Econometrica, Vol. 38, No. 3 (May, 1970)

THE ESTIMATION OF SIMULTANEOUS EQUATION MODELS WITH LAGGED ENDOGENOUS VARIABLES AND FIRST ORDER SERIALLY CORRELATED ERRORS

BY RAY C. FAIR[1]

In this paper various methods for the estimation of simultaneous equation models with lagged endogenous variables and first order serially correlated errors are discussed. The methods differ in the number of instrumental variables used. The asymptotic and small sample properties of the various methods are compared, and the variables which must be included as instruments to insure consistent estimates are derived. A suggestion on how to estimate the approximate covariance matrix of the estimators is made.

1. INTRODUCTION

RECENTLY SARGAN [**8**] has proposed various maximum likelihood estimators for the estimation of simultaneous equation models with serially correlated errors, and Amemiya [**1**] has considered the two stage least squares analogue to one of Sargan's estimators and has proposed a modified version of this analogue. Because of the large number of instrumental variables which it uses, Sargan's method (or the two stage least squares analogue) is likely to be of limited practical use, and this paper discusses which of Sargan's instrumental variables should be retained in order to insure consistent estimates. One method is proposed that is asymptotically equivalent to Sargan's method, but which uses fewer instrumental variables and may have less small sample bias. Further suggestions are made for substantially decreasing the number of instrumental variables with perhaps small loss of asymptotic efficiency. Amemiya's method is then briefly discussed and compared with the methods proposed in this paper. The paper concludes with a discussion of the asymptotic covariance matrices of the estimators.

2. THE MODEL

The model to be estimated is

$$AY + BX = U, \tag{1}$$

where

$$U = RU_{-1} + E. \tag{2}$$

Y is an $h \times T$ matrix of endogenous variables; X is a $k \times T$ matrix of predetermined (i.e., both exogenous and lagged endogenous) variables; U and E are $h \times T$ matrices of disturbance terms; and A, B, and R are $h \times h$, $h \times k$, and $h \times h$ coefficient matrices respectively. T is the number of observations. The subscript -1 for U_{-1} denotes the one period lagged values of the terms of U.

Write E as $E = (e(1)\,e(2)\ldots e(T))$, where $e(t) = (e_1(t)\,e_2(t)\ldots e_h(t))'$ is an $h \times l1$ vector of the tth value of the disturbance term. The following assumptions about

[1] I would like to thank F. M. Fisher and H. Kelejian for helpful comments on an earlier draft of this paper.

the model are made:

(i) $\mathscr{E}(E) = 0$;

(ii) $\mathscr{E}e(t)e'(t) = \Sigma, t = 1, 2, \ldots, T, \Sigma$ positive definite;

(iii) $\mathscr{E}e(t)e'(t') = 0, t, t' = 1, 2, \ldots, T, t \neq t'$;

(iv) plim $T^{-1}XE' =$ plim $T^{-1}X_{-1}E' =$ plim $T^{-1}Y_{-1}E' = 0$;

(v) the moment matrix of the endogenous, lagged endogenous, predetermined, and lagged predetermined variables is well behaved in the limit;[2]

(vi) R is a diagonal matrix of elements between minus one and one;

(vii) A has an inverse.

The estimation of the first equation in (1) is the focus of attention. Rewrite this equation as

$$(3) \qquad y_1 = -A_1 Y_1 - B_1 X_1 + u_1,$$

where

$$(4) \qquad u_1 = r_{11} u_{1_{-1}} + e_1;$$

y_1 is a $1 \times T$ vector of values of y_{1t}; Y_1 is an $h_1 \times T$ matrix of endogenous variables (other than the first) included in the first equation; X_1 is a $k_1 \times T$ matrix of predetermined variables included in the first equation; u_1 and e_1 are $1 \times T$ vectors of disturbance terms; r_{11} is the element in the first row and first column of R; and A_1 and B_1 are $1 \times h_1$ and $1 \times k_1$ vectors of coefficients corresponding to the relevant elements of A and B respectively.

From (1) and (2) the reduced form for Y is

$$(5) \qquad Y = -A^{-1}BX + A^{-1}RAY_{-1} + A^{-1}RBX_{-1} + A^{-1}E.$$

Equations (3) and (4) can be written for any value of r:

$$(6) \qquad y_1 - r y_{1_{-1}} = -A_1(Y_1 - rY_{1_{-1}}) - B_1(X_1 - rX_{1_{-1}}) + [(r_{11} - r)u_{1_{-1}} + e_1].$$

3. ESTIMATION METHODS

In (6) e_1 is correlated with Y_1, and $u_{1_{-1}}$ is correlated with $Y_{1_{-1}}$ and with the lagged endogenous variables in X_1 and $X_{1_{-1}}$. The equation can be consistently estimated, however, by the following procedure.

