THE TWO-STAGE LEAST SQUARES METHOD OF ESTIMATION

by

ALBERT CHARLES KIENTZ, JR.

B.S., KANSAS STATE UNIVERSITY, 1962

A MASTER'S REPORT

Submitted in Partial Fulfillment of the

Requirements for the Degree

MASTER OF SCIENCE

Department of Statistics and Computer Science

KANSAS STATE UNIVERSITY

MANHATTAN, KANSAS

1968

Approved by

Young A. Koh

Major Professor

TABLE OF CONTENTS

Introduction	1
Definition of Problem and Terms	3
Development of Two-Stage Least Squares Method	5
Properties of the Estimators	9
Geometric Interpretation	11
Evaluation of the Method	12
Example	14
(a) Preliminary Computations	17
(b) Computation for a Single Structural Equation	19
Acknowledgment	23
References	24

INTRODUCTION

The two-stage least square method of linear estimation of coefficients was developed by Theil (1958). Basmann (1957) independently developed a similar solution under the name of the Generalized Classical method of linear estimation which leads to equivalent estimators. Several studies have been conducted more recently on the effectiveness of the method and its limitations.

The two-stage least square method was developed to replace existing methods (Indirect least squares, Least variance ratio, or Limited-information single equation) by providing a method of more general applicability while being less expensive to apply.

The problem to which this method is to be applied differs from single equation models in two ways. First, although interest may center on a single equation of a system, the entire system of relations is considered simultaneously to obtain a solution. Second, in many instances the problem may be to estimate all the parameters in a model and to make predictions from the complete model. In other words although it may be required to analyze only a single equation of the model, information from the entire system is included in the solution. In a system of equations the variables may be classified as endogenous and exogenous. Endogenous variables are those whose values are determined by the simultaneous interaction of the relations in the model and exogenous variables are those whose values are independent of those used in the model.

In the first stage of a two-stage least squares solution ordinary least squares methodology is applied to the entire system of predetermined variables to obtain estimates for the endogenous variables. In the second stage, these estimates are substituted in the system and ordinary least squares is applied again to a particular equation or a set of equations of the system to estimate the required parameters.

The method is not only relatively expeditious in terms of time required for calculations but the estimates derived can be shown to be asymptotically unbiased, consistent, and minimum-variance.

DEFINITION OF THE PROBLEM AND ASSOCIATED TERMS

In a system of linear equations containing G endogenous variables $y_{1t},\ldots,y_{Gt}$ and K exogenous variables $x_{1t},\ldots,x_{Kt}$; a typical set of equations may be written as:

$$(1)\quad y_{1t} = \sum_{i=2}^{H} \beta_{1i}\, y_{it} + \sum_{i=1}^{K^*} \gamma_{1i} x_{it} + e_{it}\ , \qquad t=1,\ldots,T$$

where $H-1 \leq G;\ K^* \leq K$

The quantities H-1 and K* indicate the number of variables of the system present in this particular set of equations. This set of equations may in turn be written in matrix form as:

$$Y_1 = \underline{Y}_2\ \underline{\beta}_2 + \underline{X}_*\ \underline{\ }_{1*} + \underline{e}_1$$

where

$$Y_1 = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1T} \end{bmatrix} \qquad \underline{Y}_2 = \begin{bmatrix} y_{21} & \cdots & y_{H1} \\ \vdots & & \vdots \\ y_{2T} & \cdots & y_{HT} \end{bmatrix}$$

$$\underline{\beta}'_2 = \begin{bmatrix} \beta_{12} \\ \vdots \\ \beta_{1H} \end{bmatrix} \qquad \underline{X}_* = \begin{bmatrix} x_{11} & \cdots & x_{K^*1} \\ \vdots & & \vdots \\ x_{1T} & \cdots & x_{K^*T} \end{bmatrix}$$

$$\underline{\gamma}'_{1*} = \begin{bmatrix} \gamma_{11} \\ \vdots \\ \gamma_{1K^*} \end{bmatrix}$$

The stochastic errors e_{it} are assumed to have zero means and finite variances and covariances:

$$E(e_{it}e_{jt}) = \sigma_{ij} \qquad i,j = 1,2,\ldots,G$$

$$\sigma_{ij} < \infty \qquad s,t = 1,2,\ldots,T$$

$$E(e_{it}e_{js}) = 0 \qquad s \neq t$$

for all s,t

$$E(x_{kt}e_{is}) = 0 \qquad k = 1,2,\ldots,K$$

To illustrate the basis for not applying the simple least squares method to this system consider a simple example:

$$(2) \quad y_{1t} = \alpha + \beta y_{2t} + e_t$$

$$(3) \quad y_{2t} = y_{1t} + x_t$$

Substituting (2) in (3) yields

$$y_{2t} = \alpha + \beta y_{2t} + x_t + e_t,$$

or $$y_{2t} = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} x_t + \frac{e_t}{1-\beta},$$

so that y_{2t} is seen in general to be influenced by e_t.

$$E(y_{2t}) = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} X_t$$

or $$E\{e_t\,[y_{2t} - E(y_{2t})]\} = \frac{1}{1-\beta} E(e_t^2) \neq 0.$$

The random error term and the explanatory variable, y_{2t}, in the equation are thus correlated and the direct application of the simple least squares method will not yield unbiased estimates.

