML, and FIML for Klein's model I on the AES STAT system will be given along with the computer results obtained.

---

[1]The control cards are not presented as "ultimate" control cards, but merely a form of control cards which have performed considerably better than the control cards commonly used. Substantially superior forms of control cards will surely be devised in the future. Also, the control cards used in the AES STAT system are oriented toward a card input, batch-processing computer system. Since new forms of input and new concepts of processing are becoming available, it is desirable that the whole approach to specifying problems to the computer should change as well. For example, if teletypes or typewriters are used to specify a problem to a computer operating in a time sharing mode, a system of conversational controls should be devised to replace the form of control statements of the AES STAT system. The computer should query the user regarding aspects of the problem and after each user response, check for inconsistencies. The computer should notify the user of an inconsistency as quickly as it is detected, allow the user to correct the particular inconsistency detected, and proceed with the calculation of the problem.

A. Rounding Error

" . . . extreme care and precaution with accuracy of computation is not a fruitless and vain search for superfluous digits beyond the number significant in the original input. In spite of the fact that we have only two or three digits of significance in our original input variables and want only two or three-digit coefficients as an end result, we may have to carry out intermediate calculations to a very large number of places. The intermediate stages of equations systems methods of estimation are quite intricate, and if we do not carry out all of our results to many places and use the most accurate arithmetic procedures, we may find that our giant machines are spinning out masses of meaningless figures."[1]

Longley [1957] reported the results of calculating the DLS coefficients from data consisting of 16 observations with 6 independent variables and 8 dependent variables on the DLS routines which he felt were the most commonly used DLS routines in the U.S. Longley's results were startling. The most accurate routine had only 4 or 5 digits accuracy in the resulting coefficients, one routine had no digits correct for some coefficients, and one routine even had some signs of

---

[1]Klein and Nakamura [1962], pp. 298-299.

coefficients incorrect.[1]

Freund [1963] reported the results of calculating a small DLS problem on a series of computer routines and also obtained widely varying results.

If a large amount of rounding error is commonly encountered on small DLS problems, think how meaningless the results from many of even the most simple of the simultaneous stochastic equations estimating procedures must be. It is the purpose of this section to suggest computational methods for reducing rounding error so that meaningful results can be obtained.

---

[1] Longley calculated the "correct" coefficients on a hand calculator--carrying 15 digits at each step--and reported 8 places after the decimal point (plus from four to no places before the decimal point) in the resulting coefficients. The writer ran Longley's problems on the AES STAT system without making any special provision to reduce rounding error and obtained exactly the same coefficients as was reported by Longley, except for a couple of the coefficients corresponding to an overall constant in which the AES STAT overall constants disagreed at the ninth significant digit. (Subsequent experimentation with the data on the AES STAT system indicated that the AES STAT coefficients were correct to at least 15 places. Planned modification of the method of forming sums of squares and cross-products will raise the accuracy of the AES STAT system even further.)

1. Single vs. double precision

Most large scale computers currently being produced have both single precision floating point arithmetic and double precision floating point arithmetic built into the computer. The AES STAT system operates on the Control Data Corporation 3600 Computer which has a 48-bit word. Each floating point number is represented as follows in the CDC 3600 computer:

Single Precision--Each Word of Storage Contains One Number

sign bit	11 Exponent bits	36 mantissa bits ≃10 decimal digits

Double Precision--Two Consecutive Words of Storage Contain One Number

sign bit	11 Exponent bits	84 mantissa bits ≃24 decimal digits	

Each number is carried in the mantissa to a base of 2 rather than 10 and normalized so that "leading zeros" are not carried in the number. The location of the decimal point (to the base 2) is given by the exponent bits. Twice as many computer memory words are required to carry a matrix in double precision form, as to carry the matrix in single precision form. Also, for most computers, arithmetic operators require somewhat longer if they are performed in double precision than if they are performed in single precision.

As a result of the requirement of more storage if calculations are carried in double precision, and to a lesser extent as a result of the additional time required to calculate in double precision, many simultaneous stochastic equations routines are programmed in single

precision.[1] This saving of storage and time is generally a mistake. If computer time is the concern, then the programmer should be reminded that time lost in trying to find out why the coefficients "blew up" is time that should also be charged against the routine. Harder yet to evaluate is the case where (due to rounding error) none of the digits of the calculated statistics are computationally significant, but they are used anyway since the researcher is not aware of this.

If computer capacity is the problem, then the programmer should be reminded that it is on the larger problems that rounding error builds up the fastest; hence, that double precision is most needed. If there is insufficient capacity for an operation, then use of supplementary storage devices such as magnetic drums, disks, disk packs, or magnetic tapes in performing a series of operations is preferable to using single precision in order to provide sufficient storage.

Many programmers perform operations which they consider will not lead to a large amount of rounding error in single precision (e.g., a programmer may form the matrix of sums of squares and cross-products in single precision), convert the result to double precision, and then perform the more formidable computations which follow in double precision. Although little or no additional rounding error may result from such a practice in particular instances, these instances are far more rare than is generally realized.

The reason why this is generally an undesirable practice is that only a small proportion of the possible numbers convert evenly to the

---

[1]The simultaneous stochastic equations portion of the AES STAT system is programmed entirely in double precision.

power of 2 in the computer. Thus, even simple numbers like 1.1, 1.2, 1.3, 1.4, 1.6, etc. have a different representation in single precision than in double precision; that is, the double precision representation is not merely the single precision representation with a word of zeros attached. Thus, (in the case of the CDC 3600) the initial numbers will be accurate to 10 digits whether carried in single precision or double precision converted from single precision. Since rounding error starts from the last significant digit, rounding error will start to build up from 10 digits out instead of 24 digits out even if subsequent operations are performed in double-precision.[1]

One place that this is especially harmful is in the formation of a sums of squares and cross-products matrix with subsequent inversion of part of the matrix and multiplication by another part of the matrix in order to perform simple calculations such as obtaining direct least squares coefficients. Many people seem to regard the formation of a sums of squares and cross-products matrix as an operation in which little rounding error occurs (since the basic calculation is so simple); hence, they perform this operation in single precision. Since inversion of a matrix is considered to be a complicated procedure involving a lot of rounding error, the single precision matrix of sums of squares and cross-products is then converted to double precision and the inverse calculated. Unless the sums of squares and cross-products matrix is formed from variables with special characteristics

---

[1] An exception is an integral number such as 1,12,252, etc. An integral number is represented exactly by a single precision number; therefore, conversion to double precision of positive integers implies merely adding on a word of zeros. However, if division of one number by another is performed in single precision, the result will usually involve rounding to 10 places with subsequent conversion to double precision giving only 10 place accuracy.

(such as that all variables contain integral numbers, only) the rounding error battle has been lost even before inversion is started, since the rounding error will commence from somewhat less than 10 digits out during inversion rather than somewhat less than 24 digits as could have been obtained. In the case of a computer such as the IBM 360 which has only a 32 bit word (approximately 7 digits if floating point arithmetic is used) the use of single precision for some operations will have even a more pronounced effect on rounding error.

Multiplication of two matrices or a matrix and a vector is often regarded as a safe operation as compared to inversion, but one should be reminded that the formation of an element of the product requires many multiplications and additions if the matrices are large. If the numbers vary substantially in size, rounding error can be severe in even such a simple operation.

2. Standardization of variables

a. Deviations from means

The size of the mean of a variable has considerable effect on rounding error. Consider the following two variables:

x_1	x_2
100001.2763	1.2763
100002.7816	2.7816
100001.1471	1.1471
100003.0278	3.0278

If x_1 is used as an explanatory variable in a stochastic equation containing an overall constant coefficient and the coefficients of the equation are estimated by any of the simultaneous stochastic linear equations methods, the coefficient corresponding to x_1 will be the same as the coefficient which would be obtained if x_1 were replaced by x_2; only the overall constant coefficient would change if x_1 were replaced by x_2. However, sums of squares and cross-products can be calculated more accurately for x_2 than for x_1. Let us assume that the mantissa of each floating point number can contain 10 digits. Then each observation of x_1^2 will contain no digits to the right of the decimal point whereas each observation of x_2^2 will contain 9 digits to the right of the decimal point. Similarly cross-products of x_2 with other variables will contain more digits to the right of the decimal point than cross-products of x_1 with other variables.

We may regard x_2 as having been formed by subtracting 100000 from each observation of x_1. Even more accuracy would be obtained

if x_2 were formed by subtracting the mean of x_1 from x_1, since there would be even less superflous information to the left of the decimal. The new x_2 would again give the same coefficient (except for less rounding error in its calculation) as either x_1 or the original x_2. (Again the overall constant coefficient would change.)

So far the effects of the actual squaring and cross-product operations, only, have been considered. Summing operations in the formation of the sums of squares, sums of cross-products and sums of the original variables also contribute to rounding error, and again less rounding error would be produced during these summing operations if x_2 were used instead of x_1.

To illustrate the effect of summing on rounding error we will use a different example. To simplify our evaluation we will assume that each floating point number contains <u>exactly</u> 5 decimal digits in its mantissa and its exponent is carried to the base 10. Consider forming the sum of a variable, x_3, consisting of 100 observations with the first 50 observations assuming the value 10001 and the last 50 observations assuming the value 10003. The sum of the first 9 observations is 90009. When the 10^{th} observation is added, $10001 \cdot 10^1$ is obtained (i.e., the mantissa of the number is 10001 and the exponent is 1). When the 11^{th} observation is added we get $10001 \cdot 10^1 + 10001 \cdot 10^0 = 10001 \cdot 10^1 + 1000 \cdot 10^1 = 11001 \cdot 10^1$. (To add two positive floating point numbers their exponents are first equalized. This is accomplished by dividing the mantissa of the smaller number by 10^a where a is the difference between the exponents of the two numbers.) In like fashion as each of the observations which follow are added into the sum, the 1 or 3 is lost from the right hand side due to the equalization of

exponents. Thus, after adding in the 99^{th} observation we have $99001 \cdot 10^{1}$. When the 100^{th} observation is added we get $99001 \cdot 10^{1} + 10001 \cdot 10^{0} = 99001 \cdot 10^{1} + 1000 \cdot 10^{1} = 10000 \cdot 10^{2}$; whereas the exact sum is $10002 \cdot 10^{2}$.

If x_{t4} were formed as $x_{t3} - 10000$, x_4 would be 1 for the first 50 observations and 3 for the last 50 observations. The sum of x_4 would be accurately obtained as 200 (which is represented as $20000 \cdot 10^{-2}$).

The effect of summing on rounding error has been illustrated for a variable, but a similar effect is obtained if x_3 is the square of a variable or the cross-product of two variables during the formation of a sums of squares and cross-products matrix.

Let us assume that an overall constant coefficient appears in each stochastic equation in the subsystem to be estimated by any of the simultaneous stochastic linear equations methods. (We will consider the case of a stochastic equation containing no overall constant coefficient further on.)

Let M be the matrix of sums of squares and cross-products of all variables in a problem, that is:

$$M = [m_{ij}] , \quad m_{ij} = \sum_{t=1}^{T} x_{ti} x_{tj} , \tag{IX.1}$$

and let A be the matrix of sums of squares and cross-products of the deviations from their means of all variables in the problem, that is,

$$A = [a_{ij}] , \quad a_{ij} = \sum_{t=1}^{T} (x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j) \tag{IX.2}$$

$$= \sum_{t=1}^{T} x_{ti} x_{tj} - \bar{x}_i \sum_{t=1}^{T} x_{tj} .$$

[For a given problem, the two definitions of a_{ij} may not coincide due to rounding error; however, we are only defining the matrix A at this point. In section IX.A.2.c we will consider a more accurate formula than either of the formulas for a_{ij} in (IX.2).]

Also, let us define the variable x_0 as a variable which assumes the value 1 for all observations. Then the same estimates (except for rounding error) are obtained if (1) the variable x_0 is explicitly included in each equation and the M matrix is used as the Z'Z matrix in the computational formulas given in parts I and II of this paper as if (2) the variable x_0 is omitted from each equation, the A matrix is used as the Z'Z matrix in the computational formulas of parts I and II, and the overall constant coefficient for each equation is calculated as:

$$\hat{\alpha}_{\mu 0} = \bar{y}_{\mu} - \bar{Z}_{\mu}\hat{\alpha}_{\mu} = -[\bar{y}_{\mu} \vdots \bar{Z}_{\mu}]\begin{bmatrix} -1 \\ \hat{\alpha}_{\mu} \end{bmatrix} \tag{IX.3}$$

where $\hat{\alpha}_{\mu 0}$ is the overall constant coefficient for equation μ (the coefficient corresponding to variable x_0), $\bar{y}_{\mu}$ is the mean of the normalizing variable for the μ^{th} equation, $\bar{Z}_{\mu}$ is a row vector of means of the explanatory variables of the μ^{th} equation (not including x_0) and $\hat{\alpha}_{\mu}$ is the vector of coefficients of the μ^{th} equation (not including the normalizing coefficient, -1, and $\hat{\alpha}_{\mu 0}$).

If the A matrix is used as the Z'Z matrix then the disturbance variance-covariance matrix, the coefficient variance-covariance matrix, and statistics calculated from these matrices are calculated by the formulas of parts I and II except that (1) the constant coefficient is not included in the formulas and (2) if a "degrees of freedom" adjustment is made, the "degrees of freedom" should take account of

the implicit overall constant coefficient. For example the disturbance variance-covariance matrix may be estimated as:

$$S = \frac{1}{T} \hat{\alpha} A \hat{\alpha}' \quad ;$$

however, since A is calculated from deviations from means, the overall constant coefficients are omitted from the $\hat{\alpha}$ matrix. A degrees of freedom adjustment of $T/(\sqrt{T-n_{\mu}}\sqrt{T-n_{\mu'}})$ can be made in the S matrix as before, but here, n_{μ} and $n_{\mu'}$ each include the overall constant coefficient (i.e., if n_{μ} and $n_{\mu'}$ reflect the number of explanatory variables not including x_0, they are incremented by 1 in adjusting for degrees of freedom).[1]

---

[1] It is convenient to take this adjustment into account in the submatrices which output the variance-covariance matrices and related statistics rather than in the main computational section of the program. This is easily accomplished by setting a variable in COMMON to 0 if the M matrix is used and 1 if the A matrix is used and subtracting this variable from the "degrees of freedom" whenever a degrees of freedom calculation is made. It is also convenient to calculate overall constant coefficients by (IX.3) in the coefficient output subroutines also, rather than in the main program. An overall constant coefficient is calculated and printed out along with the other coefficients whenever the aforementioned variable in COMMON is 1.

b. Uniform scaling

It is sometimes (incorrectly) thought that if all numbers are carried in floating point form, the scaling of a variable up or down will have little substantive effect on the calculations, since the main effect of the scaling is the changing of exponents which designate where the decimal point occurs for each observation of the variable. The scaling of a variable does have a considerable effect on rounding error due to the effect of addition and subtraction of floating point numbers.

To add two positive floating point numbers, the exponents are first equalized. This is accomplished by shifting the mantissa of the smaller number to the right (dividing it by powers of 2) and adding to the exponent of this number until the exponents are equalized. What is left of the mantissa of the smaller number is added to the mantissa of the larger number and the resulting number is then normalized to eliminate leading zeros. Subtraction is performed in a similar manner. Thus, when addition and subtraction are performed, the sizes of the numbers involved are important. Since addition and subtraction are an integral part of the operations for all simultaneous equations calculations, rounding error is affected by the scaling of variables.

The basic information used from a set of variables is often contained in the matrix of sums of squares and cross-products of the deviations of the variables from their means [the A matrix (IX.2) of the previous section]. If some variables are scaled high and some low, the size of the elements of the A matrix may vary greatly in magnitude. Thus, if a typical element of the deviation from the mean

of one variable is about 1000 and a typical element of the deviation from the mean of another variable is about .01, the diagonal element of the A matrix of the first variable will tend to be about $(1000)^2/(.01)^2 = 10{,}000{,}000{,}000$ times as large as the diagonal element of the second variable. Control of rounding error in addition and subtraction is very difficult with such widely differing magnitudes of variables.

The above does not imply that the user of a computer program must try to scale all of his variables to the same magnitude, as this may impose a considerable burden on the user, both in setting up his data and in interpreting his results. Also, the madnitude of variables created in the computer by a prior transformation or editing procedure may be hard to predict. The above considerations suggest that the computer program should be sophisticated enough to handle automatically the scaling of data (and subsequent descaling of results) for the user.

Automatic uniform scaling of variables is easily accomplished. One method of accomplishing uniform scaling is to form the A matrix and then normalize the A matrix so that each diagonal element is 1. This is accomplished by multiplying the elements of row i and the elements of column i by d_i where

$$d_i = 1/\sqrt{a_{ii}} \quad . \tag{IX.4}$$

In matrix notation the operation may be represented as

$$A^* = DAD \tag{IX.5}$$

where A* is the A matrix normalized to 1 on the diagonals and D is a diagonal matrix whose i^{th} diagonal element is d_i .

Let $x^*_{ti} = d_i(x_{ti} - \bar{x}_i)$. Then (ignoring rounding error) A^* is the matrix of sums of squares and cross-products of the x^*_i variables. A characteristic of the A^* matrix is that each x^*_i variable has length 1--a very convenient normalization of variables.[1] Since x^*_i has mean zero and length 1, it is often referred to as a standardized variable. The A^* matrix is the usual simple (Pearson product-moment) correlations matrix.

All calculations are performed on the A^* matrix in the same manner as if it were the A matrix. Thus, many statistics such as coefficient estimates are carried in normalized form in the computer, thereby reducing rounding error in many calculations involving the coefficients. Only when statistics are printed out, must those statistics which are affected by a change of scale in the variables be denormalized. It is usually convenient to do the denormalization in output subroutines, thereby leaving the coefficients (or other statistics) in normalized form in the computer in case they are used for subsequent calculations.[2]

One might pose the question--why not first normalize the A matrix by multiplying rows and columns by powers of 10? The reasons are (1) it is as easy to normalize in the fashion indicated as by a

---

[1]The length of a vector, x_i, is defined as $\sqrt{x_i'x_i} = \sqrt{\sum_{t=1}^{T} x_{ti}^2}$.

[2]The AES STAT package uses one output subroutine for printing coefficients estimated by single equation methods, one output subroutine for printing coefficients estimated by multiple equations methods, one output subroutine for variance-covariance matrices (either disturbance or coefficient and in either denormalized form or further normalized so that 1's appear on the diagonal of the variance-covariance matrix as well), and one output subroutine to calculate and print estimated coefficient standard errors, coefficients divided by coefficient standard errors, and coefficient variances. All denormalizations are handled in the output subroutine thereby leaving the normalized coefficients and variance-covariance matrices unmodified in the computer.

power of 10, since a power of 10 is no more convenient than a number such as $1/\sqrt{\sum_{t=1}^{T} (x_{ti} - \bar{x}_i)^2}$ for the computer, (2) the knowledge that each standardized variable has length 1 (the diagonal element of the A* matrix are 1) is convenient in deriving certain computational formulas, and (3) the statistics affected by scaling variables will be completely denormalized before printing them out anyway.[1]

The normalization (or standardization) which we are imposing on the coefficients through normalization of the variables should not be confused with the normalization imposed on the coefficients of an equation by setting the coefficient of the "normalizing variable" to -1. The normalization to -1 is required for determinateness of the coefficients, whereas the normalization that we have been discussing in this section results in coefficients which are compatible with the normalized variables (and are therefore independent of scale of the original variables).[2]

If normalization of the A matrix is accomplished in the computer before computation of statistics is started, all computed statistics (before printing them out) will be uniformly scaled; that is, scaling of variables up or down by the user will have no effect on the basic A* matrix and, therefore, on the intermediate matrices and final statistics calculated. This uniform scale characteristic

---

[1] Normalization by a power of 2 is not advantageous either, since special programming would be required to adjust the exponent directly, and the extra operations would in general require more time than the use of direct multiplication for normalization and denormalization.

[2] In DLS estimation the normalized coefficients are often referred to as beta weights.

is very convenient in determining the rank of a matrix, whether a matrix being inverted is singular, whether certain variables are linear combinations of other variables during orthogonalization, the degree of convergence of an iterative procedure, etc.

Zellner and Thornber state:

> "A simple and relatively inexpensive check which we recommend is to perform calculations several times with the raw data scaled differently each run. Should the resulting sets of estimates differ, it is probable that computational errors are a problem. When faced with such results, an investigator might resort to higher precision arithmetic and/or ponder whether the information in his sample is adequate to provide even moderately good estimates of all of the parameters of his model."[1]

The writer concurs with the importance which Zellner and Thornber place on proper scaling under the assumption that the routine will not automatically scale the variables by normalizing the A matrix; however, even if several arbitrary scalings of the raw data are tried for a problem, none are likely to give as good a result as the normalization to length 1 of all deviations from means of variables which will be automatically accomplished by the procedure outlined earlier. Thus, although by normalizing the A matrix we have given up being able to affect the calculation by rescaling variables, we can expect to end up ahead rather than behind, and considerably more conveniently (from the standpoint of the user).[2]

---

[1] Zellner and Thornber [1966], p. 728.

[2] Also, to implement the suggestion of Zellner and Thornber, one must decide how much to rescale individual variables. If the alternative scalings are sufficiently pathological, differing estimates will surely be obtained.

c. Improving the estimates of sums, means, and the standardized moment matrix

In section IX.A.2.a we noted that if the mean of a variable is subtracted from each observation of the variable, a computer word can contain more meaningful information regarding the sum of the variable, the sum of squares of the variable, and the sum of cross-products of the variable with another variable. Let us also note the effect of rounding error on the calculation of means and a procedure for improving the accuracy of computed means, sums, and the standardization moment matrix.

First, let us define the set of original variables as the x_i and define a set of variables, the y_i, corresponding to the x_i as:

$$y_{ti} = x_{ti} - m_i \quad , \quad t = 1,\ldots,T \tag{IX.6}$$

where m_i is a constant subtracted from each observation of x_i in the formation of the corresponding variable y_i. A desirable choice for an m_i is the mean of the corresponding x_i, but we will not restrict an m_i to be an exact mean of the corresponding x_i. For example, an m_i might be a grossly inaccurate estimate of the mean of its corresponding x_i.

From (IX.6) we obtain:

$$x_{ti} = y_{ti} + m_i \quad , \quad t = 1,\ldots,T \quad ; \tag{IX.7}$$

hence:

$$\sum_{t=1}^{T} x_{ti} = \sum_{t=1}^{T} y_{ti} + T\, m_i \tag{IX.8}$$

$$\bar{x}_i = (1/T) \sum_{t=1}^{T} x_{ti} = \bar{y}_i + m_i \quad . \tag{IX.9}$$

If m_i is an approximation to the mean of x_i we can expect to compute $\bar{x}_i$ much more accurately as $\bar{y}_i + m_i$ than as $(1/T)\sum_{t=1}^{T} x_{ti}$, and we can expect to form the sum more accurately as

$$\sum_{t=1}^{T} x_{ti} = \sum_{t=1}^{T} y_{ti} + Tm_i \text{ , i.e., by (IX.8).}^1$$

---

[1] The arithmetic mean of the x_{ti} is, by definition, $\bar{x}_i = (1/T)\sum_{t=1}^{T} x_{ti}$. However, in calculating the mean by this definitional formula, we often do not get $\bar{x}_i$ exactly. Let m_i be the number thus actually obtained as an approximation to the mean, and let e_i be the rounding error, so we have $m_i - \bar{x}_i = e_i$, where m_i is a known (calculated) quantity but, in general, the exact values of $\bar{x}_i$ and e_i may not be known. Now consider another approximation to $\bar{x}_i$, say m_i^*, obtained as follows:

Compute (1) $y_{ti} = x_{ti} - m_i$, $t = 1,\ldots,T$;

(2) an approximation to the mean of the y_{ti}, obtained by applying the definitional formula $\bar{y}_i = (1/T)\sum_{t=1}^{T} y_{ti}$; let q_i be the result of this calculation, and f_i the corresponding rounding error, so we have $q_i - \bar{y}_i = f_i$;

and (3) $m_i^* = q_i + m_i$.

The resulting error, i.e., the difference between m_i^* and $\bar{x}_i$, is

$$\begin{aligned} m_i^* - \bar{x}_i &= q_i + m_i - \bar{x}_i \\ &= \bar{y}_i + f_i + m_i - \bar{x}_i \\ &= \bar{x}_i - m_i + f_i + m_i - \bar{x}_i \\ &= f_i \,. \end{aligned}$$

Thus, m_i^* will be a better approximation to $\bar{x}_i$ than m_i is, whenever the magnitude of f_i is less than that of e_i. Now the magnitude of a rounding error is, on the average, roughly proportional to the magnitude of the quantity being rounded off. (For example, in the decimal system, if n significant digits are carried throughout, the average magnitude of rounding error is roughly $.5 \times 10^{-n}$ times the magnitude of the number rounded off.) Since y_i is always close to zero, its magnitude is generally much smaller than that of x_i (unless x_i is already very close to zero), and hence the magnitude of f_i is generally much smaller than that of e_i, and m_i^* is generally a much better approximation to $\bar{x}_i$ than is m_i.

To illustrate the improvement which may be obtained through use of (IX.8) and (IX.9), we will return to one of the examples in section IX.A.2.a. In this example we used a variable, x_3, consisting of 100 observations, the first 50 observations assuming the value 10001 and the last 50 observations assuming the value 10003. We also assumed that each floating point number contains exactly 5 decimal digits and that its exponent is carried to the base 10. We then computed the sum of the variable to be $(\sum_{t=1}^{T} x_{t3})_{\text{1st pass}} = 10000 \cdot 10^2$. Thus,

$$m_3 = (\sum_{t=1}^{T} x_{t3})_{\text{1st pass}}/T = 10000 \cdot 10^2/100 = 10000$$

$$y_{t3} = x_{t3} - m_3 = \begin{cases} 10001 - 10000 = 1 & \text{for } t = 1,\ldots,50 \\ 10003 - 10000 = 3 & \text{for } t = 51,\ldots,100 \end{cases}$$

$$\sum_{t=1}^{100} y_{t3} = 200$$

$$\bar{y}_3 = 2$$

$$(\sum_{t=1}^{T} x_{ti})_{\text{2nd pass}} = \sum_{t=1}^{T} y_{ti} + Tm_3$$

$$= 200 + 100 \cdot 10000 = 1000200$$

$$\bar{x}_3 = \bar{y}_i + m_i = 2 + 10000 = 10002 .$$

Thus, by computing the sum and mean of x_3 in two passes, we obtained the more accurate sum 1000200 and the more accurate mean 10002.

In a Monte Carlo study, Neely [1966] used an m_i of $(1/T) \sum_{t=1}^{T} x_{ti}$ and compared computations of the means of variables through use of the formulas $\bar{y}_i + m_i$, $(1/T) \sum_{t=1}^{T} x_{ti}$, and some additional formulas. Neely found that the formula $\bar{y}_i + m_i$ performed at least as well as the

other formulas he tried and was superior to the direct computation $(1/T)\sum_{t=1}^{T} x_{ti}$.

In section IX.A.2.a we noted that the simple correlations matrix A^* is a convenient normalized moment matrix to substitute for the $Z'Z$ matrix in the formulas given in parts I and II and that this matrix has desirable properties for such a substitution. The same A^* is obtained if the y_i are substituted for the x_i except that the A^* matrix for the y_i will in general be formed more accurately.

To denormalize the statistics affected by the normalization used, we must note what normalization elements [see (IX.2) and (IX.4)] are implied by the substitution of the y_i for the x_i. If we had used the x_i directly, we would use $1/\sqrt{\sum_{t=1}^{T} x_{ti}^2 - \bar{x}_i \sum_{t=1}^{T} x_{ti}}$ as the normalization elements. Now note that

$$\text{(IX.10)} \qquad \begin{aligned} \sum_{t=1}^{T} x_{ti}^2 - \bar{x}_i \sum_{t=1}^{T} x_{ti} &= \sum_{t=1}^{T} (x_{ti} - \bar{x}_i)^2 \\ &= \sum_{t=1}^{T} [(y_{ti} + m_i) - (\bar{y}_i + m_i)]^2 \\ &= \sum_{t=1}^{T} (y_{ti} - \bar{y}_i)^2 \\ &= \sum_{t=1}^{T} y_{ti}^2 - \bar{y}_i \sum_{t=1}^{T} y_{ti} \ ; \end{aligned}$$

hence, we can use the same denormalization elements as if the y_i had been the original variables (the x_i).

In his paper, Neely [1966] also compared computations of simple correlations matrices through use of the y_i [with an m_i of $(1/T)\sum_{t=1}^{T} x_{ti}$] with computations through direct use of the x_i and with computations using other formulas. Neely found that the usual simple correlation

formula but using the y_i in place of the x_i performed at least as well as the other formulas he tried and was superior to the direct use of the x_i.

The preceding implies that, in general, accuracy can be increased by making two passes through the data.[1] In the first pass, an approximation of the mean of each variable is obtained (this approximation is improved the second pass) and in the second pass the sums of squares and cross-products matrix of variables from their approximate means is formed and the sums of the deviations of the variables from their approximate means are obtained.[2] Finally (1) the simple correlations matrix is formed from this newly formed sums of squares and cross-products matrix (ignoring the fact that the means and sums of the variables from which it was formed are approximately zero) and (2) the approximate sums and the approximate means of the x_i variables are adjusted by (IX. 8) and (IX.9) to provide more accurate computations of these quantities.

Provision should be provided for the user to specify a set of approximate means and he should be allowed to specify that only one pass be made through the data. For example, it is planned that the AES STAT system will be modified so that two passes are made through

---

[1] Longley [1967] contains a report on rounding error by many well-known DLS routines. Longley documents the improvement in accuracy obtained through making two passes through the data and the desirability of using accurately computed simple correlation matrices and accurately computed denormalization elements in the computation of DLS problems.

[2] It is convenient to incorporate the extra variable x_0 (a variable assuming the value 1 for all observations) whenever a sums of squares and cross-products matrix is formed. The sums of the variables used in forming the matrix are then automatically calculated as the elements corresponding to x_0.

the data unless the user directs that only a single pass be made through use of the ONEPASS code on the SSCP card. The user will also be permitted to specify his own approximate means (m_i) for any variables during that pass. Use of the ONEPASS code and specification of no approximate means by the user implies that only a single pass will be made through the data with all m_i equal to zero, i.e., the usual simple correlation formulas will be used.

d. Adjustments if no overall constant coefficient

If no overall constant coefficient is to be included in an equation, the sums of squares and cross-products matrix [the M matrix defined by (IX.1)] normalized in some fashion is used as Z'Z in the formulas given in parts I and II of this paper rather than the simple correlations matrix [the A* matrix defined by (IX.5), (IX.4), and (IX.2)]. From (IX.2) we recall that:

$$a_{ij} = m_{ij} - \bar{x}_i \sum_{t=1}^{T} x_{tj} \quad ; \tag{IX.11}$$

hence, m_{ij} may be formed as:

$$m_{ij} = a_{ij} + \bar{x}_i \sum_{t=1}^{T} x_{tj} \quad . \tag{IX.12}$$

If the $\bar{x}_i$ and the $\sum_{t=1}^{T} x_{tj}$ are normalized by the same normalization as is used for the A matrix in forming the A* matrix [i.e., $d_i \bar{x}_i$ and $d_j \sum_{t=1}^{T} x_{tj}$ are formed where the d_i are the normalization elements defined by (IX.4)], then the M matrix normalized (say M*) may be formed directly from the A* matrix as:

$$m^*_{ij} = a^*_{ij} + (d_i \bar{x}_i)(d_j \sum_{t=1}^{T} x_{ij}) \tag{IX.13}$$

where $M^* = [m^*_{ij}]$.

