Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations

Author(s): Arnold Zellner and H. Theil

Source: *Econometrica*, Jan., 1962, Vol. 30, No. 1 (Jan., 1962), pp. 54-78

Published by: The Econometric Society

Stable URL: https://www.jstor.org/stable/1911287

Econometrica, Vol. 30, No. 1 (January, 1962)

THREE-STAGE LEAST SQUARES: SIMULTANEOUS ESTIMATION OF SIMULTANEOUS EQUATIONS

By Arnold Zellner and H. Theil

1. Introduction
2. The Three-Stage Least Squares Method
   2.1. Description of the System
   2.2. Two-Stage Least Squares Applied to a Single Equation
   2.3. Three-Stage Least Squares Applied to a Complete System
3. The Asymptotic Moment Matrix of Three-Stage Least Squares
   3.1. Derivation of the Moment Matrix
   3.2. What Do We Gain?
4. Problems of Identification and Related Issues
   4.1. Analysis
   4.2. Summary of Rules
5. An Example; Directory of Computations
   5.1. Some Computational Simplifications
   5.2. Klein's Model I
   5.3. Computations
6. Concluding Remarks

1. INTRODUCTION

In simple though approximate terms, the two-stage least squares method of estimating a structural equation consists of two steps, the first of which serves to estimate the moment matrix of the reduced-form disturbances and the second to estimate the coefficients of one single structural equation after its jointly dependent variables are "purified" by means of the moment matrix just mentioned. The three-stage least squares method, which is developed in this paper, goes one step further by using the two-stage least squares estimated moment matrix of the structural disturbances to estimate all coefficients of the entire system simultaneously. The method has full-information characteristics to the extent that, if the moment matrix of the structural disturbances is not diagonal (that is, if the structural disturbances have nonzero "contemporaneous" covariances), the estimation of the coefficients of any identifiable equation gains in efficiency as soon as there are other equations that are over-identified. Further, the method can take account of restrictions on parameters in different structural equations. And it is very simple computationally, apart from the inversion of one big matrix.

The order of discussion is as follows. The essential ideas of the method are explained in Section 2, after which the asymptotic moment matrix of the estimator is derived in Section 3. When there are just-identified equations in the system, the computations can be simplified; this is explained in Section 4. Section 5 contains a numerical example as well as a directory of computations.

Section 6 gives some concluding remarks. The reader who is interested in results only is advised to read Section 4.2 immediately after Section 2, and then to proceed to Section 5.

2. THE THREE-STAGE LEAST SQUARES METHOD

2.1. *Description of the System.* It will be assumed throughout that we are dealing with a complete system of M linear stochastic structural equations in M jointly dependent variables and Λ predetermined variables; furthermore, that this system can be solved for the jointly dependent variables (so that the reduced form exists); finally, that the disturbances of the structural equations (the structural disturbances) have zero mean, are serially indepedent, and are "homoscedastic" in the sense that their variances and "contemporaneous" covariances are finite and constant through time. This contemporaneous covariance matrix will be supposed to be nonsingular.[1]

Let T be the number of observations. Then any structural equation, say the μth, can be written in the following form for all observations combined:

$$y_\mu = Y_\mu \gamma_\mu + X_\mu \beta_\mu + u_\mu = Z_\mu \delta_\mu + u_\mu \,. \tag{2.1}$$

where y_μ is the column vector of observations on one of the jointly dependent variables occurring in that equation (usually we write the variable "to be explained" on the left); Y_μ is the $T \times m_\mu$ matrix of values taken by the explanatory dependent variables of that equation; γ_μ is the corresponding coefficient vector; X_μ is the $T \times l_\mu$ matrix of values taken by the explanatory predetermined variables; β_μ is its coefficient vector; u_μ is the column vector of T structural disturbances; and

$$Z_\mu = [Y_\mu \quad X_\mu] \; ; \qquad \delta_\mu = \begin{bmatrix} \gamma_\mu \\ \beta_\mu \end{bmatrix} . \tag{2.2}$$

Further, we write X for the $T \times \Lambda$ matrix of values taken by all (Λ) predetermined variables, and we shall suppose that its rank is Λ. Our objective is to estimate the parameter vectors δ_μ, and for this purpose it will be supposed that all equations are identifiable.[1] This implies

$$\Lambda \geqslant n_\mu = m_\mu + l_\mu \qquad (\mu = 1, \ldots, M) \,, \tag{2.3}$$

where n_μ is the total number of coefficients to be estimated in the μth equation.

2.2. *Two-Stage Least Squares Applied to a Single Equation.* Premultiplying (2.1) by X', we obtain

$$X' y_\mu = X' Z_\mu \delta_\mu + X' u_\mu \,, \tag{2.4}$$

[1] See Section 4 for a generalization.

which is a system of Λ equations involving n_μ parameters (δ_μ) and a "disturbance vector" ($X'u_\mu$) with zero mean. In the special case $\Lambda = n_\mu$ (just-identification) it is customary to estimate δ_μ according to

$$d_\mu = (X'Z_\mu)^{-1}X'y_\mu \; ; \tag{2.5}$$

that is, one replaces δ_μ by its estimator d_μ in (2.4), and simultaneously $X'u_\mu$ by its expectation.[2] In the more usual case where $\Lambda > n_\mu$, we can also use (2.4) but in a slightly more complicated manner. Assuming that the predetermined variables are all "fixed" variables, we find for the covariance matrix of the disturbance vector $X'u_\mu$:

$$V(X'u_\mu) = E(X'u_\mu u'_\mu X) = \sigma_{\mu\mu}X'X \; , \tag{2.6}$$

where $\sigma_{\mu\mu}$ is the variance of each of the T disturbances of the μth structural equation. Then, applying Aitken's method of generalized least squares to (2.4),[3] we obtain

$$Z'_\mu X(\sigma_{\mu\mu}X'X)^{-1}X'y_\mu = Z'_\mu X(\sigma_{\mu\mu}X'X)^{-1}X'Z_\mu d_\mu \; , \tag{2.7}$$

from which we derive the two-stage least squares estimator

$$d_\mu = [Z'_\mu X(X'X)^{-1}X'Z_\mu]^{-1}Z'_\mu X(X'X)^{-1}X'y_\mu \; , \tag{2.8}$$

which reduces to (2.5) in the special case when $X'Z_\mu$ is square and nonsingular. The covariance matrix of d_μ is

$$V(d_\mu) = \sigma_{\mu\mu}[Z'_\mu X(X'X)^{-1}X'Z_\mu]^{-1} + o(1/T) \; , \tag{2.9}$$

where $o(1/T)$ stands for terms of higher order of smallness than $1/T$.

2.3. *Three-Stage Least Squares Applied to a Complete System.* The crucial

[2] It is perhaps useful to add that (2.5) is a far simpler method of computing the point estimates of the coefficients in the just-identified case than the method which is sometimes recommended and which involves estimation of reduced-form coefficients, followed by transformations from the reduced form to the structure. The only thing one has to do is to compute the second-order moments of the predetermined variables of the system on the one hand and the variables of the equation on the other, followed by the inversion of $X'Z_\mu = [X'Y_\mu \quad X'X_\mu]$ and postmultiplication by $X'y_\mu$.

[3] Aitken's method [1] works as follows. Let $r = R\theta + \varepsilon$ be the equation whose coefficient vector θ has to be estimated, r being a column vector of observations on the dependent variable, R the matrix of values taken by the explanatory variables, and ε the disturbance vector. Let $E(\varepsilon) = 0$ and $V(\varepsilon) = \Omega$ where Ω is a positive-definite matrix, than $\hat{\theta} = (R'\Omega^{-1}R)^{-1}R'\Omega^{-1}r$ is the generalized least squares estimator of θ according to Aitken, and $V(\hat{\theta}) = (R'\Omega^{-1}R)^{-1}$ if the elements of R are fixed. In the present case $r = X'y_\mu$, $R = X'Z_\mu$, $\theta = \delta_\mu$, $\varepsilon = X'u_\mu$, $\Omega = \sigma_{\mu\mu}X'X$. It is not true here that the elements of $R = X'Z_\mu$ are fixed (due to the presence of Y_μ in Z_μ); which is the reason why $(R'\Omega^{-1}R)^{-1}$ yields only the asymptotic covariance matrix (see (2.9)).