DEVELOPMENT OF TWO-STAGE LEAST SQUARES METHOD

Since the random error vector and explanatory variables cannot be assumed independent, the number of methods of estimation are reduced. However, the mutual relationships which are the cause of the complication, are used to make estimation possible.

Consider the system of simultaneous linear equations outlined in the previous section and assume this system contains $K > K^*$ predetermined variables, each of which assumes non-stochastic values. Under these conditions it is possible to base the estimation method on the independence of the random error vector, $\underline{e}$, and the set of all predetermined variables in the entire system. More precisely:

Equation (1) is one of a system of $G \geq H$ stochastic linear equations in G jointly dependent and $K \geq K^* + H-1$ predetermined variables. This system can be solved for the jointly dependent variables.

Using the above assumptions it is now possible to apply simple least squares to the system using the X matrix of values for all predetermined variables, X_* being a submatrix of X. From the above assumption, all jointly dependent variables can be written as stochastic linear functions of X.

Applying simple least squares, extimators for Y_2, the vector of (H-1) dependent variables, are obtained and are written in matrix form as:

$$\hat{Y}_2 = X\ (X'X)^{-1} X'Y_2$$

which may be rewritten as:

$$Y_2 = X\ (X'X)^{-1} X'Y_2 + V$$

where V denotes the matrix of reduced-form residuals for the (H-1) dependent variables appearing on the right-hand side of the equation (1).

Now equation (1) may be rewritten as:

$$y_1 = (\underline{Y}_2 - \underline{V})\ \underline{\beta}'_2 + \underline{X}_*\ \underline{\gamma}'_{1*} + (\underline{e} + \underline{V\beta}'_2)$$

or
$$y_1 = [(\underline{Y}_2 - \underline{V})\ \underline{X}_*] \begin{bmatrix} \hat{\underline{\beta}}'_2 \\ \hat{\gamma}'_{1*} \end{bmatrix} + (e + V\beta'_2)$$

Applying simple least squares to this relation gives

$$\begin{bmatrix} \hat{\beta}'_2 \\ \hat{\gamma}'_{1*} \end{bmatrix} = (A'A)^{-1}\ A'y_1$$

where $A = [(\underline{Y}_2 - \underline{V})\ X_*]$

now $(\underline{Y}_2 - \underline{V})'\ (\underline{Y}_2 - \underline{V}) = \underline{Y}'_2\underline{Y}_2 - \underline{V}'\underline{Y}_2 - \underline{Y}'_2\underline{V} + \underline{V}'\underline{V}$

and $\underline{V}'Y_2 = \underline{V}'(\hat{\underline{Y}}_2 + \hat{\underline{V}}) = \underline{V}'\hat{\underline{Y}}_2 + \underline{V}'\underline{V}$

Since it is a property of the least squares fit that the residual is uncorrelated with the estimate value; it follows that $\underline{V}'\hat{\underline{Y}}_2 = \underline{0}$. This yields the result:

$$\underline{V}'\underline{Y}_2 = \underline{V}'\underline{V}.$$

Similarly

$$\underline{Y}_2'\underline{V} = \underline{V}'\underline{V}.$$

Hence

$$(\underline{Y}_2 - \underline{V})' \, (\underline{Y}_2 - \underline{V}) = \underline{Y}_2'\underline{Y}_2 - \underline{V}'\underline{V}. \; .$$

Moreover,

$$\begin{aligned}\underline{X}'\underline{V} &= \underline{X}' \, [\underline{Y}_2 - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}_2] \\ &= \underline{X}'\underline{Y}_2 - \underline{X}'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}_2 \\ &= \underline{X}'\underline{Y}_2 - \underline{X}'\underline{Y}_2 \\ &= \underline{0},\end{aligned}$$

which illustrates the property of least squares that the random error vector is uncorrelated with the explanatory values. Since $\underline{X}'\underline{V}$ equals zero, $\underline{X}_*' \, \underline{V}$ must equal zero since $\underline{X}_*$ is a submatrix of $\underline{X}$.