This normalization is based on the deviations of each variable from its mean having length 1 rather than each variable having length 1; hence, the elements on the diagonal of M* will not be 1 as is the case for A*. One could do an additional normalization by multiplying the i^{th} row and the i^{th} column by c_i where $c_i = 1/\sqrt{m^*_{ii}}$ so that each variable will have length 1 (1's will appear on the diagonal of the M* matrix); however, if this is done, statistics affected by the

normalization must be denormalized based on a normalization of $c_i d_i$ rather than d_i. If a special normalization is used for the M* matrix and some estimates are based on the A* matrix and other estimates are based on the M* matrix, additional bookkeeping must be kept by the computer routine regarding which normalization was used to compute a set of statistics used in a later step (e.g., DLS coefficients may be used as starting estimates for FIML) or the statistics based on one of the normalizations must be renormalized as they are stored. The advantages of using a special normalization for the M* matrix would not, in general, appear to warrant the additional programming required.

Rather than form the M* matrix directly or for that matter form it indirectly once and for all and hold it in the computer memory or on some auxillary storage device, it is much more convenient (and in many cases, saving of computer time) to form from the A* matrix the portion of the M* matrix needed to calculate a given phase of a problem. It is planned that the AES STAT system will be revised so that it is assumed that an overall constant coefficient is to be included in each equation unless the user specifies a NOCON (for no overall constant coefficient) code on the relevant STAT control card. For example, k-class estimates will be based on the A* matrix and an overall constant coefficient will automatically be calculated (hence, the x_0 variable need not be specified on a given K card) unless the NOCON code appears on the K card.[1] If the NOCON code appears on the K card, only the part of the A* matrix corresponding to variables appearing on the K card will be extracted from the A* matrix. The M* matrix

---

[1]The K card is discussed in sections IX.B and IX.K.

corresponding to these variables will then be formed in the manner given by (IX.12) and used as the $Z'Z$ matrix in the calculation of the k-class problem.

3. Use of simultaneous equations solutions

Klein and Nakamura state, "In the evaluation of $Y'X(X'X)^{-1}X'Y$ we find it more efficient and accurate, from a computational point of view, to calculate $(X'X)^{-1}X'Y$ (as the solution vectors of sets of simultaneous equations) and then premultiply the $Y'X$. Criteria for success are judged by the positive definiteness and symmetry of $[Y'Y]_{\perp X}$. Also inconsistencies in subsequent dependent calculations indicate arithmetic errors at this stage."[1]

Calculation of $Y'X(X'X)^{-1}X'Y$ as $Y'Y - [Y'Y]_{\perp X}$ where $[Y'Y]_{\perp X}$ is obtained by direct orthogonalization (see section I.D.2) is more accurate yet; however, Klein and Nakamura's basic point is well taken. If the result of an inverse times a matrix or vector is desired, this result can be computed more accurately by treating the calculation as a "solution to a set of simultaneous equations" rather than by actually forming the inverse and performing the matrix or vector multiplication.

A slight problem presents itself when the inverse of the matrix is desired in its own right such as for the calculation of variance-covariance estimates; however, this problem can be resolved by using a subroutine which calculates the inverse and simultaneous equations solution at the same time.

---

[1]Klein and Nakamura [1962], p. 287. Klein and Nakamura used the notation $M_{yx}M_{xx}^{-1}M_{xy}$ instead of $Y'X(X'X)^{-1}X'Y$, $M_{xx}^{-1}M_{xy}$ instead of $(X'X)^{-1}X'Y$, M_{yx} instead of $Y'X$, and $M_{\hat{y}\hat{y}}$ instead of the $[Y'Y]_{\perp X}$ used in the above quote.

For either inversion or simultaneous solution, the selection of the largest element of the remaining submatrix to be operated on at each step as the pivot for the step considerably reduces rounding error for many problems.[1]

---

[1]For positive definite or positive semi-definite matrices, the largest element of the remaining matrix to be operated on at each step will occur on the diagonal of the remaining sub-matrix.

4. Direct orthogonalization

The use of direct orthogonalization in various phases of the computation has been emphasized throughout this paper. (A method for accomplishing direct orthogonalization is given in section I.D.2.) Here we note that direct orthogonalization may considerably reduce rounding error as compared to calculation by the usual method, e.g., by $Y'Y - Y'X(X'X)^{-1}X'Y$. Also matrices of the form $Y'X(X'X)^{-1}X'Y$ may be calculated more accurately as $Y'Y - [Y'Y]_{\perp X}$.

The rearrangment of rows and columns at each step so that the pivot is selected as the largest element of the remaining sub-matrix (this largest element will occur on the diagonal of the remaining sub-matrix) at each step will considerably reduce rounding error for many problems.

5. Iterative techniques

There seems to be a fairly widely held view that rounding error builds up as iteration continues in the calculation of certain solutions. In particular, it is often thought that, even though rounding error is small in the computations performed for a single iteration, FIML coefficients may be subject to quite a bit of rounding error if quite a few iterations are required for convergence. Now, continued iteration may cause difficulty in some iterative procedures, and formulas could certainly be devised which would cause rounding error to build up in the calculation of FIML coefficients; however, rounding error will not build up during iteration if the formulas given in this paper are used. This is because during an single iteration, all calculations are based only on a matrix of sums of squares and cross-products (which does not change as iteration progresses) and on the set of coefficients obtained from the last iteration. In order to move to a higher likelihood, adjustments to these coefficients are then calculated, and added to the coefficients and a new iteration is started. If the increment for a coefficient is in the wrong direction, the procedure itself will correct the coefficient at some later step (assuming that rounding error is small for the computations performed in any single iteration) as the likelihood will not be maximized until the coefficient is corrected. Thus, assuming that convergence is obtained to a desired degree of accuracy, the particular coefficients obtained are a function of the matrix of sums of squares and cross-products and the assumed structure (neither of which change as iteration progresses) rather than the

intermediate coefficients and matrices obtained in the process of iterating.[1]

[1] If the likelihood has multiple local maxima, the particular intermediate coefficients will have an influence on which maximum is obtained; however, given that we end up on a particular peak, continued iteration will not cause the coefficients which maximize the likelihood in that region to be estimated less accurately.

B. Free Field Interpretive Parameters

A very common way to give instructions to a computer routine is by a series of parameter cards containing information punched by the user in essentially the following form:

Card 1:

Cols. 5- 7: Number of rows in the matrix

Cols. 8-10: Number of columns in the matrix

etc.

Cards 2-10:

Cols. 4- 5: Number of elements on the card

Cols. 7- 8: First variable number

Cols. 9-10: First column number

Cols. 11-20: First element

Cols. 21-22: Second variable number

Cols. 23-24: Second column number

Cols. 25-34: Second element

etc.

Card 11:

Col. 5: Punch a 1 if 3SLS estimates are to be calculated from LIML coefficients.
Leave blank if 3SLS estimates to be calculated from 2SLS coefficients.

etc.

The use of this type of parameter card causes an unnecessarily large number of errors. Here a number of cards must be prepared with particular information going in particular columns, the cards differing somewhat from each other, no parameter being expressed in a form natural to the user. It is very easy (and common) to get some of the

information punched in the wrong columns. Now, some of the mispunched information may be detected by the routine as being impossible. This case may cause computer time to be wasted and results to be delayed; however, this is not the most serious problem. The serious problem is that with this type of parameter card, it is very easy to mispunch a card in such a way that the parameters still make sense to the routine. An answer is obtained, but it is the answer to a different problem than the user intended. Often the user will then go ahead and use the results.

One might be comforted by the thought that the probability is very low that (1) a card will be mispunched by a careful user who double-checks everything, and (2) if a mispunch occurs, it will appear correct to the routine, and (3) if the mispunch occurs and the routine accepts it, the answer will be within a range that a user skilled in the method will accept as correct. If so, one is falsely comforted. The probability is quite high.[1]

The more usual situation is one in which a user with little knowledge of the importance of checking his parameters and data calculated a problem by one or more of the simultaneous stochastic equations estimating procedures. Often the user has had little experience in interpreting results from the methods (after all, we have to start sometime) and so he expects almost anything. The probability of getting

---

[1]The writer has had occasion to take results of problems calculated by researchers who are far above average in carefulness of preparing cards and skill in interpreting results, but he has rarely obtained the same answer on his routines. Further checking usually disclosed an invorrect number punched or a code or number punched in the wrong column(s). The researchers had already reported their incorrect results.

an answer to a different problem than the one he actually has in mind (using the above type of parameter cards) is even higher than for a user who is aware of the need for double checking and has had much experience in checking results. Often the user makes a comparison of a series of models. In this case the probability must approach 1 that he will get some incorrect results due merely to mispunched parameters.

One way to reduce the probability of misspecifying the parameters is to use an interpretive form of parameter card that is free-field, i.e., no numbers have to be punched in particular columns. In the accompanying listing of computer control cards and following explanatory paragraphs, interpretive free-field instructions to the computer are illustrated with the parameters and data which would be used to estimate Klein's model I by DLS, 2SLS, LIML, 3SLS, I3SLS, LML, and FIML on the AES STAT system. (A number of other methods are available in the AES STAT system, but this should be adequate to illustrate the parameters.) The results are reproduced in the last section (section IX.K) of this chapter.

The SSCP card causes variables X_1 through X_{15} to be formed into a sums of squares and cross-product matrix. Certain basic statistics regarding the variables and pairs of variables are automatically calculated and printed out. The TRANS code causes the data to be transformed as indicated by the transformation cards which are discussed in the next section. The RES code causes the transformed data from X_1 through X_{15} and X_{20} to be saved for later calculation and labeling of residuals. The MAXE=8 (maximum equation number is equation 8) code sets up a coefficients pool for temporary storage and retrieval of coefficients. Quite a few additional codes are permitted on the SSCP

```
JOB,661682,FIML,5,RUBLE,BILL  KLEIN MODEL I ON 2SLS,LIML, 3SLS, I3SLS, LML,FIML
DEMAND,MT
EQUIP,20=MT(STAT)
FILE,69
MAIN,21
FILE END
FTN,L,Y,X
      SUBROUTINE TRANS1(X)
      COMMON/A/NR,NT,IS,IDR,JP  $ DIMENSION X(200)
      DOUBLE PRECISION X
      X(14)=X(4)*X(5)
      X(15)=X(2)*X(12)
      X(20)=NR*1920
      RETURN
      END
ROUTINE,SSCP,K,READC,3SLS,FIML
RUN,5,35000,2
U     SET UP DATA--KLEIN MODEL I
L  X(1)=C,I,W1,P,W,E
L  X(14)=Y,K
L  X(7)=W2,R,G,TIME,P LAGGED,K LAGGED,E LAGGED
SSCP(X(1),,,X(15)/X(20))TRANS,RES,MAXE=8
FORMAT(6X,6F5,F3,2F5,F3,2F6,F5)
01 21041.90-0.2025.5012.4028.2045.602.707.7006.60-1012.7182.8044.90
02 22045.0001.9029.3016.9032.2050.102.903.9006.10-0912.4182.6045.60
03 23049.2005.2034.1018.4037.0057.202.904.7005.70-0816.9184.5050.10
04 24050.6003.0033.9019.4037.0057.103.103.8006.60-0718.4189.7057.20
05 25052.6005.1035.4020.1038.6061.003.205.5006.50-0619.4192.7057.10
06 26055.1005.6037.4019.6040.7064.003.307.0006.60-0520.1197.8061.00
```

"Form cards" available to all users in prepunched form for insertion into their decks.

"transformation cards" to calculate or edit variables or observations

"Form cards"

tells basic procedures to be used

used in labeling results

Variables labels specified by the user. Used to label output more thoroughly.

Forms moment matrix of sums of squares and cross-products

Gives format of the data

data

C I W1 P W E W2 R G TIME P LAGGED K LAGGED E LAGGED

```
                                                                  P LAGGED      E LAGGED
        C      I      W1     P      W      E      W2     R      G   TIME      K LAGGED
07 27056.2004.2037.9019.8041.5064.4053.6067.0007.60-04019.6203.4064.00
08 28057.3003.0039.2021.1042.9064.5053.7042.2007.90-03019.8207.6064.40
09 29057.8005.1041.3021.7045.3067.0044.0048.1008.10-02021.1210.6064.50
10 30055.0001.0037.9015.8042.1061.2042.2077.0009.40-01021.7215.7067.00
11 31050.90-3.4034.5011.4039.3053.4048.0750.1010.70000015.6216.7061.20
12 32045.60-6.2029.0007.0034.3044.3055.3083.3010.20001011.4213.3053.40
13 33046.50-5.1028.5011.2034.1045.1056.6054.4009.30002007.0207.1044.30
14 34048.70-3.0030.6012.3036.6049.7066.0068.8010.00003011.2202.0045.10
15 35051.30-1.3033.2014.0039.3054.4067.1072.2010.50004012.3199.0049.70
16 36057.7002.1036.8017.6044.2062.7074.0883.3010.30005014.0197.7054.40
17 37058.7002.0041.0017.3047.7065.0067.7067.7011.00006017.6199.8062.70
18 38057.50-1.9038.2015.3045.9060.9077.7074.4013.00007017.3201.8065.00
19 39061.6001.3041.6019.0049.4069.5078.8089.9014.40008015.3199.9060.90
20 40065.0003.3045.0021.1053.0075.7088.0096.6015.40009019.0201.2069.50
21 41069.7004.9053.3023.5061.8088.4085.1166.2022.30010021.1204.5075.70
 END OF DATA
U     KLEIN MODEL I--2SLS AND LIML -- EQ 1
K(1,4,5/0,11/7,,,10,12,13)
          2SLS=C,DV,SCE=1,STE,COSTE
          LIML=C,DV,COSTE,STE
U     KLEIN MODEL I--2SLS AND LIML -- EQ 2
K(2,4/0,11,12/7,,,10,13)
          2SLS=C,DV,SCE=2,STE,COSTE
          LIML=C,DV,COSTE,STE
U     KLEIN MODEL I--2SLS AND LIML -- EQ 3
K(3,6/0,13,10/7,,,9,11,12)
          2SLS=C,DV,SCE=3,STE,COSTE
          LIML=C,DV,COSTE,STE
```

data

calculate equation 1 by 2SLS and LIML

calculate equation 2 by 2SLS and LIML

calculate equation 3 by 2SLS and LIML

```
U     KLEIN MODEL I--IDENTITY EQUATIONS
READDC(14,1,2/8,9)C,DV,SCE=4
FORMAT(5F2)
-1.1.1-1.1
READDC(14,4,5)C,DV,SCE=5
FORMAT(3F2)
-1.1.1
READDC(15,2/12)C,DV,SCE=6
FORMAT(3F2)
-1.1.1
READDC(5,3/7)C,DV,SCE=7
FORMAT(3F2)
-1.1.1
READDC(6,14/8,7)C,DV,SCE=9
FORMAT(4F2)
-1.1.1.1
U  3SLS, I3SLS STARTING FROM 2SLS--KLEIN MODEL I
3SLS(1,2,3/4,,,8/X(7),,,X(9))BEG=C,LIK,RF,DVC,DVCN
                             3SLS=C,LIK,DVC,DVCN,CVC,CVCT,CVCN,RES,NR=20,COSTE
                                  RF,STE
                             I3SLS=C,LIK,DVC,DVCN,CVC,CVCT,CVCN,COSTE,STE
                             NTH=20=C,LIK
                             MI=200,CCC=1.0-10,TIMEL=10
U  LML, FIML STARTING FROM 2SLS--KLEIN MODEL I
FIML(1,2,3/4,,,8)
                  LML=C,DVC,DVCN,CVC,CVCT
                  NTH=10=C
                  FIML=C,DVC,CVC,CVCT,CVCN,DVCN,RFRES,NR=20,RF,STE,COSTE
                  LIMIT=C,TIMEL=5,PCC=1.0-10,MI=50
END OF RUN
```

specify coefficients of 1st identity equation

specify coefficients of 2nd identity equation

specify coefficients of 3rd identity equation

specify coefficients of 4th identity equation

specify coefficients of 5th identity equation

calculate system by 3SLS and I3SLS

calculate system by LML and FIML

card to control data input, cause matrices to be printed out, etc. The SSCP card like all of the other control cards is completely free-field, blanks being permitted anywhere on the card.

The first K card sets up the basic matrices for any double k-class member for equation 1 with X_1, X_4, and X_5 as jointly dependent variables; X_0 and X_{12} as predetermined variables in the equation; and X_7 through X_{10}, X_{12}, and X_{13} as additional instrumental variables. The codes following the K card specify the particular k-class members to be calculated. The 2SLS code designates that 2SLS coefficients are to be calculated and the LIML code designates that LIML coefficients are to be calculated. (Any specific k, double k, or h-class member or series of members can be calculated by specifying k, k_1 and k_2, or h. UBK coefficients can also be calculated by using an UBK code.) The particular statistics to be printed out are designated by codes (e.g., C for coefficients, STE for standard errors) following the codes telling which estimators to calculate. The SCE=1 code following the 2SLS code designates that the 2SLS coefficients are to be stored in a coefficients pool and labeled equation 1.

The other two K cards are similar to the first in specifying statistics to calculated for equations 2 and 3. The READC cards are used to input the identity equations, and the 3SLS and FIML cards are used to calculate the 3SLS, I3SLS, LML, and FIML estimates. For iterating on the 3SLS coefficients in the calculation of the I3SLS coefficients, various stopping criteria have been specified. In this case iteration will stop when the first of the following occurs:

(1) The proportional change in all coefficients is less than

.0000000001 (CCC=1.0-10 code).

(2) The number of iterations exceeds 200 (MI=200 code).

(3) Iteration proceeds for more than 10 minutes (TIMEL=10 code).

In the computer output given further on it can be noted that the 1st criterion was satisfied in 42 iterations with the total time being 12 seconds.

The NTH=20 code causes the statistics listed after the code to be printed out each 20th iteration. In this way, the progress of iteration can be examined for problems requiring a large number of iterations.

The convergence or stopping criteria given for 3SLS are available on FIML with all of the additional stopping criteria listed in section V.C.5 available for FIML as well. The NTH=10 code causes certain statistics specified by codes to be printed out each 10^{th} iteration. Additional information regarding the form of the 3SLS and FIML cards are given in a later section (section IX.D) on the "coefficients pool."

The codes (and the entire form of the control cards) are quite arbitrary. The important features of the control cards just discussed are:

(1) The control cards are free-field; hence, there is no need to punch particular information in particular columns.

(2) An individual code may be mispunched just as a number may be mispunched into an incorrect column in the form of parameters given earlier; however, with the form of the codes given here, a mispunched code may be readily detected by the computer routine whereas there may be no way that a number punched in the wrong column can be detected by the computer routine.

(3) The control cards are open ended. If a new feature is added to the routine, a new form of control card or a new code may be easily entered to permit the user to use the new feature. (With the other type of control cards, insufficient free columns may be available, hence, additional control cards may be required; however, this may cause trouble to those who don't know about the requirement of the additional control cards. Also, a set of control cards used to calculate a problem prior to the change can no longer be used without updating the set.)

No matter what form of control cards are used, they should be printed out as they are encountered to provide a permanent record of the particular control cards used to calculate the problem.

C. Data Transformation Section

It is very easy to provide a lot of facility for transforming and editing data by providing a call to a subroutine (to be specified by the user) after each observation is read into the computer. Convenience to the user is enhanced by selecting carefully the arguments transferred to the subroutine and the variables provided in COMMON blocks to the user.

In the parameter cards given above for calculating Klein's model I, the FTN, SUBROUTINE, COMMON, DOUBLE PRECISION, RETURN, and END cards are described to the non-programmer user as "transformation form cards" which must be inserted whenever a transformation subroutine is used and copies of these cards are kept on hand at all times by the MSU Computer Laboratory. The user merely helps himself to the prepunched form cards instead of punching his own. A manual giving simple transformations is available to the non-programmer user to help him accomplish his transformation and editing. (A user with knowledge of programming recognizes that the full power of the FORTRAN compiler is available in accomplishing the transformations.)

The transformation cards are inserted between the DOUBLE PRECISION form card and the RETURN form card. The first transformation card in our illustration creates X_{14} as $X_4 + X_5$; the second transformation card creates X_{15} as $X_2 + X_{12}$; and the last transformation card creates X_{20} as the raw observation number (NR) plus 1920. (X_{20} is used further on to number each residual by the year of the particular observation.)

The variables on the COMMON card aid in such things as dropping observations, stopping date input, branching to particular sections if

multiple SSCP cards are used, lagging variables, calculating moving averages, and printing out data labeled by observation number. Additional COMMON cards are available to assist the user in particular transformation tasks; however, they are rarely required.

The transformation section also provides an open ended method of reading data from miscellaneous storage devices.

D. Coefficients Pool

In the AES STAT system, a coefficients pool is automatically established within the computer into which coefficients from any single or multiple equations estimating procedure may be stored. Also stored with the coefficients are a record of the method creating the coefficients, which coefficients pertain to jointly dependent variables and which to predetermined variables and the particular variable numbers of the coefficients.[1]

Various control cards are available for retrieving the coefficients and making calculations based on them. In particular they are retrieved for the 3SLS and FIML calculations in the illustration. As examples of the types of parameters available, the SCE=1 (save coefficients as equation 1) code following the 2SLS code on the first K card designates that the 2SLS coefficients from the first equation are to be stored as equation 1. Similarly, the SCE codes on the remainder of the K (k-class estimates) cards and on the READC (read coefficients) cards are used to assign coefficients to equation numbers. READC cards are used to input identity equations for Klein's model I; however, they could be used to input any set of coefficients.

The numbers following the "3SLS(" but before the first "/" on the 3SLS card give the equation numbers of coefficients to use as starting estimates for the 3SLS estimating procedure and, by implication, the structure of the 3SLS system. The numbers between slashes designate identity equations in the system. These identity equations

---

[1] Coefficients from the pool may be reclassified or the equation renormalized at any time through the use of RECL cards.

are not used by the 3SLS estimating procedure and could be omitted except that the RF (reduced form coefficients) code designates that reduced form coefficients are to be calculated and printed and the RFRES (reduced form residuals) code designates that reduced form residuals are to be calculated and printed; hence, the identity equations are required to complete the system so that reduced form coefficients and residuals can be calculated.

The variable numbers following the slash are instrumental variables to be used in the 3SLS estimation procedure in addition to the predetermined variables in the subsystem being estimated (equations 1 through 3).

If the SCOE code had been used after the 3SLS code, the 3SLS coefficients would have replaced the 2SLS coefficients in the coefficients pool and therefore the FIML procedure would have started from the 3SLS coefficients instead of the 2SLS coefficients.

E. Special Files

1. Data files

In the AES STAT system, two data files are available under full control of both the program and the user. Either raw or transformed data may be stored in these files by user control codes. If sufficient capacity is available in the main core memory, these files will be carried in memory; otherwise the routine automatically establishes them on the magnetic drum or (at the user's option) on magnetic tape. Any number of additional data files may be established by the user in which he exercises primary control through use of FORTRAN statements and executive functions in the transformation subroutine.

2. Intermediate storage files

Two intermediate storage files are presently used by the AES STAT package to (1) store information not needed for the immediate calculation (as examples, the coefficients pool, and the full sums of squares and cross-products matrix when only a small part of this matrix is required for a particular calculation) and (2) provide additional storage capacity. These files are automatically established within memory by the AES STAT package if sufficient capacity exists and they are created on the magnetic drum if insufficient capacity exists. User options also permit establishing these files on a magnetic drum or on magnetic tapes.

3. Matrix storage files

Not presently in the package but to be added are codes which will establish and write particular matrices into files in a manner such that other packages or extensions to the AES STAT system can readily reference them. For example, a simple correlation matrix could then be created by the AES STAT system and then used by a factor analysis package. As another use, a simple correlation matrix or matrix of sums of squares and cross-products could be stored in a file and retrieved and computation continued at a later date instead of requiring the AES STAT routine to again start the calculation of a problem from the set of data.

F. Incorporation of $\hat{y}$ and $\hat{u}$ Directly into the Sums of Squares and Cross-products Matrix

(IX.14) $$\hat{y} = Z_{\mu}\hat{\delta}_{\mu}$$

and

(IX.15) $$\hat{u} = y - \hat{y} = y - Z_{\mu}\hat{\delta}_{\mu} = -{}_{+}Z_{\mu}\,{}_{+}\hat{\delta}_{\mu}$$

where ${}_{+}Z_{\mu} = [y_{\mu} \vdots Z_{\mu}]$ and ${}_{+}\hat{\delta}_{\mu} = \begin{bmatrix} -1 \\ \hat{\delta}_{\mu} \end{bmatrix}$. Also, let a sums of squares and cross-products matrix be defined as $Z'Z$ where Z includes the variables in Z_{μ}. Suppose that it is desired that the $\hat{y}$ variable defined by (IX.14) be added as an extra row and column to the $Z'Z$ matrix. It is more accurate as well as saving of computer time to accomplish this directly rather than by calculating $\hat{y}$, forming $[Z \vdots \hat{y}]$, and then forming $[Z \vdots \hat{y}]'[Z \vdots \hat{y}]$.

To incorporate $\hat{y}$ directly into the sums of squares and cross-products matrix, we form the $[Z \vdots \hat{y}]'[Z \vdots \hat{y}]$ matrix by forming

$$Z'\hat{y} \text{ as } [Z'Z_{\mu}]\hat{\delta}_{\mu},$$

and

$$\hat{y}'\hat{y} \text{ as } \hat{\delta}_{\mu}'[Z_{\mu}'Z_{\mu}]\hat{\delta}_{\mu}.$$

The new sums of squares and cross-products matrix is:

$$\begin{bmatrix} Z'Z & Z'\hat{y} \\ \hat{y}'Z & \hat{y}'\hat{y} \end{bmatrix}.$$

Similarly $\hat{u}$ may be added as an extra row and column by forming $Z'\hat{u}$ as $-[Z'{}_{+}Z_{\mu}]{}_{+}\hat{\delta}_{\mu}$, and $\hat{u}'\hat{u}$ as ${}_{+}\hat{\delta}_{\mu}'[{}_{+}Z_{\mu}'\,{}_{+}Z_{\mu}]{}_{+}\hat{\delta}_{\mu}$. Thus, the new matrix is:

$$\begin{bmatrix} Z'Z & Z'\hat{u} \\ \hat{u}'Z & \hat{u}'\hat{u} \end{bmatrix} .$$

In the AES STAT package, incorporation of the $\hat{y}$ or $\hat{u}$ corresponding to any set of coefficients directly into the sums of squares and cross-products matrix is accomplished by the REDO control card. For example, the following REDO card will retrieve the coefficients of equation 5 (they may have been created by any method or read directly into the computer), incorporate $\hat{y}$ into the SSCP matrix as variable 10 and incorporate $\hat{u}$ into the SSCP matrix as variable 14:[1]

```
REDO(5)YHSSCP=10,USSCP=14
```

---

[1] $\hat{y}$ and $\hat{u}$ may also be added directly into the transformed data file if desired. For example a YHTD=10 code used on a REDO card will incorporate $\hat{y}$ into the transformed data as variable 10 and a UTD=14 code used on a REDO card will incorporate $\hat{u}$ into the transformed data as variable 14. The variable number of a YHSSCP code need not be the same as a variable number of a YHTD code even though they are used on the same REDO card.

G. Estimated Values of Normalizing Jointly Dependent Variables, Residuals, and Related Statistics

The calculation of structural and reduced form estimated values (of the normalizing jointly dependent variable) and residuals for each observation as a user option seems very important in a simultaneous stochastic equations package. A feature of the AES STAT package which has proved quite convenient (in cases where a logarithmic transformation has been used in the creation of the normalizing jointly dependent variable) is the option of calculating the anti-logs of the actual and estimated dependent variable for each observation and the difference between these values for each observation. If the dependent variable has been transformed to logarithms, then these three calculated values correspond to the original variable, the estimated value of the original variable, and the residual in its original or natural form. This can prove very helpful in analyzing the results. All three statistics are printed for each observation as a user option in addition to or in place of the regular actual value, estimated value, and residual in logarithmic form.

It is also desirable that a new SSE (sum of squares of error) and R^2 be calculated automatically from the residuals in anti-log form whenever they are calculated (a user option in the AES STAT package permits the new SSE and R^2 to be calculated even if the residuals in anti-log form are not actually printed). The SSE and R^2 so obtained can then be compared to the SSE and R^2 obtained from estimating the model in natural numbers (i.e., from estimating the model in some sort

of linear form, quadratic form, etc.).[1]

When calculating residuals, it is simple to calculate the Durbin-Watson statistic ($= \sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2 / \sum_{t=1}^{T} \hat{u}_t^2$) and print it out. The user can then use it if it is applicable. (The Theil-Nagar statistic is the same as the Durbin-Watson statistic, and the Von Neumann-Hart statistic may be readily calculated by the user as $T/(T - 1)$ times the Durbin-Watson statistic.)[2]

When printing out residuals it is often helpful to the user if each residual is labeled by one or more numeric or alphabetic variables. In the computer output given further on, each residual is labeled by the actual year to which it applies.

---

[1] It would be preferable to generalize the residual subroutine slightly to permit calculation of any function of the dependent variable and the estimated value of the dependent variable, since the dependent variable may be transformed in other ways besides just the logarithmic transformation.

[2] Basic references regarding these statistics are: Theil and Nagar [1961], Durbin and Watson [1950], Durbin and Watson [1951], and Hart [1942].

H. Weighting of Observations

It is very easy to provide for weighting of observations as a standard part of the simultaneous equations package. By weighting of observations we mean that the ij^{th} element of the sums of squares and cross-products matrix is formed as:

$$\sum_{t=1}^{T} c_t x_{ti} x_{tj} \tag{IX.16}$$

where the c_t are weights designated by the user.[1] (If all of the c_t are 1, we have the usual unweighted sums of squares and cross-products matrix.) For the weights to give the correct "degrees of freedom" for estimates of variance or statistical tests, the sum of the weights should be T; however, this imposes an unnecessary burden on the user. All that is necessary is that the user specify the relative weights. The relative weights will be automatically adjusted so that their sum is T if the sums of squares and cross-products matrix is formed and then the entire matrix is multiplied by $T/\sum_{t=1}^{T} c_t$, i.e., the ij^{th} element of the matrix used for further calculations is formed as:

$$m_{ij} = (T/\sum_{t=1}^{T} c_t) \sum_{t=1}^{T} c_t x_{ti} x_{ij} . \tag{IX.17}$$

---

[1] Weights may be based on the number of observations in substrata of the sample as compared to the population to which inference is desired, the inverses of some estimates of variance related to the observations, the length of time since some base point, etc. Klein [1953], pp. 293, 305-313, Goldberger [1964], pp. 235-236, 239-241, 245 and Johnston [1963], pp. 207-211 contain material on weighting of observations, mostly in adjusting for heteroskedasticity.

