TWO-STEP TWO-STAGE LEAST SQUARES ESTIMATION IN MODELS WITH RATIONAL EXPECTATIONS

Maurice Obstfeld

Robert Cumby

John Huizinga

Technical Paper No. 11

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

March 1981

We are grateful for helpful discussions with Craig Ansley, Ates Dagli, Phoebus Dhrymes, Daniel McFadden, Frederic Mishkin, Christopher Sims, and John Taylor. They share no responsibility for errors. We also would like to thank Craig Ansley, Dick Meese, and Paul Ruud for various computer programs. The research reported here is part of the NBER's research program in International Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #11
March 1981

Two-Step Two-Stage Least Squares Estimation
In Models With Rational Expectations

ABSTRACT

This paper introduces a limited-information two-step estimator for models with rational expectations and serially correlated disturbances. The estimator greatly extends the area of applicability of McCallum's (1976) instrumental variables approach to rational expectations models.

Section I reviews McCallum's method and discusses in detail the problems surrounding its use in many empirical contexts. Section II presents the two-step two-stage least squares estimator (2S2SLS) and demonstrates its efficiency relative to that of McCallum (1979). Section III provides a comparison of several estimators for a two equation macroeconomic model with rational expectations due to Taylor (1979).

Professor Maurice Obstfeld
Department of Economics
Columbia University
New York, NY 10027

(212) 280-5510

Dr. Robert Cumby
International Monetary Fund
Washington, D.C. 20431

(202) 477-3365

Professor John Huizinga
Graduate School of Business
University of Chicago
1101 E. 58th St.
Chicago, IL 60637

(312) 753-4166

Introduction

This paper introduces a limited-information, two-step estimator for simultaneous equations models with autocorrelated disturbances, and shows how it can be applied to models assuming rational expectations. The estimator greatly extends the area of applicability of McCallum's (1976) instrumental variable approach to rational expectations models, which treats realizations of random variables as their expected values measured with unpredictable forecast error. It also is useful in other contexts because it can be used to correct for moving average as well as autoregressive forms of serial dependence. It is thus more versatile than the widely used method of Fair (1970), for example, which is designed to handle autoregressive dependence only.

Recent research on estimation under the rational expectations hypothesis has focused on complete systems of equations, in which expectational consistency allows the imposition of strong cross-equation contraints on parameters. The constraints arise when unobservable expectational variables entering the structural equations are eliminated through use of the system's own forecasts. Thus, the estimation procedure makes efficient use of all the system's information in estimating each equation.[1]

In view of this efficiency property of full-system estimation, are there any advantages to direct structural estimation using the error-in-variables approach of McCallum (1976)? We believe that there are at least two. First, the procedure is feasible in situations where the full model may be too complex to be estimated by a full-information technique, and second, the limited-information approach is robust to specification errors in equations other than

---

[1]Wallis (1980), Hansen and Sargent (1979), and Hayashi (1980) discuss the estimation of simultaneous systems under rational expectations.

the one of primary interest. Full-system estimation, by employing information from all the system's equations, ensures that a single misspecification in any equation leads to inconsistent estimates of all the system's parameters.[2]

Unfortunately, the problem of autocorrelated regression disturbances limits the usefulness of the instrumental variable approach in practice. Autocorrelation may arise through serial dependence of the structural system's disturbances, because of overlapping forecasts, or because of the dating of expectational variables. While solutions to this problem have been proposed--for example by McCallum (1979) and Hayashi (1980)--little attention has been paid to the efficiency of the resulting parameter estimates.[3] In contrast, the procedure suggested in this paper recognizes that the estimation problem's structure can be exploited to yield a gain in the efficiency of estimation.

The plan of the paper is as follows. Section I reviews the limited-information method of McCallum (1976), and discusses in detail the problems surrounding its implementation in many empirical contexts. Section II develops a two-step two-stage least squares (2S2SLS) estimator for an environment with serially correlated errors and in which no strictly exogenous instruments are available so that predetermined variables must be used as instruments. This environment includes rational expectations models as a special case. The 2S2SLS estimator is a generalization of the non-linear two-stage least squares (NL2SLS) estimator of Amemiya (1974). We apply both estimators to the problems raised in Section I and show that the estimates are consistent and that the 2S2SLS estimates are asymptotically more efficient than the NL2SLS estimates.

---

[2]The existence of a robust estimator suggests a simple diagnostic test for specification error in rational expectations models. The idea is developed in Section III, below.

[3]The exception to this is Hayashi and Sims (1980).

In Section III, we use the 2S2SLS and NL2SLS estimators to estimate the parameters of a two-equation macroeconomic model with rational expectations due to Taylor (1979). The estimates we obtain are, on the whole, similar to those obtained by Taylor using a full-information, minimum distance technique. While the limited-information estimates are asymptotically less precise than the full system estimates, the decline in efficiency for the 2S2SLS estimator, in this case, appears moderate. In addition, the gain in efficiency from using the 2S2SLS estimator instead of the NL2SLS estimator proves to be important empirically as well as theoretically.

Section IV offers some concluding remarks. An appendix contains proofs of the paper's main propositions.

I. Single-Equation Estimation under Rational Expectations

In this section we describe McCallum's (1976) suggestion for estimating single equations under the rational expectations hypothesis, and discuss a major complication that arises in practice when one attempts to implement this technique. Our aim is to motivate the introduction, in the next section, of an estimator that circumvents this difficulty while yielding a gain in asymptotic efficiency over some consistent estimators now in use.