Returning to the solution for $\begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_{1*}' \end{bmatrix}$,

and using the above results, the two-stage estimator may be written

$$\begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_{1*}' \end{bmatrix} = \left[\begin{array}{c|c} (\underline{Y}_2-\underline{V})'(\underline{Y}_2-\underline{V}) & (\underline{Y}_2-\underline{V})'\underline{X}_* \\ \hline \underline{X}_*'(\underline{Y}_2-\underline{V}) & \underline{X}_*' \, \underline{X}_* \end{array}\right]^{-1} \begin{bmatrix} \underline{Y}_2'-\underline{V}' \\ \underline{X}_*' \end{bmatrix} \quad y_1$$

$$= \begin{bmatrix} \underline{Y}_2'\underline{Y}_2-\underline{V}'\underline{V} & \underline{Y}_2'\underline{X}_* \\ \underline{X}_*'\underline{Y}_2 & \underline{X}_*'\underline{X}_* \end{bmatrix}^{-1} \begin{bmatrix} \underline{Y}_2'-\underline{V}' \\ \underline{X}_*' \end{bmatrix} \quad y_1$$

The need for the inequality $K \geq H\text{-}1 + K^*$ is now apparent. The matrix $\underline{A}$ defined above is of order T by $H\text{-}1 + K^*$. Hence $\underline{A}'\underline{A}$ is a square symmetric matrix of order $H\text{-}1 + K^*$ and $\rho(\underline{A}'\underline{A}) = \rho(\underline{A})$. Now

$$\underline{A} = [(\underline{Y}_2 - \underline{V})\ \underline{X}_*] = \underline{X}\ [(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}_2\ \begin{matrix} I \\ \underline{0} \end{matrix}] ,$$

where I is the identity matrix of order K^* and $\underline{0}$ is the null matrix of order $K\text{-}K^*$ by K^*. Thus the rank of $\underline{A}$ cannot be greater than the rank of $\underline{X}$, which is K. If the rank of $\underline{A}$ is less than $H\text{-}1 + K^*$, then $A'A$ is singular and no solution can be obtained. This will happen if $K < H\text{-}1 + K^*$. However, since the inequality $K \geq H\text{-}1 + K^*$ was originally assumed to hold, a solution is assured.

The two-stage least square method of estimation may be justified in the following heuristic manner. If the parent reduced-form random error distribution corresponding to $\underline{Y}$, say $\underline{\overline{V}}$, were known; simple least squares could be applied to

$$y = (\underline{Y}_2 - \underline{\overline{V}})\ \underline{\beta}_2' + X_*\ \gamma_{1*}' + (e - \overline{V}\beta_2').$$

In this case the objection that some of the right-hand variables are not independent of the random error vector is no longer valid; $(\underline{Y} - \underline{\overline{V}})$ being an exact linear function of $\underline{X}$ and hence non-stochastic. The matrix $\underline{\overline{V}}$ is not known, but it can be estimated by means of V, and the sampling error tends to zero for increasing T under appropriate conditions. The primary conditions being that the assumptions made previously about the random error vector are true. Applying least squares as before the estimates are again:

$$\begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_{1*}' \end{bmatrix} = \begin{bmatrix} \underline{Y}'\underline{Y}_2 - \underline{V}'\underline{V} & \underline{Y}_2'\underline{X}_* \\ \underline{X}_*'\underline{Y}_2 & \underline{X}_*'\underline{X}_* \end{bmatrix}^{-1} \begin{bmatrix} (\underline{Y}_2 - \underline{V})' \; y_1 \\ \underline{X}_*' y_1 \end{bmatrix}.$$

PROPERTIES OF THE ESTIMATORS

By defining the sampling error as

$$\underline{\mu} = \begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_{1*}' \end{bmatrix} - \begin{bmatrix} \beta_2' \\ \gamma_{1*}' \end{bmatrix},$$

and using the above equations it is apparent

$$\underline{\mu} = \begin{bmatrix} \underline{Y}_2'\underline{Y}_2 - \underline{V}'\underline{V} & \underline{Y}_2'\underline{X}_* \\ \underline{X}_*'\underline{Y}_2 & \underline{X}_*'\underline{X}_* \end{bmatrix}^{-1} \begin{bmatrix} \underline{Y}' - \underline{V}' \\ \underline{X}_*' \end{bmatrix} \underline{e} \, .$$

It is seen that, under the previous assumptions on $\underline{e}$ and $\underline{X}$, the estimator is not unbiased for finite samples. However, it is asymptotically unbiased, $\lim_{T\to\infty} E\,(\underline{\mu}) = \underline{0}$, provided that each row of $\underline{Y}-\underline{V}$ is asymptotically an exact and non-stochastic linear function of the corresponding row of $\underline{X}$. This involves the assumption of consistent reduced-form estimation; hence: For each pair t, t' (=1,---,T) and for each pair z, z' (=1,...,G), the parent reduced-form random errors $\bar{v}_{z'}(t)$ and $\bar{v}_{z'}(t')$ corresponding to the right hand variables y_z and $y_{z'}$, respectfully, of equation (1) have zero mean and satisfy

$$E[\bar{v}_z(t)\,\bar{v}_{z'}(t')] = \sigma_{zz'} \text{ if } t=t'$$

$$= 0 \quad \text{if } t \neq t',$$

$\sigma_{zz'}$ being independent of t and t′.