In the AES STAT series it has been convenient to form and carry the sums of the variables (either original or deviations from means) by including an extra variable, x_0 (a variable taking on the value 1 for all observations), when the sums of squares and cross-products matrix is formed. m_{00} will then be $\sum_{t=1}^{T} c_t$ and m_{0i} will be $\sum_{t=1}^{T} c_t x_{ti}$. If the entire sums of squares and cross-products matrix is then multiplied by $T/\sum_{t=1}^{T} c_t$, the weighted sums of the variables will have been correctly adjusted as well. The same adjustment procedure is used whether the variables (except x_0) are actual or deviations from means. The simple correlations matrix (A*) may then be formed as in section IX.A.2.c and the problem calculated.

In the AES STAT series, weighting of observations is automatically imposed through use of the WOB code on the SSCP card. If a WOB=10 code is used on an SSCP card, each observation is weighted by the value of X_{10} for that observation. (X_{10} may be punched on cards along with the rest of the data or calculated in the data transformation section [see section IX.C].) The sums of squares and cross-products matrix is then automatically adjusted by $T/\sum_{t=1}^{T} c_t$ before further calculations are made.

J. Checks Against Errors

As many cross-checks against errors should be built into a computer routine as possible. Following are examples of cross-checks against errors built into the AES STAT package:

(1) The error sum of squares is calculated and printed out for any coefficients read into the computer through use of the READC card. If a set of coefficients represent an identity equation, the error sum of squares should be zero. Due to punching errors in the data, the requirement that the error sum of squares be zero is often not met. Unfortunately, users are usually lax about checking that the error sum of squares for each identity equation is zero, so we plan to revise this check so that it is performed by the computer as each 3SLS and FIML system is being set up for calculation. Thus, if this check is not met for an equation specified to be an identity equation, the 3SLS or FIML system will not be calculated and a message to the user will be printed.

(2) The ranks of matrices are calculated in the process of orthogonalization and inversion. If in a later step the method requires that certain rank requirements must be met, this check is performed by the computer and a message printed out if some requirement is not met. As an example, rk X_μ and rk X_I are calculated during the calculation of the $[{}_+Y'_\mu \ {}_+Y_\mu]_{\perp X_I}$ matrix for k-class estimation. Assuming that linear restrictions are not imposed on the coefficients, if $\text{rk } X_I < n_\mu$, unique 2SLS and LIML coefficients do not exist;[1] if $\text{rk } X_I = n_\mu$, 2SLS, LIML, and many other

---

[1] The equation is under-identified.

estimators coincide;[1] and if $\text{rk } X_I > n_\mu$ (the most common case encountered), 2SLS and LIML estimates do not in general coincide.[2] The user is notified as to which case is encountered and if $\text{rk } X_I > n_\mu$, $\text{rk } X_I - n_\mu$ is printed out.

Where possible a complete check for a given condition should be made and action taken by the computer routine rather than merely noting conditions on the computer output and relying on the user to notice the condition and take action when required. Many users merely skim through the output giving almost no thought to messages printed out. Only by not calculating part or all of his results can you obtain a user's attention to many types of errors. (The analog of this is the proverbial farmer who bats his mule over the head with a 2X4 to get his attention.) For minor errors, we have used the practice of printing out many asterisks in conjunction with error messages in the hope that the user will be sufficiently curious to look up the error. (Most error messages in the package are given a number. A separate manual then explains what is wrong for each error number and for many error numbers, steps the user can do to rectify the error.) Also, the total number of minor error numbers is prominently printed out at the end of the user's results.

To assist in "checking out the control cards and data" before calculation is actually started, a SCAN card is available. When a SCAN card is used, the data is read into the computer and some basic statistics (but not a sums of squares and cross-products matrix) are calculated. All control cards and codes are then checked for

---

[1] The equation is just-identified.

[2] The equation is over-identified.

consistency. For example, dummy coefficients are saved in the coefficients pool so that the routine can check that starting coefficients have been specified for 3SLS and FIML. After all errors have been corrected, the SCAN card is removed from the deck and the problem calculated. Considerable computer time is saved when the SCAN card is used to detect errors before actually calculating the problem.

Some errors must be detected by the user with the computer routine merely printing out statistics to aid in the detection. Data errors are usually of this nature and cause much difficulty as a result. Following are some statistics printed out by the AES STAT package to aid the user in detecting data errors:

(1) First raw observations, i.e., the first observation as it was read from cards or from a file. (Usually a mispunched format card can be detected by checking the first raw observation.)

(2) First transformed observation, i.e., the value of each variable listed on the SSCP card for the first observation incorporated into the problem. (An incorrectly specified transformation can often be detected by checking the first transformed observation.)

(3) Number of observations read, number of observations dropped, and number of observations in the problem. (Considerable editing of data is often performed in the transformation section. The number of observations in the problem may depend on the transformations and the data itself since observations may be deleted by the transformation section.)

(4) Sums of the raw observations. If the sums of the raw variables are obtained from the basic data source these sums can serve as a check on transcribing as well as data punching. (Transcribing

errors are likely to be a much larger source of data errors than punching errors, especially when card punching is verified by re-punching the cards on a card verifier.) It is usually most convenient to use a hand adding machine to get the sums, with the individual sums being kept on the adding machine tape. A sum is not obtained twice, since it is checked against the computer output. If a sum does not agree with the sum on the computer output, the error is easily traced by comparing the adding machine tape to a listing of the data.

(5) Minimum value encountered in the data for each transformed variable.

(6) Maximum value encountered in the data for each transformed variable.

(7) Means of the transformed variables. Although the exact means are usually unknown, their magnitudes should be known. A mean of the wrong magnitude may reflect many possible errors.

(8) A list of variables which are constant and a list of variables which are zero for all observations. Usually a variable which is constant or zero for all observations reflects a transformation error or failure to provide a transformation.

Some of the above checks may seem cumbersome; however, the ease with which errors are made and the effect of errors on results make the returns to such checking extremely high. Those familiar with the statistical methods given in this paper will surely agree that a single mispunched data element may have a very drastic effect on the statistics calculated.

K. Computer Output

Following is a reproduction of the computer output generated by the parameters and data for Klein's model I listed earlier. Some handwritten notes have been added to call attention to particular points in the output. Also, some blank lines and mostly irrelevant material have been removed to save reproducing costs.

Due to the particular method of output used, no number which occurs on the printed output contains more than 9 significant digits. This holds even though given numbers might contain 23 significant digits while in the computer. Following are some numbers occurring in the output:

Number Printed Out	Number Should be Interpreted as
1133.89999998	1133.90000
37275.86999989	37275.8700
20.2782089394	20.2782089
1.5261866857	1.52618669

In comparing estimated variances, covariances, and statistics computed from estimated variances (such as standard errors and coefficients divided by their standard errors) in the computer output with reported results from other sources, it should be recalled that a "degrees of freedom" adjustment of $T/(\sqrt{T - n_{\mu}}\sqrt{T - n_{\mu'}})$ is used in most of these statistics in the output whereas for results reported elsewhere no "degrees of freedom" adjustment or a different degrees of freedom adjustment is often made. For example in results reported in Goldberger [1964], no "degrees of freedom" adjustment is made. In some cases coefficient variance-covariance estimates have been

printed out both with and without the degrees of freedom adjustment to aid in comparing results with other sources. For variances, covariances, and related statistics, the computer routine prints out an indication of the denominator used in the computation.

Total execution time was about 40 seconds.[1]

---

[1]About 2.5 minutes of "set-up" time was also required. The "set-up" time will be reduced drastically if many of the subroutines are transferred to the regular library file. Current charges for the CDC 3600 computer are $330 per hour plus $.01 per page of printed output. Thus, due to the large set-up time presently required, the total cost of calculating this problem on the CDC 3600 computer was about $18.00. ($3.67 for execution time plus $13.75 for set-up time plus $.65 for printed output.)

```
MSU DRUM SCOPE                                  VERSION 1.10B
SEQUENCE NUMBER 356353000000(PRIORITY 0, PROBLEM TYPE 07)
JOB,6016B2,FIML,9,RUALE,9ILL  KLEIN MODEL I ON 2SLS,LIML, 3SLS, I3SLS, LML,FIML
DEMAND,MT
EQUIP,20=MT(STAT)
FILE,69
FTN,L,*,X
```

```
FTN5,3A                                    09/28/68                    PAGE NO,      1

            SUBROUTINE TRANS1(X)
            COMMON/A/NR,NT,IS,IOR,JP  $ DIMENSION X(200)
            DOUBLE PRECISION X
            X(14)=X(4)*X(5)
            X(15)=X(2)*X(12)                    } Transformations creating variables
            X(20)=NR*1920
            RETURN
            END
ROUTINE,SSCP,*,READC,3SLS,FIML } Sections of the simultaneous stochastic equations package used to calculate the problem.

ROUTINE STARTED  AT 170905
STAT TAPE EDITION 25
ROUTINE FINISHED AT 171046
ELAPSED0    92.645 SECONDS
RUN,5,35000,2
```

SET UP DATA--KLEIN MODEL I

CURRENT TIME 1711 • 30 DATE 09/28/68
ELAPSED SINCE LAST CURRENT TIME 0.00 SECONDS

```
U     SET UP DATA--KLEIN MODEL I                                    (STAT CONTROL CARD)
D,13,21                                                             (STAT CONTROL CARD)
L  X(1)=C,I,W1,P,W,E                                                (STAT CONTROL CARD)
L  X(14)=Y,K                                                        (STAT CONTROL CARD)
L  X(7)=W2,R,G,TIME,P LAGGED,K LAGGED,E LAGGED                      (STAT CONTROL CARD)
SSCP(X(1),,,X(15)/X(20))TRANS,RES,MAXE=8                            (STAT CONTROL CARD)
FORMAT(6X,6F5,F3,2F5,F3,2F6,F5)                                     (FORMAT)
```

Used as checks on the data, format, and transformations

1ST RAW OBSERVATION

1	2	3	4	5	6
41,90000000	-0,20000000	25,50000000	12,40000000	28,20000000	45,60000000

7	8	9	10	11	12
2,70000000	7,70000000	6,60000000	-10,00000000	12,70000000	182,80000000

13
44,90000000

1ST TRANSFORMED OBSERVATION (FROM RAW OBSERV. NO. 1)

C	I	W1	P	W	E
1	2	3	4	5	6
41,90000000	-0,20000000	25,50000000	12,40000000	28,20000000	45,60000000

W2	R	G	TIME	P LAGGED	K LAGGED
7	8	9	10	11	12
2,70000000	7,70000000	6,60000000	-10,00000000	12,70000000	182,80000000

E LAGGED	Y	K	
13	14	15	20
44,90000000	48,60000000	182,60000000	1921,00000000

TOTAL RAW OBSERVATIONS= 21
TOTAL OBSERVATIONS DROPPED= 0
TOTAL OBSERVATIONS IN PROBLEM= 21

SUMS OF 21 RAW OBSERVATIONS

1	2	3	4	5	6
1133,89999998	26,60000000	763,59999999	354,70000000	871,89999999	1261,19999999

7	8	9	10	11	12
187,50000000	142,90000000	288,20000000	0,00000000	343,90000000	4218,39999998

13
1217,09999999

TABLE A -- STATISTICS ON TRANSFORMED VARIABLES

LABEL	VAR NO	MEAN	STANDARD DEVIATION	SUM	SUM OF SQUARES	SUM OF SQUARED DEVIATIONS FROM THE MEAN
C	1	53,99523809	6,86086556	1133,89999998	62166,62999916	941,42952380
I	2	1,26666667	3,55194782	26,60000000	286,02000000	252,32666666
W1	3	36,36190476	6,38448133	763,59999999	28560,85999966	794,90952380
P	4	16,89047619	4,22017829	354,70000000	6347,24999988	356,19809523
W	5	41,48095238	7,55583344	871,09999999	37275,86999989	1141,81238094
E	6	48,85714286	10,61722993	1026,19999999	77998,57999992	2254,51142853
W2	7	5,11904762	1,95668573	107,50000000	626,86999999	76,57238095
R	8	6,80476190	2,03161911	142,90000000	1054,94999999	82,54952381
G	9	9,91428571	3,91015162	208,20000000	2369,94000000	305,78571428
TIME	10	0,00000000	6,20483682	0,00000000	770,00000000	770,00000000
P LAGGED	11	16,37619048	4,02613912	343,90000000	5956,28999996	324,51809523
K LAGGED	12	200,49523809	9,91854204	4210,39999998	846132,69999695	1967,54952380
E LAGGED	13	57,98571429	8,91859214	1217,69999999	72200,02999878	1590,82571426
Y	14	58,37142857	10,73979258	1225,79999998	73858,55999947	2306,86285710
K	15	201,76190476	9,21962451	4237,00000000	856565,21998596	1700,02952379

SIMPLE CORRELATIONS (Pearson product-moment correlations)

		1 C	2 I	3 W1	4 P	5 W	6 E	7 W2	8 R	9 G	10 TIME	11 P LAGGED
C	1	1,00000										
I	2	0,49804	1,00000									
W1	3	0,97797	0,55976	1,00000								
P	4	0,71534	0,92334	0,74992	1,00000							
W	5	0,98278	0,49675	0,97625	0,83416	1,00000						
E	6	0,96316	0,67207	0,98114	0,83372	0,94224	1,00000					
W2	7	0,64375	-0,23285	0,54786	0,03260	0,71608	0,47728	1,00000				
R	8	0,51275	-0,14278	0,46652	-0,04735	0,57738	0,44954	0,72647	1,00000			
G	9	0,73037	-0,07391	0,71378	0,18620	0,82302	0,65365	0,87835	0,81420	1,00000		
TIME	10	0,67852	-0,23957	0,58912	0,03609	0,74937	0,49774	0,98016	0,69809	0,86986	1,00000	
P LAGGED	11	0,65205	0,76785	0,72233	0,76913	0,57933	0,74456	-0,89021	0,05085	0,13440	-0,02381	1,00000
K LAGGED	12	0,31288	-0,36889	0,29777	-0,20173	0,33515	0,15146	0,33477	0,28654	0,36489	0,47983	0,15421
E LAGGED	13	0,87048	0,45569	0,89870	0,61495	0,84703	0,85747	0,37524	0,41495	0,57476	0,44779	0,83636
Y	14	0,97246	0,64899	0,98151	0,83918	0,95273	0,99051	0,51660	0,38760	0,65228	0,53858	0,70981
K	15	0,52847	-0,01168	0,53600	0,13870	0,51726	0,42186	0,27044	0,25326	0,36408	0,42390	0,46173

		12 K LAGGED	13 E LAGGED	14 Y	15 K
K LAGGED	12	1,00000			
E LAGGED	13	0,47577	1,00000		
Y	14	0,15652	0,83756	1,00000	
K	15	0,93369	0,68739	0,41841	1,00000

KLEIN MODEL I--2SLS AND LIML -- EQ 1

CURRENT TIME 1711 = 34 DATE 09/28/68
ELAPSED SINCE LAST CURRENT TIME 3.45 SECONDS

```
U     KLEIN MODEL I--2SLS AND LIML -- EQ 1                    (STAT CONTROL CARD)
K(1,4,5/0,11/7,,,10,12,13)                                    (STAT CONTROL CARD)
        2SLS=C,DV,SCE=1,STE,COSTE                             (CONTINUATION TO ABOVE CARD)
        LIML=C,DV,COSTE,STE                                   (CONTINUATION TO ABOVE CARD)
```

(SET UP FOR DBL K-CLASS MEMBER)

RANKS

PRED WITHIN	ALL PRED
2 ↑ rk X_p	8 ↑ rk X_I

(2-STAGE LEAST SQUARES)

Basmann's identifiability test statistic (ϕ). See Basmann [1960 b], pg. 651, equation (1.9).

$\frac{T-rk\,X_I}{rk\,X_I-n}\hat{\phi}$ [Based on Basmann [1960 A], pg. 652, equation (1.18).]

PHI	PHI RATIO	NUM DF	DEN DF
0,71730076	2,33122747	4	13

S^2

Due to DV after 2SLS code of K card.

SSE (Error sum of squares)	S2 (T USED IN DENOMINATOR)	S2 (T-N(I) USED IN DENOMINATOR)	PROPORTION OF VARIATION EXPLAINED BY COEFFICIENTS
21,92524735	1,04405940	1,28972043	0,9767

$\frac{(y-\bar{y})'(y-\bar{y})-(y-\hat{y})'(y-\hat{y})}{(y-\bar{y})'(y-\bar{y})}$ where $\bar{y}$ denotes mean of y. (Usually called R^2.)

Due to C after 2SLS code.

COEFFICIENTS $\hat{\delta}_i$ (2-STAGE LEAST SQUARES)

C	P	W	CONSTANT	P LAGGED
1	4	5	0	11
16,0000000000	0,0173022118	0,8101826976	16,554755765n	0,2162340405

Due to STE after 2SLS code.

COEFFICIENT STANDARD ERRORS $\sqrt{\widehat{Var}(\hat{\delta}_i)}$ (2-STAGE LEAST SQUARES)

SQ RT OF T-N(I) USED AS DENOMINATOR

P	W	CONSTANT	P LAGGED
4	5	0	11
0,13120458	0,04473506	1,46797870	0,11922168

Due to COSTE after 2SLS code.

COEFFICIENTS/COEF STANDARD ERRS $\hat{\delta}_i / \sqrt{\widehat{Var}(\hat{\delta}_i)}$ (2-STAGE LEAST SQUARES)

SQ RT OF T-N(I) USED AS FACTOR

P	W	CONSTANT	P LAGGED
4	5	0	11
0,13187201	18,11068904	11,27724524	1,81371414

Due to SCE=1 after 2SLS code.

2SLS COEFS SAVED IN COMPUTER AS EQ 1

KLEIN MODEL I--2SLS AND LIML -- EQ 1

CURRENT TIME 1711 - 37 DATE 09/28/68
ELAPSED SINCE LAST CURRENT TIME 1.05 SECONDS

(LIMITED INFORMATION SINGLE EQ.) (LIML)

SSE OF REGRESSION ON PREDET VARS IN EQ / SSE OF REGRESSION ON PREDET VARS IN SYSTEM

C	P	W
1	4	5
9,31452727	2,34844505	18,96136257

$\frac{[c'\hat{q}]_{1X_\mu}}{[c'\hat{q}]_{1X_\Sigma}}$ $\frac{[p'\hat{q}]_{1X_\mu}}{[p'\hat{q}]_{1X_\Sigma}}$ $\frac{[w'w]_{1X_\mu}}{[w'w]_{1X_\Sigma}}$

smallest eigenvalue of $[\hat{Y}'\hat{Y}]^{-1}_{1X_\Sigma}[\hat{Y}'\hat{Y}]_{1X_\mu}$ matrix = k

NUM ITER 12
number of iterations required to extract eigenvalues.

ROOTS OF LI MATRIX--1ST ROOT=K1=K2

1	2	3
1,49874551	7,61755904	186,16141858

2nd smallest eigenvalue

Basmann's identifiability test statistic ($\hat{\phi}$). See Basmann[1960b], pg. 651, equation (1.10).

$\frac{T-rk}{rk}\frac{X_\Sigma}{X_\Sigma-m}\hat{\phi}$ [Based on Basmann[1960b], pg. 652, equation (1.19).]

PHI	PHI RATIO	NUM DF	DEN DF
0,49874551	1,62092289	4	13

T log ℓ_1 test statistic of Koopmans and Hood[1953], pg. 132, equation (8.22).

T LN(ROOT1)	DF	T LN(ROOT1*ROOT2)	DF
8,49719700	4	51,13677204	5

T log(ℓ_1/ℓ_2) test statistic of Koopmans and Hood [1953], pg. 184, equation (9.27).

Due to DV after LIML code of Keard.

SSE	S2 (T USED IN DENOMINATOR)	S2 (T-N(I) USED IN DENOMINATOR)	PROPORTION OF VARIATION EXPLAINED BY COEFFICIENTS
40,88418833	1,94686611	2,40495225	0,9566

See page 388 for an explanation

Due to C after LIML code

COEFFICIENTS $\hat{\delta}_i$ (LIMITED INFORMATION SINGLE EQ.)

C	P	W	CONSTANT	P LAGGED
1	4	5	0	11
-1,0000000000	-0,2225130652	0,8225586648	17,1476546223	0,3960272883

Due to STE after LIML code

COEFFICIENT STANDARD ERRORS $\sqrt{\hat{Var}(\hat{\delta}_i)}$ (LIMITED INFORMATION SINGLE EQ.)

SQ RT OF T-N(I) USED AS DENOMINATOR

P	W	CONSTANT	P LAGGED
4	5	0	11
0,22423014	0,06154943	2,04537389	0,19294311

Due to COSTE after LIML code.

COEFFICIENTS/COEF STANDARD ERRS $\hat{\delta}_i/\sqrt{\hat{Var}(\hat{\delta}_i)}$ (LIMITED INFORMATION SINGLE EQ.)

SQ RT OF T-N(I) USED AS FACTOR

P	W	CONSTANT	P LAGGED
4	5	0	11
-0,99234234	13,36419693	8,38362840	2,05255984

KLEIN MODEL I--2SLS AND LIML -- EQ 2

CURRENT TIME 1711 - 38 DATE 09/28/68
ELAPSED SINCE LAST CURRENT TIME 0.93 SECONDS

```
U     KLEIN MODEL I--2SLS AND LIML -- EQ 2                    (STAT CONTROL CARD)
K(2,4/0,11,12/7,,,10,13)                                      (STAT CONTROL CARD)
      2SLS=C,DV,SCE=2,STE,COSTE                               (CONTINUATION TO ABOVE CARD)
      LIML=C,DV,COSTE,STE                                     (CONTINUATION TO ABOVE CARD)
```

(SET UP FOR DBL K-CLASS MEMBER)

RANKS	
PRED WITHIN	ALL PRED
3	8

(2-STAGE LEAST SQUARES)

PHI	PHI RATIO	NUM DF	DEN DF
0,09460319	0,30746037	4	13

See page 388 for an explanation

Due to DV after 2SLS code of K card

SSE	S2 (T USED IN DENOMINATOR)	S2 (T-N(I) USED IN DENOMINATOR)	PROPORTION OF VARIATION EXPLAINED BY COEFFICIENTS
29,04685846	1,38318374	1,70863873	0,8849

Due to C after 2SLS code

COEFFICIENTS $\hat{\delta}_i$ (2-STAGE LEAST SQUARES)

I	P	CONSTANT	P LAGGED	K LAGGED
2	4	0	11	12
=1,0000000000	0,1502218239	20,2782089394	0,6159435773	-0,1577876365

Due to STE after 2SLS code

COEFFICIENT STANDARD ERRORS--- $\sqrt{\widehat{Var}(\hat{\delta}_i)}$ (2-STAGE LEAST SQUARES)

SQ RT OF T-N(I) USED AS DENOMINATOR

P	CONSTANT	P LAGGED	K LAGGED
4	0	11	12
0,19253359	8,38324890	0,18092585	0,04015207

Due to COSTE after 2SLS code

COEFFICIENTS/COEF STANDARD ERRS--- $\hat{\delta}_i / \sqrt{\widehat{Var}(\hat{\delta}_i)}$ (2-STAGE LEAST SQUARES)

SQ RT OF T-N(I) USED AS FACTOR

P	CONSTANT	P LAGGED	K LAGGED
4	0	11	12
0,78023695	2,41889620	3,40439791	-3,92979106

Due to SCE=2 after 2SLS code

2SLS COEFS SAVED IN COMPUTER AS EQ 2

KLEIN MODEL I--2SLS AND LIML -- EQ 2

CURRENT TIME 1711 - 59 DATE 69/28/68
ELAPSED SINCE LAST CURRENT TIME 0.69 SECONDS

(LIMITED INFORMATION SINGLE EQ.)

SSE OF REGRESSION ON PREDET VARS IN EQ / SSE OF REGRESSION ON PREDET VARS IN SYSTEM

I	P
2	4
1.89191898	1.74483819

NUM ITER
11

ROOTS OF LI MATRIX--1ST ROOT=K1*K2

1	2
1.88595285	4.56889355

PWI	PWI RATIO	NUM DF	DEN DF
0.88595285	0.27934675	4	13

T LN(ROOT1)	DF	T LN(ROOT1*ROOT2)	DF
1.73161388	4	33.63038614	5

See page 389 for an explanation

Due to DV after LIML code & K card

SSE	S2 (T USED IN DENOMINATOR)	S2 (T-N(I) USED IN DENOMINATOR)	PROPORTION OF VARIATION EXPLAINED BY COEFFICIENTS
34.99648655	1.66649936	2.85801086	0.8613

See page 388 for an explanation

Due to C after LIML code

COEFFICIENTS $\hat{\delta}_i$ (LIMITED INFORMATION SINGLE EQ.)

I	P	CONSTANT	P LAGGED	K LAGGED
	4	8	11	12
2	0.8751847588	22.5988254446	0.6883863833	-0.1682643562
-1.0000000000				

Due to STE after LIML code

COEFFICIENT STANDARD ERRORS $\sqrt{\widehat{Var}(\hat{\delta}_i)}$ (LIMITED INFORMATION SINGLE EQ.)

SQ RT OF T-N(I) USED AS DENOMINATOR

P	CONSTANT	P LAGGED	K LAGGED
4	8	11	12
0.22471169	9.49814481	0.28914465	0.84534452

Due to COSTE after LIML code

COEFFICIENTS/COEF STANDARD ERRS $\hat{\delta}_i / \sqrt{\widehat{Var}(\hat{\delta}_i)}$ (LIMITED INFORMATION SINGLE EQ.)

SQ RT OF T-N(I) USED AS FACTOR

P	CONSTANT	P LAGGED	K LAGGED
4	8	11	12
0.33458321	2.37844588	3.25318575	-3.71879812

KLEIN MODEL I--2SLS AND LIML -- EQ 3

CURRENT TIME 1711 . 40 DATE 09/28/68
ELAPSED SINCE LAST CURRENT TIME 0.79 SECONDS

```
U     KLEIN MODEL I--2SLS AND LIML -- EQ 3                    (STAT CONTROL CARD)
K(3,6/8,13,10/7,,,9,11,12)                                    (STAT CONTROL CARD)
        2SLS=C,DV,SCE=3,STE,COSTE                             (CONTINUATION TO ABOVE CARD)
        LIML=C,DV,COSTE,STE                                   (CONTINUATION TO ABOVE CARD)
```

(SET UP FOR DBL K-CLASS MEMBER)

RANKS

PRED WITHIN	ALL PRED
3	8

(2-STAGE LEAST SQUARES)

See page 388 for an explanation

PHI	PHI RATIO	NUM DF	DEN DF
1,46919978	4,77489904	4	13

Due to DV after 2SLS code of K card.

SSE	S2 (T USED IN DENOMINATOR)	S2 (T-N(I) USED IN DENOMINATOR)	PROPORTION OF VARIATION EXPLAINED BY COEFFICIENTS
10,00496397	0,47642686	0,38852729	0,9874

Due to C after 2SLS code

(2-STAGE LEAST SQUARES) COEFFICIENTS $\hat{\delta}_i$

d1	E	CONSTANT	E LAGGED	TIME
3	6	0	13	10
+1,0000000000	0,4388590651	1,5002968860	0,1466738215	0,1303956872

Due to STE after 2SLS code

(2-STAGE LEAST SQUARES) COEFFICIENT STANDARD ERRORS $\sqrt{\widehat{Var}(\hat{\delta}_i)}$ -- SQ RT OF T-N(I) USED AS DENOMINATOR

E	CONSTANT	E LAGGED	TIME
6	0	13	10
0,03960266	1,27568637	0,04316395	0,03238839

Due to COSTE after 2SLS code

(2-STAGE LEAST SQUARES) COEFFICIENTS/COEF STANDARD ERRS $\hat{\delta}_i / \sqrt{\widehat{Var}(\hat{\delta}_i)}$ -- SQ RT OF T-N(I) USED AS FACTOR

E	CONSTANT	E LAGGED	TIME
6	0	13	10
11,08155481	1,17607033	3,39806312	4,02600104

Due to SCE=3 after 2SLS code

2SLS COEFS SAVED IN COMPUTER AS EQ 3

KLEIN MODEL I--2SLS AND LIML -- EQ 3

CURRENT TIME 1711 . 40 DATE 09/28/68
ELAPSED SINCE LAST CURRENT TIME 0.68 SECONDS

(LIMITED INFORMATION SINGLE EQ.)

SSE OF REGRESSION ON PREDET VARS IN EQ / SSE OF REGRESSION ON PREDET VARS IN SYSTEM

#1	E
3	6
2.99926668	3.02717731

NUM ITER
11

ROOTS OF LI MATRIX--1ST ROOT=K1<K2

1	2
2.46858257	3.02835339

PHI	PHI RATIO	NUM DF	DEN DF
1.46858257	4.77289334	4	13

T LN(ROOT1)	DF	T LN(ROOT1*ROOT2)	DF
10.97852665	4	42.24492645	5

See page 189 for an explanation

Due to DV after LIML code of K card

SSE	S2 (T USED IN DENOMINATOR)	S2 (T-N(I) USED IN DENOMINATOR)	PROPORTION OF VARIATION EXPLAINED BY COEFFICIENTS
18.02192874	0.47723432	0.50992479	0.9874

See page 188 for an explanation

Due to C after LIML code

COEFFICIENTS $\hat{\delta}_i$ (LIMITED INFORMATION SINGLE EQ.)

#1	E	CONSTANT	E LAGGED	TIME
3	6	0	13	18
=1.0000000000	0.4339413995	1.52618668>7	0.1513206755	0.1315931213

Due to STE after LIML code

COEFFICIENT STANDARD ERRORS $\sqrt{\text{VAR}(\hat{\delta}_i)}$ (LIMITED INFORMATION SINGLE EQ.)

SQ RT OF T-N(I) USED AS DENOMINATOR

E	CONSTANT	E LAGGED	TIME
6	0	13	18
0.07558748	1.32883786	0.07452678	0.03599549

Due to COSTE after LIML code

COEFFICIENTS/COEF STANDARD ERRS $\hat{\delta}_i / \sqrt{\text{VAR}(\hat{\delta}_i)}$ (LIMITED INFORMATION SINGLE EQ.)

SQ RT OF T-N(I) USED AS FACTOR

E	CONSTANT	E LAGGED	TIME
6	0	13	18
5.74788464	1.15546861	2.03842024	3.65582284

KLEIN MODEL I--IDENTITY EQUATIONS

CURRENT TIME 1711 - 41 DATE 09/28/68
ELAPSED SINCE LAST CURRENT TIME 0.79 SECONDS

U KLEIN MODEL I--IDENTITY EQUATIONS
READC(14,1,2/8,9)C,DV,SCE=4 [STAT CONTROL CARD]

(READ COEFFICIENTS) [FORMAT]

FORMAT(5F2)

Due to C code on READC card.

