

Office of Tax Analysis
U.S. Treasury Department
Washington, D.C. 20220
Issued: September 1976

ESTIMATION OF A SIMULTANEOUS SYSTEM OF
EQUATIONS WHEN THE SAMPLE IS UNDERSIZED

K.R. Kadiyala
Purdue University

James R. Nunns
U.S. Treasury Department

OTA Paper 14

August 1976

TABLE OF CONTENTS

	<u>page</u>
I. Introduction	1
II. The Problem	2
III. Alternative Approaches to the Problem	3
IV. A Proposed Class of Estimators	6
V. Estimation of Klein's Model I - An Illustration	8

I. INTRODUCTION

In most large and many medium-sized econometric models, the number of predetermined variables exceeds the number of observations on each variable. Estimation procedures such as two-stage least squares and other k -class ($k > 0$) procedures, as well as three-stage least squares and certain other full-information procedures are therefore inapplicable. In this paper, a class of modified two-stage least squares estimators is derived which exhibits several desirable properties in comparison to alternative estimators which have been proposed for models with under-sized samples.

II. THE PROBLEM

The j^{th} structural equation of a linear simultaneous equation system may be written as:

$$y_j = Y_j \gamma_j + X_j \beta_j + \xi_j \quad (1)$$

or more conveniently as:

$$y_j = Z_j \delta_j + \xi_j \quad (2)$$

where $Z_j = (Y_j : X_j)$, $\delta_j' = (\gamma_j' : \beta_j')$, y_j is the $n \times 1$ vector of observations on the j^{th} dependent variable, Y_j is the $n \times L_j$ matrix of observations on the

jointly dependent variables which are explanatory in the j^{th} equation, X_j is the $n \times K_j$ matrix of observations on the predetermined variables entering the j^{th} equation, γ_j and β_j are parameter vectors to be estimated, and ξ_j is an $n \times 1$ vector of disturbances. The system contains L jointly dependent variables, and K ($> K_j$) predetermined variables; X is the $n \times K$ matrix of observations on all predetermined variables in the system. It is assumed throughout the paper that the predetermined variables are "fixed", ξ_j has a zero mean and covariance matrix $\sigma_{jj}I$ ($0 < \sigma_{jj} < \infty$), the j^{th} equation is identified, and the rank of $X'Z_j$ is $L_j + K_j$ which requires $\min(K, n) \geq L_j + K_j$.

Multiplying equation (2) by X' gives:

$$X'y_j = X'Z_j\delta_j + X'\xi_j \quad (3)$$

The transformed disturbance vector $X'\xi_j$ has mean zero and covariance matrix $\sigma_{jj}X'X$. Assuming X has rank K (which requires $n \geq K$), the two-stage least squares estimator of δ_j , $\hat{\delta}_j$, is derived from (3) by applying Aitken's theorem, giving:

$$\hat{\delta}_j = (Z_j'EZ_j)^{-1}Z_j'Ey_j \quad (4)$$

where $E = X(X'X)^{-1}X'$. When the rank of X is less than K , $X'X$ is singular and the two-stage least squares estimator (as well as all other estimators which depend on the inverse of $X'X$) fails to exist. The rank of X is always less than K when $n < K$, i.e., when the sample is undersized.

III. ALTERNATIVE APPROACHES TO THE PROBLEM

A number of estimation procedures have been proposed that do not rely on the inverse of $X'X$, and are therefore at least potentially applicable when the sample is undersized. These procedures will only be discussed briefly here; more extensive discussions may be found in Theil [9] and Dutta and Lyttkens [1]. Our primary interest is in those procedures which are truly "limited-information" - requiring only specification of the j^{th} equation and the list of predetermined variables occurring in the system. Other procedures, while usually more efficient, have the undesirable property of requiring a more detailed knowledge of the entire system. Estimation of the j^{th} equation is therefore sensitive to misspecification in the remainder of the system.

Among the limited information procedures, the following three are widely known and illustrate the difficulties of estimation when the sample is undersized.