This assumption is satisfied as soon as each of the G original equations of the system have random errors that satisfy a similar condition.

In order to calculate the asymptotic standard errors, it is necessary to find

$$\lim_{T\to\infty} \quad E(T\mu\mu') \quad =$$

$$\lim_{T\to\infty} \left\{ T \begin{bmatrix} \underline{Y}_2'\underline{Y}_2-\underline{V}'\underline{V} & \underline{Y}_2'\underline{X}_* \\ \underline{X}_*'\underline{Y}_2 & \underline{X}_*'\underline{X}_* \end{bmatrix}^{-1} \begin{bmatrix} \underline{Y}_2'-\underline{V}' \\ \underline{X}_*' \end{bmatrix} \underline{e}\ \underline{e} \begin{bmatrix} \underline{Y}_2-\underline{V} \\ \underline{X}_* \end{bmatrix} \begin{bmatrix} \underline{Y}_2'\underline{Y}_2-\underline{V}'\underline{V} & \underline{Y}_2'\underline{X}_* \\ \underline{X}_*'\underline{Y}_2 & \underline{X}_*'\underline{X}_* \end{bmatrix}^{-1} \right.$$

Since $(\underline{e}\ \underline{e}')$ equals σ^2

$$\text{and} \quad \begin{bmatrix} \underline{Y}_2'-\underline{V}' \\ \underline{X}_*' \end{bmatrix} \begin{bmatrix} \underline{Y}-\underline{V} \\ \underline{X}_* \end{bmatrix} \begin{bmatrix} Y_2^2Y-V'V & \underline{Y}_2'\underline{X}_* \\ \underline{X}_*'\underline{Y}_2 & \underline{X}_*'\underline{X}_* \end{bmatrix}^{-1'} = I$$

and the result is

$$\sigma^2 \operatorname*{plim}_{T\to\infty} T \begin{bmatrix} \underline{Y}_2'\underline{Y}_2-\underline{V}'\underline{V} & \underline{Y}_2'\underline{X}_* \\ \underline{X}_*'\underline{Y}_2 & \underline{X}_*'\underline{X}_* \end{bmatrix}^{-1}$$

Thiel (1958) has shown that although two-stage least squares has a larger variance than ordinary squares; if the bias of ordinary least squares is corrected, two-stage least squares becomes the smaller. In fact, he has shown his method to be a minimum variance unbiased estimator under the given assumptions. Basman (1957) in his development of the method shows the estimators are best linear unbiased in addition to being consistent.

GEOMETRIC INTERPRETATION

In order to further clarify the method of two-stage least squares the following geometric picture is presented. Consider a T dimensional cartesian space; along the first axis measure the first observation on each variable, along the second axis the second observation, etc. The values assumed by each variable are then represented by a point in this space; or alternatively by a vector from the origin 0 to this point. This leads to one point Y_1 corresponding to the left hand dependent variable of the equation; to G-1 points $Y_2,\ldots,Y_G$ corresponding to the right hand dependent variables; and to K points $X_1,\ldots,X_{K*},\ldots,X_K$ corresponding to the predetermined variables.

The first stage of two-stage least squares amounts to replacing $Y_2,\ldots,Y_G$ by their reduced values. This gives G-1 points $Y_{2*},\ldots,Y_{G*}$ which are the projections of $Y_2,\ldots,Y_G$ respectively, in the K dimensional plane determined by the K + 1 points 0, $X_1,\ldots,X_K$. In Fig. 1 this is illustrated for the case T=3, G=2, K=2.

The second step is the application of ordinary least squares with the left hand Y, as dependent variable and the reduced right hand Y's and K* of the X's as independent variables. This implies projecting Y_1 onto the (G-1 + K*) dimensional plane determined by 0, $Y_{2*}, \ldots, Y_{G*}$, $X_1, \ldots, X_{K*}$, which leads to a point y_{1*}. After this, the decomposition of the vector Oy_{*1} in terms of the vectors $OY_{1*}, \ldots, OX_{K*}$ gives the estimated coefficients according to two-stage least squares.

