DERIVING ESTIMATES OF STRUCTURAL PARAMETERS FROM ESTIMATES OF REDUCED FORM PARAMETERS

Donald T. Sant

ECONOMETRIC RESEARCH PROGRAM
Research Memorandum No. 181
July 1975

The research described in this paper is supported by NSF Grant SOC74-11937. This paper was typed by Alice Furth.

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, New Jersey

DERIVING ESTIMATES OF STRUCTURAL PARAMETERS FROM ESTIMATES OF REDUCED FORM PARAMETERS

Donald Sant

I. Introduction

Structural equations are generally used in econometric research because they are more "understandable" in terms of economic theory. The causal relationships and the parameter restrictions that are derived from economic theory are more easily expressed and interpreted in terms of the structural equations. However, it is the coefficients of the reduced form which are the parameters of the stochastic process generating the observed random variables. Knowing the reduced form parameters permits one to make probabilistic statements about the interaction of economic phenomena, but the structural parameters are generally needed to test implications derived from economic theory. It is basically this distinction that makes the identification problem important in the linear simultaneous equation model. Under the usual assumptions, the reduced form parameters are always estimable, but only when there are restrictions on the structural coefficients does there exist a function giving the structural coefficients from the reduced form coefficients. This paper derives the particular function used implicitly in the usual single equation methods of estimating a single identified (over identified) equation. The reduced form coefficients estimated by least squares will generally not satisfy all the restrictions from an over identified system, so the inverse transformation, the transformation of the reduced form coefficients to the structural parameters, varies with the estimating technique used. From these transformations, it is easy to see that the usual single equation estimating methods are asymptotically equivalent. It is also easy to see that for an exactly identified equation, LIML, TSLS, and indirect least squares are identical.

II. The Model

Consider the single structural equation,

$$(1) \qquad y = Z\delta + \varepsilon = [Y \quad X_1]\begin{pmatrix}\gamma \\ \beta\end{pmatrix} + \varepsilon$$

where $\delta' = \begin{pmatrix}\gamma \\ \beta\end{pmatrix}$ is an unknown parameter vector, and Y is correlated with ε. This single equation is part of a larger system of equations which has the reduced form

$$(2) \qquad [y \quad Y] = [X_1 \quad X_2]\begin{bmatrix}\pi_1 & \Pi_1 \\ \pi_2 & \Pi_2\end{bmatrix} + V$$

where y is $n\times1$, Y is $n\times G$, X_1 is $n\times K_1$, X_2 is $n\times K_2$, π_1 is $K_1\times1$, π_2 is $K_2\times1$, Π_1 is $K_1\times G$, Π_2 is $K_2\times G$, γ is $G\times1$, β is $K_1\times1$, and δ is $(G+K_1)\times1$.

If we define

$$A = \begin{bmatrix}\Pi_1 & I_{K_1\times K_1} \\ \Pi_2 & O_{K_2\times K_1}\end{bmatrix}$$

the relationship between structural coefficients and reduced form coefficients is

$$A\delta = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} .$$

The necessary and sufficient condition for identification is that

$$\text{rank}[\pi_2 \quad \Pi_2] = G ,$$

Since least squares will yield "good" estimates of the reduced form $\begin{pmatrix} \pi_1 & \Pi_1 \\ \pi_2 & \Pi_2 \end{pmatrix}$, the estimation problem is to find good estimates of δ .

III. Two-Stage Least Squares

Consider first the two-stage least squares estimator.

$$\hat{\delta} = (Z'X(X'X)^{-1}X'Z)^{-1}Z'X(X'X)^{-1}X'y$$

where $X = (X_1 \quad X_2)$. The estimator $\hat{\delta}$ is also the result of the minimization problem

(3) $$\min_{\delta}(p - \hat{A}\delta)'X'X(p - \hat{A}\delta)$$

where $p = (X'X)^{-1}X'y$ is the unrestricted least squares estimator of the reduced form

$$\pi = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}$$

and $\hat{A} = \begin{pmatrix} P & I \\ & O \end{pmatrix}$ where $P = (X'X)^{-1}X'Y$

is the unrestricted least squares estimator of the reduced form

$$\Pi = \begin{pmatrix} \Pi_1 \\ \Pi_2 \end{pmatrix} .$$

Proof: The δ which minimizes expression (3) is

$$\hat{\delta} = (\hat{A}'X'X\hat{A})^{-1}\hat{A}'X'Xp$$

but $\hat{A}'X'X = \begin{bmatrix} Y'X(X'X)^{-1} \\ I \quad\quad O \end{bmatrix} X'X$

$$= \begin{bmatrix} Y'X \\ X_1'X \end{bmatrix} = Z'X$$

so substitution yields the desired result.