1. Kloeck and Mennes [3] suggested replacing X with $T = (X_j \vdots P)$ where P is a matrix of principal components of some linear combination of some or all of the columns of X . This leads to the estimator:

$$\delta_j^* = (Z_j' T (T' T)^{-1} T' Z_j')^{-1} Z_j' T (T' T)^{-1} T' y_j \quad . \quad (5)$$

A major disadvantage of this procedure is that the size of P , the columns of X from which the full set of principal components is derived, and the normalization chosen are all arbitrary. Thus, δ_j^* may be highly sensitive

to the P matrix used in its estimation. A lesser disadvantage is that the procedure requires considerably greater computational effort than the two-stage least squares procedure. Further, as is the case with all other limited-information procedures which we are aware of, short of specifying and estimating the entire system no estimates of the reduced form coefficients is possible using this procedure. Thus, projections of the dependent variables included in Y_j cannot be obtained simply on the basis of projections of the predetermined variables.

From Takeuchi's results [7] it is known that in certain cases, if P is of rank r then the even moments of order less than $r - L_j + 3$ of δ_j^* exist, but little else is known about its small sample properties. It has the desirable large sample property of consistency.

2. Applying a generalization of Aitken's theorem to equation (3), Swamy and Holmes [6] and Fischer and Wadycki [2] obtain the estimator:

$$\delta_j^- = (Z_j' E^- Z_j)^{-1} Z_j' E^- y_j \quad (6)$$

where $E^- = X(X'X)^-X'$ and $(X'X)^-$ is any (weak) generalized inverse of $X'X$. Normally when the sample is undersized, the rank of X is n in which case $X(X'X)^-X = I$ so that $\delta_j^- = (Z_j'Z_j)^{-1} Z_j' y_j$, the ordinary least squares estimator for δ_j . Since δ_j^- becomes the two-stage least squares estimator when $n > K$ (assuming that the rank of X is then K), it does not share the property of inconsistency with the ordinary least squares estimator. Consistency, however, is a large sample property; it is the small sample properties of δ_j^- which are relevant in the present context. Mariano [4]

has shown that in the general case, the even moments of order less than $n-(K_j+L_j)+1$ of the ordinary least squares estimator exist. However, Sawa [5] has shown that, for an equation with $L_j=1$, the ordinary least squares estimator has a lower mean square error than other k -class ($0 < k \leq 1$) estimators only in rather specialized circumstances. Reduced form parameters cannot be directly computed following this procedure. An advantage of the procedure, however, is its computational simplicity.

3. Partitioning X as $(X_j : \bar{X}_j)$, where \bar{X}_j is the $n \times (K-K_j)$ matrix of observations on the predetermined variables excluded from the j^{th} equation, equation (3) may be written:

$$\begin{pmatrix} X_j' y_j \\ \bar{X}_j' y_j \end{pmatrix} = \begin{pmatrix} X_j' Z_j \delta_j + X_j' \xi_j \\ \bar{X}_j' Z_j \delta_j + \bar{X}_j' \xi_j \end{pmatrix} \quad (3')$$

Theil's D_j -class estimator (d_j^*) is based on constrained estimation from the second subset of (3'), using some positive definite matrix $\sigma_{jj} D_j$ in place of $\sigma_{jj} \bar{X}_j' \bar{X}_j$ which is singular when $n < (K-K_j)$; see Theil [9]. The constraint, from the systematic part of the first subset of (3'), is $X_j' y_j = X_j' Z_j d_j^*$. Defining $C_j = \bar{X}_j D_j^{-1} \bar{X}_j'$, d_j^* is obtained by solving:

$$\begin{pmatrix} Z_j' C_j Z_j & : & Z_j' X_j \\ X_j' Z_j & : & -0 \end{pmatrix} \begin{pmatrix} d_j^* \\ \lambda_j \end{pmatrix} = \begin{pmatrix} Z_j' C_j y_j \\ X_j' y_j \end{pmatrix} \quad (7)$$

where λ_j is a vector of Lagrangian multipliers. In practice, Theil suggests that D_j be diagonal, with diagonal elements taken from the diagonal of $\bar{X}_j' \bar{X}_j$. There are several disadvantages to the D_j -class

estimators. The choice of D_j is arbitrary and d_j^* is sensitive to this choice; Theil's suggested choice disposes of a fair amount of information contained in $X'X$. The reduced form is explicitly bypassed. The computational burden is roughly the same as for the two-stage least squares estimator. The small sample properties of d_j^* are unknown; Theil shows that it is a consistent estimator, but since its asymptotic covariance matrix differs from that of two-stage least squares it is not efficient (in the limited information sense).