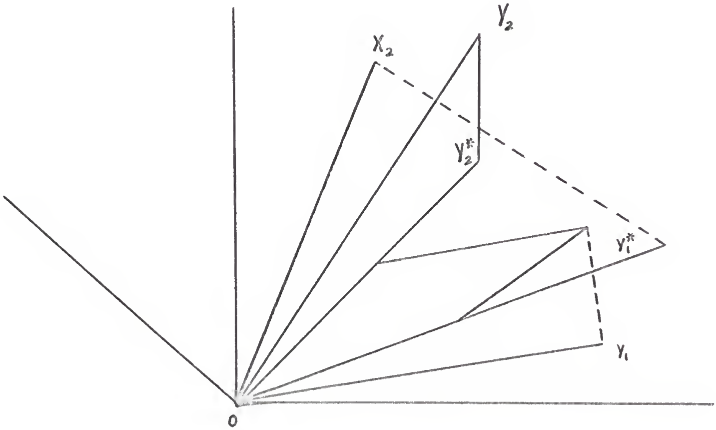

Figure 1. GEOMETRICAL ILLUSTRATION OF TWO-STAGE LEAST SQUARES

AN EVALUATION OF THE METHOD

While the two-stage least squares estimator may be shown to be asymptotically unbiased, consistent and minimum variance, this does not give an accurate picture of their performance when working with small samples of data. The advent of the computer has made it possible to conduct Monte Carlo studies on the small sample properties of the estimators in relation to other available methods. The sample size was generally chosen in the range of 15 to 40, to reflect the sample sizes which are typically found in practice.

The Monte Carlo studies by Basmann (1961), Nagar (1960), Summers (1965), and Wagner (1958) give insight on the choice of the best method of estimation for structural parameters under the restriction of small sample sizes. The evidence from these studies appears to indicate that the full-information maximum-likelyhood method is the best available. However, it has serious disadvantages. The computational burden is very heavy and the optional properties of the estimator depend heavily upon the correctness of the specification of the model. In light of these disadvantages, this method is not considered to be of practical use.

Of the remaining methods available two-stage least squares becomes the best method which can be practically applied to a system of linear equations. While the Basmann study shows the method to be superior by a more pronounced margin, all of the studies indicate the preferability of two-stage least squares.

A criticism of two-stage least squares given by G.C Chow (1964) is that the choice of a dependent variable, say Y_1, for the first equation, etc., in the second stage seems arbitrary. The estimates will differ according to the choice made. In other words, it has not yet been specified in which directions the sum of squares should be minimized in the second stage. Also a second criticism is that the method does not adequately take into account the interdependence of the e_i in different equations.

While it is generally conceded that this method is not the ultimate for finding the estimates under all conditions, it is the most universally applicable method and certainly the shortest computationally for the validity of the results. Until a better method is developed it is certain more and more applications will be found for two-stage least squares.

EXAMPLE

The following illustrates how to find the estimates of coefficients using the two-stage least square method. The model used in that of the Girshick-Haavelmo economic model composed of the following structural equations:

(1) $y_{1t} = \beta_{12} y_{2t} + \beta_{13} y_{3t} + \gamma_{18} x_{8t} + \gamma_{19} x_{9t} + \gamma_{10} + e_{1t}$

(2) $y_{2t} = \beta_{22} y_{2t} + \beta_{24} y_{4t} + \gamma_{28} x_{8t} + \gamma_{20} + e_{2t}$

(3) $y_{3t} = \gamma_{37} x_{7t} + \gamma_{39} x_{9t} + \gamma_{30} e_{3t}$

(4) $y_{4t} = \beta_{45} y_{5t} + \gamma_{46} x_{6t} + \gamma_{48} x_{8t} + \gamma_{40} + e_{4t}$

(5) $y_{5t} = \beta_{52} y_{2t} + \gamma_{58} x_{8t} + \gamma_{50} + e_{5t}$.

$t = 1,2, \ldots 20$ sample observations,

where y_{it}, $i = 1,\ldots,5$, denote the endogenous variables

x_{jt}, $j = 6,\ldots,9$, denote the exogenous variables

β_{ij}, $i,j = 1,\ldots,5$, denote the coefficients of endogenous variables

γ_{ik}, $k = 6,\ldots,9$, denote coefficients of exogenous variables

γ_{io} denotes the intercept of the i^{th} equation

It is assumed the endogenous variables are jointly distributed according to the following reduced form equations:

(6) $y_{it} = \Pi_{i6}x_{6t} + \ldots + \Pi_{9}x_{9t} + \Pi_{io} + \eta_{it}$

$i = 1,\ldots,5$

GENERALIZED CLASSICAL ESTIMATES

TABLE I*

(Table continued on following page)