idea of three-stage least squares is that (2.4) can be written in the following form for all equations combined:

$$(2.10)\qquad \begin{bmatrix} X'y_1 \\ X'y_2 \\ \vdots \\ X'y_M \end{bmatrix} = \begin{bmatrix} X'Z_1 & 0 & \dots & 0 \\ 0 & X'Z_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & X'Z_M \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{bmatrix} + \begin{bmatrix} X'u_1 \\ X'u_2 \\ \vdots \\ X'u_M \end{bmatrix},$$

which is a system of ΛM equations involving

$$(2.11)\qquad n = \sum_{\mu=1}^{M} n_\mu$$

parameters. Let us write δ for the n-element column vector of parameters on the right of (2.10). Then we can apply generalized least squares to (2.10) to estimate all elements of δ simultaneously. For this purpose we need the covariance matrix of the disturbance vector of (2.10):

$$(2.12)\qquad V\begin{bmatrix} X'u_1 \\ X'u_2 \\ \vdots \\ X'u_M \end{bmatrix} = \begin{bmatrix} \sigma_{11}X'X & \sigma_{12}X'X & \dots & \sigma_{1M}X'X \\ \sigma_{21}X'X & \sigma_{22}X'X & \dots & \sigma_{2M}X'X \\ \vdots & \vdots & & \vdots \\ \sigma_{M1}X'X & \sigma_{M2}X'X & \dots & \sigma_{MM}X'X \end{bmatrix},$$

where $\sigma_{\mu\mu'}$ is the contemporaneous covariance of the structural disturbances of the μth and the μ'th equation:

$$(2.13)\qquad E(u_\mu u'_{\mu'}) = \begin{bmatrix} \sigma_{\mu\mu'} & 0 & \dots & 0 \\ 0 & \sigma_{\mu\mu'} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma_{\mu\mu'} \end{bmatrix} = \sigma_{\mu\mu'} I\ ,$$

I being the unit matrix of order T. Also, we need the inverse of the covariance matrix (2.12):

$$(2.14)\qquad V^{-1}\begin{bmatrix} X'u_1 \\ X'u_2 \\ \vdots \\ X'u_M \end{bmatrix} = \begin{bmatrix} \sigma^{11}(X'X)^{-1} & \sigma^{12}(X'X)^{-1} & \dots & \sigma^{1M}(X'X)^{-1} \\ \sigma^{21}(X'X)^{-1} & \sigma^{22}(X'X)^{-1} & \dots & \sigma^{2M}(X'X)^{-1} \\ \vdots & \vdots & & \vdots \\ \sigma^{M1}(X'X)^{-1} & \sigma^{M2}(X'X)^{-1} & \dots & \sigma^{MM}(X'X)^{-1} \end{bmatrix},$$

where $\sigma^{\mu\mu'}$ is an element of the inverse of the contemporaneous covariance matrix of the structural disturbances:

$$(2.15)\qquad [\sigma^{\mu\mu'}] = [\sigma_{\mu\mu'}]^{-1}\ .$$

A straightforward application of generalized least squares gives then the following result: the two-stage column vector $Z'_\mu X(\sigma_{\mu\mu}X'X)^{-1}X'y_\mu$ on the left of (2.7) is replaced by

$$\begin{bmatrix} \sigma^{11}Z_1'X(X'X)^{-1}X'y_1 + \ldots + \sigma^{1M}Z_1'X(X'X)^{-1}X'y_M \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \sigma^{M1}Z_M'X(X'X)^{-1}X'y_1 + \ldots + \sigma^{MM}Z_M'X(X'X)^{-1}X'y_M \end{bmatrix};$$

and the two-stage $n_\mu \times n_\mu$ matrix $Z_\mu'X(\sigma_{\mu\mu}X'X)^{-1}X'Z_\mu$ on the right of (2.7) is replaced by the $n \times n$ matrix,

$$\begin{bmatrix} \sigma^{11}Z_1'X(X'X)^{-1}X'Z_1 & \ldots & \sigma^{1M}Z_1'X(X'X)^{-1}X'Z_M \\ \vdots & & \vdots \\ \sigma^{M1}Z_M'X(X'X)^{-1}X'Z_1 & \ldots & \sigma^{MM}Z_M'X(X'X)^{-1}X'Z_M \end{bmatrix}.$$

These matrices involve σ's which are generally unknown. We recommend that they be replaced by their two-stage least squares estimates, to be denoted by $s^{\mu\mu'}$. Thus, the three-stage least squares estimator is defined as

$$(2.16)\quad \hat{\delta} = \begin{bmatrix} s^{11}Z_1'X(X'X)^{-1}X'Z_1 & \ldots & s^{1M}Z_1'X(X'X)^{-1}X'Z_M \\ \vdots & & \vdots \\ s^{M1}Z_M'X(X'X)^{-1}X'Z_1 & \ldots & s^{MM}Z_M'X(X'X)^{-1}X'Z_M \end{bmatrix}^{-1} \times \begin{bmatrix} \sum s^{1\mu}Z_1'X(X'X)^{-1}X'y_\mu \\ \vdots \\ \sum s^{M\mu}Z_M'X(X'X)^{-1}X'y_\mu \end{bmatrix},$$

where the summations in the right hand column vector extend over $\mu = 1, \ldots, M$. As will be shown in Section 3, the moment matrix of $\hat{\delta}$ is

$$(2.17)\quad V(\hat{\delta}) = \begin{bmatrix} s^{11}Z_1'X(X'X)^{-1}X'Z_1 & \ldots & s^{1M}Z_1'X(X'X)^{-1}X'Z_M \\ \vdots & & \vdots \\ s^{M1}Z_M'X(X'X)^{-1}X'Z_1 & \ldots & s^{MM}Z_M'X(X'X)^{-1}X'Z_M \end{bmatrix}^{-1} + o(1/T).$$

It will be observed that there is a gain in asymptotic efficiency compared with two-stage least squares only if $[\sigma_{\mu\mu'}]$ is not diagonal; if it is diagonal, two-and three-stage least squares are identical, because the non-diagonal blocks of (2.17) are then zero and $\sigma^{\mu\mu} = 1/\sigma_{\mu\mu}$.

3. THE ASYMPTOTIC MOMENT MATRIX OF THREE-STAGE LEAST SQUARES

3.1. *Derivation of the Moment Matrix.* The three-stage least squares estimator $\hat{\delta}$ as defined in (2.16) differs in two respects from a straightforward application of Aitken's procedure. First, $\hat{\delta}$ is based on the $s^{\mu\mu'}$, which are the two-stage least squares estimates of the $\sigma^{\mu\mu'}$ which we would really need for this procedure. Secondly, some of the "explanatory variables" of the relation (2.10) to which this procedure is applied are correlated with the disturbances of that relation. This is so because each $X'Z_\mu$ contains a $X'Y_\mu$ as a submatrix, and the Y_μ of this product is correlated with the u's which enter into the disturbance vector of (2.10).

To handle these two difficulties we shall employ a special notation (which will be used in this section only). We shall write (2.10) and (2.12) in the form:

$$v = W\delta + \varepsilon\ ; \qquad E(\varepsilon\varepsilon') = \Omega\ , \tag{3.1}$$

where v is the left hand column vector of (2.10), W the $\Lambda M \times n$ matrix on the right, ε the disturbance vector of (2.10), and Ω its covariance matrix as specified in (2.12). It is easily verified that the elements Ω are of $O(T)$, and those of v and W are of $O(T)$ in probability. Since Ω is unknown due to the σ's on which it depends, it is replaced in the three-stage procedure by

$$\Omega_e = \Omega + \Delta_1\ , \tag{3.2}$$

where the Ω_e-elements are of $O(T)$ and those of Δ_1 are of $O(T)^{\frac{1}{2}}$ in probability. The last statement follows from the fact that the two-stage least squares estimates, $s^{\mu\mu'}$, on which Ω_e is based, differ from the corresponding $\sigma^{\mu\mu'}$ by an amount which is of $O(T^{-\frac{1}{2}})$ in probability. The $\sigma^{\mu\mu'}$ themselves being of $O(T^0) = O(1)$.

In the present notation, the three-stage least squares estimator takes the form[4]

$$\hat{\delta} = (W'\Omega_e^{-1}W)^{-1}W'\Omega_e^{-1}v\ ,$$

from which we see immediately that its sampling error is

$$\hat{\delta} - \delta = (W'\Omega_e^{-1}W)^{-1}W'\Omega_e^{-1}\varepsilon\ . \tag{3.3}$$

Applying (3.2), we find that the inverse Ω_e^{-1} can be written as

$$\Omega_e^{-1} = (\Omega + \Delta_1)^{-1} = \Omega^{-1} + \Delta_2\ , \tag{3.4}$$

where Δ_2 is of $O(T^{-1\frac{1}{2}})$ in probability because Ω is of $O(T)$ and Δ_1 is of $O(T^{\frac{1}{2}})$ in probability.