	Y	C	I	K	G
	14	1	2	8	9
COEFFICIENTS	-1.0000000000	1.0000000000	1.0000000000	-1.0000000000	1.0000000000

(SAVED IN COMPUTER AS EQ, 4) Due to SCE=4 code (READ COEFFICIENTS)

Due to DV code on READC card.

SSE	BIASED S2	INTUIT, S2
0.00000000	0.00000000	0.00000000

Verifies that this was an identity equation.

READC(14,4,5)C,DV,SCE=5 [STAT CONTROL CARD]

(READ COEFFICIENTS) [FORMAT]

FORMAT(5F2)

Due to C code on READC card.

	Y	P	W
	14	4	5
COEFFICIENTS	-1.0000000000	1.0000000000	1.0000000000

(SAVED IN COMPUTER AS EQ, 5) Due to SCE=5 code (READ COEFFICIENTS)

Due to DV code on READC card.

SSE	BIASED S2	INTUIT, S2
0.00000000	0.00000000	0.00000000

READC(15,2/12)C,DV,SCE=6 [STAT CONTROL CARD]

(READ COEFFICIENTS) [FORMAT]

FORMAT(5F2)

Due to C code on READC card.

	K	I	K LAGGED
	15	2	12
COEFFICIENTS	-1.0000000000	1.0000000000	1.0000000000

(SAVED IN COMPUTER AS EQ, 6) Due to SCE=6 code (READ COEFFICIENTS)

Due to DV code on READC card.

SSE	BIASED S2	INTUIT, S2
-0.00000000	-0.00000000	-0.00000000

KLEIN MODEL I--IDENTITY EQUATIONS

CURRENT TIME 1711 = 43 DATE 09/28/68
ELAPSED SINCE LAST CURRENT TIME 0.20 SECONDS

READC(5,3F7)C,DV,SCE=7 (STAT CONTROL CARD)

FORMAT(3F2) (FORMAT)

(READ COEFFICIENTS)

Due to C code on READC card.	C	COEFFICIENTS	(SAVED IN COMPUTER AS EQ, 7)	Due to SCE=7 code.	(READ COEFFICIENTS)
	5	M1 3	M2 7		
	=1.0000000000	1.0000000000	1.0000000000		
Due to DV code on READC card.	SSE	BIASED S2	INTUIT, S2		
	=0.00000000	=0.00000000	=0.00000000		

READC(6,14F8,7)C,DV,SCE=8 (STAT CONTROL CARD)

FORMAT(4F2) (FORMAT)

(READ COEFFICIENTS)

Due to C code on READC card.	C	COEFFICIENTS	V	(SAVED IN COMPUTER AS EQ, 8)		Due to SCE=8 code. (READ COEFFICIENTS)
	6	14	R 8	M2 7		
	=1.0000000000	1.0000000000	1.0000000000	=1.0000000000		
Due to DV code on READC card.	SSE	BIASED S2	INTUIT, S2			
	0.00000000	0.00000000	0.00000000			

3SLS, I3SLS STARTING FROM 2SLS--KLEIN MODEL I

CURRENT TIME 1711 = 44 DATE 09/28/68
ELAPSED SINCE LAST CURRENT TIME 0.21 SECONDS

```
U 3SLS, I3SLS STARTING FROM 2SLS--KLEIN MODEL I                          (STAT CONTROL CARD)
3SLS(1,2,3/4,,,8/X(7)...X(9))BE3=C,LIK,RF,DVC,DVCN                       (STAT CONTROL CARD)
                             3SLS=C,LIK,DVC,DVCN,CVC,CVCT,CVCN,RES,NR=20,COSTE  (CONTINUATION TO ABOVE CARD)
                                 RF,STE                                  (CONTINUATION TO ABOVE CARD)
                             I3SLS=C,LIK,DVC,DVCN,CVC,CVCT,CVCN,COSTE,STE  (CONTINUATION TO ABOVE CARD)
                             NTW=20=C,LIK                                (CONTINUATION TO ABOVE CARD)
                             MI=200,CCC=1.0,10,TIMEL=10                  (CONTINUATION TO ABOVE CARD)


                                                (THREE STAGE LEAST SQUARES)
      1, COEFS FOR EQ  1 PREVIOUSLY SAVED FROM   2SLS
      2, COEFS FOR EQ  2 PREVIOUSLY SAVED FROM   2SLS
      3, COEFS FOR EQ  3 PREVIOUSLY SAVED FROM   2SLS
      4, COEFS FOR EQ  4 PREVIOUSLY SAVED FROM   READC
      5, COEFS FOR EQ  5 PREVIOUSLY SAVED FROM   READC
      6, COEFS FOR EQ  6 PREVIOUSLY SAVED FROM   READC
      7, COEFS FOR EQ  7 PREVIOUSLY SAVED FROM   READC
      8, COEFS FOR EQ  8 PREVIOUSLY SAVED FROM   READC

 ENDOG VARS, IN SUBSYSTEM 1
C         P         W          I         W1         E
  1         4         5         2         3         6

 ENDOG VARS, IN SUBSYSTEM 2 BUT NOT 1
Y         K
 14        15

  PRED VARS, IN SUBSYSTEM 1
CONSTANT  P LAGGED  K LAGGED  E LAGGED  TIME
   0         11        12        13       10

  PRED VARS, IN SUBSYSTEM 2 BUT NOT 1
R         G         W2
  8         9         7

PRED, VARS, IN 3 STAGE SYSTEM
CONSTANT  P LAGGED  K LAGGED  E LAGGED  TIME       W2        R         G
   0         11        12        13       10        7         8         9

            NUM ITER/
 ITER MAX     PRINT      CONVERGENCE LEVEL        ITER TIME LIMIT (MIN)      MIN TO STOP ITER BEFORE
                                                                            JOB OR EXEC TIME LIMIT EXPIRES
      200         20        1.00000-010              10.00000000                1.00000000

                                          rk X_Z          number of variables in X_Z
                                            ↓                    ↓
RANK OF PRED, VARS, IN 3 STAGE SYSTEM= 8  ORDER= 8
```

jointly dependent variables occurring in the stochastic equations to be estimated

jointly dependent variables which do not occur in the stochastic equations to be estimated but do occur in equations in the system

predetermined variables occurring in the stochastic equations to be estimated

predetermined variables which do not occur in the stochastic equations to be estimated but do occur in equations in the system

predetermined variables in the X_z matrix

COEFFICIENTS

(STARTING ESTIMATES--PRE 3SLS)

EQUATION 1

C		P		W	CONSTANT	P LAGGED
1		4		5	8	11
-1.0000000000		0.8173822118		0.8181826976	16.5547557650	0.2162348485

EQUATION 2

I		P		CONSTANT	P LAGGED	K LAGGED
2		4		8	11	12
-1.0000000000		0.1502218239		20.2782889394	0.6159435773	-0.1577876365

EQUATION 3

d1		E		CONSTANT	E LAGGED	TIME
3		6		8	13	10
-1.0000000000		0.4388598651		1.5002968860	0.1466738219	0.1383956872

EQUATION 4

V		C		I		R		G
14		1		2		8		9
-1.0000000000		1.0000000000		1.0000000000		-1.0000000000		1.0000000000

EQUATION 5

V		P		W
14		4		5
-1.0000000000		1.0000000000		1.0000000000

EQUATION 6

K		I		K LAGGED
15		2		12
-1.0000000000		1.0000000000		1.0000000000

EQUATION 7

d		W1		W2
5		3		7
-1.0000000000		1.0000000000		1.0000000000

EQUATION 8

E		Y		R		W2
6		14		8		7
-1.0000000000		1.0000000000		1.0000000000		-1.0000000000

Due to C after BEG code

2SLS coefficients

Identity equations

DISTURBANCE VAR-COV MATRIX

(STARTING ESTIMATES--PRE 3SLS)

--SQ RT OF(T-N(J))(T-N(I))USED AS THE DENOMINATOR

Due to DVC after BEG code

C	1	1.28972843			
I	2	0.54087875	1.78863873		
W1	3	-0.47586935	0.23792536	0.38892729	
		1	2	3	
		C	I	W1	

S calculated from 2SLS coefficients

DISTURBANCE VAR-COV MATRIX NORMALIZED

(STARTING ESTIMATES--PRE 3SLS)

Due to DVCN after BEG code

C	1	1.00000		
I	2	0.36435	1.00000	
W1	3	-0.54621	0.23726	1.00000
		1	2	3
		C	I	W1

S_{2SLS} normalized

Due to LIK after BEG code

DET GAM= 1.613622-001 DET S= 5.665463-013 LIK= 4.995877+010

$\det \hat{\Gamma}_{2SLS}$ computed from coefficients corresponding to normalized variables (presently normalized to length 1)

$\det S_{2SLS}$ computed from coefficients corresponding to normalized variables (presently normalized to length 1)

$\det^2 \Gamma_{2SLS} / \det S_{2SLS}$

REDUCED FORM COEFS (STARTING ESTIMATES--PRE 3SLS)

Due to RF after BEG code

1. DEP. VAR.= 1 C

CONSTANT 0	P LAGGED 11	K LAGGED 12	E LAGGED 13	TIME 10	R 6
42,02804481	0,76849717	-0,10470549	0,17884542	0,15899682	-0,12846916
8 9	W2 7				
0,66358805	0,68422220				

2. DEP. VAR.= 4 P

CONSTANT 0	P LAGGED 11	K LAGGED 12	E LAGGED 13	TIME 10	R 6
37,03169233	0,84835668	-0,16085532	-0,05058001	-0,04496655	-1,17078101
8 9	W2 7				
1,01944183	-0,19350770				

3. DEP. VAR.= 5 W

CONSTANT 0	P LAGGED 11	K LAGGED 12	E LAGGED 13	TIME 10	R 6
31,03552979	0,66348576	-0,12580229	0,22182721	0,19720841	-0,13356501
8 9	W2 7				
0,79728863	0,84866082				

4. DEP. VAR.= 2 I

CONSTANT 0	P LAGGED 11	K LAGGED 12	E LAGGED 13	TIME 10	R 6
25,84117730	0,74338926	-0,18195102	-0,00759822	-0,00675496	-0,17587686
8 9	W2 7				
0,15314241	-0,02906908				

Reduced form coefficients calculated from 2SLS coefficients

Due to RF after BCS code

5, DEP. VAR.= 3 W1

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	0
31.03592979	0.66348576	-0.12588229	0.22182721	0.19728841	-0.13356581

0	W2
9	7
0.79728863	-0.15133918

6, DEP. VAR.= 6 E

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	0
68.66722212	1.51184243	-0.28665701	0.17124728	0.19224186	-0.30434682

0	W2
9	7
1.81673847	-0.34484688

7, DEP. VAR.= 14 Y

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	0
68.66722212	1.51184243	-0.28665701	0.17124728	0.19224186	-1.30434682

0	W2
9	7
1.81673847	0.65515312

8, DEP. VAR.= 15 K

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	0
25.84117738	0.74338526	0.81804838	-0.80759822	-0.88675496	-0.17587686

0	W2
9	7
0.19314241	-0.02986988

Reduced form coefficients continued

ITER NO, 1 MAX ABS CHANGE RATIO FROM PREV COEFS= 6.218180+000

$$\max_j \text{abs}\left[\frac{\hat{a}_j^{(\text{max})} - \hat{a}_j^{(\text{max-1})}}{\hat{a}_j^{(\text{max-1})}}\right] = 6.218180$$

Due to C after 3SLS code

COEFFICIENTS $\hat{\delta}_i$ (THREE STAGE LEAST SQUARES)

EQUATION 1

C	P	W	CONSTANT	P LAGGED
1	4	5	0	11
-1.0000000000	0.1248904748	0.7900809364	16.4407908642	0.1631440928

EQUATION 2

I	P	CONSTANT	P LAGGED	K LAGGED
2	4	0	11	12
-1.0000000000	-0.0130791824	28.1778468676	0.7557239621	-0.1948482493

EQUATION 3

W1	E	CONSTANT	E LAGGED	TIME
3	6	0	13	10
-1.0000000000	0.4804918798	1.7972177277	0.1812910150	0.1496741151

Due to STE after 3SLS code

COEFFICIENT STANDARD ERRORS $\sqrt{\widehat{Var}(\hat{\delta}_i)}$ (THREE STAGE LEAST SQUARES)

SQ RT OF T-N(I) USED AS DENOMINATOR

EQUATION 1

P	W	CONSTANT	P LAGGED
4	5	0	11
0.12017872	0.04216562	1.4499248	0.11163081

EQUATION 2

P	CONSTANT	P LAGGED	K LAGGED
4	0	11	12
0.17993761	7.55085338	0.16997567	0.03615585

EQUATION 3

E	CONSTANT	E LAGGED	TIME
6	0	13	10
0.03535863	1.24020347	0.03796536	0.03104828

Due to COSTE after 3SLS code

COEFFICIENTS/COEF STANDARD ERRS $\hat{\delta}_i / \sqrt{\widehat{Var}(\hat{\delta}_i)}$ (THREE STAGE LEAST SQUARES)

SQ RT OF T-N(I) USED AS FACTOR

EQUATION 1

P	W	CONSTANT	P LAGGED
4	5	0	11
1.03920625	18.73756045	11.33906334	1.46146116

EQUATION 2

P	CONSTANT	P LAGGED	K LAGGED
4	0	11	12
-0.07268732	3.73174335	4.44607148	-5.38912158

EQUATION 3

E	CONSTANT	E LAGGED	TIME
6	0	13	10
11.32656587	1.44913135	4.77516954	4.82068953

Due to cvc after 3SLS code

COEF VAR-COV MATRIX — SQ RT OF(T-N(J))(T-N(I))USED AS THE DENOMINATOR — (THREE STAGE LEAST SQUARES)

P	4	0.01444292					
W	5	-0.00144339	0.00177794				
CONSTANT	0	-0.02097682	-0.03793805	2.10228216			
P LAGGED	11	-0.00995945	-0.00069815	-0.00689123	0.01246144		
P	4	0.00752736	0.00109421	-0.00312366	-0.00729144	0.03237754	
CONSTANT	0	-0.01933679	-0.07668003	2.42677103	0.06755864	-0.90454202	57.01538683
P LAGGED	11	-0.00610975	-0.00106024	0.01891639	0.00782797	-0.02768463	0.74673460
K LAGGED	12	-0.00803898	0.00037687	-0.00904515	-0.00036208	0.00403866	-0.26875723
E	6	-0.00167313	-0.00004568	0.01228997	0.00109869	0.00004008	0.04684681
CONSTANT	0	0.00104928	0.00341395	-0.76008260	0.03530044	-0.00207132	0.34161665
E LAGGED	13	0.00171480	-0.00001165	-0.00001170	-0.00173843	-0.00000571	-0.05421012
TIME	10	0.00078992	-0.00063866	0.00419101	0.00054708	0.00060876	-0.04077922
		4	5	0	11	4	0
		P	W	CONSTANT	P LAGGED	P	CONSTANT
P LAGGED	11	0.02889173					
K LAGGED	12	-0.00375875	0.00130725				
E	6	0.00021905	-0.00025481	0.00125023			
CONSTANT	0	-0.01826928	0.00001902	-0.00976670	1.53810465		
E LAGGED	13	0.00008916	0.00026358	-0.00112646	-0.01592667	0.00144137	
TIME	10	-0.00081949	0.00021896	-0.00034925	0.01737095	0.00006215	0.00096400
		11	12	6	0	13	10
		P LAGGED	K LAGGED	E	CONSTANT	E LAGGED	TIME

Due to cvcT after 3SLS code

COEF VAR-COV MATRIX — T USED AS THE DENOMINATOR — (THREE STAGE LEAST SQUARES)

P	4	0.01169189					
W	5	-0.00116846	0.00143928				
CONSTANT	0	-0.01698124	-0.03071175	1.70184746			
P LAGGED	11	-0.00806241	-0.00056917	-0.00597861	0.01008783		
P	4	0.00609357	0.00088579	-0.04300487	-0.00590260	0.02621039	
CONSTANT	0	-0.01565359	-0.06207431	1.96492893	0.05469032	-0.73224830	46.15931315
P LAGGED	11	-0.00494275	-0.00085829	0.01531326	0.00633693	-0.02234677	0.60449943
K LAGGED	12	-0.00803155	0.00030509	-0.00732227	-0.00029311	0.00326939	-0.21756538
E	6	-0.00135444	-0.00003691	0.00994902	0.00088294	0.00003238	0.03791875
CONSTANT	0	0.00084935	0.00276368	-0.01530512	0.02857655	-0.00167678	0.27654682
E LAGGED	13	0.00138817	-0.00000943	-0.00000947	-0.00140730	-0.00000462	-0.04388438
TIME	10	0.00063946	-0.00051701	0.00339272	0.00044287	0.00049281	-0.03301175
		4	5	0	11	4	0
		P	W	CONSTANT	P LAGGED	P	CONSTANT
P LAGGED	11	0.02338854					
K LAGGED	12	-0.00304288	0.00105825				
E	6	0.00017651	-0.00020627	0.00101209			
CONSTANT	0	-0.01478618	0.00001540	-0.00790638	1.24513234		
E LAGGED	13	0.00007218	0.00021337	-0.00091190	-0.01289302	0.00116682	
TIME	10	-0.00066256	0.00017725	-0.00028272	0.01406220	0.00005041	0.00078038
		11	12	6	0	13	10
		P LAGGED	K LAGGED	E	CONSTANT	E LAGGED	TIME

Due to CVCN after 3SLS code

COEF VAR-COV MATRIX NORMALIZED (THREE STAGE LEAST SQUARES)

P	4	1,00000										
W	5	-0,28484	1,00000									
CONSTANT	0	-0,12038	-0,62054	1,00000								
P LAGGED	11	-0,74238	-0,14832	-0,04258	1,00000							
P	4	0,34809	0,14422	-0,20362	-0,36300	1,00000						
CONSTANT	0	-0,02131	-0,24084	0,22166	0,08015	-0,66575	1,00000					
P LAGGED	11	-0,29890	-0,14793	0,07675	0,41255	-0,90256	0,98181	1,00000				
K LAGGED	12	-0,00897	0,24720	-0,17254	-0,08971	0,62078	-0,98443	-0,61162	1,00000			
E	6	-0,39374	-0,03058	0,23972	0,27633	0,00629	0,17544	0,03628	-0,19931	1,00000		
CONSTANT	0	0,00704	0,06528	-0,42269	0,25498	-0,00928	0,03648	-0,08665	0,00042	-0,22272	1,00000	
E LAGGED	13	0,37584	-0,00728	-0,00021	-0,41019	-0,00084	-0,18910	0,01382	0,19202	-0,83914	-0,33826	1,00000
TIME	10	0,21170	-0,48783	$0.^{9}310$	0,15784	0,10896	-0,17394	$-0.^{1}5508$	$0.^{1}9505$	-0,31812	$0.^{4}5112$	$0.^{0}5272$
		4	5	0	11	4	0	11	12	6	0	13
		P	W	CONSTANT	P LAGGED	P	CONSTANT	P LAGGED	K LAGGED	E	CONSTANT	E LAGGED
TIME	10	1,00000										
		10										
		TIME										

Due to DVCN after 3SLS code

DISTURBANCE VAR-COV MATRIX (THREE STAGE LEAST SQUARES)

(-SQ RT OF(T-N(J))(T-N(I))USED AS THE DENOMINATOR

C	1	1,10158567			
I	2	0,58809972	2,58552816		
W1	3	-0,48622972	0,49788022	0,64238586	
		1	2	3	
		C	I	W1	

Due to DVCN after 3SLS code

DISTURBANCE VAR-COV MATRIX NORMALIZED (THREE STAGE LEAST SQUARES)

C	1	1,00000		
I	2	0,30187	1,00000	
W1	3	-0,57801	0,38632	1,00000
		1	2	3
		C	I	W1

Due to RES after 3SLS code

Numbering of residuals by year due to NR= 20 code.

RESIDUALS (THREE STAGE LEAST SQUARES)

OB		y for eqn 1 C: Y = X(1)	$\hat{y}$ for eqn 1 ESTIMATED Y	$\hat{u}$ for eqn 1 Y - ESTIMATED Y	y for eqn 2 I: Y = X(2)	$\hat{y}$ for eqn 2 ESTIMATED Y	$\hat{u}$ for eqn 2 Y - ESTIMATED Y
1921	1	41,90000000	42,34164434	-0,44164434	-0,20000000	1,99509936	-2,19509936
1922	2	45,00000000	46,01503199	-1,01503199	1,90000000	1,74849550	0,15150450
1923	3	49,20000000	50,72890462	-1,52890462	5,20000000	4,75942288	0,44057712
1924	4	50,60000000	51,09851123	-0,49851123	3,00000000	4,86671874	-1,86671874
1929	5	52,60000000	52,61320815	-0,01320815	5,10000000	5,02874253	0,07125747
1926	6	55,10000000	54,32413375	0,77586625	5,60000000	4,57056282	1,02943718
1927	7	56,20000000	54,89960455	1,30039545	4,20000000	3,09893481	1,10106519
1928	8	57,30000000	56,20070429	1,09929571	3,00000000	2,41471402	0,58528598

1929	9	57,80000000	58,38392015	-0,58392015
1930	10	55,80000000	55,19171571	-0,19171571
1931	11	50,90000000	51,45977013	-0,55977013
1932	12	45,60000000	46,27464216	-0,67464217
1933	13	46,50000000	45,92333196	0,57666804
1934	14	48,70000000	48,72111902	-0,02111902
1935	15	51,30000000	51,24610985	0,05389015
1936	16	57,70000000	55,84445711	1,85554289
1937	17	58,70000000	59,15959198	-0,45959198
1938	18	57,50000000	57,43872212	0,06127788
1939	19	61,60000000	60,33981196	1,26018803
1940	20	65,00000000	64,05000648	0,94999352
1941	21	69,70000000	71,64505845	-1,94505845

SUM Y $\leftarrow \sum_{t=1}^{T} y_t$: 1133,89999998

SUM RES $\leftarrow \sum_{t=1}^{T} \hat{u}_t$: -0,00000000

SSY $\leftarrow \sum_{t=1}^{T} y_t^2$: 62166,62999916

SS(Y - MEAN Y) $\leftarrow \sum_{t=1}^{T} (y_t - \bar{y})^2$: 941,42952380

SS RES $\leftarrow \sum_{t=1}^{T} \hat{u}_t^2$: 18,72699635

R2 $\leftarrow R^2$: 0,98010796

D.W. STAT ← Durbin-Watson test statistic : 1,42493901

SS(R - PREV R) $\leftarrow \sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2$: 26,68477063

Due to RES after 3SLS code.

		5,10000000	2,80476291	2,29523709
		1,00000000	2,34425423	-1,34425423
		-3,40000000	-2,40557762	-0,99442238
		-6,20000000	-4,85958581	-1,34041419
		-5,10000000	-7,03164467	1,93164467
		-3,00000000	-2,87826506	-0,12173494
		-1,30000000	-1,48465856	0,18465856
		2,10000000	0,00628984	2,09371016
		2,00000000	2,32163854	-0,32163854
		-1,90000000	1,73138322	-3,63138322
		1,30000000	0,54175399	0,75824601
		3,30000000	3,05716364	0,24283636
		4,90000000	3,96979470	0,93020530

SUM Y : 26,60000000

SUM RES : -0,00000000

SSY : 286,02000000

SS(Y - MEAN Y) : 252,32666666

SS RES : 43,95397874

R2 : 0,82580526

D.W. STAT : 1,99588410

SS(R - PREV R) : 87,72784729

(THREE STAGE LEAST SQUARES)

Numbering of residuals by year due to NR=20 code.

R E S I D U A L S

	OB	y for eqn 3 (W1) Y = Y(3)	$\hat{y}$ for eqn 3 ESTIMATED Y	$\hat{u}$ for eqn 3 Y - ESTIMATED Y
1921	1	25,50000000	26,70287287	-1,20287287
1922	2	29,30000000	28,78166415	0,51833585
1923	3	34,10000000	32,59064018	1,50935982
1924	4	33,90000000	33,90743131	-0,00743131
1925	5	35,40000000	35,68089466	-0,28089466
1926	6	37,40000000	37,73907937	-0,33907937
1927	7	37,90000000	38,59282328	-0,69282328
1928	8	39,20000000	38,85506299	0,34493701
1929	9	41,30000000	40,02409591	1,27590409
1930	10	37,90000000	38,30414466	-0,40414466
1931	11	34,50000000	34,27849422	0,22150578
1932	12	29,00000000	29,36962232	-0,36962232
1933	13	28,50000000	28,18994170	0,31005830
1934	14	30,60000000	30,32691127	0,27308873
1939	15	33,20000000	33,19283589	0,00716411
1936	16	36,80000000	37,51866038	-0,71866038
1937	17	41,00000000	40,09418124	0,90581876
1938	18	38,20000000	39,01880798	-0,81880799
1939	19	41,60000000	41,86941910	-0,26941911
1940	20	45,00000000	46,00124560	-1,00124560
1941	21	53,30000000	52,42117088	0,87882911

SUM Y : 763,59999999

SUM RES : 0,00000000

SSY : 28568,85999966

SS(Y - MEAN Y) : 794,90952380

SS RES : 10,92059968

R2 : 0,98626188

D.W. STAT : 2,15904575

SS(R - PREV R) : 23,53438569

Due to L.I.K. after 3SLS code

DET GAM= 1.807419-001 DET S= 5.572580-013 LIK= 9.862207+010

↑ det P_{3SLS} computed from normalized coefficients

↑ det S_{3SLS} computed from normalized coefficients

↑ det P_{3SLS} / det S_{3SLS}

REDUCED FORM COEFS (THREE STAGE LEAST SQUARES)

Due to RF after 3SLS code

1. DEP. VAR.= 1 C

CONSTANT 0	P LAGGED 11	K LAGGED 12	E LAGGED 13	TIME 10	R 6
46.72729776	0.74630692	-0.12366112	0.19863271	0.16399144	-0.19585198
8	W2				
9	7				
0.03465358	0.05685587				

2. DEP. VAR.= 4 P

CONSTANT 0	P LAGGED 11	K LAGGED 12	E LAGGED 13	TIME 10	R 6
42.77363341	0.89347391	-0.18946336	-0.00172918	-0.03896833	-1.10872124
8	W2				
9	7				
0.97236367	-0.20411767				

3. DEP. VAR.= 5 W

CONSTANT 0	P LAGGED 11	K LAGGED 12	E LAGGED 13	TIME 10	R 6
31.97206787	0.59687186	-0.12656799	0.20116912	0.21561829	-0.07262950
8	W2				
9	7				
0.04997211	0.06364243				

4. DEP. VAR.= 2 I

CONSTANT 0	P LAGGED 11	K LAGGED 12	E LAGGED 13	TIME 10	R 6
27.61048271	0.74403805	-0.19237022	0.00886731	0.00866052	0.01450117
8	W2				
9	7				
-0.01271772	0.00266969				

5. DEP. VAR.= 3 W1

CONSTANT 0	P LAGGED 11	K LAGGED 12	E LAGGED 13	TIME 10	R 6
31.97206787	0.59687186	-0.12656799	0.20116912	0.21561829	-0.07262950
8	W2				
9	7				
0.04997211	-0.13639757				

Due to RF after 3SLS code

6. DEP. VAR.= 6 E

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	8
74.34578847	1.49834497	-0.31603135	0.19944082	0.16465796	-0.18135074

8	M2
9	7
1.62193578	-0.34847524

7. DEP. VAR.= 14 Y

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	8
74.34578847	1.49834497	-0.31603135	0.19944082	0.16465796	-1.18135074

8	M2
9	7
1.62193578	0.65952476

8. DEP. VAR.= 15 K

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	8
27.81848271	0.74403805	0.80762978	0.00088731	0.00066652	0.01450117

8	M2
9	7
-0.81271772	0.00266969

Iteration started toward I3SLS coefficients due to I3SLS code on 3SLS card.

```
ITER NO,   2      MAX ABS CHANGE RATIO FROM PREV COEFS=     1,048822-001
ITER NO,   3      MAX ABS CHANGE RATIO FROM PREV COEFS=     5,763712-001
ITER NO,   4      MAX ABS CHANGE RATIO FROM PREV COEFS=     2,171606-001
ITER NO,   5      MAX ABS CHANGE RATIO FROM PREV COEFS=     1,027887-001
ITER NO,   6      MAX ABS CHANGE RATIO FROM PREV COEFS=     5,286157-002
ITER NO,   7      MAX ABS CHANGE RATIO FROM PREV COEFS=     2,829082-002
ITER NO,   8      MAX ABS CHANGE RATIO FROM PREV COEFS=     1,549538-002
ITER NO,   9      MAX ABS CHANGE RATIO FROM PREV COEFS=     8,618531-003   <- max abs_j [ (a_j^(ω) - a_j^(i-1)) / a_j^(i-1) ] = .00861853 1
ITER NO,  10      MAX ABS CHANGE RATIO FROM PREV COEFS=     4,845825-003
ITER NO,  11      MAX ABS CHANGE RATIO FROM PREV COEFS=     2,749297-003
ITER NO,  12      MAX ABS CHANGE RATIO FROM PREV COEFS=     1,563106-003
ITER NO,  13      MAX ABS CHANGE RATIO FROM PREV COEFS=     8,927143-004
ITER NO,  14      MAX ABS CHANGE RATIO FROM PREV COEFS=     5,106840-004
ITER NO,  15      MAX ABS CHANGE RATIO FROM PREV COEFS=     2,923557-004
ITER NO,  16      MAX ABS CHANGE RATIO FROM PREV COEFS=     1,674028-004
ITER NO,  17      MAX ABS CHANGE RATIO FROM PREV COEFS=     9,585056-005
ITER NO,  18      MAX ABS CHANGE RATIO FROM PREV COEFS=     5,487439-005
ITER NO,  19      MAX ABS CHANGE RATIO FROM PREV COEFS=     3,141145-005
ITER NO,  20      MAX ABS CHANGE RATIO FROM PREV COEFS=     1,797897-005
```

```
                          COEFFICIENTS                                                         (NTH ITER--ITERATIVE 3SLS)
EQUATION  1
           C                     P                     W                    CONSTANT              P LAGGED
           1                     4                     5                       0                     11
    -1,0000000000          0.1645088321          0.7658016213          16.5589838852          0.1765637200
EQUATION  2
           I                     P                                          CONSTANT              P LAGGED              K LAGGED
           2                     4                                             0                     11                    12
    -1,0000000000         -0.3565236984         42,8959233453           1,0112934778         -0,2601983806
EQUATION  3
           d1                    E                    CONSTANT              E LAGGED              TIME
           3                     6                       0                     13                    10
    -1,0000000000          0.3747796305          2,6247474822          0.1936505156          0.1679257251

DET GAM=            2,442085-001     DET S=          1.002644-012     LIK=           5.948051-010
```

Due to C after NTH=20 code

Due to LIK after NTH=20 code

Coefficients requested at the end of each 20 iterations by the NTH=20 code on the 3SLS card.

```
ITER NO, 21     MAX ABS CHANGE RATIO FROM PREV COEFS=     1,029008-005
ITER NO, 22     MAX ABS CHANGE RATIO FROM PREV COEFS=     5,889300-006
ITER NO, 23     MAX ABS CHANGE RATIO FROM PREV COEFS=     3,370614-006
ITER NO, 24     MAX ABS CHANGE RATIO FROM PREV COEFS=     1,929119-006
ITER NO, 25     MAX ABS CHANGE RATIO FROM PREV COEFS=     1,104116-006
ITER NO, 26     MAX ABS CHANGE RATIO FROM PREV COEFS=     6,319399-007
ITER NO, 27     MAX ABS CHANGE RATIO FROM PREV COEFS=     3,616932-007
ITER NO, 28     MAX ABS CHANGE RATIO FROM PREV COEFS=     2,070176-007
ITER NO, 29     MAX ABS CHANGE RATIO FROM PREV COEFS=     1,184882-007
ITER NO, 30     MAX ABS CHANGE RATIO FROM PREV COEFS=     6,781777-008
ITER NO, 31     MAX ABS CHANGE RATIO FROM PREV COEFS=     3,881608-008
ITER NO, 32     MAX ABS CHANGE RATIO FROM PREV COEFS=     2,221670-008
ITER NO, 33     MAX ABS CHANGE RATIO FROM PREV COEFS=     1,271590-008
ITER NO, 34     MAX ABS CHANGE RATIO FROM PREV COEFS=     7,278042-009
ITER NO, 35     MAX ABS CHANGE RATIO FROM PREV COEFS=     4,165641-009
ITER NO, 36     MAX ABS CHANGE RATIO FROM PREV COEFS=     2,384235-009
ITER NO, 37     MAX ABS CHANGE RATIO FROM PREV COEFS=     1,364634-009
ITER NO, 38     MAX ABS CHANGE RATIO FROM PREV COEFS=     7,810581-010
ITER NO, 39     MAX ABS CHANGE RATIO FROM PREV COEFS=     4,470442-010
ITER NO, 40     MAX ABS CHANGE RATIO FROM PREV COEFS=     2,958690-010
```

$$\max_j \text{abs} \left[\frac{\hat{a}_j^{(i)} - \hat{a}_j^{(i-1)}}{\hat{a}_j^{(i-1)}} \right] = .000001104116$$

```
                          COEFFICIENTS                                                         (NTH ITER--ITERATIVE 3SLS)
EQUATION  1
           C                     P                     W                    CONSTANT              P LAGGED
           1                     4                     5                       0                     11
    -1,0000000000          0,1645097662          0,7658010837          16,5589839816          0,1765641125
EQUATION  2
           I                     P                                          CONSTANT              P LAGGED              K LAGGED
           2                     4                                             0                     11                    12
    -1,0000000000         -0,3565322766         42,8963092873           1,0112993676         -0,2602000639
EQUATION  3
           d1                    E                    CONSTANT              E LAGGED              TIME
           3                     6                       0                     13                    10
    -1,0000000000          0,3747791090          2,6247708408          0,1936506529          0,1679263592

DET GAM=            2,442101-001     DET S=          1,002660-012     LIK=           5,948037-010
```

Due to C after NTH=20 code

Due to LIK after NTH=20 code

Statistics requested at the end of each 20 iterations due to using the NTH=20 code on the 3SLS card.