TABLE I*

Data Used in This Study

Year	y_1	y_2	y_3	y_4	y_5	y_6	x_7	x_8	x_9
1922	98.6	100.2	87.4	108.5	99.1	98.0	92.9	1	77.4
1923	101.2	101.6	97.6	110.1	99.1	99.1	142.9	2	87.4
1924	102.4	100.5	96.7	100.4	98.9	99.1	100.0	3	97.6
1925	100.9	106.0	98.2	104.3	110.8	98.9	123.8	4	96.7
1926	102.3	108.7	99.8	107.2	108.2	110.8	111.9	5	98.2
1927	101.5	106.7	100.5	105.8	105.6	108.2	121.4	6	99.8
1928	101.6	106.7	103.2	107.8	109.8	105.6	107.1	7	100.5
1929	101.6	108.2	107.8	103.4	108.7	109.8	142.9	8	103.2
1930	99.8	105.5	96.6	102.7	100.6	108.7	92.9	9	107.8
1931	100.3	95.6	88.9	104.1	81.0	100.6	97.6	10	96.6
1932	97.6	88.6	75.1	99.2	68.6	81.0	52.4	11	88.9
1933	97.2	91.0	76.9	99.7	70.9	68.6	40.5	12	75.1
1934	97.3	97.9	84.6	102.0	81.4	70.9	64.3	13	76.9
1935	96.0	102.3	90.6	94.3	102.3	81.4	78.6	14	84.6
1936	99.2	102.2	103.1	97.7	105.0	102.3	114.3	15	90.6
1937	100.3	102.5	105.1	101.1	110.5	105.0	121.4	16	103.1
1938	100.3	97.0	96.4	102.3	92.5	110.5	78.6	17	105.1
1939	104.1	95.8	104.4	104.4	89.3	92.5	109.5	18	96.4
1940	105.3	96.4	110.7	108.5	93.0	89.3	128.6	19	104.4
1941	107.6	100.3	127.1	111.3	106.6	93.0	238.1	20	110.7
Sum	2015.1	2013.7	1950.7	2084.8	1941.9	1933.3	2159.7	210	1901.0
Mean	100.755	100.685	97.535	104.240	97.095	96.665	107.985	10.500	95.050

TABLE I (continued)

TABLE I* (continued)

* The following series were used for the model:

y_1 is food consumption per capita published by the Bureau of Agricultural Economics. (An adjustment has been made in the official series for 1934 to exclude the quantity of meat purchased by the Government for relief purposes and distributed through noncommercial channels.)

y_2 is retail prices of food products (BAE), deflated by the Index of Consumer Prices for Moderate Income Families in Cities, published by the Bureau of Labor Statistics.

y_3 is disposable income per capita (Dept. of Commerce), deflated by the BLS Consumer Price Index.

y_4 is production of agricultural food products per capita (BAE).

y_5 is prices received by farmers for food products (BAE), deflated by BLS Consumer Price Index.

$x_6 = y_{5,t-1}$ is prices received by farmers for food products, lagged one year.

x_7 is net investment per capita, i.e., disposable income minus consumers' expenditures, based on Dep. of Commerce data, deflated by BLS Consumer Price Index.

$x_8 = t$ is Time.

$x_9 = y_{3,t-1}$ is disposable income per capita lagged one year.

All the data are expressed in terms of index numbers (1935-39 = 100) except for time, x_8, which has the values, 1,2,...,20. The analysis covers the period 1922 through 1941.

(a) Preliminary Computations

Compute:

$$M_{yy} = \sum_{t=1}^{20} (y_{it} - \bar{y}_i)(y_{it} - \bar{y}_i) \qquad i=1,2,4,5$$

where

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T=20} y_{it}$$

This will form a matrix of sums of squares and crossproducts of deviations over all the y's appearing in all the equations to be estimated. In this example it is required to estimate only equations (2), (4) and (5). Since y_{3t} does not appear in any of these, the sum of squares and crossproducts need not be taken.

From the data in Table I:

$$M_{yy} = \begin{bmatrix} 151.7295 & 62.4765 & 180.8860 & 226.1155 \\ & 583.2285 & 108.8320 & 1231.2685 \\ & & 391.6480 & 320.2040 \\ & & & 3164.9495 \end{bmatrix}$$

Compute:

$$M_{yx} = \sum_{t=1}^{20} (y_{jt} - \bar{y}_j)(x_{kt} - \bar{x}_K) \qquad \begin{matrix} j=1,2,4,5 \\ K=6,7,8,9 \end{matrix}$$

$$M_{yx} = \begin{bmatrix} 257.7885 & 1793.9165 & 72.2500 & 401.6250 \\ 920.2995 & 1870.0155 & -257.7500 & 430.5750 \\ 364.6480 & 2073.5520 & -172.7000 & 306.1900 \\ 2169.4165 & 6297.5185 & -306.8500 & 1290.2350 \end{bmatrix}$$