Next, we consider W, the matrix of values taken by the "explanatory variables" of (2.10). Each of its diagonal blocks ($X'Z_\mu$, say) can be decomposed into a part involving predetermined variables only ($X'X_\mu$) and a part containing jointly dependent variables ($X'Y_\mu$). Hence, if we write $\bar{W}$ for the conditional expectation of W (given the X-elements), and write

$$W = \bar{W} + \Delta_3\ , \tag{3.5}$$

[4] Note that $(W'C^{-1}W)\ W'C^{-1}v$ defines a linear class of estimators. With $C = I$, we have the two-stage least squares estimator, and, with $C = \Omega_e$, we obtain the three-stage least squares estimator which, as is shown below, is linear in v to within an error which is of $O(T^{-1})$ in probability.

then Δ_3 is a matrix which consists largely of zeros except that it contains submatrices in the diagonal blocks of the type X' multiplied by the matrix of reduced-form disturbances corresponding to Y_μ. Hence the Δ_3-elements are of $O(T^{\frac{1}{2}})$ in probability, as against the $\bar{W}$-elements which are of $O(T)$. Combining (3.4) and (3.5), we then find

$$W'\Omega_e^{-1} = (\bar{W} + \Delta_3)'(\Omega^{-1} + \Delta_2) = \bar{W}'\Omega^{-1} + \Delta_4 , \tag{3.6}$$

where $\bar{W}'\Omega^{-1}$ is of $O(1)$ and Δ_4 is of $O(T^{-\frac{1}{2}})$ in probability. Similarly:

$$W'\Omega_e^{-1}W = (\bar{W}'\Omega^{-1} + \Delta_4)(\bar{W} + \Delta_3) = \bar{W}'\Omega^{-1}\bar{W} + \Delta_5 ,$$

where $\bar{W}'\Omega^{-1}\bar{W}$ is of $O(T)$ and Δ_5 is of $O(T^{\frac{1}{2}})$ in probability. Hence:

$$(W'\Omega_e^{-1}W)^{-1} = (\bar{W}'\Omega^{-1}\bar{W} + \Delta_5)^{-1} = (\bar{W}'\Omega^{-1}\bar{W})^{-1} + \Delta_6 , \tag{3.7}$$

where $(\bar{W}'\Omega^{-1}\bar{W})^{-1}$ is of $O(T^{-1})$ and Δ_6 is of $O(T^{-1\frac{1}{2}})$ in probability. On combining (3.6) and (3.7) we find

$$\begin{aligned}(W'\Omega_e^{-1}W)^{-1}W'\Omega_e^{-1} &= [(\bar{W}'\Omega^{-1}\bar{W})^{-1} + \Delta_6](\bar{W}'\Omega^{-1} + \Delta_4)\\ &= (\bar{W}'\Omega^{-1}\bar{W})^{-1}\bar{W}'\Omega^{-1} + \Delta_7 ,\end{aligned}$$

where $(\bar{W}'\Omega^{-1}\bar{W})^{-1}\bar{W}'\Omega^{-1}$ is of $O(T^{-1})$ and Δ_7 is of $O(T^{-1\frac{1}{2}})$ in probability. On combining this with (3.3), we finally obtain the following expression for the sampling error of $\hat{\delta}$:

$$\hat{\delta} - \delta = (\bar{W}'\Omega^{-1}\bar{W})^{-1}\bar{W}'\Omega^{-1}\varepsilon + \Delta_7\varepsilon . \tag{3.8}$$

The proof of the covariance result (2.17) is now virtually complete, for the first term on the right of (3.8) is the leading term of the sampling error and of $O(T^{-\frac{1}{2}})$ in probability (due to the fact that ε is a vector of random variables which are of $O(T^{\frac{1}{2}})$ in probability, ε having zero mean and covariance matrix Ω which is of $O(T)$), and the second term is of $O(T^{-1})$ in probability.[5] If we then postmultiply the sampling error by its own transpose, the leading term is

$$(\bar{W}'\Omega^{-1}\bar{W})^{-1}\bar{W}'\Omega^{-1}\varepsilon\varepsilon'\Omega^{-1}\bar{W}(\bar{W}'\Omega^{-1}\bar{W})^{-1} ,$$

the expectation of which is $(\bar{W}'\Omega^{-1}\bar{W})^{-1}$. This is of $O(T^{-1})$ and all other terms are of higher order of smallness. Hence $(\bar{W}'\Omega^{-1}\bar{W})^{-1}$ is the asymptotic covariance matrix of $\hat{\delta}$. In (2.17) the expression $(W'\Omega_e^{-1}W)^{-1}$ has been in-

[5] Since $T^{\frac{1}{2}}(\bar{W}'\Omega^{-1}\bar{W})^{-1}\bar{W}'\Omega^{-1}\varepsilon$ is asymptotically normally distributed under rather general stochastic conditions and since plim $T^{\frac{1}{2}}\Delta_7\varepsilon = 0$, the asymptotic distribution of $T^{\frac{1}{2}}(\hat{\delta} - \delta)$ is normal; see the convergence theorem in [**2**, p. 254]. A similar argument can be employed to establish the asymptotic normality of the two-stage least squares estimator.

troduced as the leading term, but it follows immediately from (3.7) that these two matrices are identical to $O(T^{-1})$.[6]

3.2. *What Do We Gain?* The mere fact that two-stage least squares is not identical with three-stage least squares implies that the former method is less efficient. It is instructive, however, to obtain an explicit expression for the gain in efficiency obtained by the three-stage method. Since we have only an illustrative purpose in mind, we shall confine ourselves to a two-equation system, in which case the covariance matrix of three-stage least squares is equal to the inverse of

$$\begin{bmatrix} \sigma^{11}Z_1'X(X'X)^{-1}X'Z_1 & \sigma^{12}Z_1'X(X'X)^{-1}X'Z_2 \\ \sigma^{21}Z_2'X(X'X)^{-1}X'Z_1 & \sigma^{22}Z_2'X(X'X)^{-1}X'Z_2 \end{bmatrix} \tag{3.9}$$

(apart from higher-order terms), whereas the covariance matrix of the two-stage least squares estimator of the first equation is equal to the inverse of

$$(1/\sigma_{11})Z_1'X(X'X)^{-1}X'Z_1 \tag{3.10}$$

(again, apart from higher-order terms). Hence our objective is to compare the inverse of (3.10) with the leading $n_1 \times n_1$ submatrix of the inverse of (3.9), this submatrix being the moment matrix of the three-stage estimator of the coefficient vector of the first equation.

Since X has rank Λ, there exists a matrix K which is square of order and rank Λ such that $K'K = (X'X)^{-1}$. We can then write (3.9) and (3.10) in the form

$$\begin{bmatrix} \sigma^{11}A_1'A_1 & \sigma^{12}A_1'A_2 \\ \sigma^{21}A_2'A_1 & \sigma^{22}A_2'A_2 \end{bmatrix} \quad \text{and} \quad (1/\sigma_{11})A_1'A_1\,,$$

respectively, where $A_\mu = KX'Z_\mu$ for $\mu = 1, 2$. Hence A_1 is of order $\Lambda \times n_1$ and rank n_1, and A_2 of order $\Lambda \times n_2$ and rank n_2. Furthermore, we write the $\sigma_{\mu\mu'}$-matrix in the form

$$\begin{bmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

[6] Note that this proof is so straightforward only because the number of rows and columns of the fundamental matrices involved (W, Ω, and ε) is independent of T. If these numbers would depend on T, the situation becomes entirely different: the elements of a product such as AB then have an order of magnitude in T which is no longer determined by the orders of the A- and B-elements, but also by the number of columns of A (or rows of B), which then depends on T, too.

so that its inverse becomes

$$\begin{bmatrix} \dfrac{1}{\sigma_1^2(1-\varrho^2)} & \dfrac{-\varrho}{\sigma_1\sigma_2(1-\varrho^2)} \\ \dfrac{-\varrho}{\sigma_1\sigma_2(1-\varrho^2)} & \dfrac{1}{\sigma_2^2(1-\varrho^2)} \end{bmatrix}.$$

If we apply the notation of the preceding paragraph to (3.9) and (3.10), we observe, first, that (3.10) becomes $Q = (1/\sigma_1^2)A_1'A_1$. Hence, the moment matrix of two-stage least squares is $Q^{-1} = \sigma_1^2(A_1'A_1)^{-1}$. Secondly, we note that (3.9) becomes

$$\begin{bmatrix} \dfrac{1}{\sigma_1^2(1-\varrho^2)} A_1'A_1 & \dfrac{-\varrho}{\sigma_1\sigma_2(1-\varrho^2)} A_1'A_2 \\ \dfrac{-\varrho}{\sigma_1\sigma_2(1-\varrho^2)} A_2'A_1 & \dfrac{1}{\sigma_2^2(1-\varrho^2)} A_2'A_2 \end{bmatrix},$$

so that the leading $n_1 \times n_1$ submatrix of its inverse is itself the inverse of

$$(3.11) \qquad Q^* = \frac{1}{\sigma_1^2(1-\varrho^2)} A_1'A_1 - \frac{\varrho^2}{\sigma_1^2(1-\varrho^2)} A_1'A_2(A_2'A_2)^{-1}A_2'A_1 .$$

Thus, the three-stage moment matrix of the first equation is Q^{*-1}, whereas the two-stage moment matrix is $Q^{-1} = \sigma_1^2(A_1'A_1)^{-1}$.