We initially consider an equation of the form

$$y_t = [{}_{t-1}z_t \;\; Z_t]\delta + u_t \qquad t = 1, \ldots, T \tag{1}$$

where y_t is a scalar random variable, u_t is a disturbance with mean zero, δ is an $(n+1) \times 1$ vector of unknown parameters, Z_t is a vector consisting of k endogenous random variables correlated with u_t and $n - k$ predetermined variables, and ${}_{t-1}z_t$ is the expected value of the endogenous variable z_t, conditional on information available at time $t - 1$,

$${}_{t-1}z_t = E[z_t | I_{t-1}]. \tag{2}$$

We also assume $E[u_t | I_{t-1}] = 0$. We regard (1) as one of a system of simultaneous equations that jointly determine the values of the endogenous variables, and assume that z_t is not an element of Z_t.[4]

---

[4]However, Z_t may contain lagged values of z_t. The discussion is easity extended to the case in which ${}_{t-1}z_t$ is a vector, but equation (1) is designed to be illustrative and not a general formulation.

As the expectation ${}_{t-1}z_t$ is typically unobservable, direct estimation of the structural form (1) cannot be accomplished using standard simultaneous-equation techniques. However, an instrumental-variables technique proposed by McCallum can be employed to obtain consistent estimates of δ. To implement this approach, we note that by (2) we may write the realization z_t as its expected value plus a mean-zero forecast error uncorrelated with any variable in the conditioning set I_{t-1},

$$z_t = {}_{t-1}z_t + \eta_t \qquad E[\eta_t | I_{t-1}] = 0 \tag{3}$$

For the remainder of the paper we will use the assumption that I_{t-1} contains past observations of all the structural system's variables and of η_{t-1}, so that η_t is serially uncorrelated. Using (3), we may eliminate the expectational variable from (1), obtaining

$$y_t = [z_t \; Z_t]\delta + u_t - \delta_1\eta_t = Q_t\delta + \varepsilon_t \tag{4}$$

where δ_1 is the first element of δ. We now stack the T observations satisfying equation (4) to obtain the regression model

$$y = Q\delta + \varepsilon. \tag{5}$$

Estimation of (5) is a standard errors-in-variables problem. Because of the rational expectations assumption and the assumed distribution of the $\{u_t\}$, any exogenous or endogenous variables lagged one or more periods are uncorrelated with the composite error, ε_t, and are (under mild assumptions) eligible instruments for consistent estimation of δ.[5] If the processes

[5]Use of contemporaneous values of exogenous variables is of questionable validity. These are, by definition, uncorrelated with the structural disturbance u_t, but will in general be correlated with the forecast error η_t unless they can be forecast correctly with probability one.

generating the simultaneous system are stationary, as we shall assume, our previous assumptions ensure that $E[\varepsilon\varepsilon']$ is equal to a scalar covariance matrix with diagonal elements $\sigma_w^2 = (\sigma_u^2 - 2\delta_1\sigma_{u\eta} - \delta_1^2\sigma_\eta^2)$. Letting X denote the matrix of instruments and $\hat{\delta}$ denote the instrumental-variable estimate of δ, it follows that

$$D = \hat{\sigma}_w^2 \ (Q'X(X'X)^{-1}X'Q)^{-1} \tag{6}$$

yields a consistent estimate of the asymptotic variance of $\hat{\delta}$ (conditional on X) provided that $E[\varepsilon\varepsilon'|X] = E[\varepsilon\varepsilon']$. Here, $\hat{\sigma}_w^2$ is obtained from estimated residuals in the usual manner. The formula (6) is valid when X contains lagged endogenous variables, so that estimation of δ under the rational expectations assumption poses no special problem for standard computer packages provided one is willing to make the assumption concerning the conditional variance of ε.[6]

While McCallum's procedure is appropriate for estimating (5), a major difficulty often arises in practice which makes this procedure inapplicable. The difficulty is the critical assumption that ε is serially uncorrelated. There are many interesting rational expectations models where this will not be the case. The first is associated with a dating of expectations different from that of equation (1). The dating scheme assumed in equation (1) requires that expectational variables be predetermined, and excludes the plausible possibility that current expectations of future events may influence currently realized economic variables. For example, if (1) is replaced by

$$y_t = [{}_tz_{t+1} \ Z_t]\delta + u_t$$

---

[6] Assumptions concerning conditional variances are frequently more difficult to justify on economic grounds than are assumttions concerning conditional means. In certain models D is a consistent estimate of the asymptotic variance of $\hat{\delta}$ under weaker assumptions than those stated above.

use of McCallum's errors-in-variables approach yields the equation

$$y_t = [z_{t+1}\ Z_t]\delta + u_t - \delta_1\eta_{t+1} = Q_t\delta + \varepsilon_t \tag{7}$$

where η_{t+1} is the error of forecast between periods t and $t + 1$.[7] The important difference between (7) and (4), as Hayashi (1980) has pointed out, is that the composite error $\varepsilon_t = u_t - \delta_1\eta_{t+1}$ will not in general be serially uncorrelated. The forecast error between periods $t - 1$ and t, η_t, and the innovation, u_t, occurring at time t are likely to be correlated, causing ε_t and ε_{t-1} to be correlated. A second, related problem arises when the relevant forecast horizon is greater than the sampling interval, so that successive forecast periods overlap.[8] In this case serial correlation of ε is due to non-zero correlation between η's across time. A third example occurs when the structural disturbance u is not a white-noise process.