Compute:

$$M_{xx} = \sum_{t=1}^{20} (x_{kt} - \bar{x}_K)(x_{nt} - \bar{x}_n) \qquad k,n=6,7,8,9$$

$$M_{xx} = \begin{bmatrix} 3071.7255 & 3963.8095 & 415.2500 & 1714.9250 \\ & 32367.3055 & 658.1500 & 4956.0350 \\ & & 665.0000 & 317.7000 \\ & & & 2067.0700 \end{bmatrix}$$

Compute by means of the forward Doolittle method, the matrix

$$M^*_{yy} = M_{yx} \cdot M^{-1}_{xx} \cdot M'_{yx}$$

(b) Computation for a Single Structural equation.

For expository purposes suppose it is required to compute estimates of the coefficients β_{ij} and γ_{ik} appearing in the structural equation.

$$(7) \quad -y_{1t} + \sum_{j=2}^{G} y_{jt}\beta_{1j} + \sum_{k=1}^{K} X_{kt}\gamma_{1k} + e_{1t} = 0$$

The variable y_1 appears with coefficient $\beta_{11} = -1$. Let y_Δ denote the vector of $H - 1 \leq G$ endogenous variables (except y_1) appearing in (7) with non-zero coefficients, and let x_Δ denote the vector of $K^* \leq K$ exogenous variables appearing in (7) with non-zero coefficients. It is assumed that necessary condition for the identification of (7) are met; i.e. $K \geq K^* + H$.

Define:

$\underline{M_{y\Delta x\Delta}}$ as the submatrix of M_{yx} involving only the sums of cross products of endogenous and exogenous variables appearing with non-zero coefficients in (7).

$\underline{M_{x\Delta x\Delta}}$ as the submatrix of M_{yx} involving only the sums of squares and cross products of the exogenous variables appearing with non-zero coefficients in (7).

$\underline{M^*_{y\Delta y\Delta}}$ as the submatrix of M^*_{yy} corresponding to the endogenous variables appearing among the y .

$\underline{M^*_{y1y\Delta}}$ as the 1 x(H-1) submatrix of M^*_{yy} corresponding to y_1 and y.

$\underline{M_{y1x\Delta}}$ as the 1 x K* submatrix of M_{yx} involving only the sums of cross products of y_1 and the endogenous variables appearing with non-zero coefficients in (7).

$\underline{M_{y1y\Delta}}$ as the corresponding submatrix of M_{yy}.

Form the compound matrices:

$$S_\Delta = \left[\begin{array}{c|c} M^*_{y\Delta y\Delta} & M_{y\Delta x\Delta} \\ \hline M_{y\Delta x\Delta} & M_{x\Delta x\Delta} \end{array}\right]$$

which is square of order H-1 + K* and

$$M_{y1z} = [M^*_{y1y\Delta} \quad M_{y1x\Delta}] \quad , \quad 1 \text{ x } (H-1 + K^*)$$

Arrange the above matrices for Doolittle computation

$$[S_\Delta][M'_{y1z}][I_{H-1+K^*}]$$

and compute the vector of sample estimates, (b_Δ, c_Δ), of non-zero coefficients β_Δ, and γ_Δ appearing in (7), where

$$\begin{bmatrix} b_\Delta \\ c_\Delta \end{bmatrix} = S_\Delta^{-1} M_{y1z}$$

Compute the sample variance $\hat{\omega}_{11}$ of the residual e_{11} in equation (7):

$$\hat{\omega} = \frac{1}{T^*}\{M_{y1}^2 - 2\, M'_{y\Delta y1}\, b_\Delta + 2\, M_{y1x\Delta}\, M_{x\Delta x\Delta}^{-1}\, M'_{y\Delta x\Delta b\Delta}$$

$$+ b'\, [M_{y\Delta y\Delta} - M_{y\Delta x\Delta}\, M_{x\Delta x\Delta}^{-1}\, M'_{y\Delta x\Delta}]\, b_\Delta\}$$

where $T^* = T - K - 1$.

Compute the sample variance-covariance matrix of the estimated coefficients b_Δ and c_Δ

$$\widehat{\text{Var - Cov}}\,(b_\Delta,\, c_\Delta) = \omega_{11}\, S^{-1}$$

To illustrate the numerical application of the steps outlined in Section (b), the method is applied to Equation 2. The appropriate Doolittle layout is exhibited in Table II.