(a) *First stage regression*: Choose a set of instrumental variables which are uncorrelated with e_1 and which at least include $y_{1_{-1}}$, $Y_{1_{-1}}$, X_1, and $X_{1_{-1}}$; regress each row of Y_1 on this set and calculate the predicted values of Y_1 (denoted as $\hat{Y}_1$) from these regressions.

(b) *Second stage regression*: For a given r estimate equation (6) by ordinary least squares, using $\hat{Y}_1 - rY_{1_{-1}}$ in place of $Y_1 - rY_{1_{-1}}$, and calculate the sum of squared residuals of the regression.

[2] See Christ [**2**, p. 354]. It should be noted that in general some of the same variables are included in both the predetermined and lagged endogenous matrices (X and Y_{-1}), but in the moment matrix referred to in assumption (v) these variables are obviously included only once. Likewise, the constant term is included only once, even though strictly speaking it is included in both X and X_{-1}.

(c) *Scanning or iterative procedure*: Repeat (b) for various values of r between minus one and one (or use an iterative procedure),[3] and choose that r and the corresponding estimates of A_1 and B_1 which yield the smallest sum of squared residuals of the second stage regression.

Consistency of this procedure can be seen heuristically as follows. Let $\hat{V}_1 = Y_1 - \hat{Y}_1$. Then the equation estimated in the second stage regression is

$$(7) \qquad y_1 - ry_{1_{-1}} = -A_1(\hat{Y}_1 - rY_{1_{-1}}) - B_1(X_1 - rX_{1_{-1}}) + [(r_{11} - r)u_{1_{-1}} + e_1 - A_1\hat{V}_1].$$

From (3), $u_{1_{-1}} = y_{1_{-1}} + A_1 Y_{1_{-1}} + B_1 X_{1_{-1}}$, and since $y_{1_{-1}}$, $Y_{1_{-1}}$, and $X_{1_{-1}}$ are used as instruments in the first stage regression, by the property of least squares $u_{1_{-1}}$ and $\hat{V}_1$ are orthogonal. By assumption, $u_{1_{-1}}$ and e_1 are uncorrelated. Therefore, the minimum sum of squared residuals of (7) occurs at the point where r equals r_{11}, leaving as the error term $e_1 - A_1\hat{V}_1$, which is uncorrelated with $\hat{Y}_1 - rY_{1_{-1}}$ and $X_1 - rX_{1_{-1}}$.[4]

It is now clear why $y_{1_{-1}}$, $Y_{1_{-1}}$, and $X_{1_{-1}}$ must be used as instruments in the first stage regression: unless $\hat{V}_1$ is orthogonal to $u_{1_{-1}}$, the minimum sum of squared residuals does not necessarily occur at the point where r equals r_{11}. Another way of looking at this is the following. Rewrite equation (7) as

$$(8) \qquad y_1 = r_{11}y_{1_{-1}} - A_1\hat{Y}_1 + r_{11}A_1Y_{1_{-1}} - B_1X_1 + r_{11}B_1X_{1_{-1}} + (e_1 - A_1\hat{V}_1).$$

The general estimation method outlined above consists of choosing estimates of r_{11}, A_1, and B_1 (say $\hat{r}_{11}$, $\hat{A}_1$, and $\hat{B}_1$) such that the sum of squared residuals in (8) is at a minimum. The case where r_{11} is assumed to be zero corresponds to the ordinary two stage least squares method. The error term $e_1 - A_1\hat{V}_1$ in (8) has zero expected value ($\hat{V}_1$ has zero mean value by the property of least squares) and is not correlated with $y_{1_{-1}}$, $\hat{Y}_1$, $Y_{1_{-1}}$, X_1, and $X_{1_{-1}}$ ($\hat{V}_1$ is orthogonal to these variables since $y_{1_{-1}}$, $Y_{1_{-1}}$, X_1, and $X_{1_{-1}}$ are used as instruments in the first stage regression). Equation (8) can thus be considered a nonlinear equation with an additive error term whose properties are sufficient for insuring consistent estimates by minimizing the sum of squared residuals.[5]

[3] An iterative procedure which can be used is the following. From initial estimates of A_1 and B_1 (say $A_1^{(0)}$ and $B_1^{(0)}$), calculate

$$r^{(1)} = \frac{(y_{1_{-1}} + A_1^{(0)}Y_{1_{-1}} + B_1^{(0)}X_{1_{-1}})(y_1 + A_1^{(0)}Y_1 + B_1^{(0)}X_1)'}{(y_{1_{-1}} + A_1^{(0)}Y_{1_{-1}} + B_1^{(0)}X_{1_{-1}})(y_{1_{-1}} + A_1^{(0)}Y_{1_{-1}} + B_1^{(0)}X_{1_{-1}})'};$$

use this value of $r^{(1)}$ to compute new estimates, $A_1^{(1)}$ and $B_1^{(1)}$, of A_1 and B_1; use these estimates to compute $r^{(2)}$; and so on until two successive estimates of r are within a prescribed tolerance level. In practice, this technique has been found to converge quite rapidly.