Suppose now that the second equation is just-identified. This implies $\Lambda = n_2$, so that A_2 is square. Supposing that it is also nonsingular, we can write

$$A_1'A_2(A_2'A_2)^{-1}A_2'A_1 = A_1'A_2A_2^{-1}A_2'^{-1}A_2'A_1 = A_1'A_1 ,$$

from which it follows that (3.11) can be simplified to $Q^* = (1/\sigma_1^2)A_1'A_1 = Q$. In words, this means that if the second equation is just-identified, the three-stage moment matrix of the first equation is identical with its two-stage moment matrix. Hence we gain nothing for the first equation in this case.

But suppose that the second equation is over-identified, so that $\Lambda > n_2$ in which case A_2 has more rows than columns. In that case we can write

$$A_1'A_2(A_2'A_2)^{-1}A_2'A_1 = A_1'A_1 - A^* .$$

A^* is a positive semi-definite matrix, for the left hand matrix product can be interpreted as the matrix of sums of squares and products of the "systematic part" of the least squares regression of A_1 on A_2, while $A_1'A_1$ is the matrix of sums of squares and products of the "dependent variables" (arranged in the A_1-matrix) themselves; and, therefore, A^* is the matrix of sums of squares and products of the corresponding least squares residuals. It follows that (3.11) can be written in the form

$$(3.12) \qquad Q^* = \frac{1}{\sigma_1^2} A_1' A_1 + \frac{\varrho^2}{\sigma_1^2(1-\varrho^2)} A^* = Q + \frac{\varrho^2}{\sigma_1^2(1-\varrho^2)} A^* .$$

This shows that Q^* is obtained by adding a positive semi-definite matrix to Q. Since Q and Q^* are symmetric and Q is positive-definite, this implies that the determinant of Q^* is larger than or equal to that of Q.[7] Hence the determinant of Q^{*-1} is smaller than (or equal to) that of Q^{-1}. But the determinants of Q^{*-1} and Q^{-1} are the generalized variances of the three-stage and two-stage least squares estimators, respectively; therefore, the former generalized variance is smaller than the latter except when $\varrho = 0$ or $A^* = 0$, in which case the two are equal.[8]

4. PROBLEMS OF IDENTIFICATION AND RELATED ISSUES

4.1. *Analysis.* It has been assumed that the matrix $[\sigma_{\mu\mu'}]$ is nonsingular; but when there are definitional equations or identities or, more generally, equations that hold exactly with disturbances that are identically zero, this assumption is no longer tenable since each such equation causes one row and one column of $[\sigma_{\mu\mu'}]$ to be zero. This is not a serious problem, however, because the coefficients of such equations are in general known (usually 1, —1, or 0), so that there is no need to estimate them. The recommendation is simply to delete such equations in the three-stage procedure.[9]

Another problem is that of the under-identified equations. Their two-stage least squares estimates do not exist, and, therefore, each such equation implies that one row and one column of $[\sigma_{\mu\mu'}]$ cannot be estimated. Our recommendation is to delete such equations and to confine the three-stage procedure to identifiable equations.[10]

Furthermore, it is useful to distinguish between just and over-identified equations. The analysis of Section 3.2 shows for the two-equation case that the estimation of the coefficients of one equation is improved if the other equation is over-identified, and is not improved if the latter is just-identified; and a result of this kind holds in fact more generally. Let us write p for the number of over-identified equations and hence $M-p$ for that of the just-

[7] See, e.g., Rao [**4**, p. 26].

[8] In a similar way, it can be shown that the variances of the three-stage coefficient estimators are smaller than the corresponding two-stage variances.

[9] 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.

[10] But see Section 6 under (2).

identified equations. Further, let us arrange them such that the former equations have numbers $1,\ldots,p$ and the latter $p+1,\ldots,M$; for reasons of typographical simplification we write q for $p+1$. Then the moment matrix of the three-stage least squares estimates for all M equations combined can be written in partitioned form:

$$\begin{bmatrix} G_1 & G_3 \\ G_3' & G_2 \end{bmatrix}^{-1} = \begin{bmatrix} H_1 & H_3 \\ H_3' & H_2 \end{bmatrix}, \tag{4.1}$$

where the index 1 refers to the set of over-identified equations, 2 to the just-identified equations, and 3 to a mixture of both sets:

$$G_1 = \begin{bmatrix} \sigma^{11}A_1'A_1 & \ldots & \sigma^{1p}A_1'A_p \\ \ldots & \ldots & \ldots \\ \sigma^{p1}A_p'A_1 & \ldots & \sigma^{pp}A_p'A_p \end{bmatrix}, \quad G_2 = \begin{bmatrix} \sigma^{qq}A_q'A_q & \ldots & \sigma^{qM}A_q'A_M \\ \ldots & \ldots & \ldots \\ \sigma^{Mq}A_M'A_q & \ldots & \sigma^{MM}A_M'A_M \end{bmatrix},$$

$$G_3 = \begin{bmatrix} \sigma^{1q}A_1'A_q & \ldots & \sigma^{1M}A_1'A_M \\ \ldots & \ldots & \ldots \\ \sigma^{pq}A_p'A_q & \ldots & \sigma^{pM}A_p'A_M \end{bmatrix}, \tag{4.2}$$

the A_μ being equal to $KX'Z_\mu$ where K is square and such that $K'K=(X'X)^{-1}$ as in Section 3.2. By partitioned multiplication we obtain from (4.1):

$$\begin{aligned} H_1 &= (G_1 - G_3G_2^{-1}G_3')^{-1}\ ; \\ H_3' &= -G_2^{-1}G_3'H_1\ ; \\ H_2 &= G_2^{-1} + G_2^{-1}G_3'H_1G_3G_2^{-1}\ . \end{aligned} \tag{4.3}$$

Our first objective will be H_1, which is the moment matrix of the three-stage coefficients of the over-identified equations. We shall show that *we shall arrive at precisely the same moment matrix if the three-stage estimation procedure is confined to this group of over-identified equations,* i.e., if the estimation is carried out under complete disregard of the just-identified equations. Consider then G_2^{-1}, which plays a role in the H_1-equation of (4.3). Taking account of the fact that all A_μ are square for $\mu \geqslant q$, we find

$$G_2^{-1} = \begin{bmatrix} \tau_{qq}A_q^{-1}A_q'^{-1} & \ldots & \tau_{qM}A_q^{-1}A_M'^{-1} \\ \ldots & \ldots & \ldots \\ \tau_{Mq}A_M^{-1}A_q'^{-1} & \ldots & \tau_{MM}A_M^{-1}A_M'^{-1} \end{bmatrix}, \tag{4.4}$$

where

$$[\tau_{\mu\mu'}] = \begin{bmatrix} \sigma^{qq} & \ldots & \sigma^{qM} \\ \ldots & \ldots & \ldots \\ \sigma^{Mq} & \ldots & \sigma^{MM} \end{bmatrix}^{-1}. \tag{4.5}$$

By postmultiplying G_2^{-1} by G_3' we obtain

$$(4.6)\qquad G_2^{-1}G_3' = \begin{bmatrix} v_{q1}A_q^{-1}A_1 & \dots & v_{qp}A_q^{-1}A_p \\ \dots & \dots & \dots \\ v_{M1}A_M^{-1}A_1 & \dots & v_{Mp}A_M^{-1}A_p \end{bmatrix},$$

where

$$(4.7)\qquad [v_{\mu\mu'}] = \begin{bmatrix} \sigma^{qq} & \dots & \sigma^{qM} \\ \dots & \dots & \dots \\ \sigma^{Mq} & \dots & \sigma^{MM} \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{q1} & \dots & \sigma^{qp} \\ \dots & \dots & \dots \\ \sigma^{M1} & \dots & \sigma^{Mp} \end{bmatrix}.$$

We go on in this way by premultiplying $G_2^{-1}G_3'$ by G_3 and subtracting the resulting product from G_1. According to (4.3), this gives the inverse of H_1:

$$(4.8)\qquad H_1^{-1} = \begin{bmatrix} \varphi_{11}A_1'A_1 & \dots & \varphi_{1p}A_1'A_p \\ \dots & \dots & \dots \\ \varphi_{p1}A_p'A_1 & \dots & \varphi_{pp}A_p'A_p \end{bmatrix},$$

where

$$(4.9)\qquad [\varphi_{\mu\mu'}] = \begin{bmatrix} \sigma^{11} & \dots & \sigma^{1p} \\ \dots & \dots & \dots \\ \sigma^{p1} & \dots & \sigma^{pp} \end{bmatrix}$$

$$- \begin{bmatrix} \sigma^{1q} & \dots & \sigma^{1M} \\ \dots & \dots & \dots \\ \sigma^{pq} & \dots & \sigma^{pM} \end{bmatrix} \begin{bmatrix} \sigma^{qq} & \dots & \sigma^{qM} \\ \dots & \dots & \dots \\ \sigma^{Mq} & \dots & \sigma^{MM} \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{q1} & \dots & \sigma^{qp} \\ \dots & \dots & \dots \\ \sigma^{M1} & \dots & \sigma^{Mp} \end{bmatrix}.$$