TABLE II

$$\begin{array}{c} y_2 \\ y_4 \\ x_8 \end{array} \begin{bmatrix} 342.036418 & 201.496935 & -257.7500 \\ & 221.303177 & 172.7000 \\ & & 665.0000 \end{bmatrix} \begin{bmatrix} 97.187553 \\ 115.531281 \\ 72.2500 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Results of Doolittle computations applied to Table II are exhibited in Table III along with the corresponding estimated standard deviation given in parenthesis. The sample variances of residuals have been computed according to Section (b) and are exhibited in the last column.

TABLE III

(Table continued on following page)

TABLE III

ESTIMATES OF EQUATION (2)

y_2	y_4	x_8	ω_{22}
0.1633	0.6366	0.3372	1.3873
(.0997)	(.1168)	(.0545)	

ACKNOWLEDGMENT

The writer wishes to express his sincere appreciation to his major professor, DR. Y.O. KOH, for suggesting this topic and for his advice and assistance during the preparation of this report.

The writer would also like to express his graditude to DR. A.M. FEYERHERM and DR. G.V.L. NARASIMHAM for their assistance in the literature review.

REFERENCES

Basmann, R.L. (1957)
"A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," *Econometrica*, *25*, 77-83.

Basmann, R.L. (1959)
"A Computation of Generalized Classical Estimates of Coefficients in a Structural Equation," *Econometrica*, *27*, 72-81.

Basmann, R.L. (1963)
"A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Over Identified Cases," *J. Am. Statist. Assoc.*, 56, 619-636.

Chow, Gregory C. (1964)
"A Comparison of Alternative Estimators for Simultaneous Equations," *Econometrica*, *32*, 532-553.

Johnston, J. (1963)
"*Econometric Methods*," 231-273, McGraw-Hill Book Company, Inc.

Klein, L.R. (1956)
"On the Interpretation of Theil's Method of Estimating Economic Relationships," *Metroeconomica*, *7*, 145-153.

Nagar, A.L. (1960)
"A Monte Carlo Study of Alternative Simultaneous Equation Estimators," *Econometrica*, *28*, 573-590.

Summers, Robert (1965)
"A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators," *Econometrica*, *33*, 1-41.

Theil, H. (1958)
"*Economic Forecasts and Policy*," 223-227, 336-360, North-Holland Publishing Company.

Wagner, H. (1958)
"A Monte Carlo Study of Estimates of Simultaneous Linear Structural Equations," *Econometrica*, *26*, 117-133.

THE TWO-STAGE LEAST SQUARES METHOD OF ESTIMATION

by

ALBERT CHARLES KIENTZ, JR.

B.S., KANSAS STATE UNIVERSITY, 1962

AN ABSTRACT OF A MASTER'S REPORT

Submitted in Partial Fulfillment of the

Requirements for the Degree

MASTER OF SCIENCE

Department of Statistics and Computer Science

KANSAS STATE UNIVERSITY
MANHATTAN, KANSAS

1968

The two-stage least squares method of linear estimation of coefficients was developed to replace existing methods by providing a method of more general applicability while being less expensive to apply. The basic idea of two-stage least squares is to apply simple least squares to the entire system of predetermined variables to obtain estimates for the dependent variables. Then, with these estimates substituted in the system, simple least squares may be applied again to a particular equation or a set of equations of the system to estimate the required parameters.

A typical set of equations of a linear system containing G dependent variables, y_{it} and K predetermined variables, X_{it}, may be written as

$$y_{1t} = \sum_{i=2}^{H} \beta_{1i}\, y_{it} + \sum_{i=1}^{K^*} \gamma_{1i}\, X_{it} + e_{1t}, \quad t=1,\dots,T$$

where $H-1 \leq G$; $K^* \leq K$

In addition to the usual assumptions on the random error vector, the assumption of $K \geq K^* + H-1$ will assure a solution of the system by the two-stage least squares method.

By successive application of simple least squares the parameter estimates are

$$\begin{bmatrix} \hat{\beta}_2' \\ \hat{\gamma}_{1*}' \end{bmatrix} = \begin{bmatrix} Y_2'Y_2 - V'V & Y_2'X_* \\ X_*'Y_2 & X_*'X_* \end{bmatrix}^{-1} \begin{bmatrix} (Y_2 - V)' y_1 \\ X_*'y_1 \end{bmatrix}$$

Where V denotes the matrix of reduced-form residuals found in the first application of simple least squares. These estimators are asymptotically unbiased, consistent and minimum-variance.

Several Monte Carlo studies have been made to determine the performance of the two-stage least square method when working with small samples of data. The studies made to date indicate two-stage least squares is the best available method which can be applied practically to a system of linear equations. While it is generally conceded that this method is not the ultimate for finding the estimates under all conditions, it is the most universally applicable method and certainly the shortest computationally for the validity of the results.