[4] For the single equation case (i.e., where $A_1 = 0$) see Malinvaud [**6**, p. 469, n. ††] for an outline of the proof that a procedure as in (c) yields consistent and, if e_1 is normally distributed, asymptotically efficient estimates.

[5] Minimizing the sum of squared residuals of (8) with respect to r_{11}, A_1, and B_1 yields the following equation for $\hat{r}_{11}$:

$$\hat{r}_{11} = \frac{(y_{1_{-1}} + \hat{A}_1Y_{1_{-1}} + \hat{B}_1X_{1_{-1}})(y_1 + \hat{A}_1\hat{Y}_1 + \hat{B}_1X_1)'}{(y_{1_{-1}} + \hat{A}_1Y_{1_{-1}} + \hat{B}_1X_{1_{-1}})(y_{1_{-1}} + \hat{A}_1Y_{1_{-1}} + \hat{B}_1X_{1_{-1}})'}.$$

Since $\hat{Y}_1 = Y_1 - \hat{V}_1$ and since $\hat{V}_1$ is orthogonal to $y_{1_{-1}}$, $Y_{1_{-1}}$, and $X_{1_{-1}}$, this equation can be written:

$$\hat{r}_{11} = \frac{(y_{1_{-1}} + \hat{A}_1Y_{1_{-1}} + \hat{B}_1X_{1_{-1}})(y_1 + \hat{A}_1Y_1 + \hat{B}_1X_1)'}{(y_{1_{-1}} + \hat{A}_1Y_{1_{-1}} + \hat{B}_1X_{1_{-1}})(y_{1_{-1}} + \hat{A}_1Y_{1_{-1}} + \hat{B}_1X_{1_{-1}})'},$$

which is the formula used to calculate successive values of r in the iterative technique described in footnote 2.

In Sargan's method all of the predetermined and lagged variables in the model are used as instruments[6] (i.e., all of the variables in X, X_{-1}, and Y_{-1}). From (5) it is seen that these are all of the variables which enter the reduced form for Y_1. Some lagged endogenous variables are included in both Y_{-1} and X, but they are obviously counted only once as instruments. The disadvantage with Sargan's method (denoted, following Amemiya [**1**], as S2SLS) is the large number of instruments which are used. All predetermined and lagged predetermined variables are used, as well as all the lagged endogenous variables which are not already included among the predetermined variables. For even moderately sized models the number of instruments proposed by S2SLS is likely to exceed the number of observations. In addition, if the predetermined variables are strongly correlated with their lagged values or with each other, the matrix to be inverted in the first stage regression may be nearly singular and pose computational difficulties.

Because the (diagonal) elements of R can be consistently estimated, the number of instruments proposed by S2SLS can be decreased with no loss of asymptotic efficiency. While the technique which will now be described is of limited usefulness itself, it does suggest a way in which the number of instruments can be substantially decreased with perhaps small loss of asymptotic efficiency. The technique can best be described by an example. Assume that (1) consists of two equations:

$$\text{(1a)} \quad y_{1t} = -a_{12}y_{2t} - b_{11}x_{1t} - b_{12}x_{2t} - b_{14}y_{1t-1} + u_{1t} \quad (t = 1, 2, \ldots, T),$$

$$\text{(1b)} \quad y_{2t} = -a_{21}y_{1t} - b_{22}x_{2t} - b_{23}x_{3t} - b_{25}y_{2t-1} + u_{2t} \quad (t = 1, 2, \ldots, T),$$

where

$$\text{(2a)} \quad u_{1t} = r_{11}u_{1t-1} + e_{1t} \quad (t = 1, 2, \ldots, T),$$

$$\text{(2b)} \quad u_{2t} = r_{22}u_{2t-1} + e_{2t} \quad (t = 1, 2, \ldots, T).$$

If equation (1a) is to be estimated, then analogous to equation (6) it can be written

$$\text{(6a)} \quad \begin{aligned} y_{1t} - ry_{1t-1} = {} & -a_{12}(y_{2t} - ry_{2t-1}) - b_{11}(x_{1t} - rx_{1t-1}) \\ & -b_{12}(x_{2t} - rx_{2t-1}) - b_{14}(y_{1t-1} - ry_{1t-2}) \\ & +[(r_{11} - r)u_{1t-1} + e_{1t}] \qquad (t = 1, 2, \ldots, T). \end{aligned}$$