ITER NO. 41 MAX ABS CHANGE RATIO FROM PREV COEFS= 1.464485-010

ITER NO. 42 MAX ABS CHANGE RATIO FROM PREV COEFS= 8.382085-011 ← $\max_i$ abs $\left[\frac{\delta_i^{(k)} - \delta_i^{(k-1)}}{\delta_i^{(k-1)}}\right]$

of .00000000 0838 2085 meets coefficient convergence criterion of .00000001 set by the CCC 3SLS-10 code on the 3SLS card

COEFFICIENTS δ_i (CONVERGED ITERATIVE 3 STAGE LS)

Due to C after 3SLS code

EQUATION 1

C	P	W	CONSTANT	P LAGGED
1	4	5	0	11
-1.0000000000	0.1645097662	0.7658010857	16.5589839816	0.1765641125

EQUATION 2

I	P	CONSTANT	P LAGGED	K LAGGED
2	4	0	11	12
-1.0000000000	-0.3565322767	42.8963092918	1.0112993676	-0.2682888639

EQUATION 3

M1	E	CONSTANT	E LAGGED	TIME
3	6	0	13	10
-1.0000000000	0.3747791898	2.6247788418	0.1936506529	0.1679263592

COEFFICIENT STANDARD ERRORS $\sqrt{V(\delta_i)}$ (CONVERGED ITERATIVE 3 STAGE LS)

SQ RT OF T-N(I) USED AS DENOMINATOR

Due to STE after 3SLS code

EQUATION 1

P	W	CONSTANT	P LAGGED
4	5	0	11
0.10891792	0.03863350	1.36884681	0.10814067

EQUATION 2

P	CONSTANT	P LAGGED	K LAGGED
4	0	11	12
0.28914848	11.77442895	0.27649778	0.05653823

EQUATION 3

E	CONSTANT	E LAGGED	TIME
6	0	13	10
0.03456876	1.32879133	0.0360126l	0.03215287

COEFFICIENTS/COEF STANDARD ERRS $\hat{\delta}_i / \sqrt{V(\delta_i)}$ (CONVERGED ITERATIVE 3 STAGE LS)

SQ RT OF T-N(I) USED AS FACTOR

Due to COSTE after 3SLS code

EQUATION 1

P	W	CONSTANT	P LAGGED
4	5	0	11
1.53865471	19.82220183	12.16815415	1.76316082

EQUATION 2

P	CONSTANT	P LAGGED	K LAGGED
4	0	11	12
-1.23304218	3.64317535	3.65753166	-4.60219683

EQUATION 3

E	CONSTANT	E LAGGED	TIME
6	0	13	10
10.84195552	1.97538782	5.37730117	5.22274794

Due to CVC after IBSLS code

COEF VAR-COV MATRIX — SQ RT OF(T-N(J))(T-N(I))USED AS THE DENOMINATOR — (CONVERGED ITERATIVE 3 STAGE LS)

P	4	0.01143144					
W	5	-0.00109559	0.00149255				
CONSTANT	0	-0.01803473	-0.03367482	1.85190185			
P LAGGED	11	-0.00791402	-0.00059430	-0.00509911	0.01802815		
P	4	0.01612216	0.00079177	-0.07425973	-0.01399808	0.00360485	
CONSTANT	0	-0.13816587	-0.07997615	3.31601422	0.13476795	-2.04730956	138.63717701
P LAGGED	11	-0.01333092	-0.00100119	0.02441296	0.01479483	-0.07043679	1.64864944
K LAGGED	12	0.00041978	0.00039739	-0.01208894	-0.00070135	0.00892109	-0.65207858
E	6	-0.00130449	-0.00012486	0.01302749	0.00081736	0.00008411	0.12970084
CONSTANT	0	0.00184949	0.00693078	-0.86136053	0.03158895	-0.00995401	0.94959991
E LAGGED	13	0.00131926	0.00000979	0.00009547	-0.00139133	0.00008455	-0.14996549
TIME	10	0.00088213	-0.00068735	0.00544743	0.00049859	0.00155677	-0.11115933
		4 P	5 W	0 CONSTANT	11 P LAGGED	4 P	0 CONSTANT
P LAGGED	11	0.07645182					
K LAGGED	12	-0.00851948	0.00819657				
E	6	0.00083018	-0.00072180	0.00119508			
CONSTANT	0	-0.06184214	0.00130368	-0.01398339	1.76568639		
E LAGGED	13	0.00019287	0.00072510	-0.00099654	-0.01535305	0.00129691	
TIME	10	-0.00226636	0.00060837	-0.00041611	0.01848243	0.00011362	0.00103381
		11 P LAGGED	12 K LAGGED	6 E	0 CONSTANT	13 E LAGGED	10 TIME

Due to CVCT after IBSLS code

COEF VAR-COV MATRIX — T USED AS THE DENOMINATOR — (CONVERGED ITERATIVE 3 STAGE LS)

P	4	0.00925402					
W	5	-0.00088691	0.00120825				
CONSTANT	0	-0.01459954	-0.02726057	1.49915864			
P LAGGED	11	-0.00640659	-0.00048110	-0.00477547	0.00811803		
P	4	0.01305127	0.00064857	-0.06011502	-0.01133178	0.00768173	
CONSTANT	0	-0.11184856	-0.06150451	2.68439246	0.10909786	-1.65734583	112.23009567
P LAGGED	11	-0.01079178	-0.00081049	0.01976287	0.01197677	-0.05702026	1.33219241
K LAGGED	12	0.00033983	0.00032169	-0.00978628	-0.00056776	0.00722184	-0.52787314
E	6	-0.00105602	-0.00010108	0.01119365	0.00066167	0.00006809	0.10499592
CONSTANT	0	0.00149397	0.00561063	-0.69745377	0.02597201	-0.00805849	0.76872374
E LAGGED	13	0.00106798	0.00000793	0.00007729	-0.00112632	0.00006845	-0.12140063
TIME	10	0.00071418	-0.00055643	0.00440983	0.00040362	0.00126025	-0.08998289
		4 P	5 W	0 CONSTANT	11 P LAGGED	4 P	0 CONSTANT
P LAGGED	11	0.06188992					
K LAGGED	12	-0.00689991	0.00658770				
E	6	0.00067209	-0.00058431	0.00096738			
CONSTANT	0	-0.04941507	0.00105936	-0.01151989	1.42936518		
E LAGGED	13	0.00015613	0.00058698	-0.00080672	-0.01242866	0.00104988	
TIME	10	-0.00183467	0.00049249	-0.00083685	0.01489721	0.00009198	0.00083689
		11 P LAGGED	12 K LAGGED	6 E	0 CONSTANT	13 E LAGGED	10 TIME

COEF VAR-COV MATRIX NORMALIZED (CONVERGED ITERATIVE 3 STAGE LS)

Due to OVER-ID after ZSLS code

P	4	1.00000											
W	5	-0.26524	1.00000										
CONSTANT	8	-0.12305	-0.84052	1.00000									
P LAGGED	11	-0.73916	-0.15361	-0.84329	1.00000								
P	4	0.52150	0.06738	-0.10872	-0.48343	1.00000							
CONSTANT	8	-0.10975	-0.16782	0.20695	0.11430	-0.88134	1.00000						
P LAGGED	11	-0.45194	-0.09373	0.06488	0.53433	-0.88102	0.90548	1.00000					
K LAGGED	12	0.86944	0.10193	-0.15712	-0.12387	0.94570	-0.97953	-0.54491	1.00000				
E	6	-0.35294	-0.09349	0.29393	0.23611	0.88842	0.91865	0.88686	-0.36931	1.00000			
CONSTANT	8	0.81299	0.13981	-0.47645	0.23739	-0.82591	0.86869	-0.16614	0.81735	-0.38442	1.00000		
E LAGGED	13	0.34263	0.08784	0.08195	-0.38588	0.88812	-0.95367	0.01937	0.35612	-0.88849	-0.32884	1.00000	
TIME	10	0.29648	-0.58334	0.12458	0.15485	0.16745	-0.29361	-0.25493	0.33466	-0.37438	0.43872	0.89812	1.00000
		4	5	8	11	4	8	11	12	6	8	13	
		P	W	CONSTANT	P LAGGED	P	CONSTANT	P LAGGED	K LAGGED	E	CONSTANT	E LAGGED	

TIME	10	1.00000
		10
		TIME

DISTURBANCE VAR-COV MATRIX (CONVERGED ITERATIVE 3 STAGE LS)

(SQ RT OF(T-N(J))(T-N(I))USED AS THE DENOMINATOR

Due to OVER-ID after ZSLS code

C	1	1.13818163		
I	2	0.79273541	5.62728482	
W1	3	-0.53733378	0.98732882	0.74815482
		1	2	3
		C	I	W1

DISTURBANCE VAR-COV MATRIX NORMALIZED (CONVERGED ITERATIVE 3 STAGE LS)

Due to OVER-ID after ZSLS code

C	1	1.00000		
I	2	0.31435	1.00000	
W1	3	-0.58435	0.44220	1.00000
		1	2	3
		C	I	W1

Due to LIK after ZSLS code

DET SAM= 2.442181-001 DET S= 1.882660-012 LIK= 5.948837-010

$\det \hat{\Gamma}_{ZSLS}$ computed from normalized coefficients

$\det S_{ZSLS}$ computed from normalized coefficients

$\det \hat{\Gamma}_{ZSLS} / \det S_{ZSLS}$

LML, FIML STARTING FROM 2SLS--KLEIN MODEL I

CURRENT TIME 1712 . 12 DATE 09/28/68
ELAPSED SINCE LAST CURRENT TIME 2,39 SECONDS

```
U  LML, FIML STARTING FROM 2SLS--KLEIN MODEL I                    (STAT CONTROL CARD)
FIML(1,2,3/4,,,8)                                                 (STAT CONTROL CARD)
                LML=C,DVC,DVCN,CVC,CVCT                           (CONTINUATION TO ABOVE CARD)
                NT4=18=C                                          (CONTINUATION TO ABOVE CARD)
                FIML=C,DVC,CVC,CVCT,CVCN,DVCN,RES,NR=20,RF,STE,COSTE   (CONTINUATION TO ABOVE CARD)
                LIMIT=C,TIMEL=9,PCC=1,0=10,MI=50                  (CONTINUATION TO ABOVE CARD)
```

(BEGINNING ESTIMATES--PRE FI)

1. COEFS FOR EQ 1 PREVIOUSLY SAVED FROM 2SLS
2. COEFS FOR EQ 2 PREVIOUSLY SAVED FROM 2SLS
3. COEFS FOR EQ 3 PREVIOUSLY SAVED FROM 2SLS
4. COEFS FOR EQ 4 PREVIOUSLY SAVED FROM READC
5. COEFS FOR EQ 5 PREVIOUSLY SAVED FROM READC
6. COEFS FOR EQ 6 PREVIOUSLY SAVED FROM READC
7. COEFS FOR EQ 7 PREVIOUSLY SAVED FROM READC
8. COEFS FOR EQ 8 PREVIOUSLY SAVED FROM READC

ENDOG VARS, IN ENTIRE SYSTEM

C	P	W	I	W1	E	Y	K
1	4	5	2	3	6	14	15

← jointly dependent variables in the system

PREDO VARS, IN ENTIRE SYSTEM

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R	G	W2
0	11	12	13	10	8	V	7

← predetermined variables in the system

VARS IN SUBSYSTEM 1

C	P	W	CONSTANT	P LAGGED	I	K LAGGED	W1	E	E LAGGED	TIME
1	4	5	0	11	2	12	3	6	13	10

← variables in the stochastic equations

ITER MAX	NUM ITER/ PRINT	CONVERGENCE LEVELS COEF	PARTIAL	LIK	ITER TIME LIMIT	MIN TO STOP ITER BEFORE JOB OR TIME LIMIT EXPIRES
50	10	0,000-000	1,000-010	0,000-000	9,00	1,00

OLD LIK= 0,000-000 MAX ABS RHS ELT= 0,000-000 MAX ABS COEF CHANGE IF STEP OF 1= 0,000-000

ITER	STATE	STEP	DET S	DET GAMMA	NEW LIK	RATIO-1	MAX COEF CHANGE
0	17171	0,000-000	9,665-013	1,614-001	4,996-010	BEG LIK STATS	

$\det^{\circ} \hat{P}_{2SLS} / \det S_{2SLS} = 95,96\,900\,900$

SEE statistics (regarding the 2SLS coefficients) could have been output here.

ITER 1, EIGVALS OF METRIC--MAX= 7,78-885, MIN= 1,57-881, MIN ABS= 1,57-881

← smallest eigenvalue is positive; therefore, for the first iteration, the coefficients are in a positive definite region

OLD LIK= 4,596-818 MAX ABS RHS ELT= 2,411-882 MAX ABS COEF CHANGE IF STEP OF 1= 5,764-888

ITER	STATE	STEP	DET S	DET GAMMA	NEW LIK	RATIO-1	MAX COEF CHANGE
1	66474	1,888-888	8,117-813	2,239-881	6,177-818	5,441-811	5,764-888

step size of 1 imposed the 1st iteration due to LIML code on the FIXML card

$f'(\alpha_{max}) = \det^2 \Gamma_{new} / \det^2 \Gamma_{old}$ = 6.197 approx

$\frac{f'(\alpha_{max})}{f'(\alpha_{max})} - 1 = .344$

max abs $\left[\frac{a_i^{(max)} - a_i^{(const)}}{a_i^{(const)}}\right] = 5.769$

(LINEARIZED MAXIMUM LIKELIHOOD)

COEFFICIENTS (LINEARIZED MAXIMUM LIKELIHOOD)

Due to C after LIML code on FIXML card

EQUATION 1

C	P	W	CONSTANT	P LAGGED
1	4	5	0	11
-1,0000000000	0,1178384349	0,7798674714	16,4178483589	0,2885498484

EQUATION 2

I	P	CONSTANT	P LAGGED	K LAGGED
2	4	0	11	12
-1,0000000000	-0,1562837381	32,2488646918	0,8656799788	-0,2128328885

EQUATION 3

W1	E	CONSTANT	E LAGGED	TIME
3	6	0	13	10
-1,0000000000	0,3365197917	2,3687984143	0,2378291225	0,1818642230

Due to CVC after LIML code

COEF VAR-COV MATRIX (LINEARIZED MAXIMUM LIKELIHOOD)

← SQ RT OF (T-N(J))(T-N(I)) USED AS THE DENOMINATOR

P	4	0,01213348						
W	5	-0,00169659	0,00176316					
CONSTANT	0	-0,01086852	-0,03573268	2,01692578				
P LAGGED	11	-0,00755337	-0,00053424	-0,01789096	0,01022418			
P	4	0,00204981	0,00090633	-0,01467025	-0,00351408	0,02769019		
CONSTANT	0	0,13591448	-0,07723769	1,21234265	-0,01699749	-0,79109925	53,23174727	
P LAGGED	11	-0,00153366	-0,00088539	-0,01402642	0,00468310	-0,02317325	0,62709385	
K LAGGED	12	-0,00072515	0,00038120	-0,00393674	-0,00000169	0,00350576	-0,24967051	
E	6	-0,00194696	0,00008148	0,01209628	0,00106307	0,00293312	-0,08056975	
CONSTANT	0	0,00835918	0,00192039	-0,76590584	0,03189975	-0,02629182	1,50070665	
E LAGGED	13	0,00187235	-0,00011750	0,00028934	-0,00165118	-0,00258448	0,05776272	
TIME	10	0,00112341	-0,00071979	0,00403192	0,00041834	-0,00067918	0,02156124	
		4	5	0	11	4	0	
		P	W	CONSTANT	P LAGGED	P	CONSTANT	
P LAGGED	11	0,02472278						
K LAGGED	12	-0,00319485	0,00121088					
E	6	-0,00224678	0,00033827	0,00168488				
CONSTANT	0	0,00275761	-0,00343881	-0,01269336	1,59849498			
E LAGGED	13	0,00227948	-0,00025656	-0,00132617	-0,01324711	0,00188914		
TIME	10	0,00029531	-0,00007444	-0,00052881	0,01861879	0,00021852	0,00104540	
		11	12	6	0	13	10	
		P LAGGED	K LAGGED	E	CONSTANT	E LAGGED	TIME	

COEF VAR-COV MATRIX ---------- T USED AS THE DENOMINATOR (LINEARIZED MAXIMUM LIKELIHOOD)

Due to CVCT after LML code

P	4	0.00982234					
W	5	-0.00137348	0,00142732				
CONSTANT	0	-0.00879832	-0,02892645	1,63274938			
P LAGGED	11	-0.00611464	-0,00043248	-0,01432126	0,00827665		
P	4	0.00165937	0,00073369	-0,01187591	-0,00284473	0,02241987	
CONSTANT	0	0.11802601	-0.06252575	0.98142024	-0.01375987	-0,64041368	43,09236684
P LAGGED	11	-0.00124315	-0.00071675	-0.01135472	0.00379108	-0,01875930	0,50764740
K LAGGED	12	-0.00058702	0.00030859	-0.00286308	-0.00000137	0.00283800	-0,20211422
E	6	-0.00157611	0,00006596	0,00979223	0.00086058	0.00237443	-0,06522313
CONSTANT	0	0.00676689	0.00155461	-0,62001901	0,02582361	-0,02128385	1,21485777
E LAGGED	13	0.00151371	-0.00009512	0,00023422	-0,00133667	-0,00209220	0,04676029
TIME	10	0.00090943	-0.00058269	0,00326393	0,00033866	-0,00054981	0,01745434
		4	5	0	11	4	0
		P	W	CONSTANT	P LAGGED	P	CONSTANT

P LAGGED	11	0.02001368					
K LAGGED	12	-0.00258630	0.00098024				
E	6	-0.00181982	0.00027384	0.00136395			
CONSTANT	0	0.00223235	-0.00448284	-0.01027558	1.26163879		
E LAGGED	13	0.00184930	-0.00020769	-0.00123547	-0.01072385	0.00146454	
TIME	10	0.00023986	-0.00006026	-0.00042161	0.01507236	0.00017674	0.00084628
		11	12	6	0	13	10
		P LAGGED	K LAGGED	E	CONSTANT	E LAGGED	TIME

DISTURBANCE VAR-COV MATRIX (LINEARIZED MAXIMUM LIKELIHOOD)
SQ RT OF(T-N(J))(T-N(I))USED AS THE DENOMINATOR

Due to DVC after LML code

C	1	1.15076738		
I	2	0.70010032	3.60933151	
W1	3	-0.50901081	0.98341883	0.90250129
		1	2	3
		C	I	W1

DISTURBANCE VAR-COV MATRIX NORMALIZED (LINEARIZED MAXIMUM LIKELIHOOD)

Due to DVCN after LML code

C	1	1.00000		
I	2	0.34352	1.00000	
W1	3	-0.48365	0.52762	1.00000
		1	2	3
		C	I	W1

Iteration toward the maximum of the likelihood commences

smallest eigenvalue is negative, therefore the coefficients have moved into a non-positive definite region (even though the starting coefficients are in a positive definite region)

ITER 2, EIGVALS OF METRIC--MAX= 1,36+006, MIN= -2,35-003 MIN ABS= 2,35-003

OLD LIK= 6,177+010 MAX ABS RHS ELT= 2,631+002 MAX ABS COEF CHANGE IF STEP OF 1= 6,145+000

$\left[\dfrac{\text{new } f^*(a) - \text{old } f^*(a)}{\text{old } f^*(a)}\right]$ where "new" means trial value and "old" means end of last iteration

ITER	STATE	STEP	DET S (new det S)	DET GAMMA (new det $\hat{\Gamma}$)	NEW LIK (new $f^*(a) = \det^* \hat{\Gamma}/\det S$)	RATIO-1	MAX COEF CHANGE
2	46090	5,000-001	1,681-010	2,784-001	4,609+008	-9,925-001	3,072+000
2	46145	2,500-001	9,093-012	2,518-001	6,972+009	-8,871-001	1,536+000
2	46145	1,250-001	1,216-012	2,360-001	4,659+010	-2,457-001	7,681-001
2	46145	6,250-002	9,042-013	2,310-001	6,636+010	7,429-002	3,840-001
2	46200	4,303-002	7,935-013	2,268-001	6,597+010	6,807-002	2,644-001
2	46210	6,250-002	9,042-013	2,310-001	6,636+010	7,429-002	3,840-001

trial step sizes for iteration 2

step size of .06250 finally selected for iteration 2

det S and det $\hat{\Gamma}$ are computed from coefficients corresponding to normalized variables (presently normalized to length 1)

smallest eigenvalue is positive, therefore the coefficients have moved back into a positive definite region

ITER 3, EIGVALS OF METRIC--MAX= 1,37+006, MIN= 4,42-002, MIN ABS= 4,42-002

OLD LIK= 8,630+010 MAX ABS RHS ELT= 5,863+001 MAX ABS COEF CHANGE IF STEP OF 1= 7,724+001

ITER	STATE	STEP	DET S	DET GAMMA	NEW LIK	RATIO-1	MAX COEF CHANGE
3	46050	5,000+001	6,761+013	2,190+001	7,093+010	8,891+002	3,862+001
3	46063	1,000+000	8,176+013	2,073+001	5,256+010	+2,080+001	7,724+001
3	46208	8,496+001	6,969+013	2,226+001	7,108+010	7,118+002	2,701+001

step size of .8496 selected for iteration 3

ITER 4, EIGVALS OF METRIC--MAX= 1,48+006, MIN= 1,14+001, MIN ABS= 1,14+001

OLD LIK= 7,108+010 MAX ABS RHS ELT= 1,776+001 MAX ABS COEF CHANGE IF STEP OF 1= 1,653+000

ITER	STATE	STEP	DET S	DET GAMMA	NEW LIK	RATIO-1	MAX COEF CHANGE
4	46050	5,000+001	9,596+013	2,916+001	7,366+010	3,623+002	8,265+001
4	46063	1,000+000	1,856+012	2,818+001	7,521+010	5,819+002	1,653+000
4	46063	2,000+000	1,574+012	3,457+001	7,391+010	6,798+002	3,306+000
4	46063	4,000+000	3,461+012	4,870+001	6,852+010	+3,683+002	6,612+000
4	46208	6,791+001	1,449+012	3,319+001	7,682+010	8,958+002	2,960+000

step size of 1.791 selected for iteration 4

ITER 5, EIGVALS OF METRIC--MAX= 1,98+006, MIN= 5,39-002, MIN ABS= 5,39-002

OLD LIK= 7,682+010 MAX ABS RHS ELT= 2,847+001 MAX ABS COEF CHANGE IF STEP OF 1= 1,329+001

ITER	STATE	STEP	DET S	DET GAMMA	NEW LIK	RATIO-1	MAX COEF CHANGE
5	46050	5,000-001	1,656+012	3,570+001	7,698+010	1,256+002	6,647+000
5	46063	1,000+000	1,895+012	3,827+001	7,730+010	1,679+002	1,329+001
5	46063	2,000+000	2,526+012	4,362+001	7,531+010	+9,405+003	2,659+001
5	46208	8,996-001	1,839+012	3,770+001	7,729+010	1,667+002	1,184+001
5	46210	1,000+000	1,895+012	3,827+001	7,730+010	1,679+002	1,329+001

step size of 1 selected for iteration 5

OLD LIK= 7,730+010 MAX ABS RHS ELT= 2,390+001 MAX ABS COEF CHANGE IF STEP OF 1= 8,503+001

ITER	STATE	STEP	DET S	DET GAMMA	NEW LIK	RATIO-1	MAX COEF CHANGE
6	46050	5,000-001	2,126+012	4,062+001	7,759+010	3,767+003	4,251+001
6	46063	1,000+000	2,377+012	4,298+001	7,773+010	5,590+003	8,503+001
6	46063	2,000+000	2,938+012	4,777+001	7,766+010	4,710+003	1,701+000
6	46208	1,364+000	2,572+012	4,472+001	7,775+010	5,899+003	1,160+000

step size of 1.364 selected for iteration 6

OLD LIK= 7,775+010 MAX ABS RHS ELT= 9,545+000 MAX ABS COEF CHANGE IF STEP OF 1= 2,081+001

ITER	STATE	STEP	DET S	DET GAMMA	NEW LIK	RATIO-1	MAX COEF CHANGE
7	46050	5,000-001	2,702+012	4,584+001	7,778+010	4,018+004	1,041+001
7	46063	1,000+000	2,837+012	4,698+001	7,779+010	5,072+004	2,081+001
7	46063	2,000+000	3,124+012	4,928+001	7,773+010	+3,110+004	4,162+001
7	46208	9,027-001	2,805+012	4,671+001	7,779+010	5,108+004	1,837+001

step size of .9027 selected for iteration 7

OLD LIK= 7,779+010 MAX ABS RHS ELT= 2,167+001 MAX ABS COEF CHANGE IF STEP OF 1= 2,428+002

ITER	STATE	STEP	DET S	DET GAMMA	NEW LIK	RATIO-1	MAX COEF CHANGE
8	46050	5,000-001	2,822+012	4,686+001	7,779+010	5,512+006	1,214+002
8	46063	1,000+000	2,840+012	4,700+001	7,779+010	7,433+006	2,428+002
8	46063	2,000+000	2,875+012	4,729+001	7,779+010	7,921+007	4,857+002
8	46208	1,028+000	2,841+012	4,701+001	7,779+010	7,436+006	2,497+002

step size of 1.028 selected for iteration 8

OLD LIK= 7,779•010 MAX ABS RHS ELT= 6,440•003 MAX ABS COEF CHANGE IF STEP OF 1= 8,275•005

ITER	STATE	STEP	DET S	DET GAMMA	NEW LIK	RATIO-1	MAX COEF CHANGE
9	46050	5,000•001	2,841•012	4,701•001	7,779•010	1,170•009	4,137•005
9	46063	1,000•000	2,841•012	4,701•001	7,779•010	1,560•009	8,275•005
9	46063	2,000•000	2,841•012	4,702•001	7,779•010	•3,208•013	1,655•004
9	46200	9,999•001	2,841•012	4,701•001	7,779•010	1,560•009	8,274•005

← step size of .9999 selected for iteration 9

OLD LIK= 7,779•010 MAX ABS RHS ELT= 1,275•006 MAX ABS COEF CHANGE IF STEP OF 1= 3,979•008

ITER	STATE	STEP	DET S	DET GAMMA	NEW LIK	RATIO-1	MAX COEF CHANGE
10	46050	5,000•001	2,841•012	4,701•001	7,779•010	1,627•017	1,790•008
10	46063	1,000•000	2,841•012	4,701•001	7,779•010	2,169•017	3,979•008
10	46063	2,000•000	2,841•012	4,701•001	7,779•010	•6,853•022	7,158•008
10	46200	1,000•000	2,841•012	4,701•001	7,779•010	2,169•017	3,979•008

← step size of 1 selected for iteration 10

Due to C after NTH=10 code on FIML card

COEFFICIENTS (NTH ITER OF FULL INFO MAX LIK)

EQUATION 1

C	P	W	CONSTANT	P LAGGED
1	4	5	0	11
•1,0000000000	•0.2323887662	0,80184433v1	18,3432721831	0,3856730901

EQUATION 2

I	P	CONSTANT	P LAGGED	K LAGGED
2	4	0	11	12
•1,0000000000	•0.8010060259	27,2638657629	1,0518521413	•0,1480990630

EQUATION 3

d1	E	CONSTANT	E LAGGED	TIME
3	6	0	13	10
•1,0000000000	0.2341176398	5,7942875804	0,2846766802	0,2348346971

The NTH=10 code on the FIML card indicates that statistics are to be calculated and printed every 10 iterations.

max abs(γ_i) of .000000 000 009783 satisfies partial derivative convergence criterion of .000000001 set by EPS=1.0-10 code on FIML card.

routine checks that final coefficients are in a positive definite region, therefore has converged to a local maximum rather than a saddle point