We proceed to consider the τ's, v's and φ's in more detail. They are all obtained by multiplying and inverting submatrices of $[\sigma^{\mu\mu'}]$, and it is easily seen that more convenient expressions are available. Let us write

$$(4.10)\qquad \begin{bmatrix} \sigma^{11} & \dots & \sigma^{1p} & \sigma^{1q} & \dots & \sigma^{1M} \\ \vdots & & \vdots & \vdots & & \vdots \\ \sigma^{p1} & \dots & \sigma^{pp} & \sigma^{pq} & \dots & \sigma^{pM} \\ \sigma^{q1} & \dots & \sigma^{qp} & \sigma^{qq} & \dots & \sigma^{qM} \\ \vdots & & \vdots & \vdots & & \vdots \\ \sigma^{M1} & \dots & \sigma^{Mp} & \sigma^{Mq} & \dots & \sigma^{MM} \end{bmatrix} = \begin{bmatrix} R_1 & R_3 \\ R_3' & R_2 \end{bmatrix} = \begin{bmatrix} S_1 & S_3 \\ S_3' & S_2 \end{bmatrix}^{-1}.$$

Then we have

$$(4.11)\qquad [\tau_{\mu\mu'}] = R_2^{-1}; \qquad [v_{\mu\mu'}] = R_2^{-1}R_3'; \qquad [\varphi_{\mu\mu'}] = R_1 - R_3R_2^{-1}R_3',$$

so that $[\varphi_{\mu\mu'}] = S_1^{-1}$. But S_1 is nothing else than the moment matrix of the disturbances of the over-identified equations; hence

$$(4.12) \qquad [\varphi_{\mu\mu'}] = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \dots & \dots & \dots \\ \sigma_{p1} & \dots & \sigma_{pp} \end{bmatrix}^{-1} .$$

On combining (4.8) and (4.12) and remembering that H_1 is the moment matrix of the three-stage least squares estimates of the over-identified equations, we find that it is indeed true that this matrix is the same as the one which is obtained when the three-stage procedure is confined to the set of over-identified equations. The latter method is preferable because it is computationally efficient, since the matrix to be inverted to obtain the estimates is of smaller order.

Let us now write $\hat{\delta}_A$ for the three-stage estimator when referring to the over-identified equations, and $\hat{\delta}_B$ for the just-identified equations. Assume that $\hat{\delta}_A$ has been computed as indicated above. Disregarding for a moment the differences between the σ's and the s's, we can write the estimation equations corresponding to (2.16) in the following form:

$$\begin{bmatrix} G_1 & G_3 \\ G_3' & G_2 \end{bmatrix} \begin{bmatrix} \hat{\delta}_A \\ \hat{\delta}_B \end{bmatrix} = \begin{bmatrix} A_1' KX' \sum \sigma^{1\mu} y_\mu \\ \vdots \\ A_M' KX' \sum \sigma^{M\mu} y_\mu \end{bmatrix},$$

so that the coefficient vector $\hat{\delta}_B$ of the just-identified equations can be written as

$$(4.13) \qquad \hat{\delta}_B = G_2^{-1} \begin{bmatrix} A_q' KX' \sum \sigma^{q\mu} y_\mu \\ \vdots \\ A_M' KX' \sum \sigma^{M\mu} y_\mu \end{bmatrix} - G_2^{-1} G_3' \hat{\delta}_A .$$

Hence $\hat{\delta}_B$ is the sum of two vectors. Applying (4.4) and (4.5), we find that the first can be written as

$$\begin{bmatrix} A_q^{-1} KX' \sum_{r=q}^{M} \sum_{\mu=q}^{M} \tau_{qr} \sigma^{r\mu} y_\mu \\ \vdots \\ A_M^{-1} KX' \sum_{r=q}^{M} \sum_{\mu=q}^{M} \tau_{Mr} \sigma^{r\mu} y_\mu \end{bmatrix} = \begin{bmatrix} A_q^{-1} KX' y_q \\ \vdots \\ A_M^{-1} KX' y_M \end{bmatrix} = \begin{bmatrix} (X'Z_q)^{-1} X' y_q \\ \vdots \\ (X'Z_M)^{-1} X' y_M \end{bmatrix} = d_B ,$$

i.e., the two-stage least squares estimator of the coefficients of the just-identified equations. For the second vector on the right of (4.13) we find, by applying (4.6),

$$G_2^{-1} G_3' \hat{\delta}_A = \begin{bmatrix} v_{q1} A_q^{-1} A_1 \dots v_{qp} A_q^{-1} A_p \\ \dots\dots\dots \\ v_{M1} A_M^{-1} A_1 \dots v_{Mp} A_M^{-1} A_p \end{bmatrix} \hat{\delta}_A = \begin{bmatrix} (X'Z_q)^{-1} X' \sum v_{q\mu} X' Z_\mu \hat{\delta}_\mu \\ \vdots \\ (X'Z_M)^{-1} X' \sum v_{M\mu} X' Z_\mu \hat{\delta}_\mu \end{bmatrix},$$

where the summation on the right extends over $\mu = 1, \ldots, p$. Hence the three-stage least squares estimator for the just-identified equations is

$$\hat{\delta}_B = d_B - \begin{bmatrix} (X'Z_q)^{-1}X' \sum_1^{\mu} v_{q\mu} X'Z_\mu \hat{\delta}_\mu \\ \vdots \\ (X'Z_M)^{-1}X' \sum_1^{\mu} v_{M\mu} X'Z_\mu \hat{\delta}_\mu \end{bmatrix}, \tag{4.14}$$

which shows that the three-stage estimate is obtained by adding to the two-stage estimate a linear combination of the three-stage estimates of the over-identified equations. The v's can be derived from (4.7); but it is computationally more efficient to obtain them from

$$[v_{\mu\mu'}] = -S_3'S_1^{-1} = -\begin{bmatrix} \sigma_{q1} & \cdots & \sigma_{qp} \\ \cdots & \cdots & \cdots \\ \sigma_{M1} & \cdots & \sigma_{Mp} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \cdots & \cdots & \cdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix}^{-1}, \tag{4.15}$$

(see (4.10) and (4.11)).

The three-stage moment matrix of the just-identified equations is H_2, as defined in (4.3). It is the sum of two matrices, the first of which is G_2^{-1}, as specified in (4.4). It is useful to write matrices such as $A_q^{-1}A_q'^{-1}$ in the form

$$(A_q'A_q)^{-1} = (Z_q'XK'KX'Z_q)^{-1} = (X'Z_q)^{-1}X'X(Z_q'X)^{-1}.$$

The second matrix on the right of (4.3) is H_1 premultiplied by $G_2^{-1}G_3'$ and postmultiplied by $G_3G_2^{-1}$. Now H_1 is available when the computations for the over-identified equations are completed, and $G_2^{-1}G_3'$ is specified in (4.6). But it is useful to write such matrices as $A_q^{-1}A_1$ in the form

$$(KX'Z_q)^{-1}KX'Z_1 = (X'Z_q)^{-1}K^{-1}KX'Z_1 = (X'Z_q)^{-1}X'Z_1.$$

Hence the three-stage moment matrix of the just-identified equations is

$$H_2 = \begin{bmatrix} \tau_{qq}(X'Z_q)^{-1}X'X(Z_q'X)^{-1} & \cdots & \tau_{qM}(X'Z_q)^{-1}X'X(Z_M'X)^{-1} \\ \cdots & \cdots & \cdots \\ \tau_{Mq}(X'Z_M)^{-1}X'X(Z_q'X)^{-1} & \cdots & \tau_{MM}(X'Z_M)^{-1}X'X(Z_M'X)^{-1} \end{bmatrix} \tag{4.16}$$
$$+ \begin{bmatrix} v_{q1}(X'Z_q)^{-1}X'Z_1 & \cdots & v_{qp}(X'Z_q)^{-1}X'Z_p \\ \cdots & \cdots & \cdots \\ v_{M1}(X'Z_M)^{-1}X'Z_1 & \cdots & v_{Mp}(X'Z_M)^{-1}X'Z_p \end{bmatrix}$$
$$\times H_1 \begin{bmatrix} v_{1q}Z_1'X(Z_q'X)^{-1} & \cdots & v_{1M}Z_1'X(Z_M'X)^{-1} \\ \cdots & \cdots & \cdots \\ v_{pq}Z_p'X(Z_q'X)^{-1} & \cdots & v_{pM}Z_p'X(Z_M'X)^{-1} \end{bmatrix},$$

where the v's are specified in (4.15) and the τ's in (4.5). A more convenient expression for the τ's is:

$$(4.17)\quad [\tau_{\mu\mu'}] = S_2 - S_3' S_1^{-1} S_3 = S_2 + [v_{\mu\mu'}]S_3 = \begin{bmatrix} \sigma_{qq} & \cdots & \sigma_{qM} \\ \cdots & \cdots & \cdots \\ \sigma_{Mq} & \cdots & \sigma_{MM} \end{bmatrix}$$

$$- \begin{bmatrix} \sigma_{q1} & \cdots & \sigma_{qp} \\ \cdots & \cdots & \cdots \\ \sigma_{M1} & \cdots & \sigma_{Mp} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \cdots & \cdots & \cdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{1q} & \cdots & \sigma_{1M} \\ \cdots & \cdots & \cdots \\ \sigma_{pq} & \cdots & \sigma_{pM} \end{bmatrix},$$

because this does not require inversion of the complete matrix $[\sigma_{\mu\mu'}]$ of all equations. In the same way, the "cross-moment" matrix of the just- and over-identified equations is

$$(4.18)\quad H_3' = - \begin{bmatrix} v_{q1}(X'Z_q)^{-1}X'Z_1 & \cdots & v_{qp}(X'Z_q)^{-1}X'Z_p \\ \cdots & \cdots & \cdots \\ v_{M1}(X'Z_M)^{-1}X'Z_1 & \cdots & v_{Mp}(X'Z_M)^{-1}X'Z_p \end{bmatrix} H_1 .$$