The reduced form for y_{2t} (analogous to (5)) is

$$\text{(5b)} \quad \begin{aligned} y_{2t} = {} & \frac{1}{1 - a_{21}a_{12}}[(r_{22} - a_{21}a_{12}r_{11})y_{2t-1} - b_{25}(y_{2t-1} - r_{22}y_{2t-2}) \\ & +a_{21}(r_{22} - r_{11})y_{1t-1} + a_{21}b_{14}(y_{1t-1} - r_{11}y_{1t-2}) \\ & +a_{21}b_{11}(x_{1t} - r_{11}x_{1t-1}) + (a_{21}b_{12} - b_{22})x_{2t} \\ & -(a_{21}b_{12}r_{11} - b_{22}r_{22})x_{2t-1} - b_{23}(x_{3t} - r_{22}x_{3t-1}) \\ & -a_{21}e_{1t} + e_{2t}] \qquad (t = 1, 2, \ldots, T). \end{aligned}$$

[6] See Sargan [**8**, p. 422].

In estimating (6a), S2SLS would use as instruments y_{2t-1}, y_{2t-2}, y_{1t-1}, y_{1t-2}, x_{1t}, x_{1t-1}, x_{2t}, x_{2t-1}, x_{3t}, and x_{3t-1}. As was seen above, y_{1t-1}, y_{1t-2}, y_{2t-1}, x_{1t}, x_{1t-1}, x_{2t}, and x_{2t-1} must be used as instruments to insure consistent estimates. Notice, however, that x_{3t} and x_{3t-1} do not enter as separate variables in the reduced form (5b), but only as $x_{3t} - r_{22}x_{3t-1}$. If a consistent estimate of r_{22} were available (say $\hat{r}_{22}$), then knowledge of this restriction could be used, and x_{3t} and x_{3t-1} need enter the first stage regression only as $x_{3t} - \hat{r}_{22}x_{3t-1}$. This suggests the following procedure. First estimate each equation separately by S2SLS, and then re-estimate each equation using knowledge of the reduced form and of the estimates of the r_{ii} coefficients to decrease the number of instruments used in the first stage regression. Providing it converges, this procedure can be repeated until the estimates of the r_{ii} coefficients from two successive trials are within a prescribed tolerance level. This iterative procedure will be denoted as I2SLS.[7] Notice from the example just given that I2SLS saves instruments only to the extent that a given exogenous variable appears in only one equation of the model. In macroeconomic models, however, with income identities, the possibilities for decreasing the number of instrumental variables used for any one equation (given estimates of the r_{ii} coefficients) are usually greater, as an examination of the reduced form will reveal.

Both S2SLS and I2SLS yield consistent estimates. With respect to asymptotic efficiency, the difference between S2SLS and I2SLS is that S2SLS in effect adds instruments which (in the limit) do not add anything to the explanation of the endogenous variables in the reduced form and which are uncorrelated with the reduced form error term. Instruments which add nothing to the explanation of the endogenous variables in the reduced form and which are uncorrelated with the reduced form error term will be referred to as "unnecessary" instruments. It is shown in the Appendix that adding unnecessary instruments in the two stage least squares technique does not change the asymptotic covariance matrix of the estimator. This implies, therefore, that S2SLS and I2SLS have the same asymptotic efficiency. Even though S2SLS fails to account for certain restrictions in the reduced form, this has no detrimental effect on its asymptotic efficiency.

With respect to small sample properties, the Appendix shows, using a theorem of Nagar [7], that adding unnecessary instruments in the two stage least squares technique increases the bias, to the order T^{-1}, of the estimator.[8] This result is not too surprising, since for small samples adding unnecessary instruments uses up degrees of freedom and does not seem likely to be of any positive benefit. Since S2SLS in effect adds unnecessary instruments only in the limit, it does not necessarily follow from this result that I2SLS has less small sample bias that S2SLS. In the above example, only if r_{22} were known (as opposed to a consistent estimate being available), could the result in the Appendix be directly applied to conclude (footnote 8 aside) that I2SLS had less small sample bias than S2SLS. Intuitively,

[7] I2SLS can be considered to be a special case of the iterative method developed by Nagar and discussed in Theil [9, pp. 354–355].

[8] The theorem of Nagar used in the Appendix has only formally been proven for the case where there are no lagged endogenous variables among the predetermined variables in the model.

however, it would seem that with respect to small sample properties the advantage of saving a degree of freedom by using I2SLS would outweigh the disadvantage of having only a consistent estimate of r_{22} available.