The serial correlation in ε presents several familiar problems. First, the instrumental variables estimator proposed by McCallum (1976) is inefficient among the class of limited information estimators. Second, the variance-covariance matrix reported by standard regression packages will be incorrect.[9] Third, indiscriminate use of lagged endogenous variables as instruments will lead to inconsistent parameters estimates if the serial correlation comes from serially correlated structural errors.

---

[7] This formulation often arises in models dealing with real interest rates. See for example Mishkin (1980).

[8] Kennan (1979) confronts a problem of this sort in estimating partial-adjustment models. Hansen and Hodrick (1980) and Huizinga (1981) discuss a similar problem in a different context. The estimation of macro models of the type employed by Fischer (1977) and by Phelps and Taylor (1977) would also involve overlapping forecast errors.

[9] McCallum (1979) discusses how, in some cases, the correct standard errors may be obtained.

While these problems are standard, the conventional corrections for serial correlation are now known to lead to inconsistent parameter estimates in the context of rational expectations models. This has been pointed out for the general case by Bernanke (1977) and Hansen (1979). Flood and Garber (1980) have described in detail how the widely used method of Fair (1970) becomes inapplicable. Intuitively, the inconsistency of generalized least squares (GLS) arises because the procedure transforms the model by replacing variables with linear combinations of their past, present, and future values.[10] Thus, even if the original instruments and error term are contemporaneously uncorrelated, the transformed instruments and the transformed error term need not be. This problem arises specifically in rational expectations models and not in the standard simultaneous equations models because in the standard models it is assumed that the instruments are "strictly" exogenous. That is, the instruments are uncorrelated with the error term at all leads and lags. This assumption is untenable in the rational expectations formulation because part of the error term is a forecast error likely to be contemporaneously correlated with all variables, both endogenous and exogenous.

To understand this inconsistency more fully, it is perhaps useful to consider the case of Theil's (1961) generalized two stage least squares estimator (G2SLS) and the model of equation (7).[11]

For simplicity assume

$$E(u_t\, u_{t-j}) = 0 \qquad j \neq 0$$
$$E(u_t\, \eta_t) \neq 0$$
$$E(u_t\, \eta_{t-j}) = 0 \qquad j \neq 0$$

---

[10] Because Fair's method is not a GLS procedure, the following explanation is not directly relevant. See Flood and Garber (1980) for a discussion of why Fair's method leads to inconsistent estimates.

[11] In the standard simultaneous equation framework this estimator is efficient among the class of limited information estimators.

These assumptions are sufficient to guarantee that the composite error term ε, is a first order moving average process. If we use Σ to denote the variance-covariance matrix of ε, the generalized two stage least squares estimator is given by[12]

$$\hat{\delta}_{G2SLS} = \delta + Q'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\varepsilon \tag{8}$$

where X denotes the instrument matrix, assumed to include exogenous variables.[13] In order for this estimator to be consistent, it is necessary that $1/T\ (X'\Sigma^{-1}\varepsilon)$ converge in probability to zero as T (the sample size) goes to infinity. This is unlikely in general because a typical element of the column vector $X'\Sigma^{-1}\varepsilon$ is

$$X'\Sigma\varepsilon_i = \sum_{k=1}^{T} \sum_{\ell=1}^{T} X_{ki}\Sigma^{-1}_{k\ell}\ \varepsilon_\ell \tag{9}$$

where X_{ki} is the value of the i^{th} instrumental variable at time k. Now Σ, being the variance-covariance matrix of a first order moving average process is a band symmetric matrix with zeros on all but the main and first off-diagonals. However, Σ^{-1} will contain all non-zero elements.[14] Therefore equation (9) will contain terms involving the product of $X_{t+1,i}$ and ε_t. These need not converge to zero in probability (after dividing by T) because part of

---

[12] In practice $\hat{\Sigma}^{-1}$ would be used in place of Σ^{-1} but this difference is immaterial here.

[13] Exogenous variables are those which are uncorrelated at all leads and lags with the structural error u.

[14] In fact, for this particular case, Σ^{-1} will be of the same form as the variance-covariance matrix for a first-order autoregressive process.

ε_t is the forecast error in predicting z_{t+1} which is in general correlated with $X_{t+1,i}$.[15,16]

The inconsistency of standard GLS procedures in the rational expectations framework discussed here essentially leaves one with two options. First, one can continue to use the errors in variables approach of McCallum (1976) without attempting to correct for the serial correlation in ε and adjust the standard errors of the estimates accordingly.[17] Secondly, one could search for an estimator which corrects for the serial correlation in ε and still preserves consistency of the parameter estimates. In line with this second option, the next section turns to the proposed two-step two-stage least squares estimator. It is shown there that this yields more efficient estimates than following strategy one.[18]

---

[15] If it is known that certain X's do not appear in the reduced form for z_t, or the particular form of Σ is such that it contains zeros in strategic locations, a version of generalized two-stage least square with a carefully chosen list of instruments may be applicable. These can only be special cases, however.

[16] In a recent article, Revankar (1980) proposes a three-step GLS estimator for estimating models where the error term is the sum of a white-noise and an autoregressive part. That estimator, like that of Fair (1970) and Theil (1961), is inconsistent for the model discussed above. Inconsistency in the rational expectations framework is also a problem for the estimator of Dhrymes, Berner and Cummins (1974).

[17] This option is discussed in McCallum (1979), along with a discussion of overcoming the problem of using lagged endogenous variables as instruments in the case of serially correlated structural errors.

[18] Hayashi (1980) proposes an estimator which, like two-step two-stage least squares, corrects for serial correlation and maintains consistency. Hayashi's estimator cannot be shown to be more efficient than option one, however.