IV. A PROPOSED CLASS OF ESTIMATORS

In partitioned form, we have:

$$X'X = \begin{pmatrix} X_j'X_j & : & X_j'\bar{X}_j \\ \bar{X}_j'X_j & : & \bar{X}_j'\bar{X}_j \end{pmatrix} \quad (8)$$

Since $X_j'X_j$ is positive definite by assumption, if we "disturb" $\bar{X}_j'\bar{X}_j$ slightly by adding to it any (symmetric) positive definite matrix A_j (so that in Theil's notation, $D_j = \bar{X}_j'\bar{X}_j + A_j$), a comparison of the quadratic forms associated with $X'X$ and

$$V_j = \begin{pmatrix} X_j'X_j & : & X_j'\bar{X}_j \\ \bar{X}_j'X_j & : & D_j \end{pmatrix} \quad (9)$$

shows that V_j is positive definite.

The partitioned inverse of V_j may be written:

$$V_j^{-1} = \begin{pmatrix} (X_j'X_j - X_j'C_jX_j)^{-1} & : & -(X_j'X_j - X_j'C_jX_j)^{-1}X_j'\bar{X}_jD_j^{-1} \\ -(D_j - \bar{X}_j'E_j\bar{X}_j)^{-1}\bar{X}_j'X_j(X_j'X_j)^{-1} & : & (D_j - \bar{X}_j'E_j\bar{X}_j)^{-1} \end{pmatrix} \quad (10)$$

where $E_j = X_j(X_j'X_j)^{-1}X_j'$ and (retaining Theil's notation) $C_j = \bar{X}_j D_j^{-1} \bar{X}_j'$.

Using (10), we define:

$$N_j = XV_j^{-1}X' = [X_j(X_j'X_j - X_j'C_jX_j)^{-1}X_j'] [I - C_j] + [\bar{X}_j(D_j - \bar{X}_j'E_j\bar{X}_j)^{-1}\bar{X}_j'] [I - E_j]. \quad (11)$$

It follows immediately from equation (11) that N_j is symmetric and that

$N_jX_j = X_j$; therefore $X_j'N_j = X_j'$. The estimator for δ_j based on N_j ($\tilde{\delta}_j$), is obtained by simply replacing E with N_j in equation (4), giving:

$$\tilde{\delta}_j = (Z_j'N_jZ_j)^{-1}Z_j'N_jy_j \quad (12)$$

The estimator $\tilde{\delta}_j$ has several desirable properties. It is a true limited-information estimator. In terms of computational difficulty, it is equivalent to two-stage least squares. Under the usual assumptions (see, for example, Theil [8, Chapter 10]), it is also asymptotically equivalent to two-stage least squares, assuming $\text{plim } n^{-1}A_j = 0$ since then $\text{plim } n^{-1}V_j = \text{plim } n^{-1}X'X$. Thus, $\tilde{\delta}_j$ is consistent, asymptotically efficient (in the limited information sense), and asymptotically normally distributed with mean δ_j and a covariance matrix which is consistently estimated by:

$$\tilde{S}_{jj} (Z_j'N_jZ_j)^{-1} Z_j'N_jN_jZ_j (Z_j'N_jZ_j)^{-1} \quad (13)$$

where

$$\tilde{S}_{jj} = \frac{1}{n - K_j - L_j} (y_j - Z_j\tilde{\delta}_j)' (y_j - Z_j\tilde{\delta}_j) \quad (14)$$

is; by the above, a consistent estimator for σ_{jj} .

Further, a consistent (but biased) estimate of the reduced form parameters of the system (II) is obtained from:

$$\tilde{\Pi} = V_j^{-1} X'Y \quad . \quad (15)$$

Let $\tilde{\Pi}_j$ represent the columns of $\tilde{\Pi}$ corresponding to Y_j . Note that $X\tilde{\Pi}_j = N_j Y_j = \tilde{Y}_j$, so $N_j Z_j = (\tilde{Y}_j : X_j)$. Given projections of the pre-determined variables of the system, $X^P = (X_j^P : \bar{X}_j^P)$, we may project Y_j from

$$y_j^P = X^P \tilde{\Pi}_j \quad (16)$$

and then, defining $Z_j^P = (Y_j^P : X_j^P)$, project y_j from $Z_j^P \tilde{\delta}_j$.

While the proposed estimator $\tilde{\delta}_j$ is defined for any suitable choice of A_j , in practice we suggest specifying $A_j = aI$ where $0 < a < \infty$. Simplicity is, of course, a major advantage of this specification. In addition, our (quite limited) experience with this specification, reported below, suggests that the elements of $\tilde{\delta}_j$ are reasonably stable over fairly large ranges of a . Our current research is directed in part toward finding the "optimal" value of a for a given equation. A second direction for research is the small sample properties of $\tilde{\delta}_j$.