ITER 11, EIGVALS OF METRIC=-MAX= 1,97•006, MIN= 4,75•002 MIN ABS= 4,75•002

OLD LIK= 7,779•010 MAX ABS RHS ELT= 9,783•012 MAX ABS COEF CHANGE IF STEP OF 1= 2,736•013

$\max_i \left| \frac{\gamma_i^{(i)} - \gamma_i^{(i-1)}}{\gamma_i^{(i-1)}} \right| = .000000 000 000 2734$

ITER	STATE	STEP	DET S	DET GAMMA	NEW LIK	RATIO-1	MAX COEF CHANGE
11	46050	5,000•001	2,841•012	4,701•001	7,779•010	•1,881•022	1,368•013
11	46145	2,500•001	2,841•012	4,701•001	7,779•010	•1,357•021	6,839•014
11	46145	1,250•001	2,841•012	4,701•001	7,779•010	•5,238•022	3,419•014
11	46145	6,250•002	2,841•012	4,701•001	7,779•010	3,187•021	1,710•014
11	46200	6,013•002	2,841•012	4,701•001	7,779•010	1,064•021	1,649•014
11	46218	6,250•002	2,841•012	4,701•001	7,779•010	3,187•021	1,710•014

COEFFICIENTS $\hat{\delta}_i$ (CONVERGED FULL INFO MAX LIK)

Due to 2 after FIML code

EQUATION 1

C	P	W	CONSTANT	P LAGGED
1	4	5	0	11
+1.0000000000	+0.2323887662	0.8018443391	18.3432721831	0.3854738981

EQUATION 2

I	P	CONSTANT	P LAGGED	K LAGGED
2	4	0	11	12
+1.0000000000	+0.8018860259	27.2638657629	1.8518521413	-0.1488990630

EQUATION 3

W1	E	CONSTANT	E LAGGED	TIME
3	6	0	13	10
+1.0000000000	0.2341176398	5.7942875864	0.2846766802	0.2348346971

Due to STE after FIML code

COEFFICIENT STANDARD ERRORS $\sqrt{\text{var}(\hat{\delta}_i)}$ (CONVERGED FULL INFO MAX LIK)

SQ RT OF T-N(I) USED AS DENOMINATOR

EQUATION 1

P	W	CONSTANT	P LAGGED
4	5	0	11
0.64532378	0.04945290	5.1411320	0.33537188

EQUATION 2

P	CONSTANT	P LAGGED	K LAGGED
4	0	11	12
0.93381694	10.59712128	0.47165111	0.05201870

EQUATION 3

E	CONSTANT	E LAGGED	TIME
6	0	13	10
0.10560123	3.60175026	0.06986828	0.06282732

Due to COSTE after FIML code

COEFFICIENTS/COEF STANDARD ERRS $\hat{\delta}_i / \sqrt{\text{var}(\hat{\delta}_i)}$ (CONVERGED FULL INFO MAX LIK)

SQ RT OF T-N(I) USED AS FACTOR

EQUATION 1

P	W	CONSTANT	P LAGGED
4	5	0	11
-0.36011189	16.21430507	3.56794361	1.14998636

EQUATION 2

P	CONSTANT	P LAGGED	K LAGGED
4	0	11	12
-0.85777629	2.57276151	2.23014877	-2.84747386

EQUATION 3

E	CONSTANT	E LAGGED	TIME
6	0	13	10
2.21699723	1.60874218	4.07447608	3.73777985

COEF VAR-COV MATRIX — (CONVERGED FULL INFO MAX LIK)

SQ RT OF(T-N(J))(T-N(I))USED AS THE DENOMINATOR

Due to CVC after FIML code

P	4	0.41644278					
W	5	-0.01531831	0.00244559				
CONSTANT	0	-3.87558757	0.07476785	20.48124725			
P LAGGED	11	-0.20291114	0.00503904	1.37634105	0.11247430		
P	4	0.56530786	-0.01700786	-4.30799014	-0.27691584	0.87201407	
CONSTANT	0	-1.72500778	-0.04977104	19.75394380	0.70281679	-4.88039103	112.29897935
P LAGGED	11	-0.23999873	0.00355545	1.59241729	0.14129000	-0.37426670	1.83175698
K LAGGED	12	-0.01941715	0.00137069	0.13546910	0.00828265	-0.01865021	-0.29651874
E	6	0.04545362	-0.00286894	-0.34397436	-0.01860950	0.00106098	0.07711893
CONSTANT	0	-1.19739756	0.08907527	9.56835400	0.43695101	-1.85192278	10.41171826
E LAGGED	13	-0.02642749	0.00150425	0.19173858	0.01173880	-0.03130470	-0.25551725
TIME	10	-0.01977396	0.00042420	0.17737251	0.00847713	-0.02775034	-0.01473327
		4	5	0	11	4	0
		P	W	CONSTANT	P LAGGED	P	CONSTANT

P LAGGED	11	0.22245477					
K LAGGED	12	0.00422366	0.00270511				
E	6	-0.01479435	-0.00432027	0.01115162			
CONSTANT	0	0.19916710	0.08894701	-0.30543802	12.97260493		
E LAGGED	13	0.01188809	0.00294065	-0.00628424	0.09435207	0.00488158	
TIME	10	0.00383274	0.00209822	-0.00546591	0.18315204	0.00250268	0.00394727
		11	12	6	0	13	10
		P LAGGED	K LAGGED	E	CONSTANT	E LAGGED	TIME

COEF VAR-COV MATRIX — (CONVERGED FULL INFO MAX LIK)

T USED AS THE DENOMINATOR

Due to CVCT after FIML code

P	4	0.33712835					
W	5	-0.01240853	0.00197976				
CONSTANT	0	-2.48976137	0.06052636	21.39672397			
P LAGGED	11	-0.16426140	0.00407922	1.11418085	0.09105062		
P	4	0.45763817	-0.01376827	-3.48742059	-0.22416996	0.70591616	
CONSTANT	0	-1.39643481	-0.03705275	19.99096403	0.56894692	-3.93460226	90.90869757
P LAGGED	11	-0.19428469	0.00287822	1.28909971	0.11437762	-0.30297781	1.48285089
K LAGGED	12	-0.01571365	0.00110961	0.10966546	0.00670500	-0.01509779	-0.24003898
E	6	0.03679579	-0.00232247	-0.27845544	-0.01506484	0.04943032	0.06242961
CONSTANT	0	-0.96932183	0.06807045	7.74581038	0.35372225	-1.49917598	8.42853383
E LAGGED	13	-0.02139369	0.00121772	0.15521695	0.00950284	-0.02534190	-0.20684730
TIME	10	-0.01600749	0.00034340	0.14374917	0.00686244	-0.02246456	-0.01192693
		4	5	0	11	4	0
		P	W	CONSTANT	P LAGGED	P	CONSTANT

P LAGGED	11	0.18088243					
K LAGGED	12	0.00341916	0.00218985				
E	6	-0.01197638	-0.00349736	0.00902750			
CONSTANT	0	0.16123051	0.07200472	-0.24717840	10.50163256		
E LAGGED	13	0.00962369	0.00238053	-0.00508725	0.07638025	0.00395175	
TIME	10	0.00310269	0.00169856	-0.00442479	0.14826594	0.00202591	0.00319541
		11	12	6	0	13	10
		P LAGGED	K LAGGED	E	CONSTANT	E LAGGED	TIME

COEF VAR-COV MATRIX NORMALIZED (CONVERGED FULL INFO MAX LIK)

Due to CVCN after FIML code

P	4	1.00000											
W	5	-0.48000	1.00000										
CONSTANT	8	-0.92703	0.29488	1.00000									
P LAGGED	11	-0.93757	0.30383	0.79825	1.00000								
P	4	0.93889	-0.36838	-0.89733	-0.88422	1.00000							
CONSTANT	8	-0.25225	-0.88734	0.36258	0.19776	-0.49116	1.00000						
P LAGGED	11	-0.78852	0.15243	0.65672	0.89323	-0.84976	0.36649	1.00000					
K LAGGED	12	-0.57852	0.53291	0.50663	0.47484	-0.38400	-0.93799	0.17218	1.00000				
E	6	0.66699	-0.94936	-0.63358	-0.52546	0.61920	0.08891	-0.29703	-0.78659	1.00000			
CONSTANT	8	-0.51517	0.47764	0.51673	0.36174	-0.35861	0.27279	0.11724	0.47482	-0.80278	1.00000		
E LAGGED	13	-0.58614	0.43536	0.53379	0.50098	-0.47981	-0.34911	0.36075	0.80923	-0.85173	0.37494	1.00000	
TIME	18	-0.48772	0.13653	0.54975	0.48232	-0.47388	-0.02213	0.12934	0.64211	-0.82384	0.88937	0.97012	
		4	5	8	11	4	8	11	12	6	8	13	
		P	W	CONSTANT	P LAGGED	P	CONSTANT	P LAGGED	K LAGGED	E	CONSTANT	E LAGGED	
TIME	18	1.00000											
		18											
		TIME											

DISTURBANCE VAR-COV MATRIX (CONVERGED FULL INFO MAX LIK)

Due to DVC after FIML code

SQ RT OF(T-N(J))(T-N(I))USED AS THE DENOMINATOR

C	1	2.59924719		
I	2	4.79172989	15.77060874	
W1	3	0.59503625	4.76511926	2.22498894
		1	2	3
		C	I	W1

DISTURBANCE VAR-COV MATRIX NORMALIZED (CONVERGED FULL INFO MAX LIK)

Due to DVCN after FIML code

C	1	1.00000		
I	2	0.74828	1.00000	
W1	3	0.24744	0.90429	1.00000
		1	2	3
		C	I	W1

REDUCED FORM COEFS (CONVERGED FULL INFO MAX LIK)

Due to RF after FIML code

1, DEP. VAR,= 1 C

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	8
24,67760754	0,39440675	-0,00089977	0,29759618	0,24549217	0,23866714
8		W2			
9		7			
0,08667548		0,80064845			

2, DEP. VAR,= 4 P

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	8
21,86431237	0,68650783	-0,07072693	-0,03517437	-0,02901594	-0,50648963
8		W2			
9		7			
0,47756228		-0,09463167			

3, DEP. VAR,= 5 W

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	8
14,00451980	0,20985415	-0,02161994	0,35094543	0,29775005	0,15085800
8		W2			
9		7			
0,14998293		0,97107266			

4, DEP. VAR,= 2 I

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	8
18,39122462	0,50195523	-0,09144469	0,02817488	0,02324194	0,40570124
8		W2			
9		7			
-0,38253826		0,07580054			

Due to RF after FIML code

5. DEP. VAR.= 3 W1

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	0
14.88491988	0.28985415	-0.82161994	0.36894543	0.29775085	0.15889888

G		W2	
9		7	
0.14998293		-0.02892734	

6. DEP. VAR.= 6 E

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	0
35.86883217	0.89636198	-0.89234646	0.32577186	0.20873412	0.64436838

G		W2	
9		7	
0.02354522		-0.12355901	

7. DEP. VAR.= 14 Y

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	0
35.86883217	0.89636198	-0.89234646	0.32577186	0.20873412	-0.35563162

G		W2	
9		7	
0.02354522		0.87644099	

8. DEP. VAR.= 15 K

CONSTANT	P LAGGED	K LAGGED	E LAGGED	TIME	R
0	11	12	13	10	0
18.39122462	0.58195523	0.90855331	0.02817488	0.02324194	0.40570124

G		W2	
9		7	
-0.38253026		0.07588854	

Due to RFRES after FINL code

REDUCED FORM RESIDUALS

y for reduced form eqn. for variable C.

OBS		Y = X(1)
1921	1	41,90000000
1922	2	45,00000000
1923	3	49,20000000
1924	4	50,60000000
1925	5	52,60000000
1926	6	55,10000000
1927	7	56,20000000
1928	8	57,30000000
1929	9	57,80000000
1930	10	55,80000000
1931	11	50,90000000
1932	12	45,60000000
1933	13	46,90000000
1934	14	48,70000000
1935	15	51,30000000
1936	16	57,70000000
1937	17	58,70000000
1938	18	57,50000000
1939	19	61,60000000
1940	20	65,00000000
1941	21	69,70000000

$\hat{y}$ for reduced form equation for variable C

ESTIMATED Y
44,46880581
44,05462848
47,60092776
50,49707972
51,58971033
53,70599578
54,81670187
54,74155367
55,72090871
57,99314116
54,54634353
51,40539258
46,75549794
49,55880723
51,78835721
55,40636394
58,60181628
60,39151650
59,07633083
63,67255598
67,90800469

$\hat{\theta}$ for reduced form eqn. for variable C

Y - ESTIMATED Y
-2,46880581
0,94537152
1,59907223
0,10292028
1,01028967
1,39400422
1,38329813
2,25844633
2,07949129
-2,99314116
-3,64634353
-5,80539258
-0,25549794
-0,85880723
-0,48835721
2,29363606
0,09818372
-2,89151650
2,52366917
1,32744402
2,19199531

SUM Y $\sum_{t=1}^{T} y_t$ = 1133,89999998

SUM RES $\sum_{t=1}^{T} \hat{\theta}_t$ = 0,00000000

SSY $\sum_{t=1}^{T} y_t^2$ = 62166,62999916

SS(Y - MEAN Y) $\sum_{t=1}^{T} (y_t - \bar{y})^2$ = 941,42952380

SS RES $\sum_{t=1}^{T} \hat{\theta}_t^2$ = 109,37278036

R2 R^2 = 0,88382266

D.W. STAT = 1,21345017 (Durbin-Watson test statistic)

SS(R - PREV R) $\sum_{t=2}^{T} (\hat{\theta}_t - \hat{\theta}_{t-1})^2$ = 132,71841912

variable P:

(CONVERGED FULL INFO MAX LIK)

p

$\hat{p}$

Y = X(4)	ESTIMATED Y	Y - ESTIMATED Y
12,40000000	14,56141838	-2,16141838
16,90000000	15,98292644	0,91707356
18,40000000	18,15431405	0,24568595
19,40000000	19,40426427	-0,00426427
20,10000000	18,93484226	1,16515774
19,60000000	18,16705210	1,43294490
19,80000000	17,89431326	1,90568674
21,10000000	19,09150731	2,00849269
21,70000000	19,90767541	1,79232459
15,60000000	18,56981597	-2,96981597
11,40000000	15,15173699	-3,75173699
7,00000000	12,06292977	-5,06292977
11,20000000	10,78249498	0,41750502
12,30000000	13,55673316	-1,25673316
14,00000000	14,35997543	-0,35997543
17,60000000	14,64897553	2,95102447
17,30000000	17,86183397	-0,56183397
15,30000000	17,91046173	-2,61046173
19,00000000	16,68641504	2,31358496
21,10000000	18,90712721	2,19287279
23,50000000	22,10318373	1,39681627

SUM Y = 354,70000000

SUM RES = 0,00000000

SSY = 6347,24999988

SS(Y - MEAN Y) = 356,19809523

SS RES = 98,22921062

R2 = 0,72423993

D.W. STAT = 1,24973717

SS(R - PREV R) = 122,75569626

REDUCED FORM RESIDUALS

W

OBS		Y = X(3)
1921	1	28,20000000
1922	2	32,20000000
1923	3	37,00000000
1924	4	37,00000000
1925	5	38,60000000
1926	6	40,70000000
1927	7	41,90000000
1928	8	42,90000000

ESTIMATED Y
30,69348280
30,73322505
33,62078858
36,87343574
37,61905317
39,69911921
41,24579417
41,40260664

Y - ESTIMATED Y
-2,49348280
1,46677495
3,37921142
0,12656426
0,98094683
1,00088079
0,65420583
1,49739336

(CONVERGED FULL INFO MAX LIK)

I

Y = X(2)	ESTIMATED Y	Y - ESTIMATED Y
-0,20000000	1,88609537	-2,08609537
1,90000000	0,46152301	1,43847699
5,20000000	3,17417487	2,02582513
3,00000000	2,98062029	0,01937971
5,10000000	3,96418510	1,13581490
5,60000000	4,56017853	1,03982147
4,20000000	3,42336556	0,77663444
3,00000000	2,05276028	0,94723972

1929	9	45,30000000	42,23494861	3,06505139	5,10000000	2,32211938	2,77788470
1930	10	42,10000000	44,39288002	-2,29288002	1,80000000	3,26955484	-2,26955484
1931	11	39,30000000	42,83766611	-2,73766611	-3,40000000	-0,55694843	-2,84305957
1932	12	34,30000000	39,24539340	-4,94539340	-6,20000000	-1,99702941	-4,20297059
1933	13	34,10000000	35,19167430	-1,09167430	-5,10000000	-4,68132865	-0,41867135
1934	14	36,60000000	37,47164813	-0,87164813	-3,80000000	-1,73042594	-1,26957406
1935	15	39,30000000	39,95588848	-0,65588848	-1,30000000	-0,77249338	-0,52750662
1936	16	44,20000000	43,73408171	0,46591829	2,10000000	0,97669338	1,12330662
1937	17	47,70000000	46,91881632	0,78118368	2,80000000	1,67883401	0,12116599
1938	18	45,90000000	49,30918388	-3,40918388	-1,90000000	1,22812911	-3,12812911
1939	19	49,40000000	48,27619762	1,12380238	1,30000000	0,38628182	0,91371818
1940	20	53,80000000	52,87223090	0,12776910	3,30000000	2,30680214	0,99319786
1941	21	61,80000000	57,57172514	4,22827485	4,90000000	1,46690419	3,43309581
		SUM Y		SUM RES	SUM Y		SUM RES
		871,09999999		0,00000000	26,60000000		0,00000000
		SSY	SS(Y - MEAN Y)	SS RES	SSY	SS(Y - MEAN Y)	SS RES
		37275,86999999	1141,81238894	104,66693247	286,02800000	252,32666060	79,74708128
		R2	D.W. STAT	SS(R - PREV R)	R2	D.W. STAT	SS(R - PREV R)
		0,90833263	1,34912095	141,28839162	0,68399382	1,27867419	181,33255792

Due to AR(2) after FINAL code

REDUCED FORM RESIDUALS (CONVERGED FULL INFO MAX LIK)

	DW	W1 Y = X(3)	ESTIMATED Y	Y - ESTIMATED Y	E Y = X(6)	ESTIMATED Y	Y - ESTIMATED Y
1921	1	25,50000000	27,99348288	-2,49348288	45,60000000	50,25490118	-4,65490118
1922	2	29,30000000	27,83322505	1,46677495	50,10000000	47,71619149	2,38384851
1923	3	34,10000000	30,72078858	3,37921142	57,20000000	53,57510263	3,62489737
1924	4	33,90000000	33,77343574	0,12656426	57,10000000	56,97770001	0,12229999
1925	5	35,40000000	34,41905317	0,98094683	61,00000000	58,85389543	2,14610457
1926	6	37,40000000	36,39911921	1,00088079	64,00000000	61,56617431	2,43382569
1927	7	37,90000000	37,64575417	0,25424583	64,40000000	62,24006743	2,15993257
1928	8	39,20000000	37,70280664	1,49719336	64,50000000	60,99431394	3,50568606
1929	9	41,30000000	38,23494861	3,06505139	67,00000000	62,14262401	4,85737599
1930	10	37,90000000	40,19288002	-2,29288002	61,20000000	66,46269599	-5,26269599
1931	11	34,50000000	37,23766611	-2,73766611	53,40000000	59,88940310	-6,48940310
1932	12	29,80000000	33,94539340	-4,94539340	44,30000000	54,30832317	-10,00832317
1933	13	28,50000000	29,59167430	-1,09167430	43,10000000	45,77416928	-0,67416928
1934	14	30,60000000	31,47164813	-0,87164813	49,70000000	51,82838129	-2,12838129
1935	15	33,20000000	33,85588848	-0,65588848	54,40000000	55,41586392	-1,01586392
1936	16	36,80000000	36,33408171	0,46591829	62,70000000	59,28305725	3,41694275
1937	17	41,00000000	40,21881632	0,78118368	65,00000000	64,78065029	0,21934971
1938	18	38,20000000	41,60918388	-3,40918388	60,90000000	66,91964561	-6,01964561
1939	19	41,60000000	40,47619762	1,12380238	69,50000000	66,06261266	3,43738734
1940	20	45,00000000	44,87223090	0,12776910	75,70000000	73,37935811	2,32064189
1941	21	53,30000000	49,87172514	4,22827485	88,40000000	82,77490888	5,62509112
		SUM Y		SUM RES	SUM Y		SUM RES
		763,59999999		0,00000000	1261,19999999		0,00000000
		SSY	SS(Y - MEAN Y)	SS RES	SSY	SS(Y - MEAN Y)	SS RES
		28568,85999966	794,98952380	104,66693247	77998,57999992	2254,51142853	364,37794529
		R2	D.W. STAT	SS(R - PREV R)	R2	D.W. STAT	SS(R - PREV R)
		0,86832858	1,34912095	141,28839162	0,83837831	1,23066077	448,42783817

REDUCED FORM RESIDUALS

OB		Y Y • X(14)	ESTIMATED Y	Y • ESTIMATED Y
1921	1	40,60000000	45,25490118	-4,62490118
1922	2	49,10000000	46,71615149	2,38384851
1923	3	55,40000000	51,77510263	3,62489737
1924	4	56,40000000	56,27770001	0,12229999
1925	5	58,70000000	56,55309543	2,14690457
1926	6	60,30000000	57,86617431	2,43382569
1927	7	61,30000000	59,14006743	2,15993257
1928	8	64,80000000	60,49431394	3,30568606
1929	9	67,00000000	62,14252401	4,85737599
1930	10	57,70000000	62,96269599	-5,26269599
1931	11	50,70000000	57,18940310	-6,48940310
1932	12	41,30000000	51,30832317	-10,00832317
1933	13	45,30000000	45,37416928	-0,07416928
1934	14	48,90000000	51,02838129	-2,12838129
1935	15	53,30000000	54,31586392	-1,01586392
1936	16	61,80000000	58,38305725	3,41694275
1937	17	65,00000000	64,78065029	0,21934971
1938	18	61,20000000	67,21964561	-6,01964561
1939	19	68,40000000	64,96261266	3,43738734
1940	20	74,10000000	71,77935811	2,32064189
1941	21	85,30000000	79,67490888	5,62509112

SUM Y		SUM RES
1225,79999998		0,00000000
SSY	SS(Y • MEAN Y)	SS RES
73858,55999947	2386,86285710	364,37794529
R2	D,W, STAT	SS(R • PREV R)
0,84284612	1,23006677	448,42783017

Due to RFREQ after FIML code

END OF RUN

(CONVERGED FULL INFO MAX LIK)

K Y • X(15)	ESTIMATED Y	Y • ESTIMATED Y
182,00000000	184,08609537	-2,08609537
184,50000000	183,06152501	1,43847699
189,70000000	187,67417487	2,02582513
192,70000000	192,68062029	0,01937971
197,80000000	196,66418510	1,13581490
203,40000000	202,36017852	1,03982147
207,60000000	206,82336556	0,77663444
210,60000000	209,65276027	0,94723972
215,70000000	212,92211530	2,77788470
216,70000000	218,96955483	-2,26955484
213,30000000	216,14305957	-2,84305957
207,10000000	211,30297059	-4,20297059
202,00000000	202,41867134	-0,41867135
199,00000000	200,26957406	-1,26957406
197,70000000	198,22750670	-0,52750670
199,80000000	198,67669330	1,12330670
201,80000000	201,67883481	0,12116999
199,90000000	203,02812911	-3,12812911
201,20000000	200,28628182	0,91371818
204,50000000	203,50680213	0,99319786
209,40000000	205,96698419	3,43309581

SUM Y		SUM RES
4237,00000000		0,00000000
SSY	SS(Y • MEAN Y)	SS RES
856565,21998596	1780,02952479	79,74708128
R2	D,W, STAT	SS(R • PREV R)
0,95389877	1,27067419	101,33255792

(STAT CONTROL CARD)

END OF RUN CARD TERM,

EXECUTE	0	0 HOURS	0 MINUTES	40,434 SECONDS
SETUP	0	0 HOURS	0 MINUTES	35,763 SECONDS
ROUTINE	0	0 HOURS	1 MINUTES	37,360 SECONDS
FTN	0	0 HOURS	0 MINUTES	02,266 SECONDS
FILE	0	0 HOURS	0 MINUTES	01,097 SECONDS
LOADER	0	0 HOURS	0 MINUTES	29,589 SECONDS
*OTHER	0	0 HOURS	0 MINUTES	00,000 SECONDS

← actual execution time

"Setup" time. If routines were in the regular library file, most of this time would be saved.

ELAPSED TIME: 0 HOURS 3 MINUTES 26,509 SECONDS
2491 WRITE OPERATIONS ON LU 61 WERE DONE DURING EXECUTION.

APPENDICES

APPENDIX A

COMPUTATION OF $[Z_1'Z_2]_{\perp Z_3}$ AND THE RANK OF Z_3 AS AN INTERMEDIATE STEP IN THE COMPUTATION OF $[Z_1'Z_2]_{\perp}[Z_3 : Z_4]$

The method given in this appendix is very general in that it may be used to calculate a matrix of the form $[Z_1'Z_2]_{\perp Z_3}$ and the rank of Z_3 as an intermediate step in the computation of a matrix of the form $[Z_1'Z_2]_{\perp}[Z_3 : Z_4]$ in which:

(1) Z_1, Z_2, Z_3, and Z_4 contain jointly dependent or predetermined variables or both.

(2) Variables in any of the matrices may occur in any or all of the other three matrices as well.

(3) Z_1, Z_2, Z_3, and Z_4 may have less than full column rank.

In this paper, Z_1 and Z_2 will most commonly be $_+Y_\mu$ (the matrix of jointly dependent variables in the μ^{th} equation), Z_3 will most commonly be X_μ (the matrix of predetermined variables in the μ^{th} equation), and $[Z_3 : Z_4]$ will most commonly be X_I (a matrix of instruments containing X_μ) or X (the matrix of predetermined variables in the entire system).

Matrices of the form $[Z_1'Z_2]_{\parallel Z_3}$ and $[Z_1'Z_2]_{\parallel [Z_3 : Z_4]}$ are calculated as $[Z_1'Z_2] - [Z_1'Z_2]_{\perp Z_3}$ and $[Z_1'Z_2] - [Z_1'Z_2]_{\perp [Z_3 : Z_4]}$, respectively.

A computational procedure for calculating $[Z_1'Z_2]_{\perp Z_3}$ and rk Z_3 as an intermediate step in the calculation of $[Z_1'Z_2]_{\perp}[Z_3 : Z_4]$ follows:

(1) Let Z be a $T \times N$ matrix containing all of the variables which occur in Z_1 or Z_2. If desired, Z could be defined as $Z = [Z_1 : Z_2]$; however, there is no need to repeat variables

common to both Z_1 and Z_2.[1] If $Z_1 = Z_2$, then $Z = Z_1 = Z_2$. Z may contain variables in addition to those in Z_1 and Z_2 if desired.

Calculate the moment matrix (sums of squares and cross-products matrix) of $[Z_3 \vdots Z_4 \vdots Z]$, i.e., calculate:

$$(A.1)\quad [Z_3 \vdots Z_4 \vdots Z]'[Z_3 \vdots Z_4 \vdots Z] = \begin{bmatrix} Z_3'Z_3 & Z_3'Z_4 & Z_3'Z \\ N_3{}^{X}N_3 & N_3{}^{X}N_4 & N_3{}^{X}N \\ Z_4'Z_3 & Z_4'Z_4 & Z_4'Z \\ N_4{}^{X}N_3 & N_4{}^{X}N_4 & N_4{}^{X}N \\ Z'Z_3 & Z'Z_4 & Z'Z \\ N{}^{X}N_3 & N{}^{X}N_4 & N{}^{X}N \end{bmatrix}.$$

(Required computer capacity may be reduced by forming only the upper or lower triangular part of the above matrix, since all operations which follow may be performed on only a triangular part.)

(2) Calculate

$$(A.2)\quad \{[Z_4 \vdots Z]'[Z_4 \vdots Z]\}_{\perp Z_3} = \begin{bmatrix} Z_4'Z_4 & Z_4'Z \\ Z'Z_4 & Z'Z \end{bmatrix}_{\perp Z_3}$$

$$= \begin{bmatrix} [Z_4'Z_4]_{\perp Z_3} & [Z_4'Z]_{\perp Z_3} \\ [Z'Z_4]_{\perp Z_3} & [Z'Z]_{\perp Z_3} \end{bmatrix} = \begin{bmatrix} [Z_4]'_{\perp Z_3}[Z_4]_{\perp Z_3} & [Z_4]'_{\perp Z_3}Z_{\perp Z_3} \\ Z'_{\perp Z_3}[Z_4]_{\perp Z_3} & Z'_{\perp Z_3}Z_{\perp Z_3} \end{bmatrix}$$

$$= \{[Z_4]_{\perp Z_3} \vdots Z_{\perp Z_3}\}'\{[Z_4]_{\perp Z_3} \vdots Z_{\perp Z_3}\}.$$

---

[1] Repeating variables in the Z matrix causes no computational difficulty.

in the manner given by section I.D.2.[1] The matrix given by (A.1) will have been transformed to:

$$\text{(A.3)} \qquad \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ N_3 \times N_3 & N_3 \times N_4 & N_3 \times N \\ 0 & [Z_4'Z_4]_{\perp Z_3} & [Z_4'Z]_{\perp Z_3} \\ N_4 \times N_3 & N_4 \times N_4 & N_4 \times N \\ 0 & [Z'Z_4]_{\perp Z_3} & [Z'Z]_{\perp Z_3} \\ N \times N_3 & N \times N_4 & N \times N \end{bmatrix} .$$

The rank of Z_3 is the number of diagonal elements used as pivots (number of columns treated) before the maximum diagonal element becomes less than ϵ (see section I.D.2).

If $Z = {}_{+}Y_{\mu}$, $Z_3 = X_{\mu}$, and LIML coefficients are to be calculated, then the $[Z'Z]_{\perp Z_3} = [{}_{+}Y_{\mu}' \; {}_{+}Y_{\mu}]_{\perp X_{\mu}}$ matrix should be saved aside at this point, since it is a basic matrix used in LIML calculations.

(3) The computation of $[Z'Z]_{\perp}[Z_3 : Z_4]$ is completed by performing elementary row operations on the matrix given by (A.2) which is a submatrix of (A.3) in the manner given by section I.D.2;[2] that is, do elementary row operations on the matrix given by (A.2) until the first N_4 columns are reduced to zeros below the diagonal. (It is advisable to select the pivot as the largest diagonal

---

[1] Do elementary row operations on the matrix given by (A.1) until the first N_3 columns are reduced to zeros below the diagonal. Thus, the pivot element for each step is selected from among the first N_3 columns, only. (It is advisable to select the pivot for each step as the largest diagonal element to reduce rounding error.)