II. Two Step Two-Stage Least Squares

In this section we describe a two step procedure for estimating the unknown $h \times 1$ parameter vector δ in the nonlinear model

$$y = Qf(\delta) + \varepsilon \tag{10}$$

where f is a differentiable, one-to-one function taking the vector δ into a space of higher dimension.[19] It is assumed that

$$E(\varepsilon_t | N_{t-m}) = 0, \qquad t = 1, 2, \ldots \tag{11}$$

where N_{t-m} is any arbitrary subset of the set $\{W_t, W_{t-1}, W_{t-2}, \ldots, \varepsilon_{t-m}, \varepsilon_{t-m-1}, \ldots\}$ and the W_t's are observable vector valued random variables to be used as instruments. As the notation suggests, the W_t's will usually be known at time $t - m$ and should be thought of as lagged variables, i.e., $W_t = X_{t-m}$. This particular formulation of the equation to be estimated includes the rational expectations models discussed in Section I as a special case and makes it apparent that our technique is applicable in a general simultaneous equation setting with predetermined, rather than strictly exogenous, instruments.[20]

To see how the models discussed in the previous section can be expressed in terms of equations (10) and (11), return to the specification,

$$Y_t = [{}_t z_{t+1} \; Z_t] g(\delta) + u_t \tag{12}$$

---

[19] This section describes a non-linear extension of the estimator proposed in Obstfeld (1978).

[20] Certain other standard assumptions, including stationarity and correlations between W and Q, are described in detail in the Appendix where proofs of all propositions can be found.

and suppose u_t follows a first order autoregressive scheme,

$$u_t = \rho u_{t-1} + v_t, \qquad E(v_t | I_{t-1}) = 0.^{21} \tag{13}$$

After imposing rational expectations and quasi-differencing, equation (11) becomes,

$$y_t = \rho y_{t-1} + [z_{t+1}\ Z_t]g(\delta) - [z_t\ Z_{t-1}]\rho g(\delta) + v_t - g_1(\delta)\eta_{t+1} + \rho g_1(\delta)\eta_t \tag{14}$$

where as before η_t is the forecast error and $g_1(\delta)$ represents the first element of $g(\delta)$. By choosing

$$Q_t = [y_{t-1}\ z_{t+1}\ Z_t\ z_t\ Z_{t-1}]$$

$$f(\delta) = \rho + g(\delta) - \rho g(\delta) \qquad \text{and}$$

$$\varepsilon_t = v_t - g_1(\delta)\eta_{t+1} + \rho g_1(\delta)\eta_t$$

it is clear that equation (14) becomes equation (10). Note also, that while the composite error, ε, is correlated with its own value lagged once, this correlation dies out after one period.[22] Furthermore, $E(\varepsilon_t | I_{t-1}) = 0$ by construction so that endogenous or exogenous variables lagged one or more periods may be used in forming the vector W_t. Thus all conditions for putting the model in terms of equations (10) and (11) are met. If u_t in equation

---

[21] I_{t-1} is assumed to include all the system's exogenous and endogenous variables dated t-1 and before so that this assumption is, of course, stronger than the assumption that $\{v_t\}$ is a white-noise process. It states that the system's lagged endogenous variables are predetermined with respect to the innovations v_t.

[22] This is true provided that $E(v_t | \eta_{t-j}) = 0$ for $j > 0$ which seems to be a reasonable assumption. Recall that $E(\eta_t \eta_s) = 0$ for $s \neq t$ and $E(\eta_t v_{t-j}) = 0$ for $j > 0$ by rational expectations.

(12) is not a first order autoregressive process but rather a first order moving average process,

$$u_t = v_t + \rho v_{t-1} \qquad E(v_t | I_{t-1}) = 0 \ , \tag{15}$$

we can again satisfy the formulation of equations (10) and (11). In this case no quasi-differencing is needed since the composite error $v_t + \rho v_{t-1} - g_1(\delta)\eta_{t+1}$ is uncorrelated with itself after two periods and

$$E(v_t + \rho_{t-1} + g_1(\delta)\eta_{t+1} | I_{t-2}) = 0. \tag{16}$$

We merely need to use variables lagged at least twice to form W_t.

Obviously, then, allowing u_t in equation (12) to be a general ARMA process provides no problems for fitting the rational expectations model to the form of equations (10) and (11). A higher order autoregressive component can be handled by more quasi-differencing and a higher order moving average component can be handled by moving the information set back in time.[23] In addition, it should be clear that the dating of the expectations provides no problems. Replacing ${}_tz_{t+1}$ in equation (12) with the arbitrary ${}_{t-j}z_{t+k}$ requires raising only two points. First, when forecasts are made for several periods ahead and data are collected every period rational expectations does not guarantee serially uncorrelated forecast errors in successive periods. Thus the forecast error component of the composite error, ε, need not always be white noise. However, if the longest forecast horizon appearing in equation (12) is k periods, rational expectations does guarantee that the forecasts are

---

[23]In the present setting, consistent estimation does require knowledge of a maximum length for non-zero structural disturbance correlations. Hatanaka (1975), Sims (1979), and others have argued that this identifying information will frequently be unavailable. In practice, therefore, one must view hypothesis tests as tests of joint hypotheses including assumptions about lags. Salemi and Sargent (1979) adopt this pragmatic approach.

uncorrelated after k lags. Thus multi-period forecasts will only affect the value of m in equation (11). The second point is that when ${}_tz_{t+1}$ is replaced by an arbitrarily dated expectation in equation (12), if the earliest expectation is ${}_{t-j}z_{t+k}$, the matrix W_t must have all variables dated at time $t - j - 1$ or before. Rational expectations guarantees that any such information is uncorrelated with the present forecast error component of the composite error.