V. ESTIMATION OF KLEIN'S MODEL I - AN ILLUSTRATION

Although the sample underlying Klein's Model I is not undersized ($n=21, K=8$), it has the advantage that it has been estimated using all of the alternative procedures previously discussed, including two-stage least squares, so that a numerical comparison of the various procedures is possible. The model consists of three behavioral equations:

$$C_t = \gamma_1 P_t + \gamma_2 (W_t + W'_t) + \beta_1 P_{t-1} + \beta_0 + \xi_t \quad (17)$$

$$I_t = \gamma'_1 P_t + \beta'_1 P_{t-1} + \beta'_2 K_{t-1} + \beta'_0 + \xi'_t \quad (18)$$

$$W_t = \gamma''_1 X_t + \beta''_1 X_{t-1} + \beta''_2 (t-1931) + \beta''_0 + \xi''_t \quad (19)$$

where t is measured in calendar years, C is consumption, P profits, W the private wage bill, W' the government wage bill, I net investment, K capital stock at the end of the year, and X the output of the private sector. The six endogenous variables are C , P , I , W , X and K ; the model is closed by three definitional equations. The eight predetermined variables consist of three lagged endogenous variables, P_{-1} , K_{-1} , X_{-1} and t , W' , 1 (the constant), T (business taxes), and G (government nonwage expenditure). In (17), $W+W'$ is considered one endogenous variable. The underlying data is available in Theil [8, page 456].

Point estimates of coefficients, their asymptotic standard errors, and estimated variances are shown in the accompanying table. For the procedure proposed in this paper, coefficient point estimates are from equation (12), standard errors are square roots of the diagonals from equation (13), variances are from equation (14), and we have specified $A_j = aI$.

Using the full sample ($n=21$), the proposed estimator with $a=1$ gives results which are virtually identical to two-stage least squares. This result is to be expected, since when $n > K$, the proposed procedure converges to the two-stage least squares procedure as $a \rightarrow 0$. With $a=21$, coefficients on the highly correlated variables P and P_{-1} in equations (17)

and (18) and X and X_{-1} in equation (19) diverge somewhat from the two-stage least squares estimates. Standard errors, however, are quite similar. For Theil's D_j -class procedure, standard errors tend to be larger in all equations, and the divergence of coefficients on P and P_{-1} in equations (17) and (18) from the two-stage least squares estimates is greater than for the proposed procedure, but there is no divergence for any coefficient in equation (19). The Kloek and Mennes principal components procedure performs quite well in equation (18), but the coefficient of P in equation (17) and of X and X_{-1} in equation (19) diverge somewhat from the two-stage least squares estimates. These results, of course, are no more than suggestive of the relative merits of the alternative procedures.

To illustrate the proposed procedure when the sample is undersized, Klein's Model I was estimated for $n=7$, where the observations are those for 1922, 25, 28, 31, 34, 37 and 1940. These years are fairly representative of the full 21 year observation period. Note that when $n < K$, as $a \rightarrow 0$ the proposed procedure converges to ordinary least squares, which normally coincides with the procedure of Swamy and Holmes [6] and Fischer and Wadycki [2].