In conclusion, we observe that if $[\sigma_{\mu\mu'}]$ is block-diagonal, so that it can be written in the form

$$(4.19)\quad [\sigma_{\mu\mu'}] = \begin{bmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \Sigma_k \end{bmatrix},$$

the three-stage estimation procedure can be conveniently carried out block by block; i.e., the estimation of any structural equation is then essentially based only on the over-identifying restrictions of the equations whose disturbances are correlated with its own disturbances. It will be noted that the procedure recommended for systems containing definitions (see the first paragraph of this section) is a special case of (4.19).

4.2. *Summary of Rules.* The results derived in Section 4.1 can be summarized as follows:

(1) If the system contains definitional equations (and possibly other equations whose coefficients are known and whose disturbances are zero), then these are to be disregarded, i.e., the M of (2.16) and (2.17) is to be interpreted as the total number of equations excluding this category.

(2) If the system contains under-identified equations, these are to be disregarded too.[11]

(3) If the system contains just-identified as well as over-identified equations, it is computationally efficient to apply the three-stage procedure sep-

[11] But see Section 6 under (2).

arately to these two groups. For the over-identified equations, one should proceed in the manner described in Section 2 but interpret the M of (2.16) and (2.17) as the number of these equations, i.e., the total number of equations excluding the definitions and the just- and under-identified equations. (If there is only one over-identified equation, this procedure does of course not lead to additional efficiency beyond that of two-stage least squares, so that it is not to be applied in that case.)

(4) For just-identified equations one proceeds as follows. Suppose that the system contains at least one over-identified equation (if there is no such equation, the procedure does not lead to additional efficiency). Write $\mu = 1, \ldots, p$ for the over-identified equations and $\mu = q, \ldots, M$ for the just-identified group, where $q = p + 1$. Then the three-stage least squares estimator of the coefficient vector of the just-identified equations is $\hat{\delta}_B$ as specified in (4.14), where on the right d_B is the two-stage least squares estimator of this vector and $\hat{\delta}_1, \ldots, \hat{\delta}_p$ is the three-stage least squares estimators of the over-identified equations (which are computed as stated under (3)). The asymptotic covariance matrix of $\hat{\delta}_B$ is H_2 as given in (4.16), and the matrix of the asymptotic covariances of δ_B, on the one hand, and $\delta_1, \ldots, \delta_p$, on the other, is H_3' as given in (4.18), where H_1 is the asymptotic covariance matrix of $\hat{\delta}_1, \ldots, \hat{\delta}_p$ and is computed as indicated under (3). The τ's and v's occurring in these expressions are given in (4.17) and (4.15), respectively. For the actual calculation one should replace the σ's of these formulae (the variances and covariances of the structural disturbances) by corresponding s's (their two-stage least squares estimates).

(5) If the covariance matrix of the structural disturbances is block-diagonal as in (4.19), the three-stage estimation procedure should be carried out separately for each group of equations corresponding with such a block.

5. AN EXAMPLE; DIRECTORY OF COMPUTATIONS[12]

5.1. *Some Computational Simplifications*. There are two groups of numerical results which we need. One of them deals with such matrices as $Z_1' X(X'X)^{-1} X'Z_1$, $Z_1' X(X'X)^{-1} X'y_1$, etc., and the other with the elements $s^{\mu\mu'}$ of the inverse of the two-stage least squares estimated contemporaneous covariance matrix of the structural disturbances. Starting with the former group, we observe that it is useful to introduce the partioned matrix

$$Z = [Y \quad X], \tag{5.1}$$

where Y is the matrix of values taken by all relevant jointly dependent

[12] The example contains no just-identified equations; reference is made to point (4) of Section 4.2 for the appropriate calculations when there are such equations.

variables in such a way that y_μ, Y_μ are all submatrices of Y. It is then easy to see that the matrices $Z_1'X(X'X)^{-1}X'Z_1$, $Z_1'X(X'X)^{-1}X'y_1$, etc., are all submatrices of the symmetric matrix

$$Z'X(X'X)^{-1}X'Z = \begin{bmatrix} Y'X(X'X)^{-1}X'Y & Y'X \\ X'Y & X'X \end{bmatrix}. \tag{5.2}$$

Proceeding to the $s^{\mu\mu'}$, we start by writing the M equations (2.1) in the form

$$[y_1\ y_2 \dots y_M] = [Y_1\ Y_2 \dots Y_M] \begin{bmatrix} \gamma_1 & 0 & \dots & 0 \\ 0 & \gamma_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \gamma_M \end{bmatrix} + [X_1\ X_2 \dots X_M] \begin{bmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \beta_M \end{bmatrix} + [u_1\ u_2 \dots u_M]. \tag{5.3}$$

Let us write Y_L for the matrix of values taken by the left hand jointly dependent variables and similarly Y_R, X_R for those referring to the right hand dependent and predetermined variables respectively:

$$\begin{aligned} Y_L &= [y_1\quad y_2\ \dots\ y_M], \\ Y_R &= [Y_1\quad Y_2\ \dots\ Y_M], \\ X_R &= [X_1\quad X_2\ \dots\ X_M]. \end{aligned} \tag{5.4}$$

Then the two-stage least squares estimate of (5.3) takes the form

$$Y_L = Y_R C + X_R B + \hat{U}, \tag{5.5}$$

where $\hat{U}$ is the $T \times M$ matrix of two-stage least squares estimated structural disturbances, and

$$C = \begin{bmatrix} c_1 & 0 & \dots & 0 \\ 0 & c_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & c_M \end{bmatrix}, \quad B = \begin{bmatrix} b_1 & 0 & \dots & 0 \\ 0 & b_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & b_M \end{bmatrix}, \tag{5.6}$$

the c's and b's being two-stage least squares coefficient estimates.

The separation in terms of Y_R and X_R is chosen because it is useful to take account of the condition, $X_R'\hat{U} = 0$. Doing so, we find

$$[Ts_{\mu\mu'}] = \hat{U}'\hat{U} = Y_L'Y_L - Y_L'Y_R C - C'Y_R'Y_L + C'Y_R'Y_R C - B'X_R'X_R B. \tag{5.7}$$

The first matrix on the right ($Y_L'Y_L$) is simply a submatrix of Y', Y and is derived as such. The second is computed by postmultiplying such submatrices by $c_1, c_2, \ldots$, giving

$$(5.8) \qquad Y_L'Y_RC = [Y_L'Y_1c_1 \quad Y_L'Y_2c_2 \ \ldots \ Y_L'Y_Mc_M] .$$

The third matrix on the right of (5.7) is the transpose of the second. The fourth is a symmetric matrix which can be partitioned as follows:

$$(5.9) \qquad C'Y_R'Y_RC = \begin{bmatrix} c_1'Y_1'Y_1c_1 & \ldots & c_1'Y_1'Y_Mc_M \\ \vdots & & \vdots \\ c_M'Y_M'Y_1c_1 & \ldots & c_M'Y_M'Y_Mc_M \end{bmatrix},$$

which implies that we have to premultiply such matrices as $Y_1'Y_2$ by c_1' and postmultiply by c_2. The last matrix on the right of (5.7) is completely analogous:

$$(5.10) \qquad B'X_R'X_RB = \begin{bmatrix} b_1'X_1'X_1b_1 & \ldots & b_1'X_1'X_Mb_M \\ \vdots & & \vdots \\ b_M'X_M'X_1b_1 & \ldots & b_M'X_M'X_Mb_M \end{bmatrix}.$$

5.2. *Klein's Model I.* The example which we shall use is Klein's Model I [**3**], which consists of three behavioral equations and three definitions. The former category includes a consumption function which describes consumption (C) linearly in terms of profits (Π), profits lagged one year (Π_{-1}), and the total wage bill ($W_1 + W_2$), where W_1 refers to the private wage bill and W_2 to the government wage bill. Thus,

$$(5.11) \qquad C = \alpha_0 + \alpha_1\Pi + \alpha_2(W_1 + W_2) + \alpha_3\Pi_{-1} + u_1 .$$

Furthermore, there 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}):

$$(5.12) \qquad I = \beta_0 + \beta_1\Pi + \beta_2\Pi_{-1} + \beta_3K_{-1} + u_2 .$$

Finally, there is a demand-for-labor equation which describes the private wage bill linearly in terms of national income (Y) plus indirect taxes (T) minus the government wage bill, the same variable lagged one year, and time (measured in calendar years):

$$(5.13) \qquad W_1 = \gamma_0 + \gamma_1(Y + T - W_2) + \gamma_2(Y + T - W_2)_{-1} + \gamma_3t + u_3 .$$

The system is completed by three definitions:

$$(5.14) \qquad \begin{aligned} Y + T &= C + I + G , \\ Y &= W_1 + W_2 + \Pi , \\ K &= K_{-1} + I , \end{aligned}$$

where G is government expenditure on goods and services.