[2] I.e., calculate $[Z_{\perp Z_3}' Z_{\perp Z_3}]_{\perp}([Z_4]_{\perp Z_3}) = [Z'Z]_{\perp}[Z_3 : Z_4]$ in the manner given by section I.D.2.

element at each step to reduce rounding error.) The matrix given by (A.1) will have become:

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ N_3 \times N_3 & N_3 \times N_4 & N_3 \times N \\ 0 & A_{22} & A_{23} \\ N_4 \times N_3 & N_4 \times N_4 & N_4 \times N \\ 0 & 0 & [Z'Z]_{\perp[Z_3 : Z_4]} \\ N \times N_3 & N \times N_4 & N \times N \end{bmatrix} \tag{A.4}$$

The rank of $[Z_4]_{\perp Z_3}$ is the number of diagonal elements used as pivots (number of columns treated) before the maximum diagonal element of A_{22} becomes less than ϵ (see section I.D.2). Also

$$\text{rk}[Z_3 : Z_4] = \text{rk } Z_3 + \text{rk } [Z_4]_{\perp Z_3} . \tag{A.5}$$

APPENDIX B

COMPUTATION BY DIRECT ORTHOGONALIZATION OF A MOMENT MATRIX OF VARIABLES EACH OF WHICH IS ORTHOGONAL TO A DIFFERENT SUBSET OF VARIABLES

The computation by direct orthogonalization of a moment matrix of the form:[1]

$$\text{(B.1)} \quad \{[y_1]_{\perp Z_1} \cdots [y_m]_{\perp Z_m}\}'\{[y_1]_{\perp Z_1} \cdots [y_m]_{\perp Z_m}\} = \begin{bmatrix} [y_1]'_{\perp Z_1}[y_1]_{\perp Z_1} & \cdots & [y_1]'_{\perp Z_1}[y_m]_{\perp Z_m} \\ \vdots & & \vdots \\ [y_m]'_{\perp Z_m}[y_1]_{\perp Z_1} & \cdots & [y_m]'_{\perp Z_m}[y_m]_{\perp Z_m} \end{bmatrix}$$

by direct orthogonalization will be illustrated by showing how to

---

[1] A moment matrix of the form $\{[y_1]_{\shortparallel Z_1} \cdots [y_m]_{\shortparallel Z_m}\}'\{[y_1]_{\shortparallel Z_1} \cdots [y_m]_{\shortparallel Z_m}\}$ may be calculated as $[y_1 \cdots y_m]'[y_1 \cdots y_m] - \{[y_1]_{\perp Z_1} \cdots [y_m]_{\perp Z_m}\}'\{[y_1]_{\perp Z_1} \cdots [y_m]_{\perp Z_m}\}$.

compute a moment matrix of this form for $m = 3$.[1,2]

(1) Let y_1, y_2, and y_3 be $T \times 1$ vectors; Z_1 be a $T \times N_1$ matrix of variables, Z_2 be a $T \times N_2$ matrix of variables, and Z_3 be a $T \times N_3$ matrix of variables.[3]

Calculate the moment matrix (sums of squares and cross-products matrix) of $[Z_1, Z_2, Z_3, y_1, y_2, y_3]$, i.e., calculate:

---

[1]The matrix noted in (B.1) could be calculated by first calculating the m $T \times 1$ vectors $[y_1]_{\perp Z_1}, \cdots, [y_m]_{\perp Z_m}$ in the manner indicated in footnotes 1 and 2 of page 43 (i.e., as m sets of residuals from m least squares calculations) and then forming the matrix of sums of squares and cross-products of these calculated vectors (residuals). A more accurate method of computation (which also requires less computer time) is to (1) calculate the m sets of least squares coefficients which give the m $[y_i]_{\perp Z_i}$ vectors in the manner noted in footnote 1 of page 59; i.e., calculate $a_i^* = [Z_i^{*\prime} Z_i^*]^{-1} Z_i^{*\prime} y_i$ for i=1,...,m (where Z_i^* is a matrix of variables from Z_i with rk Z_i^* = rk Z_i--see p. 47), (2) form a moment matrix, Z'Z, using all of the y_i and all of the variables occuring in any Z_i^* (Z may be expanded to include all of the variables in Z_i, if desired), (3) form a_i from a_i^* by rearranging the coefficients of a_i^* to the same order as Z'Z, inserting a "-1" into the a_i vector in the position corresponding to y_i, and inserting 0's into the positions corresponding to all of the remaining variables of Z'Z. The ij^{th} element of the desired moment matrix is then calculated as $[y_i]'_{\perp Z_i} [y_j]_{\perp Z_j} = a_i' Z' Z a_j$.

The method given in this appendix is more accurate than either of the above methods and requires slightly less computer time than the second method (which requires considerably less computer time than the first method).

[2]A verification that the computational procedure produces the correct matrix is given at the end of this appendix.

[3]In this appendix, y_1, y_2, and y_3 denote any variables (not necessarily jointly dependent variables) and Z_1, Z_2, and Z_3 denote matrices of variables--not the explanatory variables of equations 1, 2, and 3. Z_1, Z_2, and Z_3 may contain variables in common or the variables in any two of the matrices may be linearly independent. The matrices Z_1, Z_2, and Z_3 need not have full column rank.

(B.2) $[Z_1, Z_2, Z_3, y_1, y_2, y_3]'[Z_1, Z_2, Z_3, y_1, y_2, y_3] =$

$$\left[\begin{array}{c|ccc|cc|c}
\begin{matrix} Z_1'Z_1 \\ N_1 \times N_1 \end{matrix} & \multicolumn{3}{c|}{\begin{matrix} Z_1'Z_2 & Z_1'Z_3 \\ N_1 \times N_2 & N_1 \times N_3 \end{matrix}} & \begin{matrix} Z_1'y_1 \\ N_1 \times 1 \end{matrix} & \begin{matrix} Z_1'y_2 \\ N_1 \times 1 \end{matrix} & \begin{matrix} Z_1'y_3 \\ N_1 \times 1 \end{matrix} \\
\hline
\begin{matrix} Z_2'Z_1 \\ N_2 \times N_1 \end{matrix} & \multicolumn{3}{c|}{\begin{matrix} Z_2'Z_2 & Z_2'Z_3 \\ N_2 \times N_2 & N_2 \times N_3 \end{matrix}} & \begin{matrix} Z_2'y_1 \\ N_2 \times 1 \end{matrix} & \begin{matrix} Z_2'y_2 \\ N_2 \times 1 \end{matrix} & \begin{matrix} Z_2'y_3 \\ N_2 \times 1 \end{matrix} \\
\begin{matrix} Z_3'Z_1 \\ N_3 \times N_1 \end{matrix} & \multicolumn{3}{c|}{\begin{matrix} Z_3'Z_2 & Z_3'Z_3 \\ N_3 \times N_2 & N_3 \times N_3 \end{matrix}} & \begin{matrix} Z_3'y_1 \\ N_3 \times 1 \end{matrix} & \begin{matrix} Z_3'y_2 \\ N_3 \times 1 \end{matrix} & \begin{matrix} Z_3'y_3 \\ N_3 \times 1 \end{matrix} \\
\begin{matrix} y_1'Z_1 \\ 1 \times N_1 \end{matrix} & \multicolumn{3}{c|}{\begin{matrix} y_1'Z_2 & y_1'Z_3 \\ 1 \times N_2 & 1 \times N_3 \end{matrix}} & \begin{matrix} y_1'y_1 \\ 1 \times 1 \end{matrix} & \begin{matrix} y_1'y_2 \\ 1 \times 1 \end{matrix} & \begin{matrix} y_1'y_3 \\ 1 \times 1 \end{matrix} \\
\begin{matrix} y_2'Z_1 \\ 1 \times N_1 \end{matrix} & \multicolumn{3}{c|}{\begin{matrix} y_2'Z_2 & y_2'Z_3 \\ 1 \times N_2 & 1 \times N_3 \end{matrix}} & \begin{matrix} y_2'y_1 \\ 1 \times 1 \end{matrix} & \begin{matrix} y_2'y_2 \\ 1 \times 1 \end{matrix} & \begin{matrix} y_2'y_3 \\ 1 \times 1 \end{matrix} \\
\begin{matrix} y_3'Z_1 \\ 1 \times N_1 \end{matrix} & \multicolumn{3}{c|}{\begin{matrix} y_3'Z_2 & y_3'Z_3 \\ 1 \times N_2 & 1 \times N_3 \end{matrix}} & \begin{matrix} y_3'y_1 \\ 1 \times 1 \end{matrix} & \begin{matrix} y_3'y_2 \\ 1 \times 1 \end{matrix} & \begin{matrix} y_3'y_3 \\ 1 \times 1 \end{matrix}
\end{array}\right]$$

(Required computer capacity may be reduced by forming only the upper or lower triangular part of the above matrix, since all operations which follow may be performed on only a triangular part.)

(2) Let us designate the part of (B.2) below and to the right of the dashed line as (B.3). (B.3) is saved aside at this point for use in later calculations.

(3) Calculate $[(Z_2, Z_3, y_1, y_2, y_3)'(Z_2, Z_3, y_1, y_2, y_3)]_{\downarrow Z_1}$ in the manner given in section I.D.2.[1] The matrix given by (B.2) will

---

[1] Do elementary row operations on the matrix given by (B.2) until the first N_1 columns are reduced to zeros below the diagonal. Thus, the pivot element for each step is selected from among the first N_1 columns, only. (It is advisable to select the pivot for each step as the largest diagonal element to reduce rounding error.)

have been transformed to:

$$
(B.4)\qquad
\begin{bmatrix}
A_{11} & A_{12} & A_{13} & A_{1,y_1} & A_{1,y_2} & A_{1,y_3} \\
0 & \multicolumn{5}{l}{\multirow{5}{*}{$\begin{bmatrix} z_2'z_2 & z_2'z_3 & z_2'y_1 & z_2'y_2 & z_2'y_3 \\ z_3'z_2 & z_3'z_3 & z_3'y_1 & z_3'y_2 & z_3'y_3 \\ y_1'z_2 & y_1'z_3 & y_1'y_1 & y_1'y_2 & y_1'y_3 \\ y_2'z_2 & y_2'z_3 & y_2'y_1 & y_2'y_2 & y_2'y_3 \\ y_3'z_2 & y_3'z_3 & y_3'y_1 & y_3'y_2 & y_3'y_3 \end{bmatrix}_{\perp Z_1}$}} \\
0 & & & & & \\
0 & & & & & \\
0 & & & & & \\
0 & & & & &
\end{bmatrix}
$$

where A_{11}, A_{12}, etc. stand for matrices of no further interest to us where $Z_1'Z_1$, $Z_1'Z_2$, etc. occurred before. The entire lower right hand submatrix is the moment matrix of the part of the variables inside it orthogonal to Z_1, (e.g., the submatrix in the position $Z_3'y_1$ is $[Z_3'y_1]_{\perp Z_1}$). rk Z_1 is the number of pivots used (see section I.B.2).

(4) Retrieve (B.3) and replace the row and column of (B.3) corresponding to y_1 by the corresponding elements of (B.4). (B.3) will have become:

$$
(B.5)\qquad
\left[\begin{array}{c|cccc}
z_2'z_2 & z_2'z_3 & [z_2'y_1]_{\perp Z_1} & z_2'y_2 & z_2'y_3 \\
\hline
z_3'z_2 & z_3'z_3 & [z_3'y_1]_{\perp Z_1} & z_3'y_2 & z_3'y_3 \\
[y_1'z_2]_{\perp Z_1} & [y_1'z_3]_{\perp Z_1} & [y_1'y_1]_{\perp Z_1} & [y_1'y_2]_{\perp Z_1} & [y_1'y_3]_{\perp Z_1} \\
y_2'z_2 & y_2'z_3 & [y_2'y_1]_{\perp Z_1} & y_2'y_2 & y_2'y_3 \\
y_3'z_2 & y_3'z_3 & [y_3'y_1]_{\perp Z_1} & y_3'y_2 & y_3'y_3
\end{array}\right].
$$

Let us designate the part of (B.5) below and to the right of the dashed lines as (B.6). (B.6) is saved aside at this point so that it may be used for later calculations. (B.2) will already have been overwritten and we are finished with (B.3) and (B.4).

(5) Calculate $\{(Z_3, [y_1]_{\perp Z_1}, y_2, y_3)'(Z_3, [y_1]_{\perp Z_1}, y_2, y_3)\}_{\perp Z_2}$ in the manner given in section I.D.2; that is, perform elementary row operations on the (B.5) matrix until all elements below the diagonal of the first N_2 columns are reduced to zeros. Thus, the pivot element for each step is selected from the first N_2 columns. As before it is advisable to rearrange rows and columns at each step to improve accuracy. After all of the first N_2 diagonal elements have been used as pivots or the largest diagonal element has become less than ϵ, the matrix will have been transformed to:

(B.7)
$$\left[\begin{array}{c|cccc} A_{22} & A_{23} & A_{2,y_1} & A_{2,y_2} & A_{2,y_3} \\ \hline 0 & [Z_3'Z_3]_{\perp Z_2} & A_{3,y_1} & [Z_3'y_2]_{\perp Z_2} & [Z_3'y_3]_{\perp Z_2} \\ 0 & A_{y_1,3} & A_{y_1,y_1} & [y_1]_{\perp Z_1}'[y_2]_{\perp Z_2} & A_{y_1,y_3} \\ 0 & [y_2'Z_3]_{\perp Z_2} & [y_2]_{\perp Z_2}'[y_1]_{\perp Z_1} & [y_2'y_2]_{\perp Z_2} & [y_2'y_3]_{\perp Z_2} \\ 0 & [y_3'Z_3]_{\perp Z_2} & A_{y_3,y_1} & [y_3'y_2]_{\perp Z_2} & [y_3'y_3]_{\perp Z_2} \end{array}\right].$$

where the A_{ij} are submatrices of no further interest to us.

As before, rk Z_2 is the number of pivots used.

(6) Retrieve (B.6) and replace the row and column of (B.6) corresponding to y_2 by the corresponding elements of (B.7). (B.6) will have become:

$$
\text{(B.8)}\qquad
\left[\begin{array}{c|ccc}
z_3'z_3 & [z_3'y_1]_{\perp z_1} & [z_3'y_2]_{\perp z_2} & z_3'y_3 \\
\hline
[y_1'z_3]_{\perp z_1} & [y_1'y_1]_{\perp z_1} & [y_1]_{\perp}'z_1[y_2]_{\perp z_2} & [y_1'y_3]_{\perp z_1} \\
[y_2'z_3]_{\perp z_2} & [y_2]_{\perp}'z_2[y_1]_{\perp z_1} & [y_2'y_2]_{\perp z_2} & [y_2'y_3]_{\perp z_2} \\
y_3'z_3 & [y_3'y_1]_{\perp z_1} & [y_3'y_2]_{\perp z_2} & y_3'y_3
\end{array}\right]
$$

Let us designate the part of (B.8) below and to the right of the dashed lines as (B.9). (B.9) is saved aside at this point for further calculations. (B.2) through (B.7) are not used for further calculations.

(7) Calculate $\{([y_1]_{\perp z_1}, [y_2]_{\perp z_2}, y_3)'([y_1]_{\perp z_1}, [y_2]_{\perp z_2}, y_3)\}_{\perp z_3}$ in the manner given in section I.D.2; that is, perform elementary row operations on the (B.8) matrix until all elements below the diagonal of the first N_3 columns are reduced to zeros. Thus, the pivot element for each step is selected from the first N_3 columns. (As before, it is advisable to rearrange rows and columns at each step to improve accuracy.) After all of the first N_3 diagonal elements have been used as pivots or the largest diagonal element has become less than ϵ, the (B.8) matrix will have been transformed to:

$$
\text{(B.10)}\qquad
\left[\begin{array}{cccc}
A_{33} & A_{3,y_1} & A_{3,y_2} & A_{3,y_3} \\
0 & A_{y_1,y_1} & A_{y_1,y_2} & [y_1]_{\perp}'z_1[y_3]_{\perp z_3} \\
0 & A_{y_2,y_1} & A_{y_2,y_2} & [y_2]_{\perp}'z_2[y_3]_{\perp z_3} \\
0 & [y_3]_{\perp}'z_3[y_1]_{\perp z_1} & [y_3]_{\perp}'z_3[y_2]_{\perp z_2} & [y_3'y_3]_{\perp z_3}
\end{array}\right]
$$

where the A_{ij} are submatrices of no further interest to us. As before, rk Z_3 is the number of pivots used.

(8) Retrieve (B.9) and replace the row and column of (B.9) corresponding to y_3 by the corresponding elements of (B.10). (B.9) will have become:

$$\text{(B.11)} \quad \begin{bmatrix} [y_1'y_1]_{\perp Z_1} & [y_1]'_{\perp Z_1}[y_2]_{\perp Z_2} & [y_1]'_{\perp Z_1}[y_3]_{\perp Z_3} \\ [y_2]'_{\perp Z_2}[y_1]_{\perp Z_1} & [y_2'y_2]_{\perp Z_2} & [y_2]'_{\perp Z_2}[y_3]_{\perp Z_3} \\ [y_3]'_{\perp Z_3}[y_1]_{\perp Z_1} & [y_3]'_{\perp Z_3}[y_2]_{\perp Z_2} & [y_3'y_3]_{\perp Z_3} \end{bmatrix}$$

which is the desired moment matrix.

Modifications and Generalizations of the Preceding Procedure

The preceding procedure can be modified and generalized in several ways. Following are some of them:

(1) Only the upper triangular or lower triangular part of the initial moment matrix need be formed and all of the calculations given previously can be performed within this triangular part of the matrix, thereby saving computer storage.

(2) The procedure given for $m=3$ may, of course, be used for m matrices Z_i and corresponding y_i. Thus $[Z_1 \cdots Z_m, y_1 \cdots y_n]'[Z_1 \cdots Z_m, y_1 \cdots y_m]$ is formed and then an orthogonalization performed for each Z_i. Before orthogonalizing with respect to a given Z_i, the part of the moment matrix corresponding to $Z_{i+1} \cdots Z_m, y_1 \cdots y_m$, (say M_{ii}) is saved. After orthogonalizing with respect to Z_i, M_{ii} is retrieved and the row and column corresponding to y_i replaced by the corresponding row and column of the just orthogonalized matrix.

The desired matrix is obtained by replacing the m^{th} row and column of $M_{i-1,i-1}$ by the m^{th} row and column of the orthogonalized matrix of the m^{th} step.

(3) If it is desired that some of the jointly dependent variables (say $y_1^\dagger \cdots y_n^\dagger$) not be adjusted, i.e., that a moment matrix of the form $\{[y_1]_{\perp Z_1} \cdots [y_m]_{\perp Z_m}, y_1^\dagger \cdots y_n^\dagger\}'\{[y_1]_{\perp Z_1} \cdots [y_m]_{\perp Z_m}, y_1^\dagger \cdots y_n^\dagger\}$ be formed, this can readily be accomplished by starting with $[Z_1 \cdots Z_m, y_1 \cdots y_m, y_1^\dagger \cdots y_n^\dagger]'[Z_1 \cdots Z_n, y_1 \cdots y_m, y_1^\dagger \cdots y_m^\dagger]$ and then stopping after m orthogonalizations as before. (This is correct because $[y_i]'_{\perp Z_i} y_j^\dagger = [y_i' y_j^\dagger]_{\perp Z_i}$ [see (I.56)].) Of the jointly dependent variables of an equation, the normalizing variable is the most likely one not to be specially adjusted, i.e., to be designated a $y_i^\dagger$.)

(4) More than one y_i may be adjusted by the same Z_i. For example if $\{[y_1]_{\perp Z_1}, [y_2]_{\perp Z_2}, [y_3]_{\perp Z_2}, [y_4]_{\perp Z_4}\}'\{[y_1]_{\perp Z_1}, [y_2]_{\perp Z_2}, [y_3]_{\perp Z_2}, [y_4]_{\perp Z_4}\}$ is desired, this can be accomplished by starting with $[Z_1, Z_2, Z_4, y_1, y_2, y_3, y_4]'[Z_1, Z_2, Z_4, y_1, y_2, y_3, y_4]$ and (letting the matrix saved just before orthogonalizing by X_2 be denoted M_{11} and the matrix obtained by orthogonalizing by X_2 be denoted P_{22}) replacing the rows and columns corresponding to both y_2 and y_3 of the M_{11} matrix by the corresponding elements of P_{22}. As a last step, the orthogonalization by X_4 would be as usual.

(5) Usually there will be a set of variables common to all of the Z_i matrices. (For example, the variables in the matrix X_μ will usually be contained as instruments in all of the matrices of instruments used for an equation.) If so, we can orthogonalize with

respect to the set of variables initially and then omit them from the Z_i matrices. The following example illustrates this:

Suppose the variables in the $T \times N_0$ matrix Z_0 had been common to Z_1, Z_2, and Z_3 in the example given previously. Then we could form the moment matrix

$[Z_0, Z_1^-, Z_2^-, Z_3^-, y_1, y_2, y_3]'[Z_0, Z_1^-, Z_2^-, Z_3^-, y_1, y_2, y_3]$ where Z_1^- is the matrix of $N_1 - N_0$ variables of Z_1 not in Z_0, Z_2^- is the matrix of $N_2 - N_0$ variables of Z_2 not in Z_0, and Z_3^- is the matrix of $N_3 - N_0$ variables of Z_3 not in Z_0. This matrix would then be orthogonalized with respect to Z_0 giving us:

$M = [Z_1^-, Z_2^-, Z_3^-, y_1, y_2, y_3]'[Z_1^-, Z_2^-, Z_3^-, y_1, y_2, y_3]_{\perp Z_0}$.

The $(N_1 + N_2 + N_3 - 3N_0 + 3) \times (N_1 + N_2 + N_3 - 3N_0 + 3)$ matrix M is then substituted for (B.2). All calculations then proceed the same as before (steps 2 through 8) until the desired matrix (B.11) is obtained. The only difference in the procedure is that in place of the N_1 rows and columns corresponding to Z_1, we have $N_1 - N_0$ rows and columns corresponding to $[Z_1^-]_{\perp Z_0}$; in place of the N_2 rows and columns corresponding to Z_2, we have $N_2 - N_0$ rows and columns corresponding to $[Z_2^-]_{\perp Z_0}$; and in place of the N_3 rows and columns corresponding to Z_3, we have $N_3 - N_0$ rows and columns corresponding to $[Z_3^-]_{\perp Z_0}$. Also, $\text{rk } Z_1 = \text{rk } Z_0 + \text{rk } [Z_1^-]_{\perp Z_0}$, $\text{rk } Z_2 = \text{rk } Z_0 + [Z_2^-]_{\perp Z_0}$, and $\text{rk } Z_3 = \text{rk } Z_0 + \text{rk } [Z_3^-]_{\perp Z_0}$.

Verification that the Procedure Produces the Desired Matrix

That (B.4) and any submatrix of the form $[Z_i'y_j]_{\perp Z_k}$, $[y_i'Z_j]_{\perp Z_k}$, or $[y_i'y_j]_{\perp Z_k}$ are as claimed may be verified by comparing the calculations producing them with the calculations given in section I.D.2.

(The calculations given in section I.D.2 are verified at the end of section I.D.2.)

We will now verify that the elements $[y_i]'_{\perp Z_i}[y_j]_{\perp Z_j}$ are as claimed.

The relevant submatrices used in computing $[y_i]'_{\perp Z_i}[y_j]_{\perp Z_j}$ are (assuming $i < j$):

$$\text{(B.12)} \qquad \begin{bmatrix} \underset{N_1 \times N_1}{Z_j'Z_j} & \underset{N_1 \times 1}{Z_j'y_j} \\ \underset{1 \times N_1}{[y_i'Z_j]_{\perp Z_i}} & \underset{1 \times 1}{[y_i'y_j]_{\perp Z_i}} \end{bmatrix}$$

Let us initially assume that Z_j has full column rank.

Performing elementary row operations on (B.12) to reduce the first N_1 columns to zero below the diagonal is equivalent to pre-multiplying (B.12) by a nonsingular matrix $\begin{bmatrix} E_{11} & 0 \\ E_{21} & I \end{bmatrix}$ such that:

$$\begin{bmatrix} E_{11} & 0 \\ E_{21} & I \end{bmatrix} \begin{bmatrix} Z_j'Z_j & Z_j'y_j \\ [y_i'Z_j]_{\perp Z_i} & [y_i'y_j]_{\perp Z_i} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} .$$

Thus:

$$E_{21}[Z_j'Z_j] + [y_i'Z_j]_{\perp Z_i} = 0 \quad \text{or} \quad E_{21} = -[y_i'Z_j]_{\perp Z_i}[Z_j'Z_j]^{-1}$$

and:

$$E_{21}[Z_j'y_j] + [y_i'y_j]_{\perp Z_i} = A_{22} .$$

Substituting for E_{21} in the last equation we get:

$$\text{(B.13)} \qquad -[y_i'Z_j]_{\perp Z_i}[Z_j'Z_j]^{-1}[Z_j'y_j] + [y_i'y_i]_{\perp Z_i} = A_{22} .$$

Since $[y_i'Z_j]_{\perp Z_i} = y_i'[I - Z_i^*(Z_i^{*\prime}Z_i^*)^{-1}Z_i^{*\prime}]Z_j$ and $[y_i'y_j]_{\perp Z_i} = y_i'[I - Z_i^*(Z_i^{*\prime}Z_i^*)^{-1}Z_i^{*\prime}]y_j$ where Z_i^* is a subset of the variables in Z_i, Z_i^* having full column rank which is the same rank as Z_i (see section I.D.1), we may rewrite (B.13) as:

(B.14) $A_{22} =$

$$-y_i'[I - Z_i^*(Z_i^{*\prime}Z_i^*)^{-1}Z_i^{*\prime}]Z_j[Z_j'Z_j]^{-1}[Z_j'y_j] + y_i'[I - Z_i^*(Z_i^{*\prime}Z_i^*)^{-1}Z_i^{*\prime}]y_j$$

$$= y_i'[I - Z_i^*(Z_i^{*\prime}Z_i^*)^{-1}Z_i^{*\prime}][I - Z_j(Z_j'Z_j)^{-1}Z_j']y_j$$

$$= [y_i]_{\perp Z_i}'[y_j]_{\perp Z_j} \quad \text{[see (I.47)]}.$$

Thus, in the case of a Z_j having full column rank, the desired element is obtained.

In the case of a Z_j having rank $N_j^* < N_j$, row operations are performed on the columns corresponding to N_j^* of the variables in Z_j before the diagonal elements corresponding to the remaining variables become less than ϵ. (The orthogonalization stops at this point.) This is equivalent to performing row operations on the following submatrix of (B.12) (letting Z_j^* be a submatrix of Z_j containing the variables corresponding to the N_j^* diagonal elements used as pivots):

(B.15)
$$\begin{bmatrix} Z_j^{*\prime}Z_j^* & Z_j^{*\prime}y_j \\ N_j^* \times N_j^* & N_j^* \times 1 \\ [y_i'Z_j^*]_{\perp Z_i} & [y_i'y_i]_{\perp Z_i} \\ 1 \times N_j^* & 1 \times 1 \end{bmatrix}.$$

The same derivation may now be performed on (B.15) as was performed with (B.12), the only difference in the intermediate matrices obtained is that Z_j^* will occur in place of Z_j wherever Z_j

presently occurs. Thus, (B.14) becomes:

$$(B.16)\quad A_{22} = y_i'[I - Z_i^*(Z_i^{*\prime}Z_i^*)^{-1}Z_i^{*\prime}][I - Z_j^*(Z_j^{*\prime}Z_j^*)^{-1}Z_j^{*\prime}]y_2$$

$$= [y_i]_{\perp Z_i}[y_j]_{\perp Z_j} \qquad [\text{see (I.47)}].$$

Thus, the desired element is obtained even in the case of a Z_j having less than full column rank.

APPENDIX C

TENTATIVE PROOFS REGARDING THE CONSISTENCY OF $\hat{\delta}_{k_1,k_2}$ AND $\hat{\sigma}^2_{k_1,k_2}$

Consistency and the concept of a probability limit (plim) are discussed in section I.B. For a discussion of the matrix algebra of the plim operator see Goldberger [1964], especially pp. 115-120 and Christ [1966].[1]

In this appendix we will need to distinguish between the matrices Y and Y_μ. As in our initial notation (section I.C), Y will refer to the $T \times G$ matrix of all of the jointly dependent variables in the system, and Y_μ will refer to the $T \times m_\mu$ submatrix of Y corresponding to the m_μ "explanatory" jointly dependent variables in equation μ. Since it will cause no notational conflict, the y_μ vector (the $T \times 1$ submatrix of Y corresponding to the normalizing jointly dependent variable in equation μ) will be written as y and the u_μ vector (the $T \times 1$ vector of disturbances of the μ^{th} equation) will be written as u. Also, m_μ, ℓ_μ, and n_μ will be shortened to m, ℓ, and n respectively.

From preceding assumptions or derivations we have:

(C.1) $\operatorname{plim}(1/T)U'U = \Sigma$ where U is $T \times M$ and Σ is $M \times M$.[2] The diagonal element of Σ corresponding to the μ^{th} equation is σ^2; hence, $\operatorname{plim}(1/T)u'u = \sigma^2$.

(C.2) $\operatorname{plim}(1/T)X'X = \Omega_{XX}$ where X is $T \times \Lambda$ and Ω_{XX} is $\Lambda \times \Lambda$.[3] The submatrix of X consisting of the predetermined variables in

---

[1] As before in this paper, plim stand for $\operatorname{plim}_{T \to \infty}$.

[2] Assumption 2, section I.C.3.

[3] Assumption 3, section I.C.3.

the μ^{th} equation is the $T\times\ell$ matrix X_μ ; hence,

$\text{plim}(1/T)X_\mu'X_\mu = \Omega_{X_\mu X_\mu}$, an $\ell\times\ell$ matrix.

(C.3) $\text{plim}(1/T)X'U = 0$ where 0 is a $\Lambda\times M$ matrix.[1] Thus,

$\text{plim}(1/T)X_\mu'U = 0$ (with 0 an $\ell\times M$ matrix).

(C.4) $\text{plim}(1/T)X'V = 0$ where V is $T\times G$ and, therefore, 0 is $\Lambda\times G$.[2]

(C.5) $\text{plim}(1/T)X_I'X_I = \Omega_{X_I X_I}$ where X_I is a $T\times K$ matrix of instruments and $\Omega_{X_I X_I}$ is $K\times K$.[3] We will assume that the variables in X_μ are contained in X_I (hence, $\Omega_{X_\mu X_\mu}$ is a submatrix of $\Omega_{X_I X_I}$ as well as a submatrix of Ω_{XX}).

Let $\text{rk}\,\Omega_{X_I X_I} = p$. Since for all T, $(1/T)X_I'X_I$ is a moment matrix, $\Omega_{X_I X_I}$ is a moment matrix and there exists a non-singular $p\times p$ submatrix $\Omega_{X_I^* X_I^*} = \text{plim}(1/T)X_I^{*\prime}X_I^*$. X_I^* is a matrix of variables from X_I . If $\Omega_{X_I X_I}$ is $p\times p$, then $\Omega_{X_I^* X_I^*}$ is the entire matrix $\Omega_{X_I X_I}$ and $X_I^* = X_I$. We will assume that $\text{rk}\,X_I^* = \text{rk}\,X_I$ for all T and $\text{rk}\,X_I^* = p$ for T sufficiently large; however, X_I^* and X_I may have rank less than p for small T.