ALTERNATIVE PARAMETER ESTIMATES OF KLEIN'S MODEL I

Estimation Procedure	Equation 17 (C)					Equation 18 (I)					Equation 19 (W)				
	γ_1 (P)	γ_2 (W+W')	β_1 (P ₋₁)	β_0 (1)	σ_{ϵ}^2	γ_1 (P)	β_1 (P ₋₁)	β_2 (K ₋₁)	β_0 (1)	σ_{ϵ}^2	γ_1 (X)	β_1 (X ₋₁)	β_2 (t-1931)	β_0 (1)	σ_{ϵ}^2
1. Two-stage least squares (n=21)	.02 (.13)	.81 (.04)	.22 (.12)	16.6 (1.5)	1.29	.15 (.19)	.62 (.18)	-.16 (.04)	20.3 (8.4)	1.71	.44 (.04)	.15 (.04)	.13 (.03)	1.5 (1.3)	.59
2. Proposed estimator (n=21, a=1)	.02 (.13)	.81 (.04)	.21 (.12)	16.5 (1.5)	1.28	.14 (.19)	.62 (.18)	-.16 (.04)	20.5 (8.5)	1.73	.44 (.04)	.15 (.04)	.13 (.03)	1.5 (1.3)	.59
3. Proposed estimator (n=21, a=21)	.05 (.13)	.81 (.04)	.19 (.12)	16.4 (1.4)	1.20	.12 (.21)	.64 (.20)	-.16 (.04)	21.3 (8.9)	1.85	.41 (.04)	.17 (.05)	.14 (.03)	1.6 (1.3)	.61
4. Theil's D _j -class ^{1/} (n=21)	.09 (.19)	.82 (.04)	.15 (.16)	16.1 (1.4)	1.12	.06 (.32)	.69 (.29)	-.17 (.06)	23.0 (11.9)	2.12	.44 (.09)	.15 (.09)	.13 (.03)	1.5 (1.3)	.59
5. Kloek and Mennes ^{2/} (n=21)	-.00 (.19)	.81 (.05)	.23 (.16)		1.35	.15 (.21)	.62 (.19)	-.16 (.04)		1.72	.40 (.05)	.18 (.05)	.14 (.03)		.64
6. Proposed estimator (n=7, a=1)	.12 (.13)	.83 (.06)	.26 (.18)	13.4 (2.1)	.65	.21 (.06)	.59 (.06)	-.18 (.01)	23.2 (2.8)	.04	.36 (.07)	.19 (.07)	.15 (.06)	3.9 (2.4)	.54
7. Proposed estimator (n=7, a=7)	.08 (.16)	.82 (.07)	.30 (.20)	13.6 (2.3)	.76	.14 (.13)	.66 (.13)	-.19 (.03)	26.0 (5.6)	.07	.37 (.07)	.19 (.07)	.15 (.06)	3.8 (2.4)	.55
8. Ordinary least squares ^{3/} (n=7)	.19 (.12)	.83 (.06)	.19 (.16)	13.1 (1.9)	.59	.23 (.05)	.57 (.06)	-.17 (.01)	22.3 (2.5)	.04	.34 (.07)	.20 (.07)	.15 (.06)	4.1 (2.4)	.53

^{1/} From Theil [9], pages 123 and 124. Variances and standard errors have been corrected for degrees of freedom.

^{2/} From Kloek and Mennes [3], page 59. The results are those using two principal components for all three equations and their method 4, in which principal components are computed for all predetermined variables. Variables were measured as deviations from means and therefore no constant was reported. Note that they report σ rather than σ^2 .

^{3/} Since $n < K$, OLS corresponds to the procedure of Swamy and Holmes [6] and Fischer and Wadycki [2]. Standard errors computed as if Z_j contained only predetermined variables.

REFERENCES

1. Dutta, M. and E. Lyttkens, "Iterative Instrumental Variables Method and Estimation of a Large Simultaneous System", Journal of the American Statistical Association, 69,348 (December 1974), 977-986.
2. Fischer, Walter D. and Walter J. Wadycki, "Estimating a Structural Equation in a Large System", Econometrica, 39,3 (May 1971), 461-465.
3. Kloek, T. and L.B.M. Mennes, "Simultaneous Equations Estimation Based on Principal Components of Predetermined Variables", Econometrica, 28,1 (January 1960), 45-61.
4. Mariano, Roberto S., "The Existence of Moments of the Ordinary Least Squares and Two-Stage Least Squares Estimators", Econometrica, 40,4 (July 1972), 643-652.
5. Sawa, Takamitsu, "Finite-Sample Properties of the k-Class Estimators", Econometrica, 40,4 (July 1972), 653-680.
6. Swamy, P.A.V.B. and James Holmes, "The Use of Undersized Samples in the Estimation of Simultaneous Equation Systems", Econometrica, 39,3 (May 1971), 455-459.
7. Takeuchi, K., "Exact Sampling Moments of the Ordinary Least Squares, Instrumental Variable, and Two-Stage Least Squares Estimators", International Economic Review, 11,1 (February 1970), 1-12.
8. Theil, Henri, Principles of Econometrics, New York: John Wiley and Sons, 1971.
9. Theil, Henri, "A Simple Modification of the Two-Stage Least-Squares Procedure for Undersized Samples", in Arthur S. Goldberger and Otis Dudley Duncan (eds.), Structural Equation Models in the Social Sciences, New York: Seminar Press, 1973, 113-129.