The system (5.11)-(5.14) is a complete one in six jointly dependent variables (C, Π, W_1, I, Y, K) and eight predetermined variables ($W_2, T, G, t, \Pi_{-1}, K_{-1}, (Y+T-W_2)_{-1}$, and the variable corresponding to the constant terms). Our objective is to estimate the α's, β's, and γ's on the basis of American data for the period 1921–1941.

5.3. *Computations*. The three definitions (5.14) are to be disregarded according to point (1) of Section 4.2. Furthermore, the three behavioral equations are all over-identified, hence we should proceed along the lines of Section 2 for $M = 3$. The detailed steps are as follows:

STEP ONE: *Arranging the variables*. Arrange the jointly dependent variables falling under y_μ, Y_μ for $\mu = 1,\ldots,M$ into the matrix Y; similarly, arrange all predetermined variables under X and form the partitioned matrix $Z = [Y\ X]$, see (5.1).

For Klein's Model I the arrangement of Table I is convenient. It defines one variable for $W_1 + W_2$ because this has one single coefficient in (5.11). This variable is a mixture of jointly dependent and predetermined variables and therefore listed as dependent (because it is not independent of the disturbances). The same applies to $Y+T-W_2$. The variable representing the constant term is indicated by $x_8 \equiv 1$.

TABLE I

ARRANGEMENT OF VARIABLES

Jointly dependent variables					
y_1 C	y_2 I	y_3 W_1	y_4 Π	y_5 W_1+W_2	y_6 $Y+T-W_2$

Predetermined variables							
x_1 W_2	x_2 T	x_3 G	x_4 t	x_5 Π_{-1}	x_6 K_{-1}	x_7 $(Y+T-W_2)_{-1}$	x_8 1

STEP TWO: *Computing the moments*. Compute the sums of squares and products of all variables listed in Step One: $Y'Y$, $X'X$, $X'Y$, and also $Y'X(X'X)^{-1}X'Y$ (which is the matrix of the sums of squares and products of the computed value of Y in its regression on X).

These four matrices in our numerical case are specified in Table II.[13]

STEP THREE: *Computing the two-stage least squares estimates*. Select the variables falling under y_μ, Y_μ, X_μ (and hence $Z_\mu = [Y_\mu \quad X_\mu]$) for each

[13] All data are taken from [**3**, Appendix, p. 135].

TABLE II

MOMENT MATRICES USED IN 3SLS ESTIMATION OF KLEIN'S MODEL I

A. *Jointly dependent*, $Y'Y$*:

	y_1	y_2	y_3	y_4	y_5	y_6
y_1	62166.63	1679.01	42076.78	19566.35	48054.11	69501.99
y_2		286.02	1217.92	726.10	1321.72	2104.42
y_3			28560.86	13296.61	32604.93	47173.09
y_4				6347.25	15117.72	22049.39
y_5					37275.87	53827.54
y_6						77998.58

B. *Predetermined*, $X'X$*:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	626.87	789.27	1200.19	238.00	1746.22	21683.18	6364.43	107.50
x_2		1054.95	1546.11	176.00	2348.48	28766.25	8436.53	142.90
x_3			2369.94	421.70	3451.86	42026.14	12473.50	208.20
x_4				770.00	−11.90	590.60	495.60	0.00
x_5					5956.29	69073.54	20542.22	343.90
x_6						846132.70	244984.77	4210.40
x_7							72200.03	1217.70
x_8								21.00

C. *Jointly dependent–predetermined*, $Y'X$*:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
y_1	5977.33	7858.86	11633.68	577.70	18929.37	227767.38	66815.25	1133.99
y_2	103.80	160.40	243.19	−105.60	655.33	5073.25	1831.13	26.60
y_3	4044.07	5315.62	7922.46	460.90	12871.73	153470.56	45288.51	763.60
y_4	1821.11	2405.53	3578.05	18.90	6070.13	70946.78	21030.44	354.70
y_5	4670.94	6104.89	9122.65	698.90	14617.95	175153.74	51652.94	871.10
y_6	6654.45	8776.10	13046.62	655.80	21290.34	253183.59	74755.48	1261.20

D. *"Calculated" jointly dependent*, $Y'X(X'X)^{-1}X'Y$:

	y_1	y_2	y_3	y_4	y_5	y_6
y_1	**	**	**	19507.721	48010.448	69399.702
y_2		**	**	681.203	1283.799	2021.602
y_3			**	13255.034	32564.923	47091.507
y_4				6285.300	15076.144	21945.864
y_5					37235.862	53745.956
y_6						77813.470

* Klein's data are given accurately to one decimal place; thus these moments have two places beyond the decimal point.
** These moments are not needed in computing 3SLS estimates.

equation and compute the two-stage least squares estimates by using the formula

$$(5.15)\qquad d_\mu = \begin{bmatrix} Y'_\mu X(X'X)^{-1}X'Y_\mu & Y'_\mu X_\mu \\ X'_\mu Y_\mu & X'_\mu X_\mu \end{bmatrix}^{-1} \begin{bmatrix} Y'_\mu X(X'X)^{-1}X'y_\mu \\ X'_\mu y_\mu \end{bmatrix}.$$

In our case this means:

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_0 \end{bmatrix} = \left[\begin{array}{cc|cc} 6285.300 & 15076.144 & 6070.13 & 354.70 \\ 15076.144 & 37235.862 & 14617.95 & 871.10 \\ \hline 6070.13 & 14617.95 & 5956.29 & 343.90 \\ 354.70 & 871.10 & 343.90 & 21.00 \end{array}\right]^{-1} \left[\begin{array}{c} 19507.721 \\ 48010.448 \\ \hline 18929.37 \\ 1133.90 \end{array}\right] = \begin{bmatrix} 0.017 \\ 0.810 \\ 0.216 \\ 16.555 \end{bmatrix},$$

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_0 \end{bmatrix} = \left[\begin{array}{c|ccc} 6285.300 & 6070.13 & 70946.78 & 354.70 \\ \hline 6070.13 & 5956.29 & 69073.54 & 343.90 \\ 70946.78 & 69073.54 & 846132.70 & 4210.40 \\ 354.70 & 343.90 & 4210.40 & 21.00 \end{array}\right]^{-1} \left[\begin{array}{c} 681.203 \\ \hline 655.33 \\ 5073.25 \\ 26.60 \end{array}\right] = \begin{bmatrix} 0.150 \\ 0.616 \\ -0.158 \\ 20.278 \end{bmatrix},$$

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_0 \end{bmatrix} = \left[\begin{array}{c|ccc} 77813.470 & 74755.48 & 655.80 & 1261.20 \\ \hline 74755.48 & 72200.03 & 495.60 & 1217.70 \\ 655.80 & 495.60 & 770.00 & 0.00 \\ 1261.20 & 1217.70 & 0.00 & 21.00 \end{array}\right]^{-1} \left[\begin{array}{c} 47091.507 \\ \hline 45288.51 \\ 460.90 \\ 763.60 \end{array}\right] = \begin{bmatrix} 0.439 \\ 0.147 \\ 0.130 \\ 1.500 \end{bmatrix}.$$

STEP FOUR: *Estimating the moment matrix of the structural disturbances.* Form the matrices needed to calculate the matrix products in (5.7). For Klein's model we have

$$\begin{aligned} Y_L &= [y_1 \ \ y_2 \ \ y_3]\,, \\ Y_R &= [Y_1 \ \ Y_2 \ \ Y_3] = [y_4 \ y_5 \ y_4 \ y_6]\,, \\ X_R &= [X_1 \ \ X_2 \ \ X_3] \\ &= [x_5 \ \ x_8 \ \ x_5 \ \ x_6 \ \ x_8 \ \ x_7 \ \ x_4 \ \ x_8]\,. \end{aligned}$$