(C.6) $\text{plim}(1/T)X_I'X = \Omega_{X_I X}$ with $\Omega_{X_I X}$ a $K\times\Lambda$ matrix.[4] ($\Omega_{X_\mu X_\mu}$ is a submatrix of $\Omega_{X_I X}$ as well as a submatrix of Ω_{XX} and $\Omega_{X_I X_I}$.)

(C.7) $\text{plim}(1/T)X_I'U = 0$ where 0 is a $K\times M$ matrix.[5] Thus,

$\text{plim}(1/T)X_I^{*\prime}U = 0$ (with 0 a $\text{rk}\,X_I \times M$ matrix).

(C.8) $\text{plim}(1/T)X_I'V = 0$ where 0 is $K\times G$.[6]

---

[1] Assumption 3, section I.C.3.

[2] Follows from C.3. See (I.25).

[3] Assumption 5, section I.C.3.

[4] Assumption 5, section I.C.3.

[5] Assumption 5, section I.C.3.

[6] Follows from (C.7) See (I.26).

(C.9) $\operatorname{plim}(1/T)V'V = \Omega_{VV}$, a $G \times G$ matrix.[1]

Given assumptions (C.1) through (C.9) we will derive that:[2]

$$(C.10)\quad \operatorname{plim}\frac{1}{T}\begin{bmatrix} Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_{\perp} X_I & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix} = \begin{bmatrix} \Omega_{Y_\mu X_I^*}\Omega_{X_I^*X_I^*}^{-1}\Omega_{X_I^*Y_\mu} & \Omega_{Y_\mu X_\mu} \\ \Omega_{X_\mu Y_\mu} & \Omega_{X_\mu X_\mu} \end{bmatrix}$$

where $\Omega_{Y_\mu X_I^*}$ and $\Omega_{Y_\mu X_\mu}$ are defined by (C.13).

In addition to assumptions (C.1) through (C.9) we will require the assumption:

$$(C.11)\quad \begin{bmatrix} \Omega_{Y_\mu X_I^*}\Omega_{X_I^*X_I^*}^{-1}\Omega_{X_I^*Y_\mu} & \Omega_{Y_\mu X_\mu} \\ \Omega_{X_\mu Y_\mu} & \Omega_{X_\mu X_\mu} \end{bmatrix} \text{ is nonsingular.}$$

It will be convenient for us to derive the plim of 1/T times the sums of cross-product of certain matrices with Y before commencing the main derivations of this appendix

$$(C.12)\quad \operatorname{plim}(1/T)Y'U = \operatorname{plim}(1/T)[X\Pi' + V]'U$$

$$= \operatorname{plim}(1/T)\Pi X'U + \operatorname{plim}(1/T)V'U$$

$$= \Pi \cdot \operatorname{plim}(1/T)X'U + \operatorname{plim}(1/T)(-1)\Gamma^{-1}[U \vdots 0]'U$$

$$= \Pi \cdot 0 - \Gamma^{-1}\begin{bmatrix} \operatorname{plim}(1/T)U'U \\ 0 \end{bmatrix} = -\Gamma^{-1}\begin{bmatrix} \Sigma \\ 0 \end{bmatrix} \overset{\text{def}}{=} \Omega_{YU}\ .^{3}$$

We will use $\operatorname{plim}(1/T)Y'u = \Omega_{Y_\mu u}$, an $m \times 1$ submatrix of Ω_{YU} .

---

[1] Follows from the relationship between V and U. See (I.22).

[2] See (C.22).

[3] "def" denotes that we are defining $-\Gamma^{-1}\begin{bmatrix} \Sigma \\ 0 \end{bmatrix}$ as Ω_{YU} .

$$\text{(C.13)}\quad \operatorname{plim}(1/T)Y'X_I = \operatorname{plim}(1/T)[X\Pi' + V]'X_I$$

$$= \Pi\operatorname{plim}(1/T)X'X_I + \operatorname{plim}(1/T)V'X_I$$

$$= \Pi\Omega_{XX_I} + 0 = \Pi\Omega_{XX_I} \stackrel{\text{def}}{=} \Omega_{YX_I} .$$

We will use $\operatorname{plim}(1/T)Y_\mu'X_I = \Omega_{Y_\mu X_I}$ (an $m \times K$ submatrix of Ω_{YX_I}), $\operatorname{plim}(1/T)Y_\mu'X_\mu = \Omega_{Y_\mu X_\mu}$ (an $m \times \ell$ submatrix of Ω_{YX_I}), and $\operatorname{plim}(1/T)Y_\mu'X_I^* = \Omega_{Y_\mu X^*_I}$ (an $m \times p$ submatrix of Ω_{YX_I}).

$$\text{(C.14)}\quad \operatorname{plim}(1/T)Y'Y = \operatorname{plim}(1/T)[X\Pi' + V]'[X\Pi' + V]$$

$$= \operatorname{plim}(1/T)[\Pi X'X\Pi' + \Pi X'V + V'X\Pi' + V'V]$$

$$= \Pi[\operatorname{plim}(1/T)X'X]\Pi' + \Pi\operatorname{plim}(1/T)X'V + [\operatorname{plim}(1/T)V'X]\Pi' + \operatorname{plim}(1/T)V'V$$

$$= \Pi\Omega_{XX}\Pi' + \Pi\cdot 0 + 0\cdot\Pi' + \Omega_{VV} = \Pi\Omega_{XX}\Pi' + \Omega_{VV} \stackrel{\text{def}}{=} \Omega_{YY} .$$

We will use $\operatorname{plim}(1/T)Y_\mu'Y_\mu = \Omega_{Y_\mu Y_\mu}$, a submatrix of Ω_{YY}.

Theorem (C.15):

If $\operatorname{plim}(k_1 - 1) = \operatorname{plim}(k_2 - 1) = 0$, then

$$\hat{\delta}_{k_1,k_2} = \begin{bmatrix} Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_{\perp X_I} & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y_\mu'y - k_2[Y_\mu'y]_{\perp X_I} \\ X_\mu'y \end{bmatrix}$$

is a consistent estimator of δ.[1]

Proof of Theorem (C.15):

$$\text{(C.16)}\quad \hat{\delta}_{k_1,k_2} - \delta = \begin{bmatrix} Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_{\perp X_I} & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y_\mu' - k_2[Y_\mu]_{\perp X_I}' \\ X_\mu' \end{bmatrix} y - \delta$$

---

[1] Since $[Y_\mu'y]_{\perp X_I} = [Y_\mu]_{\perp X_I}' y$ [see (I.56)], the right hand vector may be written as $\begin{bmatrix} Y_\mu' - k_2[Y_\mu]_{\perp X_I}' \\ X_\mu' \end{bmatrix} y$.

(and since $y = [Y_\mu \vdots X_\mu]\delta + u$):

$$= \begin{bmatrix} Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_{\perp X_I} & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y_\mu' - k_2[Y_\mu]_{\perp X_I}' \\ X_\mu' \end{bmatrix} \{[Y_\mu \vdots X_\mu]\delta + u\} - \delta$$

(and since $[Y_\mu]_{\perp X_I}' Y_\mu = [Y_\mu'Y_\mu]_{\perp X_I}$ [see (I.56)] and $[Y_\mu]_{\perp X_I}' X_\mu = [Y_\mu'X_\mu]_{\perp X_I} = 0$ [see (I.56) and (I.45)]):

$$= \begin{bmatrix} Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_{\perp X_I} & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y_\mu'Y_\mu - k_2[Y_\mu'Y_\mu]_{\perp X_I} & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix} \delta$$

$$+ \begin{bmatrix} Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_{\perp X_I} & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y_\mu' - k_2[Y_\mu]_{\perp X_I}' \\ X_\mu' \end{bmatrix} u - \delta$$

$$= A\delta + d$$

where:

$$\text{(C.17)} \qquad A = \begin{bmatrix} Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_{\perp X_I} & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y_\mu'Y_\mu - k_2[Y_\mu'Y_\mu]_{\perp X_I} & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix} - I$$

and

$$\text{(C.18)} \qquad d = \begin{bmatrix} Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_{\perp X_I} & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y_\mu' - k_2[Y_\mu]_{\perp X_I}' \\ X_\mu' \end{bmatrix} u \ .$$

In evaluating plim A and plim d, we will first find it convenient to evaluate $\operatorname{plim}(1/T)\{Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_{\perp X_I}\}$ and $\operatorname{plim}(1/T)\{Y_\mu'Y_\mu - k_2[Y_\mu'Y_\mu]_{\perp X_I}\}$.

(C.19) $\operatorname{plim}(1/T)[Y_\mu'Y_\mu]_{\perp X_I} = \operatorname{plim}(1/T)\{Y_\mu'Y_\mu - [Y_\mu'Y_\mu]_{\perp X_I}\}$ and for T sufficiently large [see (C.5)] and using (I.55):

$$= \operatorname{plim}(1/T)[Y_\mu'Y_\mu] - \operatorname{plim}(1/T)Y_\mu'X_I^*(X_I^{*\prime}X_I^*)^{-1}X_I^{*\prime}Y_\mu$$

$$= \Omega_{Y_\mu Y_\mu} - [\operatorname{plim}(1/T)Y_\mu'X_I^*][\operatorname{plim}(1/T)(X_I^{*\prime}X_I^*)]^{-1}[\operatorname{plim}(1/T)X_I^{*\prime}Y_\mu]$$

$$= \Omega_{Y_\mu Y_\mu} - \Omega_{Y_\mu X_I^*}\Omega^{-1}_{X_I^*X_I^*}\Omega_{X_I^*Y_\mu} .$$

Therefore,

$$\text{(C.20)}\quad \operatorname{plim}(1/T)\{Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_{\perp X_I}\} =$$

$$\operatorname{plim}(1/T)Y_\mu'Y_\mu - \operatorname{plim} k_1 \cdot \operatorname{plim}(1/T)[Y_\mu'Y_\mu]_{\perp X_I}$$

$$= \Omega_{Y_\mu Y_\mu} - 1\cdot[\Omega_{Y_\mu Y_\mu} - \Omega_{Y_\mu X_I^*}\Omega^{-1}_{X_I^*X_I^*}\Omega_{X_I^*Y_\mu}] = \Omega_{Y_\mu X_I^*}\Omega^{-1}_{X_I^*X_I^*}\Omega_{X_I^*Y_\mu} .$$

Similarly,

$$\text{(C.21)}\quad \operatorname{plim}(1/T)\{Y_\mu'Y_\mu - k_2[Y_\mu'Y_\mu]_{\perp X_I}\} = \Omega_{Y_\mu X_I^*}\Omega^{-1}_{X_I^*X_I^*}\Omega_{X_I^*Y_\mu} .$$

Applying (C.6), (C.13), (C.20), and (C.21) we have:

$$\text{(C.22)}\quad \operatorname{plim}\frac{1}{T}\begin{bmatrix} Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_{\perp X_I} & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix} = \operatorname{plim}\frac{1}{T}\begin{bmatrix} Y_\mu'Y_\mu - k_2[Y_\mu'Y_\mu]_{\perp X_I} & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix}$$

$$= \begin{bmatrix} \Omega_{Y_\mu X_I^*}\Omega^{-1}_{X_I^*X_I^*}\Omega_{X_I^*Y_\mu} & \Omega_{Y_\mu X_\mu} \\ \Omega_{X_\mu Y_\mu} & \Omega_{X_\mu X_\mu} \end{bmatrix} ;$$

hence:

$$\text{(C.23)}\quad \operatorname{plim} A =$$

$$\operatorname{plim}\left\{\left\{\frac{1}{T}\begin{bmatrix} Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_{\perp X_I} & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix}\right\}^{-1}\left\{\frac{1}{T}\begin{bmatrix} Y_\mu'Y_\mu - k_2[Y_\mu'Y_\mu]_{\perp X_I} & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix}\right\}\right] - I$$

$$= \begin{bmatrix} \Omega_{Y_\mu X_I^*}\Omega^{-1}_{X_I^*X_I^*}\Omega_{X_I^*Y_\mu} & \Omega_{Y_\mu X_\mu} \\ \Omega_{X_\mu Y_\mu} & \Omega_{X_\mu X_\mu} \end{bmatrix}^{-1}\begin{bmatrix} \Omega_{Y_\mu X_I^*}\Omega^{-1}_{X_I^*X_I^*}\Omega_{X_I^*Y_\mu} & \Omega_{Y_\mu X_\mu} \\ \Omega_{X_\mu Y_\mu} & \Omega_{X_\mu X_\mu} \end{bmatrix} - I$$

$= I - I = 0$ (where I and 0 are $n \times n$ matrices).

To evaluate plim d, we first note that:

$$(C.24)\quad \operatorname{plim}(1/T)[Y_\mu]_\perp' X_I u = \operatorname{plim}(1/T)[Y_\mu' u - Y_\mu' X_I^* (X_I^{*\prime} X_I^*)^{-1} X_I^{*\prime} u]$$

$$= [\operatorname{plim}(1/T) Y_\mu' u] - [\operatorname{plim}(1/T) Y_\mu' X_I^*][\operatorname{plim}(1/T) X_I^{*\prime} X_I^*]^{-1}[\operatorname{plim}(1/T) X_I^{*\prime} u]$$

$$= \Omega_{Y_\mu u} - \Omega_{Y_\mu X_I^*} \Omega_{X_I^* X_I^*}^{-1} 0 \qquad \text{(where } 0 \text{ is a } K \times 1 \text{ vector)}$$

$$= \Omega_{Y_\mu u} .$$

Hence:

$$(C.25)\quad \operatorname{plim} d = \operatorname{plim}\left\{\left\{\frac{1}{T}\begin{bmatrix} Y_\mu' Y_\mu - k_1 [Y_\mu' Y_\mu]_\perp X_I & Y_\mu' X_\mu \\ X_\mu' Y_\mu & X_\mu' X_\mu \end{bmatrix}\right\}^{-1}\left\{\frac{1}{T}\begin{bmatrix} Y_\mu' u - k_2 [Y_\mu]_\perp' X_\mu u \\ X_\mu' u \end{bmatrix}\right\}\right\}$$

$$= \begin{bmatrix} \Omega_{Y_\mu X_I^*} \Omega_{X_I^* X_I^*}^{-1} \Omega_{X_I^* Y_\mu} & \Omega_{Y_\mu X_\mu} \\ \Omega_{X_\mu Y_\mu} & \Omega_{X_\mu X_\mu} \end{bmatrix}^{-1} \begin{bmatrix} \Omega_{Y_\mu u} - 1 \cdot \Omega_{Y_\mu u} \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \Omega_{Y_\mu X_I^*} \Omega_{X_I^* X_I^*}^{-1} \Omega_{X_I^* Y_\mu} & \Omega_{Y_\mu X_\mu} \\ \Omega_{X_\mu Y_\mu} & \Omega_{X_\mu X_\mu} \end{bmatrix}^{-1} 0 = 0$$

(where 0 is an $n \times 1$ vector).

Finally, using (C.16), (C.23), and (C.25) we have:

$$(C.26)\quad \operatorname{plim}(\hat{\delta}_{k_1,k_2} - \delta) = \operatorname{plim}(A\delta + d) = [\operatorname{plim} A]\delta + \operatorname{plim} d$$

$= 0\delta + 0 = 0$ (where the first 0 is an $n \times n$ matrix and the other two 0's are $n \times 1$ vectors).

Hence, $\operatorname{plim} \hat{\delta}_{k_1,k_2} = \delta$ and $\hat{\delta}_{k_1,k_2}$ is a consistent estimator of δ.

Theorem (C.27):

If $\operatorname{plim}(k_1 - 1) = \operatorname{plim}(k_2 - 1) = 0$, then $\hat{\sigma}^2_{k_1,k_2} = \hat{u}'_{k_1,k_2}\hat{u}_{k_1,k_2}/(T - n)$ is consistent estimate of σ^2.

Proof of Theorem (C.27):

We will first show that $\hat{u}'_{k_1,k_2}\hat{u}_{k_1,k_2}/T$ is a consistent estimate of σ^2.

(C.28) $$\hat{u}_{k_1,k_2} = y - Z_\mu\hat{\delta}_{k_1,k_2} = y - Z_\mu\delta - Z_\mu[\hat{\delta}_{k_1,k_2} - \delta]$$

$$= u - Z_\mu[A\delta + d] \quad \text{[see (C.16)]}.$$

Therefore,

(C.29) $$\operatorname{plim}(1/T)\hat{u}'_{k_1,k_2}\hat{u}_{k_1,k_2} = \operatorname{plim}(1/T)\{u - Z_\mu[A\delta + d]\}'\{u - Z_\mu[A\delta + d]\}$$

$$= \operatorname{plim}(1/T)\{u'u - u'Z_\mu[A\delta + d] - [A\delta + d]'Z'_\mu u + [A\delta + d]'Z'_\mu Z_\mu[A\delta + d]\}$$

$$= \operatorname{plim}(1/T)u'u - [\operatorname{plim}(1/T)u'Z_\mu]\operatorname{plim}[A\delta + d]$$

$$- \operatorname{plim}[A\delta + d]'\operatorname{plim}(1/T)Z'_\mu u + \operatorname{plim}[A\delta + d]'[\operatorname{plim}(1/T)Z'_\mu Z_\mu]\operatorname{plim}[A\delta + d]$$

$$= \sigma^2 - [\Omega_{uY_\mu} \vdots 0]0 - 0'\begin{bmatrix}\Omega_{Y_\mu u}\\ 0\end{bmatrix} + 0'\begin{bmatrix}\Omega_{Y_\mu Y_\mu} & \Omega_{Y_\mu X_\mu}\\ \Omega_{X_\mu Y_\mu} & \Omega_{X_\mu X_\mu}\end{bmatrix}0 = \sigma^2 .$$

Finally,

(C.30) $$\hat{\sigma}^2_{k_1,k_2} = \operatorname{plim}\hat{u}'_{k_1,k_2}\hat{u}_{k_1,k_2}/(T - n)$$

$$= \operatorname{plim}[T/(T - n)]\cdot\operatorname{plim}(1/T)\hat{u}'_{k_1,k_2}\hat{u}_{k_1,k_2} = 1\cdot\sigma^2 = \sigma^2 .$$

Verification that $(1/T)\operatorname{plim} T\,\hat{\mathrm{V}}\mathrm{ar}(\hat{\delta}_{k_1,k_2})$ [with $\hat{\mathrm{V}}\mathrm{ar}(\hat{\delta}_{k_1,k_2})$ given by (III.4)] Equals the Asymptotic Coefficient Variance-covariance Matrix of the 2SLS Estimator

(C.31) Let $\hat{\mathrm{V}}\mathrm{ar}(\hat{\delta}_{k_1,k_2})$ be given by (III.4) and let $\operatorname{plim}(k_1 - 1) = \operatorname{plim}(k_2 - 1) = 0$. Then

$$\frac{1}{T}\text{plim } T\ \hat{\text{Var}}(\hat{\delta}_{k_1,k_2}) = \frac{1}{T}\text{plim } T\hat{\sigma}^2 \begin{bmatrix} Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_\perp X_I & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix}^{-1}$$

$$= \frac{1}{T}\text{plim } \hat{\sigma}^2 \cdot \text{plim}\left\{\frac{1}{T}\begin{bmatrix} Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_\perp X_I & Y_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix}\right\}^{-1}$$

$$= \frac{1}{T}\sigma^2\left\{\text{plim}\frac{1}{T}\begin{bmatrix} Y_\mu'Y_\mu - k_1[Y_\mu'Y_\mu]_\perp X_I & X_\mu'X_\mu \\ X_\mu'Y_\mu & X_\mu'X_\mu \end{bmatrix}\right\}^{-1}$$

$$= \frac{1}{T}\sigma^2\begin{bmatrix} \Omega_{Y_\mu X_I^*}\Omega_{X_I^*X_I^*}^{-1}\Omega_{X_I^*Y_\mu} & \Omega_{Y_\mu X_\mu} \\ \Omega_{X_\mu Y_\mu} & \Omega_{X_\mu X_\mu} \end{bmatrix}^{-1} \qquad \text{[by (C.22)]}$$

which is the asymptotic coefficient variance-covariance matrix of the 2SLS estimator [see Goldberger [1964], p. 333].

BIBLIOGRAPHY

BIBLIOGRAPHY

Aitken, A.C. [1934-35], "On Least Squares and Linear Combination of Observation," Proceedings of the Royal Society of Edinburgh, LV (1934-35), 42-48

Anderson, T.W., and Rubin, Herman [1949], "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," Annals of Mathematical Statistics, XX, No. 1 (March,1949), 46-63.

_____ [1950], "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations," Annals of Mathematical Statistics, XXI, No. 4 (December,1950), 570-582.

Basmann, R.L. [1957], "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," Econometrica, XXV, No. 1 (January,1957), 77-83.

_____ [1960a], "On the Asymptotic Distribution of Generalized Linear Estimators," Econometrica XXVIII, No. 1 (January,1960), 97-107.

_____ [1960b], "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics," Journal of the American Statistical Association, LV, No. 292 (December,1960), 650-659.

_____ [1961], "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Overidentified Cases," Journal of the American Statistical Association, LVI, No. 295 (September,1961), 619-636.

Chernoff, Herman and Divinsky, Nathan [1953], "The Computation of Maximum-Likelihood Estimates of Linear Structural Equations," Chapter X in W.C. Hood and T.C. Koopmans, Studies in Econometric Method. New York: John Wiley and Sons, Inc., 1953, 236-302.

Chipman, John S., and Rao, M.M. [1964a], "The Treatment of Linear Restrictions in Regression Analysis," Econometrica, XXXII, No.1-2 (January-April,1964), 198-209.

_____ [1964b], "Projections, Generalized Inverses and Quadratic Forms," Journal of Mathematical Analysis and Applications, No. 9 (August,1964), 1-11.

Chow, Gregory C. [1964], "A Comparison of Alternative Simultaneous Estimators for Simultaneous Equations," Econometrica, XXXII, No. 4 (October,1964), 546-548.

Chow, Gregory C. and Ray-Chaudhuri, D.K. [1967], "An Alternative Proof of Hannan's Theorem on Canonical Correlation and Multiple Equation Systems," Econometrica, XXXV, No. 1 (January, 1967), 139-142.

Christ, Carl F. [1966], Econometric Models and Methods. New York: John Wiley and Sons, Inc., 1966.

Cragg, J.G. [1966], "On the Sensitivity of Simultaneous-Equations Estimators to the Stochastic Assumptions of the Models," Journal of the American Statistical Association, LXI, No. 313 (March, 1966), 136-151.

_____ [1967], "On the Relative Small-Sample Properties of Several Structural-Equation Estimators," Econometrica, XXXV, No. 1 (January, 1967), 89-110.

Crockett, J.B., and Chernoff, H. [1955], "Gradient Methods of Maximization," Pacific Journal of Mathematics, V, No. 5 (1955), 33-50.

Durbin, J. and Watson, G.S. [1950], "Testing for Serial Correlation in Least Squares Regression. I," Biometrica, XXXVII (1950), 409-428.

Durbin, J. and Watson, G.S. [1951], "Testing for Serial Correlation in Least Squares Regression. II," Biometrica, XXXVIII (1951), 159-178.

Dusenberry, J.S., Fromm, G., Klein, L.R., and Kuh, E. [1965], The Brookings Quarterly Econometric Model of the United States. Chicago: Rand McNally and Company, 1965.

Eisenpress, Harry [1961], "Forecasting by Econometric Systems," IB 9 FES (709 Program Description), Library Services Department, IBM Data Processing Department, 590 Madison Avenue, New York 22, New York, 1961.

Eisenpress, Harry [date unknown], "Experiments in Convergence in Full-Information Estimation," Multilithed Paper.

Eisenpress, Harry and Greenstadt, John [1964], "Non-Linear Full-Information Estimation," Multilithed Paper Presented at Meeting of Econometric Society, Chicago, Illinois, December 29, 1964, 17 pages.

Faddeeva, V.N. [1959], Computational Methods of Linear Algebra, (Translated by Curtis D. Benster). New York: Dover Publications, Inc., 1959.

Fisher, Franklin [1965], "Dynamic Structure and Estimation in Economy-wide Econometric Models," Chapter 15 in Dusenberry, J.S., Fromm, G., Klein, L.R., and Kuh, E., The Brookings Quarterly Econometric Model of the United States. Chicago: Rand McNally and Company, 1965, 588-635.

Freund, R.J. [1963], "A Warning of Roundoff Errors in Regression," The American Statistician, XVII, No. 5 (December, 1963), 13-15.

Goldberger, Arthur S. [1964], Econometric Theory. New York: John Wiley and Sons, Inc., 1964.

Hannan, E.J. [1967], "Cannonical Correlation and Multiple Equation Systems in Economics," Econometrica, XXXV, No. 1 (January, 1967), 123-138.

Hart, B.I. [1942], "Tabulation of the Probabilities for the Ratio of the Mean Square Successive Difference to the Variance," Annals of Mathematical Statistics, XIII (1942), 207-214.

Hood, Wm. C., and Koopmans, Tjalling C. [1953], Studies in Econometric Method. (Cowles Commission for Research in Economics Monograph No. 14.) New York: John Wiley and Sons, Inc., 1953.

Johnston, J. [1963], Econometric Methods. New York: McGraw-Hill Book Company, Inc., 1963.

Kendall, M.G., and Stuart, A. [1961], The Advanced Theory of Statistics, Vol. 2, New York: Hafner Publishing Company, 1961.

Klein, Lawrence R. [1950], Economic Fluctuations in the United States 1921-1941. New York: John Wiley and Sons, Inc., 1950.

Klein, Lawrence R. [1953], A Textbook of Econometrics. Evanston, Illinois: Row, Peterson and Company, 1953.

Klein, L.R., and Nakamura, Mitsugu [1962], "Singularity in the Equation System of Econometrics: Some Aspects of the Problem of Multicollinearity," International Economic Review, III, No. 3 (September, 1962), 274-299.

Kloek, T., and Mennes, L.B.M. [1960], "Simultaneous Equations Estimation Based on Principal Components of Predetermined Variables," Econometrica, XXVIII, No. 1 (January, 1960), 45-61.

Kmenta, Jan and Gilbert, Roy F. [1967], "Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regressions," Michigan State University Econometrics Workshop Paper No. 6707, (December, 1967), 41 pages. Forthcoming in the Journal of the American Statistical Association.

Koopmans, T.C., Rubin, H., and Leipnik, R.B. [1950], "Measuring the Equation Systems of Dynamic Economics," Chapter II in Tjalling C. Koopmans, Statistical Inference in Dynamic Economic Models. New York: John Wiley and Sons, Inc., 1950, 53-237.

Koopmans, Tjalling C. [1953], "Identification Problems in Economic Model Construction," Chapter II in W.C. Hood and T.C. Koopmans, Studies in Econometric Method. New York: John Wiley and Sons, Inc., 1953, 27-47.

Koopmans, Tjalling C., and Hood, Wm. C. [1953], "The Estimation of Simultaneous Linear Economic Relationships," Chapter VI in W.C. Hood and T.C. Koopmans, Studies in Econometric Method. New York: John Wiley and Sons, Inc., 1953, 112-199.

Koopmans, Tjalling C. [1950], Statistical Inference in Dynamic Economic Models. (Cowles Commission for Research in Economics Monograph No. 10.) New York: John Wiley and Sons, Inc., 1950.

Longley, James W. [1967], "An Appraisal of Least Squares Programs for the Electronic Computer from the Point of View of the User," Journal of the American Statistical Association. LXII, No. 319 (September, 1967), 819-841.

Madansky, Albert [1964], "On the Efficiency of Three-Stage Least-Squares Estimation," Econometrica, XXXII, No. 1-2 (January-April, 1964), 51-56.

Mann, H.B., and Wald, A. [1943a], "On the Statistical Treatment of Linear Stochastic Difference Equations," Econometrica, XI, Nos. 3&4 (July-October, 1943), 173-220.

_____ [1943b], "On Stochastic Limit and Order Relationships," Annals of Mathematical Statistics, XIV (1943), 217-226.

Marquardt, D.W. [1959], "Solution of Nonlinear Chemical Engineering Models," Chemical Engineering Progress, LV, No. 6 (June, 1959), 65-70.

Nagar, A.L. [1959], "The Bias and Moment Matrix of the General k-class Estimators of the Parameters in Simultaneous Equations," Econometrica, XXVII, No. 4 (October, 1959), 575-595.

_____ [1960], "A Monte Carlo Study of Alternative Simultaneous Equation Estimators," Econometrica, XXVIII, No. 3 (July, 1960), 573-587.

_____ [1961], "A Note on Residual Variance Estimation in Simultaneous Equations," Econometrica, XXIX, No. 2 (April, 1961), 238-243.

_____ [1962], "Double k-class Estimators of Parameters in Simultaneous Equations and Their Small Sample Properties," International Economic Review, III, No. 2 (May, 1962), 168-188.

Neely, Peter M. [1966], "Comparison of Several Algorithms for Computation of Means, Standard Deviations and Correlation Coefficients," Communications of the ACM, IX, No. 7 (July, 1966), 497-499.

Rothenberg, T.J., and Leenders, C.T. [1964], "Efficient Estimation of Simultaneous Equation Systems," Econometrica, XXXII, No. 1-2 (January-April, 1964), 57-76.

Rubin, Herman [1948], "Observations on the Computational Procedures for Maximum Likelihood Estimates," Cowles Commission Discussion Paper, Statistics, 317, July 17, 1948 (ditto) 2 pages.

Sargan, J.D. [1964], "Three-Stage Least-Squares and Full Maximum Likelihood Estimates," Econometrica, XXXII, No. 1-2 (January-April, 1964), 77-81.

Stroud, Arthur, and Zellner, Arnold [1962], "Program for Computing Efficient Regression Estimates and Associated Statistics," Systems Formulation and Methodology Workshop Paper 6204, Social Systems Research Institute, University of Wisconsin, May 22, 1962.

Summers, Robert [1965], "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators," Econometrica, XXXIII, No. 1 (January, 1965), 1-41.

Telser, Lester G. [1964], "Iterative Estimation of a Set of Linear Regression Equations," Journal of the American Statistical Association, LIX, No. 307 (September, 1964), 845-862.

Theil, H. [1961], Economic Forecasts and Policy. Amsterdam: North-Holland Publishing Company, 1961.

Theil, H. and Nagar, A.L. [1961], "Testing the Independence of Regression Disturbances," Journal of the American Statistical Association, LVI (December, 1961), 793-806.

Wagner, H.M. [1958], "A Monte Carlo Study of Estimates of Simultaneous Linear Structural Equations," Econometrica, XXVI, No. 1 (January, 1958), 117-133.

Zellner, Arnold, and Theil, H. [1962], "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations," Econometrica, XXX, No. 1 (January, 1962), 54-78.

Zellner, Arnold [1962], "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association, LVII, No. 298 (June, 1962), 348-368.

Zellner, A., and Thornber, H. [1966], "Computational Accuracy and Estimation of Simultaneous Equation Econometric Models," Econometrica, XXXIV, No. 3 (July, 1966), 727-729.