Unlike S2SLS, I2SLS requires knowledge of the reduced form. It is also computationally more expensive and, as mentioned above, in general saves instruments only to the extent that a given exogenous variable appears in only one equation of the model. An alternative method is thus proposed which uses substantially fewer instrumental variables and does not require knowledge of the reduced form.

From (1) and (2) for any value r_0

$$A(Y - r_0 Y_{-1}) + B(X - r_0 X_{-1}) = (R - r_0 I)U_{-1} + E, \tag{9}$$

where I is the $h \times h$ identity matrix. Therefore,

$$Y = r_0 Y_{-1} - A^{-1}B(X - r_0 X_{-1}) + [A^{-1}(R - r_0 I)U_{-1} + A^{-1}E]. \tag{10}$$

Equation (10) states that any endogenous variable, such as y_{it}, can be expressed as a function of $r_0 y_{it-1}$, of all of the predetermined variables in the form $x_{it} - r_0 x_{it-1}$, and of an error term. When all of the serial correlation coefficients in the model are equal (to r_0 say), then R equals $r_0 I$, and the error term in (10) reduces to that in (5). While it is unrealistic to assume that all of the serial correlation coefficients in the model are equal, in many cases it may not be too unrealistic to assume that they are nearly equal (to r_0 say) so that $A^{-1}(R - r_0 I)U_{-1}$ in (10) is reasonably small. If this is true, it suggests that in the estimation of equation (6) the first stage regression should consist of regressing $Y_1 - r_0 Y_{1_{-1}}$ on $X - r_0 X_{-1}$ to get $\widehat{Y_1 - r_0 Y_{1_{-1}}}$ and then computing $\hat{Y}_1$ as $\widehat{Y_1 - r_0 Y_{1_{-1}}} + r_0 Y_{1_{-1}}$. It was seen above, however, that $y_{1_{-1}}$, $Y_{1_{-1}}$, X_1, and $X_{1_{-1}}$ must be included as separate instruments to insure consistent estimates. Thus the suggestion should be modified to state that in the first stage regression Y_1 should be regressed on $y_{1_{-1}}$, $Y_{1_{-1}}$, X_1, $X_{1_{-1}}$, and $X_2 - r_0 X_{2_{-1}}$, where X_2 denotes all the variables in X which are not in X_1.[9] Since the number of variables in X_1 is likely to be small relative to the number in X_2, the number of instruments saved by using $X_2 - r_0 X_{2_{-1}}$ instead of X_2 and $X_{2_{-1}}$ separately is likely to be substantial. Notice also that the only lagged endogenous variables which are used as instruments, other than $y_{1_{-1}}$ and $Y_{1_{-1}}$, are those in X_1 and X_2.

This technique (which will be denoted as X2SLS) is asymptotically less efficient than S2SLS or I2SLS since in general the error term in (10) is larger than the one in (6).[10] Since X2SLS uses substantially fewer instrumental variables and thus substantially fewer degrees of freedom, however, it may, depending on how nearly equal

[9] The value of r_0 must be chosen in advance when using this technique. In an earlier draft of this paper the suggestion was made that for each iteration on r, $X_2 - rX_{2_{-1}}$ be used as instruments for Y_1 in the first stage regression. In this case, however, the r which minimizes the sum of squared residuals of equation (7) will not necessarily equal r_{11}, since $\hat{V}_1$ in (7) will be a function of r and there is no guarantee that $\hat{V}_1$ will be at a minimum for r equal to r_{11}.

[10] In fact, X2SLS does not yield consistent estimates of the reduced form coefficients, since U_{-1} in (10) is correlated with the lagged endogenous variables in the model. Even though the first stage estimates are inconsistent, the estimates of the coefficients of (7) in the second stage will be consistent as long as the error in the second stage is uncorrelated with all of the instrumental variables (which it is at the point where r equals r_{11} in (7)). The proofs of consistency of two stage least squares given in two leading econometric texts, Christ [2] and Goldberger [5], use the assumption that the first stage estimates are consistent, but it is easy to show that this assumption is not necessary.

R and $r_0 I$ are, have better (or at least not worse) small sample properties than S2SLS or I2SLS. From a more practical point of view, if the number of instruments proposed by S2SLS must be reduced because it exceeds or nearly exceeds the number of observations, decreasing the number in the manner suggested by X2SLS may (again depending on how nearly equal R and $r_0 I$ are) lead to a smaller efficiency loss than excluding particular variables in X_2 and $X_{2_{-1}}$.