Having shown how rational expectations models can be put in the form of equations (10) and (11), we return to the issue of estimation. One way to estimate δ would be to use the estimator $\hat{\delta}$ of δ obtained by minimizing the function

$$\phi(\delta) = (y - Qf(\delta))'W(W'W)^{-1}W'(y - Qf(\delta)). \qquad (17)$$

This estimator is the nonlinear version of that proposed by McCallum (1976, 1979) and is, of course, the non-linear two stage least squares (NL2SLS) estimator of Amemiya (1974).

The asymptotic properties of the NL2SLS estimator are given in Proposition 1 below, where the matrix Ω is defined as follows. Let W_t' be the transpose of the t^{th} row of the instrument matrix, and define $q_t \equiv \varepsilon_t W_t'$. Assume that

$$\underset{T\to\infty}{\text{plim}} \sum_t \frac{q_t q_t'}{T}$$

is finite, and set

$$R(\ell) = \underset{T\to\infty}{\text{plim}} \sum_t \frac{q_t q_{t-\ell}}{T} .$$

Note that $R(\ell) = E[q_t q_{t-\ell}'] = E[q_{t+\ell} q_t'] = E[q_t q_{t+\ell}']' = R(-\ell)'$, when all processes are jointly covariance stationary. Finally, define

$$\Omega \equiv \sum_{\ell=-m}^{m} R(\ell) \tag{18}$$

where m is as given in equation (11). Ω will necessarily be non-singular.

Proposition 1. The nonlinear two stage least squares estimate $\hat{\delta}$ is consistent for δ. Asymptotically, $\sqrt{T}(\hat{\delta} - \delta)$ has a normal distribution with mean zero and variance-covariance matrix

$$\operatorname*{plim}_{T\to\infty} T^2 \, (V'W(W'W)^{-1}W'V)^{-1}V'W(W'W)^{-1}\Omega(W'W)^{-1}W'V(V'W(W'W)^{-1}W'V)^{-1} \tag{19}$$

where

$$V \equiv Q \left.\frac{\partial f}{\partial \delta}\right|_{\delta} = Q \left(\left.\frac{\partial f}{\partial \delta_1}\right|_{\delta}, \ldots, \left.\frac{\partial f}{\partial \delta_h}\right|_{\delta} \right). \tag{20}$$

Proof of this, and of all subsequent propositions, is given in the Appendix.

Proposition 1 provides a formal justification for following the option described in section I of using McCallum's standard errors in variables approach, provided one uses the correct formula for the variance of the estimates given in equation (19).[24] The failure of NL2SLS to correct for serial correlation in ε, however, raises the possibility that using a distance metric other than $W(W'W)^{-1}W'$ may yield more efficient estimates. In particular, consider the estimator d of δ that minimizes

$$\Lambda(\delta) = (y - Qf(\delta))\,'W\,\Omega^{-1}\,W'(y - Qf(\delta)). \tag{21}$$

---

[24] It should be noted in passing that equation (24) is not what the standard regression package will report, however. The standard package will set $\Omega = W'W$ so that equation (19) becomes $\operatorname{plim} T^2(V'W(W'W)^{-1}W'V)^{-1}$. This was pointed out by McCallum (1979) where the issue of the correct variance matrix for NL2SLS is discussed.

In order to motivate this estimator, consider the following derivation which parallels the standard derivation of the linear two stage least squares estimator (see Dhrymes (1974)) and shows how d arises as an Aitken estimator. Premultiply equation (10) by W' to obtain

$$W'y = W'Q\, f(\delta) + W'\varepsilon . \qquad (22)$$

Nonlinear least squares applied to (22) would give a consistent estimate of δ, but since $\text{var}(W'\varepsilon)$ is not proportional to the identity matrix, this estimate is likely to be inefficient relative to the one given by some kind of generalized least squares. Following this line, let R be the Cholesky decomposition of the matrix Ω in equation (18) (so that $\Omega = RR'$) and consider the transformed model,

$$R^{-1}W'y = R^{-1}W'Qf(\delta) + R^{-1}W'\varepsilon \qquad (23)$$

Application of nonlinear least squares to (23) yields the estimate d which minimizes $\Lambda(\delta)$. This is readily verified in the linear case since ordinary least squares on equation (23) with $f(\delta) = \delta$ yields

$$\bar{\delta} = (Q'W\Omega^{-1}W'Q)^{-1}Q'W\Omega^{-1}W'y \qquad (24)$$

which satisfies the first order conditions for minimizing $\Lambda(\delta)$. This expression for d in the linear case is also useful because by substituting equation (10) into equation (24) one sees that consistency depends only on having $\underset{T\to\infty}{\text{plim}}(1/T)W'\varepsilon = 0$; and this follows from equation (11).