As a corollary:

The minimization problem will yield a consistent estimator of δ if $\hat{\pi}$ is any consistent estimator of π, $\hat{\Pi}$ is any consistent estimator of Π, and $\lim_{n\to\infty} \frac{X'X}{n} = M$ nonsingular.

<u>Proof</u>:

$$\text{Plim } \hat{\delta} = \text{Plim } (\hat{A}' \frac{X'X}{n} \hat{A})^{-1} \hat{A}' \frac{X'X}{n} \hat{\pi}$$

$$= (A'MA)^{-1} A'MA\delta = \delta$$

since $\pi = A\delta$.

Besides giving us the inverse transformation from the estimated reduced form to the structural estimates, this minimization has a natural statistical justification. The true parameters satisfy the relationship

$$A\delta = \pi \quad .$$

The variance-covariance matrix of the estimated π (using least-squares) is proportional to $(X'X)^{-1}$, so the above minimization problem is the classical modified minimum Chi-Square method of estimation which yields estimators with good large sample properties.1/

One further result is that TSLS based on the true Π is not necessarily better than TSLS based on the estimated Π . If σ^2 is the variance of the structural error of equation 1 and w^2 is the variance of the reduced form error in equation 2 corresponding to the random variable y , using the true Π improves asymptotic efficiency only if $w^2 \leq \sigma^2$.

The asymptotic distribution of $\sqrt{n}(\hat{\delta} - \delta)$ under the usual assumptions is $N(O, \sigma^2(A'MA)^{-1})$. Under the same assumptions, if δ^* is the TSLS estimator using the true Π , $\sqrt{n}(\delta^* - \delta) \xrightarrow{\ell} N(O, w^2(A'MA)^{-1})$, where $\xrightarrow{\ell}$ means converges in distribution.

<u>Proof</u>:

$$\delta^* = (A'X'XA)^{-1}A'X'Xp$$

$$\sqrt{n}(\delta^* - \delta) = \sqrt{n}[(A'X'XA)^{-1}A'X'Xp - (A'X'XA)^{-1}A'X'X\pi]$$

since $\pi = A\delta$.

$$\sqrt{n}(\delta^* - \delta) = \left(\frac{A'X'XA}{n}\right)^{-1}A' \frac{X'X}{n}[\sqrt{n}(p - \pi)]$$

but under the usual assumptions the least squares estimator p has a normal asymptotic distribution

$$\sqrt{n}(p - \pi) \overset{\ell}{\to} N(O, w^2M^{-1})$$

and since

$$\lim_{n\to\infty}(A'\frac{X'XA}{n})^{-1}A'\frac{X'X}{n} = (A'MA)^{-1}A'M$$

it follows that

$$\sqrt{n}(\delta^* - \delta) \overset{\ell}{\to} N(O, w^2(A'MA)^{-1}) ,$$

since it is the product of a sequence converging to $(A'MA)^{-1}A'M$ and a random vector $\sqrt{n}(p - \pi)$ which converges to a normal random vector.

It should be noted that even though asymptotic efficiency is ambiguous, the exact expectation and variance is known for δ^* . $E(\delta^*) = \delta$ and $E(\delta^* - \delta)(\delta^* - \delta)' = w^2(A'X'XA)^{-1}$.

IV. Limited Information Maximum Likelihood

The modified minimum Chi-Square method of estimation does not use the sampling variation of P in forming the estimate of δ. In imposing the rank restriction for identification, the resulting reduced form estimator has the form

$$\begin{bmatrix} P_1\hat{\gamma}+\hat{\beta} & P_1 \\ P_2\hat{\gamma} & P_2 \end{bmatrix}$$

where P_1 and P_2 correspond to the same partition of the least squares estimators as Π_1 and Π_2. The estimated reduced form has the appropriate rank

$$\text{rank}[P_2\hat{\gamma} \quad P_2] = G$$

but in some sense, TSLS has placed all the burden of the restrictions on the first column.

Defining the $K\times(G+1)$ matrices

$$\overline{P} = (p \quad P) \qquad \overline{\Pi} = (\pi \quad \Pi)$$

the minimum Chi-Square estimator is the result of the minimization problem

$$(4) \qquad \min_{\substack{\overline{\Pi} \\ \text{rank } \overline{\Pi}_2=G}} \text{tr}[(\overline{P}-\overline{\Pi})'X'X(\overline{P}-\overline{\Pi})S^{-1}]$$

where $S = n^{-1}[(y \quad Y)-X\bar{P}]'[(y \quad Y)-X\bar{P}]$

is the estimate of the variance covariance matrix of the reduced form errors.[2/]