Amemiya's modification of Sargan's method (which Amemiya [**1**] denotes as MS2SLS) consists in dropping Y_{-1} from Sargan's list of instrumental variables and in the first stage regression, for each value of r, regressing $Y_1 - rY_{1_{-1}}$ on X and X_{-1} to yield $\widehat{Y_1 - rY_{1_{-1}}}$ to be used in the second stage regression. If there are no lagged endogenous variables in X, which Amemiya implicitly assumes, then this technique will result in consistent estimates of A_1 and B_1 in (7) regardless of the value of r chosen, since neither $\widehat{Y_1 - rY_{1_{-1}}}$ nor $X_1 - rX_{1_{-1}}$ will be correlated with $u_{1_{-1}}$, e_1, and $\hat{V}_1$. Given consistent estimates $\hat{A}_1$ and $\hat{B}_1$ of A_1 and B_1, r_{11} can be consistently estimated by the second equation in footnote 4. If there are lagged endogenous variables in X, then Amemiya's method can be modified by treating all of these variables as "endogenous" as well.[11]

Amemiya's method (as just modified to include the case where there are lagged endogenous variables in X) uses fewer instrumental variables than S2SLS, but considerable loss of efficiency is likely to result by treating all of the lagged endogenous variables in the model as endogenous. Against first order serial correlation, Amemiya's method is thus likely to be much less efficient than X2SLS. It does have the one advantage of yielding consistent estimates of A_1 and B_1 under more general assumptions about the autoregressive properties of the errors in the model.

4. ASYMPTOTIC COVARIANCE MATRICES

Let C_1 equal the $1 \times (h_1 + k_1)$ vector $(A_1\ B_1)$ and let $\hat{C}_1$ equal $(\hat{A}_1\ \hat{B}_1)$, where $\hat{C}_1$ denotes the S2SLS, I2SLS, or X2SLS estimate of C_1 when r_{11} is known. When r_{11} is known the asymptotic covariance of $\sqrt{T}(\hat{C}_1 - C_1)$ is

$$\text{(11)} \qquad \operatorname{asy\,cov}[\sqrt{T}(\hat{C}_1 - C_1)', \sqrt{T}(\hat{C}_1 - C_1)] = \sigma_{11} \operatorname{plim} T(Q_1 Q_1')^{-1},$$

where Q_1' is the $T \times (h_1 + k_1)$ matrix $(\hat{Y}_1' - r_{11}Y_{1_{-1}}', X_1' - r_{11}X_{1_{-1}}')$ and σ_{11} is the element in the first row and first column of Σ. $\hat{Y}_1$ is equal to $Y_1 Z'(ZZ')^{-1}Z$, where Z is the $k_0 \times T$ matrix of instrumental variables used by the particular method. From the result in the Appendix, it follows that plim $T(Q_1Q_1')^{-1}$ is the same for S2SLS and I2SLS.

Define the $T \times T$ matrix P_1 such that

$$P_1 = \begin{pmatrix} \sqrt{1 - r_{11}^2} & -r_{11} & 0 & \cdots 0 & 0 \\ 0 & 1 & -r_{11} & \cdots 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots 1 & -r_{11} \\ 0 & 0 & 0 & \cdots 0 & 1 \end{pmatrix}.$$

[11] This technique of treating lagged endogenous variables as endogenous is used by Fisher [**4**].

Let W'_1 equal the $T \times (h_1 + k_1)$ matrix $(Z'(ZZ')^{-1}ZY'_1\ X'_1)$. Then $P'_1W'_1$ equals $(Z'(ZZ')^{-1}ZY'_1 - r_{11}Z'(ZZ')^{-1}ZY'_{1_{-1}}\ X'_1 - r_{11}X'_{1_{-1}})$ except for the first row of $P'_1W'_1$. Since the variables in $Y_{1_{-1}}$ are used as instruments and are thus in Z, $Z'(ZZ')^{-1}Y'_{1_{-1}}$ equals $Y_{1_{-1}}$, and so $P'_1W'_1$ equals Q'_1 except for the first row of $P'_1W'_1$. Therefore, $Q_1Q'_1$ approximately equals $W_1P_1P'_1W'_1$ in (11). It can be shown from the assumptions in Section 2 that $\mathscr{E}(u'_1u_1) = \sigma_{11}\Omega_1$, where $P_1P'_1 = \Omega_1^{-1}$. Therefore, the asymptotic covariance of $\sqrt{T}(\hat{C}_1 - C_1)$ in (11) approximately equals

$$\sigma_{11} \operatorname{plim} T(W_1\Omega_1^{-1}W'_1)^{-1}. \tag{12}$$[12]