The relationship between d and other GLS estimators can best be appreciated by observing that if W satisfies one additional condition, Ω

reduces to $\operatorname{plim}_{T\to\infty} W'\Sigma_T W$, $\Sigma_T = E[\varepsilon\varepsilon']$.[25] In this case there is an alternative derivation of d. This derivation begins with factoring $\Sigma_T = AA'$ and premultiplying the regression equation (11) by A^{-1} obtaining,

$$\bar{y} = \bar{Q}f(\delta) + \bar{\varepsilon}\ , \qquad E(\bar{\varepsilon}\bar{\varepsilon}') = I, \tag{25}$$

where $\bar{y} = A^{-1}y$, $\bar{Q} = A^{-1}Q$, and $\bar{\varepsilon} = A^{-1}\varepsilon$. Next, form the matrix $\tilde{W} = A'W$. The key observation to be made is that the transformations leave the correlation between the instruments and the regressors and that between the instruments and the disturbances unchanged,

$$\operatorname{plim}_{T\to\infty}(1/T)\tilde{W}'\bar{Q} = \operatorname{plim}_{T\to\infty}(1/T)W'AA^{-1}Q = \operatorname{plim}_{T\to\infty}(1/T)W'Q$$

$$\operatorname{plim}_{T\to\infty}(1/T)\tilde{W}'\bar{\varepsilon} = \operatorname{plim}_{T\to\infty}(1/T)W'AA^{-1}\varepsilon = \operatorname{plim}_{T\to\infty}(1/T)W'\varepsilon\ .$$

Thus $\tilde{W}$ may be used as an instrument matrix in estimating δ in equation (25) by nonlinear two stage least squares. This estimate, however, is obtained by minimizing $\Lambda(\delta)$ and is therefore the estimator d.[26]

We now turn to the asymtotic properties of the estimator d which minimizes $\Lambda(\delta)$ and the efficiency of d relative to NL2SLS.

---

[25] That condition is $E(\varepsilon_t\varepsilon_{t-s}|W_t, \ldots, W_{t-m}) = E(\varepsilon_t\varepsilon_{t-s})$ for all $m \geq s \geq 0$.

[26] In contrast, Theil's generalized two-stage least squares transforming the instruments by the transpose of A^{-1} rather than A'. As pointed out in section I, this creates a correlation of $\operatorname{plim}_{T\to\infty}(1/T)W'\Omega^{-1}\varepsilon$ between the transformed instruments and the transformed error. And while $\operatorname{plim}_{T\to\infty}(1/T)W'\varepsilon = 0$, $\operatorname{plim}_{T\to\infty}(1/T)W'\Omega^{-1}\varepsilon \neq 0$.

Proposition 2. The estimator d is consistent for δ. Asymptotically $\sqrt{T}(d - \delta)$ has a normal distribution with mean zero and variance covariance matrix,

$$\operatorname*{plim}_{T\to\infty} T^2(V'W\Omega^{-1}W'V)^{-1}.$$

Proposition 3. The estimator d is more efficient than the NL2SLS estimator $\hat{\delta}$ in the sense that the asymptotic variance matrix of $\hat{\delta}$ minus the asymptotic variance matrix of d is positive definite.

While Proposition 3 provides a major justification for using d, one obvious problem remains. In practice, Ω is not known a priori and must be estimated. Fortunately, as Hansen (1979) has shown, the following procedure yields a consistent estimate of Ω. Using NL2SLS, obtain the estimated residuals $\hat{\varepsilon}_t = y_t - Q_t f(\hat{\delta})$. Define $\hat{q}_t = \hat{\varepsilon}_t W_t'$, and form

$$\hat{R}(\ell) = \sum_t \frac{\hat{q}_t \hat{q}_{t-\ell}'}{T} \quad \text{for} \quad \ell < m \tag{26}$$

Finally, let

$$\hat{\Omega} = \sum_{\ell=-m}^{m} \hat{R}(\ell). \tag{27}$$

The ability to consistently estimate Ω allows us to define the <u>two-step two-stage least squares</u> (2S2SLS) estimator of δ as the value of δ that minimizes the quadratic form of equation (21) with Ω replaced by a consistent estimate, $\hat{\Omega}$. The first step involves doing NL2SLS to obtain a set of residuals used to estimate Ω. The second step involves minimizing $\Lambda(\delta)$, as given in equation (21), replacing Ω with $\hat{\Omega}$.[27] The following proposition

---

[27] Since writing this paper it has been brought to our attention that the 2S2SLS estimator has independently been proposed by White (1980) in a cross section model with heteroskasdicity. Hansen (1980) has also independently proposed a related estimator.

states the relationship between the estimator d (formed with a known Ω) and the 2S2SLS estimator.

Proposition 4. The 2S2SLS estimator and the estimator d have the same asymptotic distribution provided that $\hat{\Omega}$ is a consistent estimator of Ω.

Before closing this section several additional points should be made. First, as described earlier, there are conditions under which $\Omega = W'\Sigma_T W$, $\Sigma_T = E(\varepsilon\varepsilon')$.[29] In this case one would use $\hat{\Omega} = W'\hat{\Sigma}_T W$, where $\hat{\Sigma}$, can be found directly from the $\hat{\varepsilon}$ of NL2SLS in the usual fashion. Note that when $\Sigma_T = \sigma^2 I$, 2S2SLS and NL2SLS are the same estimator.

Second, when equation (10) is just identified, so that the number of instruments equals the number of parameters, h, to be estimated, there is no efficiency gain from minimizing $\Lambda(\delta)$ instead of $\phi(\delta)$. In this special case $V'W$ is invertible and both $\sqrt{T}(d - \delta)$ and $\sqrt{T}(\hat{\delta} - \delta)$ converge in distribution to $N(0, \operatorname*{plim}_{T\to\infty} T(V'W)^{-1}\Omega(V'W)^{-1})$.[30]

A final point which deserves emphasis is that in contrast to the usual single-equation form of GLS, 2S2SLS does not require inversion of a $T \times T$ matrix, where T is the number of observations. Instead, only the $K \times K$ matrix $\hat{\Omega}$ need be inverted, where K is the number of instruments. This enhances the computations because K is typically much smaller than T.