To derive the minimizing value of $\bar{\Pi}$, a reparameterization is useful. If we let $w = (I_n - X_1(X_1'X_1)^{-1}X_1')X_2 = M_1X_2$ (the residuals from a regression of X_2 on X_1) and define

$$\Gamma_1 = \bar{\Pi}_1 + (X_1'X_1)^{-1}X_1'X_2\bar{\Pi}_2$$

$$\Gamma_2 = \bar{\Pi}_2$$

the reduced form is now

$$(y \quad Y) = (X_1 \quad W)\begin{pmatrix}\Gamma_1 \\ \Gamma_2\end{pmatrix} + V$$

and

$$\begin{pmatrix}X_1' \\ W'\end{pmatrix}(X_1 \quad W) = \begin{pmatrix}X_1'X_1 & 0 \\ 0 & W'W\end{pmatrix} .$$

The implied constraints are

$$\Gamma_1\begin{pmatrix}1 \\ -\gamma\end{pmatrix} = \beta$$

$$\Gamma_2\begin{pmatrix}1 \\ -\gamma\end{pmatrix} = 0$$

giving rank $\Gamma_2 = G$.

Letting $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ represent the unrestricted least squares estimators of this system, the minimization problem (4) then reduces to

$$(5) \qquad \min_{\substack{\Gamma_2 \\ \text{rank } \Gamma_2 = G}} \operatorname{tr}[(\hat{\Gamma}_2-\Gamma_2)'W'W(\hat{\Gamma}_2-\Gamma_2)S^{-1}]$$

because of the orthogonality of X_1 and W . The solution to this problem is a straightforward application of the following lemma.

Lemma: If Q is a $J\times K$ matrix, $J\leq K$ and B is $J\times K$, rank $B=b\leq$ rank $Q\leq J$, the matrix B which minimizes either

$$\operatorname{tr}(Q-B)(Q-B)' \quad \text{or}$$

$$|(Q-B)(Q-B)'|$$

is $B = Q - \sum_{i=b+1}^{J} L_i L_i' Q$

where the L_i are the latent vectors corresponding to the $J-b-1$ smallest latent roots of QQ' .

The proof of this lemma follows from the Courant-Fischer min-max theorem, Bellman (1960 p. 113), but was also proven by Eckhart-Young (1936).

If we factor S^{-1} and $W'W$ such that $CC' = S^{-1}$, C a nonsingular $(G+1)\times(G+1)$ matrix and $D'D = W'W$ where D is a nonsingular $K_2\times K_2$ matrix the minimization problem is

(6) $$\min_{\substack{\Gamma_2 \\ \text{rank }(\Gamma_2)=G}} \operatorname{tr}(D\hat{\Gamma}_2C - D\Gamma_2C)(D\hat{\Gamma}_2C - D\Gamma_2C)'$$

which is the same problem as in the lemma (assuming $K_2 \geq G$) with

$$D\hat{\Gamma}_2C = Q \quad \text{and} \quad D\Gamma_2C = B \ .$$

From the lemma then, the minimizing value of $D\Gamma_2C$ is

$$D\Gamma_2C = D\hat{\Gamma}_2C - L_{G+1}L'_{G+1}D\hat{\Gamma}_2C$$

where L_{G+1} is the latent vector corresponding to the smallest latent root of

$$D\hat{\Gamma}_2S^{-1}\hat{\Gamma}'_2D' \ .$$

This root is equivalent to the smallest value of ℓ such that

$$|\hat{\Gamma}'_2W'W\hat{\Gamma}_2 - \ell S| = 0$$

which is the same as the root calculated by limited information maximum likelihood.[3/]

This minimization problem has an heuristic appeal besides the statistical property of yielding an estimator with good large sample properties. Since the least-squares estimator is a "good" estimator, and the true parameter matrix is assumed to have a certain rank, we are approximating our good estimator by another matrix with the specified rank.

V. Equivalence of the Estimator

Whatever the motivation is for choosing a particular estimating technique, using this framework permits one to easily see the asymptotic equivalence of TSLS and LIML. First, in the exactly identified case, where the dimension of $\bar{\Pi}_2$ is $G\times(G+1)$ $(K_2=G)$, expression (3) and (4) can be made identically equal to zero. The estimate of δ will then satisfy

$$\hat{A}\hat{\delta} = p$$

for both TSLS and LIML. There is no need to approximate the least squares estimators of the reduced form since they satisfy the rank condition. In the over identified case $(K_2>G)$, the estimators of δ will differ in finite samples but will be asymptotically equivalent. Under the usual assumptions, the least squares estimators will be consistent estimators of the reduced form parameters. Thus asymptotically, they will be of the appropriate rank. Expression (3) can then be made identically equal to zero by the value of δ such that $A\delta = \pi$. The estimate of δ found by LIML is the value of $(\gamma \quad \beta)$ which satisfies

$$\text{(7)} \qquad \hat{\Gamma}_1\begin{pmatrix}1\\-\gamma\end{pmatrix} = \beta$$

$$\hat{\Gamma}_2\begin{pmatrix}1\\-\gamma\end{pmatrix} = 0$$

where $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ are the reestimated reduced form parameters. Asymptotically though, $\hat{\Gamma}_2$ will converge to $\bar{\Pi}_2$. Thus the solution to (7) δ^*, and the TSLS estimator $\hat{\delta}$ are constructed in such a way that

$$\text{Plim } \sqrt{n}(\delta^* - \hat{\delta}) = 0 \ .$$

This implies (Rao 1965, p. 101) that they have the same limiting distribution.