Equation (11) was defined for r_{11} known. For purposes of this discussion let $\hat{\hat{C}}_1$ denote the S2SLS, I2SLS, or X2SLS estimates of C_1 when only a consistent estimate $\hat{r}_{11}$ of r_{11} is available. Let D_1 equal the $1 \times (h_1 + k_1 + 1)$ vector $(C_1\ r_{11})$ and let $\hat{D}_1$ equal $(\hat{\hat{C}}_1\ \hat{r}_{11})$. The asymptotic covariance of $\sqrt{T}(\hat{D}_1 - D_1)$ can be derived by approximating equation (8) by the linear terms of the Taylor series expansion about $\hat{D}_1$ and then deriving the asymptotic covariance matrix from the resulting linear equation. This produces

$$\begin{aligned} &\text{asy cov}\,[\sqrt{T}(\hat{D}_1 - D_1)', \sqrt{T}(\hat{D}_1 - D_1)] \\ &\quad = \sigma_{11} \operatorname{plim} T\begin{pmatrix} Q_1Q'_1 & Q_1u'_{1_{-1}} \\ u_{1_{-1}}Q'_1 & u_{1_{-1}}u'_{1_{-1}} \end{pmatrix}^{-1}. \end{aligned} \tag{13}$$

If the probability limit of $T^{-1}Q_1u'_{1_{-1}}$ were zero, then the asymptotic covariance of $\sqrt{T}(\hat{\hat{C}}_1 - C_1)$ in (13) would reduce to (11). But plim $T^{-1}Q_1u'_{1_{-1}}$ is not zero since Q_1 includes lagged endogenous variables.[13] It is easy to show by taking the inverse of (13) that σ_{11} plim $T(Q_1Q'_1)^{-1}$ differs from the true asymptotic covariance of $\sqrt{T}(\hat{\hat{C}}_1 - C_1)$ by a positive semidefinite matrix and thus that (11) underestimates the asymptotic covariance of $\sqrt{T}(\hat{\hat{C}}_1 - C_1)$.[14] Since plim $T^{-1}Q_1u'_{1_{-1}}$ is complicated to evaluate (note that lagged endogenous variables are included among the instrumental variables as well), in practice it probably should be assumed to be zero and the approximate covariance of $\hat{\hat{C}}_1$ estimated as $\hat{\sigma}_{11}(\hat{Q}_1\hat{Q}'_1)^{-1}$, where $\hat{Q}'_1 = \hat{Y}'_1 - \hat{r}_{11}Y'_{1_{-1}}\ X'_1 - \hat{r}_{11}X'_{1_{-1}})$, $\hat{\sigma}_{11} = T^{-1}\hat{u}_1\hat{u}'_1$, and $\hat{u}_1 = y_1 - \hat{r}_{11}y_{1_{-1}} + \hat{\hat{A}}_1(Y_1 - \hat{r}_{11}Y_{1_{-1}}) + \hat{\hat{B}}_1(X_1 - \hat{r}_{11}X_{1_{-1}})$. Since plim $T^{-1}u_{1_{-1}}u'_{1_{-1}}$ equals $\sigma_{11}/(1 - r_{11}^2)$, in practice the approximate variance of $\hat{r}_{11}$ can be estimated as $T^{-1}(1 - \hat{r}_{11}^2)$.

Princeton University

Manuscript received July, 1968; revision received February, 1969.

[12] When all of the serial correlation coefficients in the model are known and are equal (so that $\Omega_1 = \Omega_2 = \ldots = \Omega_h$), (12) is equivalent to equation (6.149) in Theil [9, p. 345], which is the asymptotic covariance matrix of Theil's "generalized two stage least squares" estimator.

[13] For Amemiya's method (as modified above) plim $T^{-1}Q_1u'_{1_{-1}}$ is zero, since for this method Q_1 includes predicted (as opposed to actual) values of the lagged endogenous variables, the predicted values being uncorrelated with $u_{1_{-1}}$.

[14] Cooper [3] in an unpublished note has derived the exact expression for the probability limit of the off-diagonal expression for the single equation model with one lagged dependent variable. He assumes that the errors are normally distributed and works with the likelihood function. The results here are essentially an extension of Cooper's results to the simultaneous equation case, except that here no simple expression for the probability limit of the off-diagonal matrix can be found. Also, due to the nature of the error term in (8), the estimates here cannot be considered to be maximum likelihood estimates.

APPENDIX

First we show that adding unnecessary instruments (i.e., instruments which add nothing to the explanation of the endogenous variables in the reduced form and which are uncorrelated with the reduced form error term) in the first stage of the two stage least squares procedure has no effect on the asymptotic covariance matrix of the estimator. For purposes of the discussion in this Appendix, assume that equation (3) is to be estimated where $u_1 = e_1$ and that the overall model is (1) where $U = E$ (i.e., no serial correlation problems). Write the reduced form for Y_1 as

(A1) $$Y_1 = \Pi_{1x}X + V_1,$$

where Π_{1x} is an $h_1 \times k$ matrix of reduced form coefficients and V_1 is a $h_1 \times T$ matrix of reduced form disturbance terms. The asymptotic covariance of $\sqrt{T}(\hat{C}_1 - C_1)$ is[15]

(A2) $$\sigma_{11} \operatorname{plim} T \begin{pmatrix} \hat{Y}_1 Y_1' & Y_1 X_1' \\ X_1 Y_1' & X_1 X_1' \end{pmatrix}^{-1},$$

where $\hat{Y}_1 = Y_1 X'(XX')^{-1}X$.