III. An Application

In this section, we use the nonlinear two stage least squares and the

---

[29] See footnote 25.

[30] This is reminiscent of the relationship between three-stage least squares and two-stage least squares. In the just identified care, three-stage least squares is no more efficient than two-stage least squares.

two-step two-stage least squares techniques to estimate Taylor's (1979) macroeconomic model of the United States. The purpose of the application is to compare the limited-information 2S2SLS and NL2SLS estimates, which do not depend on cross-equation constraints, with each other and with the full system estimates reported by Taylor. Comparison of the NL2SLS and 2S2SLS estimates allows us to determine whether the efficiency gain from doing the second step of 2S2SLS is important empirically, as well as theoretically.

Taylor's model consists of the following four equations:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3(m_t - p_t) + \beta_4(m_{t-1} - p_{t-1}) + \beta_5 \pi_t^e + \beta_6 t + \beta_0 + u_t \quad (28)$$

$$\pi_t = \pi_{t-1} + \gamma_1 y_t^e + \gamma_0 + v_t \quad (29)$$

$$u_t = \eta_t - \theta_1 \varepsilon_{t-1} \quad (30)$$

$$v_t = \varepsilon_t - \theta_2 \varepsilon_{t-1}. \quad (31)$$

Here, y_t is the deviation of the log of real expenditure from trend, m_t is the log of the nominal money supply, p_t is the log of the price level, π_t denotes $p_{t+1} - p_t$, and ε_t and η_t are mean-zero disturbances such that $var(\varepsilon_t, \eta_t)' = \Gamma$, $E[(\varepsilon_t, \eta_t)' (\varepsilon_{t-j}, \eta_{t-j})] = 0$ for $j > 0$. The superscript "e" attached to a variable denotes its expected value, conditional on information available at time $t - 1$. This information is assumed to include the following period's price level, p_t, and money stock, m_t. These two variables are therefore predetermined at time t, and so, uncorrelated with the disturbances η_t and ε_t.

Equations (29) and (30) both involve a <u>forward</u> expectation--an expectation of a future rather than current variable. Further, both-structural disturbances are assumed to be serially dependent. Thus the efficiency of instrumental-variable estimation can be improved by the two-step technique described in Section II.

The full-system estimates obtained by Taylor (1979) and the single equation estimates are reported in Table 1. Equations (i) and (iv) are the output and price equations estimated by Taylor, who applied a minimum distance technique to the rational-expectations reduced form of (28)-(31). Equations (ii) and (v) give the NL2SLS estimates and equations (iii) and (vi) give the 2S2SLS estimates. The instruments used for NL2SLS and 2S2SLS in the price equation were a constant, y_{t-2}, y_{t-3}, p_{t-1}, p_{t-2}, m_{t-1} and m_{t-2}. The instruments used in the output equation were a constant, time trend, y_{t-2}, y_{t-3}, m_{t-1}, m_{t-2}, m_{t-3}, p_{t-1}, p_{t-2}, and p_{t-3}. The data series were the ones used by Taylor.[31]

Given the magnitude of the asymptotic standard errors, the coefficients produced by NL2SLS, 2S2SLS and the full-system method are quite similar. Only the coefficients of the time trend and constant in the output equation change sign when a single equation estimator is used. The lag pattern implied by Taylor's output equation is unchanged. In particular, the negative sign on expected inflation in the output equation, indicating a "perverse" aggregate demand response, remains.

The most notable differences between the single equation and full system estimates is found in the estimated output effect of changes in the

---

[31]Thus, trend output was taken to be the CEA's revised potential output series, while other series (all seasonally adjusted) came from the Citibank data base. It should also be pointed out that the estimated Ω used here is estimated in the frequency domain, not the time domain. This is possible because Ω is the value of the spectrum of $q_t = \varepsilon_t W_t'$ and is done to ensure that $\hat{\Omega}$ is positive definite.

Table 1

Full-System and Limited Information Estimates of Taylor's Model

1954:1 - 1975:3

Output Equation

(i) System: $$y_t = \underset{(.0877)}{1.167}\, y_{t-1} - \underset{(.090)}{.324}\, y_{t-2} + \underset{(.175)}{.578}\, (m_t - p_t) - \underset{(.194)}{.484}\, (m_{t-1} - p_{t-1}) - \underset{(.391)}{.447}\pi^e_t + \underset{(.00746)}{.00843}\, t + \underset{(.0343)}{.0720}$$

(ii) NL2SLS: $$y_t = \underset{(.1443)}{1.405}\, y_{t-1} - \underset{(.140)}{.557}\, y_{t-2} + \underset{(.338)}{.308}\, (m_t - p_t) - \underset{(.364)}{.191}\, (m_{t-1} - p_{t-1}) - \underset{(.601)}{.673}\pi^e_t - \underset{(.00707)}{.00900}\, t - \underset{(.0476)}{.0877}$$

(iii) 2S2SLS: $$y_t = \underset{(.1122)}{1.404}\, y_{t-1} - \underset{(.101)}{.539}\, y_{t-2} + \underset{(.257)}{.250}\, (m_t - p_t) - \underset{(.275)}{.159}\, (m_{t-1} - p_{t-1}) - \underset{(.431)}{.653}\pi^e_t - \underset{(.00584)}{.00651}\, t - \underset{(.0380)}{.0676}$$

Price Equation

(iv) System: $$\pi_t = \pi_{t-1} + \underset{(.0058)}{.0180}\, y^e_t + \underset{(.0172)}{.0515}$$