(Check twice!) Then the following matrices are employed in calculating $[Ts_{\mu\mu'}]$ from (5.7):

$$Y'_L Y_L = \begin{bmatrix} 62166.63 & 1679.01 & 42076.78 \\ & 286.02 & 1217.92 \\ & & 28560.86 \end{bmatrix},$$

$$Y'_L Y_R = \begin{bmatrix} 19566.35 & 48054.11 & 19566.35 & 69501.99 \\ 726.10 & 1321.72 & 726.10 & 2104.42 \\ 13296.61 & 32604.93 & 13296.61 & 47173.09 \end{bmatrix},$$

$$Y'_R Y_R = \begin{bmatrix} 6347.25 & 15117.72 & 6347.25 & 22049.39 \\ & 37275.87 & 15117.72 & 53827.54 \\ & & 6347.25 & 22049.39 \\ & & & 77998.58 \end{bmatrix},$$

$$X_R'X_R = \begin{bmatrix} 5956.29 & 343.90 & 5956.29 & 69073.54 & 343.90 & 20542.22 & -11.90 & 343.90 \\ & 21.00 & 343.90 & 4210.40 & 21.00 & 1217.70 & 0.00 & 21.00 \\ & & 5956.29 & 69073.54 & 343.90 & 20542.22 & -11.90 & 343.90 \\ & & & 846132.70 & 4210.40 & 244984.77 & 590.60 & 4210.40 \\ & & & & 21.00 & 1217.70 & 0.00 & 21.00 \\ & & & & & 72200.03 & 495.60 & 1217.70 \\ & & & & & & 770.00 & 0.00 \\ & & & & & & & 21.00 \end{bmatrix},$$

$$C = \begin{bmatrix} 0.0173 & 0 & 0 \\ 0.8102 & 0 & 0 \\ 0 & 0.1502 & 0 \\ 0 & 0 & 0.4389 \end{bmatrix}, \qquad B = \begin{bmatrix} 0.2162 & 0 & 0 \\ 16.5548 & 0 & 0 \\ 0 & 0.6159 & 0 \\ 0 & -0.1578 & 0 \\ 0 & 20.2782 & 0 \\ 0 & 0 & 0.1467 \\ 0 & 0 & 0.1304 \\ 0 & 0 & 1.5003 \end{bmatrix}.$$

With these matrices the right hand side of (5.7) can be computed to yield

$$[Ts_{\mu\mu'}] = \begin{bmatrix} 21.926 & 9.966 & -5.758 \\ & 29.047 & -4.156 \\ & & 10.005 \end{bmatrix}.$$

Then by inverting $[s_{\mu\mu'}]$ we obtain

$$[s^{\mu\mu'}] = \begin{bmatrix} 1.272 & -0.353 & 0.585 \\ & 0.867 & 0.157 \\ & & 2.503 \end{bmatrix}.$$

STEP FIVE: *Calculating 3SLS estimates.* Form the matrices in (2.16) by selecting appropriate submatrices from Table II and multiplying their elements by elements of $[s^{\mu\mu'}]$. It is to be remembered that on the right side of (2.16) Z_μ represents the partitioned matrix $[Y_\mu \quad X_\mu]$. Then the first matrix in (2.16) is inverted. On comparing (2.16) and (2.17), it is seen that this inverse, given in Table III, is the estimated covariance matrix of 3SLS estimators. The vector appearing in (2.16) is then calculated. In our case we have

TABLE III

ESTIMATED COVARIANCE MATRIX OF 3SLS COEFFICIENT ESTIMATORS FOR KLEIN'S MODEL I

	a_1	a_2	a_3	a_0	b_1	b_2	b_3	b_0	c_1	c_2	c_3	c_0
a_1	.013119	—.001384	.009026	—.016376	.003928	—.002985	—.000394	.061589	—.001226	.001268	.000423	.000072
a_2		.001490	—.000493	—.030378	.001287	—.001180	.000396	—.081919	.000039	—.000059	—.000328	.001067
a_3			.010946	—.006351	—.004515	.005086	—.000110	.014973	.000780	—.001201	.000322	.022799
a_0				1.690435	—.045805	.016064	—.007990	2.135237	.006297	.000701	.001177	—.431873
b_1					.028505	—.024394	.003845	—.852912	—.000707	.000695	—.000410	.002157
b_2						.025220	—.003552	.711270	.000432	—.000714	.000611	.015454
b_3							.001202	—.247848	.000047	—.000057	—.000154	.000430
b_0								52.516473	—.004647	.011291	.027734	—.385044
c_1									.001203	—.001126	—.000299	—.006968
c_2										.001422	.000045	—.014845
c_3											.000821	.015335
c_0												1.301896

$$\begin{bmatrix} \sum_{\mu} s_1^{1\mu} Z_1' X(X'X)^{-1} X' y_\mu \\ \sum_{\mu} s^{2\mu} Z_2' X(X'X)^{-1} X' y_\mu \\ \sum_{\mu} s^{3\mu} Z_3' X(X'X)^{-1} X' y_\mu \end{bmatrix} = \begin{bmatrix} 32335.954 \\ 79687.291 \\ 31384.947 \\ 1880.124 \\ -4212.733 \\ -4091.236 \\ -51886.945 \\ -257.212 \\ 158827.353 \\ 152771.080 \\ 1475.328 \\ 2579.466 \end{bmatrix}.$$

This vector is then premultiplied by the matrix given in Table III to yield 3SLS coefficient estimates which are presented in Table IV along with their 2SLS counterparts and also with estimated variances of coefficient estimators.

TABLE IV

THREE-STAGE AND TWO-STAGE LEAST SQUARES ESTIMATES OF PARAMETERS IN KLEIN'S MODEL I

Equation	Coefficient of	3SLS		2SLS	
		Coefficient estimate	Variance of coefficient estimator	Coefficient estimate	Variance of coefficient estimator
Consumption	Π	0.0479	0.013119	0.0173	0.013936
	$W_1 + W_2$	0.8170	0.001490	0.8102	0.001620
	Π_{-1}	0.1897	0.010946	0.2162	0.011506
	1	16.1923	1.690	16.5548	1.745
Investment	Π	0.2111	0.028505	0.1502	0.030084
	Π_{-1}	0.5667	0.025220	0.6159	0.026499
	K_{-1}	−0.1472	0.001202	−0.1578	0.001305
	1	17.9210	52.516	20.2782	56.892
Demand for labor	$Y+T-W_2$	0.4282	0.001203	0.4389	0.001270
	$(Y+T-W_2)_{-1}$	0.1543	0.001422	0.1467	0.001508
	t	0.1356	0.000821	0.1304	0.000849
	1	1.6935	1.302	1.5003	1.317

6. CONCLUDING REMARKS

(1) When there are linear a priori restrictions on coefficients of different structural equations (e.g., $\alpha_1 = \frac{1}{2}\beta_1$ in (5.11)–(5.12)), one proceeds as follows. Write the three-stage least squares estimation equations in the form (3.1), $v = W\delta + \varepsilon$, and write the restrictions in the form $R\delta = r$, where R and r are matrices with known elements. The number of rows of R and r equals the number of a priori restrictions, and the number of their columns in n and 1, respectively. Then apply constrained generalized least squares (see Theil [**5**, Appendix to Chapter 8]).

(2) When the μth equation is under-identified, (2.4) is a system of $\Lambda < n_\mu$ equations. But if there is some a priori knowledge about the components of δ_μ which can be written in linear stochastic form (and which may therefore be uncertain knowledge), this raises effectively the number of estimation equations. This is the method of mixed linear estimation (see Theil and Goldberger [**6**]).

(3) 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 set up a new stage based on this estimator rather than $[s_{\mu\mu'}]$ and proceed iteratively. No reports on this method can be made as yet, but we hope to come back to it in the future.

University of Wisconsin and the Netherlands School of Economics

REFERENCES

[**1**] AITKEN, A. C.: "On Least-Squares and Linear Combination of Observations," *Proceedings of the Royal Society of Edinburgh*, vol. 55 (1934–35), pp. 42–48.

[**2**] CRAMER, H.: *Mathematical Methods of Statistics*, Princeton, Princeton University Press, 1946.

[**3**] KLEIN, L. R.: *Economic Fluctuations in the United States, 1921–1941*, New York, John Wiley and Sons, 1950.

[**4**] RAO, C. R.: *Advanced Statistical Methods in Biometric Research*, New York, John Wiley and Sons, 1952.

[**5**] THEIL, H.: *Economic Forecasts and Policy*, Second Edition, Amsterdam, North-Holland Publishing Company, 1961.

[**6**] THEIL, H., AND A. S. GOLDBERGER: "On Pure and Mixed Statistical Estimation in Economics," Report 6009 of the Econometric Institute of the Netherlands School of Economics (1960). To be published in the *International Economic Review*.