Now assume that unnecessary instruments are added to the first stage regression and let W denote the $k_2 \times T$ matrix of these instruments. W and V_1 are assumed to uncorrelated. Write (A1) as

(A3) $$Y_1 = \Pi_{1z}Z + V_1,$$

where $\Pi_{1z} = (\Pi_{1x}\, 0)$ and $Z' = (X'\ W')$. The predicted values of Y_1 from the first stage regression using Z as instruments are

(A4) $$\bar{Y}_1 = Y_1 Z'(ZZ')^{-1}Z.$$

Using $\bar{Y}_1$ in place of Y_1 in equation (3) in the second stage regression results in the following asymptotic covariance of $\sqrt{T}(\hat{C}_1 - C_1)$(using the fact that $\bar{Y}_1\bar{Y}_1' = \bar{Y}_1 Y_1' = Y_1\bar{Y}_1'$ and $X_1\bar{Y}_1' = X_1 Y_1'$):

(A5) $$\sigma_{11} \operatorname{plim} T \begin{pmatrix} \bar{Y}_1 Y_1' & Y_1 X_1' \\ X_1 Y_1' & X_1 X_1' \end{pmatrix}^{-1}.$$

It is easy to show that $\operatorname{plim} T^{-1}\hat{Y}_1 Y_1' = \operatorname{plim} T^{-1}\bar{Y}_1 Y_1 = \Pi_{1x} \operatorname{plim} T^{-1} XX'\Pi_{1x}'$, so that (A2) and (A5) are the same.

Heuristically this proof says that since the reduced form coefficient matrix is consistently estimated regardless of how many unnecessary instruments are added, nothing is changed in the limit by adding the extra instruments.

With respect to small sample properties, Nagar [7] has shown[16] that the bias, to the order T^{-1}, of the two stage least squares estimator is $(L - 1)Qq$, where

$$Q = \begin{pmatrix} \tilde{Y}_1\tilde{Y}_1' & \tilde{Y}_1 X_1' \\ X_1\tilde{Y}_1' & X_1 X_1' \end{pmatrix}^{-1} \quad \text{and} \quad q = T^{-1}\begin{pmatrix} \mathscr{E}(Y_1 e_1') \\ \mathscr{E}(X_1 e_1') \end{pmatrix}.$$

L is the total number of predetermined (instrumental) variables in the model less the number of coefficients in the equation being estimated, and $\tilde{Y}_1 = \Pi_{1x}X$. Adding unnecessary instruments increases L, but has no effect on Q or q. Therefore, adding unnecessary instruments increases the absolute value of the bias, to the order T^{-1}, of the estimator.

REFERENCES

[1] AMEMIYA, TAKESHI: "Specification Analysis in the Estimation of Parameters of a Simultaneous Equation Model with Autoregressive Residuals," *Econometrica*, XXXIV (April, 1966), 283–306.

[2] CHRIST, CARL F.: *Econometric Models and Methods*, New York: Wiley & Sons, Inc., 1966.

[3] COOPER, J. P.: "Notes on the Use of Procedures for Linear Regression in the Presence of First Order Serially Correlated Disturbances," mimeographed.

[4] FISHER, FRANKLIN M.: "The Choice of Instrumental Variables in the Estimation of Economy-Wide Econometric Models," *International Economic Review*, VI (September, 1965), 245–274.

[15] See, for example, Goldberger [5, p. 332].

[16] Nagar [7] has only formally shown this to be true when there are no lagged endogenous variables among the predetermined variables.

[5] GOLDBERGER, ARTHUR S.: *Econometric Theory*, New York: Wiley & Sons, Inc., 1964.
[6] MALINVAUD, E.: *Statististical Methods of Econometrics*, Amsterdam: North-Holland Publishing Company, 1966.
[7] NAGAR, A. L.: "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, XXVII (October, 1959), 575–595.
[8] SARGAN, J. D.: "The Maximum Likelihood Estimation of Economic Relationships With Autoregressive Residuals," *Econometrica*, XXIX (July, 1961), 414–426.
[9] THEIL, H.: *Economic Forecasts and Policy*, 2nd revised ed., Amsterdam: North-Holland Publishing Company, 1965.