VI. The k-Class Estimators

The k-class estimators have been a natural organizing framework for comparing single equation estimating methods. This section derives a relationship between the actual reduced form and the structural parameters which is true of all the k-class estimators.

From the relationship between structural parameters and reduced form parameters, β is given by

$$\beta = \pi_1 - \Pi_1 \gamma \ . \tag{8}$$

Given consistent estimates of π_1, Π_1, and γ, a consistent estimator of β could be found by using 8. The actual estimator used in the k-class is given by

Proposition: For the k-class estimators

$$\hat{\beta} = p_1 - P_1\hat{\gamma} + (X_1'X_1)^{-1}X_1'X_2(p_2 - P_2\hat{\gamma})$$

where p_1, P_1, p_2, P_2 are the least squares estimators of the reduced form parameters, and $\hat{\gamma}$ is the k-class estimator for γ.

Proof: The estimators $(\hat{\gamma} \quad \hat{\beta})$ satisfy

$$\begin{bmatrix} Y'(I_n - kM)Y & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix} \begin{bmatrix} \hat{\gamma} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} Y'(I_n - kM)y \\ X_1'y \end{bmatrix}$$

where $M = I_n - X(X'X)^{-1}X'$. The second row gives

(9) $$X_1'Y\hat{\gamma} + X_1'X_1\hat{\beta} = X_1'y \ .$$

The normal equations for the least squares estimators give

(10) $$(X_1'X_1 \quad X_1'X_2) \begin{bmatrix} p_1 & P_1 \\ p_2 & P_2 \end{bmatrix} = (X_1'y \quad X_1'Y) \ .$$

Postmultiplying (10) by $\begin{pmatrix} 1 \\ -\hat{\gamma} \end{pmatrix}$ yields

(11) $$X_1'X_1(p_1 - P_1\hat{\gamma}) + X_1'X_2(p_2 - P_2\hat{\gamma}) = X_1'y - X_1'Y\hat{\gamma} \ .$$

Inserting (11) into (9) and solving for $\hat{\beta}$ gives

$$\hat{\beta} = (p - P_1\hat{\gamma}) + (X_1'X_1)^{-1}X_1'X_2(p_2 - P_2\hat{\gamma}) \ .$$

VII. Conclusion

The sufficient conditions for the identification of structural parameters require that a certain submatrix of the reduced form have a specified rank.

In an overidentified equation, the unrestricted least squares estimators will not have the appropriate rank. This paper derives the relationships between reduced form estimates (using least squares) and the structural estimators (of the standard estimating techniques) and how the restrictions are incorporated in this transformation.

Footnotes

1/ More exactly it is a modified reduced minimum Chi-Square method since observed values of A are used. See Ferguson (1958) and Neyman (1949). See Basmann (1960) for a similar derivation of TSLS.

2/ See Rothenberg (1973) pp. 24, 81-82 for a discussion of minimum Chi-Square estimation.

3/ A related result has been shown by Goldberger and Olkin (1971) and by Malinvaud (1970) pp. 702-706, but the lemma used in the derivation here seems to have wider applicability in econometrics than has been previously used.

References

Basmann, R. L., 1960, "An Expository Note on Estimation of Simultaneous Structural Equations." Biometrics 16:464-480.

Bellman, R., 1960, Introduction to Matrix Analysis. New York: McGraw-Hill.

Eckhart, C. and Young, G., 1936, "The Approximation of one Matrix by Another of Lower Rank." Psychometrika 1:211-218.

Ferguson, T. S., 1958, "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities." Annals of Mathematical Statistics 29:1046-62.

Goldberger, A. S., 1964, Econometric Theory. New York: Wiley.

Goldberger, A. S. and Olkin, I., 1971, "A Minimum-Distance Interpretation of Limited-Information Estimation." Econometrica 40:979-1002.

Malinvaud, E., 1970, Statistical Methods of Econometrics. New York: American Elsevier.

Neyman, J., 1949, "Contribution to the Theory of the χ^2 Test." Berkeley Symposium on Mathematical Statistics and Probability. University of California Press.

Rao, C. R., 1965, Linear Statistical Inference and Its Applications. New York: Wiley.

Rothenberg, T. J., 1973, Efficient Estimation with A Priori Information. New Haven: Yale University Press.