(v) NL2SLS: $$\pi_t = \pi_{t-1} + \underset{(.0110)}{.0184}\, y^e_t + \underset{(.0337)}{.0414}$$

(vi) 2S2SLS: $$\pi_t = \pi_{t-1} + \underset{(.0079)}{.0189}\, y^e_t + \underset{(.0196)}{.0519}$$

Standard Errors in parentheses

real money stock and changes in π^e. The 2S2SLS estimates of β_3 and β_4 are smaller in absolute value than those found by Taylor, and are not significant at the 5 percent level. However, their sum $\beta_3 + \beta_4 (= .091)$ is close to being significant (the t-ratio is 1.76), and is similar to the one found by Taylor. The 2S2SLS estimates imply a long-run output elasticity of .676 with respect to real money. On the basis of Taylor's parameter estimates, this elasticity is calculated to be .599. With regards to the effect of expected inflated, the Taylor estimates are a short run elasticity of -.447 and a long-run elasticity of -2.85. The 2S2SLS estimates are -.653 and -4.83, respectively, which are noticeably larger.

Turning to the issue of efficiency, the entries in Table 1 indicate that for this model, 2S2SLS produces a noticeable increase in efficiency over NL2SLS. The average ratio of 2S2SLS standard errors to NL2SLS standard errors is .76 in the output equation and .65 in price equation. The better relative performance of 2S2SLS in the price equation is presumably due to the larger number of over-identifying restrictions. Examination of the individual standard errors shows that the efficiency gain attributable to the second step of 2S2SLS is fairly evenly spread across all parameters in the output equation, and relatively more important for the constant term in the price equation. As regards the relative efficiency of full and limited information estimators, the results presented in Table 1 indicate that the decline in efficiency due to the limited information nature of 2S2SLS is, in this case, only moderate. The average ratio of full system to 2S2SLS standard errors is .82 in the output equation and .81 in the price equation.[32]

---

[32]None of the above calculations do not include the standard error on the time trend in equation (29) since it is smaller than the reported standard error for the full system estimate. This would, of course, be reversed in larger samples.

The last issue to be discussed in this section is this issue of misspecification. As we have argued above, an advantage of the instrumental-variable technique is its robustness to specification error in other equations. This fact may be exploited to obtain a diagnostic test of specification. For example, if there is reason to suspect a specification error in the output equation, one may estimate the price equation in isolation by 2S2SLS, and compare the resulting estimates of γ_0 and γ_1 to those produced by full-information estimation. If these differ greatly, it is evidence of some misspecification in the output equation which is spread to the price equation through imposition of the rational expectations hypothesis on the system's reduced-form coefficients.

More precisely, we can proceed as follows. Let $\hat{\gamma} = [\hat{\gamma}_1, \hat{\gamma}_2]$ be the estimate obtained by applying an efficient estimation procedure under the null hypothesis to the system. Let $g = [g_1, g_2]$ be the coefficients obtained by estimating the price equation directly using a limited-information rational expectations estimator of the previous section. Of course, g is inefficient under the null hypothesis that the output equation is not misspecified, but is consistent under either hypothesis. Now form the test statistic

$$H = (g - \hat{\gamma})' \, [\text{var}(g) - \text{var}(\hat{\gamma})]^{-1} (g - \hat{\gamma}). \tag{32}$$

As shown by Hausman (1978), the asymptotic distribution of H is χ^2 with two degrees of freedom under the null hypothesis.

Those familiar with the Wald test, will note that the test statistic given in equation (32) is exactly that, a Wald Test, since $\text{var}(g - \hat{\gamma}) = \text{var}(g) - \text{var}(\hat{\gamma})$ asymptotically. The test is testing whether $(g - \hat{\gamma})$ is "significantly" different from zero. This insight makes clear why 2S2SLS is to be preferred to NL2SLS for carrying out these specification tests. By using 2S2SLS to get the g estimates, one will get a smaller $\hat{\text{var}}(g)$ and a tighter estimate of $g - \hat{\gamma}$; this leads to a more powerful test.

Unfortunately, for the model dealt with here, the specification test described by equation (32) is infeasible. This results from the fact that while $\text{var}(g) - \text{var}(\hat{\gamma})$ is guaranteed to be positive semi-definite asymptotically, there is no guarantee that this will hold in finite samples. For the Taylor model and our finite sample, the off-diagonal terms of $\text{var}(g)$ exceed the off-diagonal terms of the $\text{var}(\hat{\gamma})$ by an amount which is sufficient to make $\text{var}(g) - \text{var}(\hat{\gamma})$ nonpositive definite, even though it has positive diagonal terms.[33]

IV. Conclusion

In this paper, we have developed a flexible, two-step extension of nonlinear two-stage least squares that is more efficient than that technique when equation disturbances are serially correlated and there are no strictly exogenous variables. The estimator, called two-step two-stage least squares, is attractive because it is computationally tractable and may be applied when disturbances exhibit moving-average as well as autoregressive dependence.

Our main interest here has been the application of the two-step estimator to equations containing expectational variables among the regressors. McCallum (1976) has shown how consistent parameter estimates can be obtained under the rational expectations hypothesis by exploiting the fact that realized variables equal their conditional expectations plus a forecast error orthogonal to all variables included in the conditioning set. The two-step two-stage least squares

---

[33] There may also be a problem because Taylor's standard errors are not exactly correct. See Taylor (1979) footnote 7.

estimator enhances the applicability of McCallum's errors-in-variables approach by extending it to equations in which (i) expectations of future events (or "forward" expectations) influence current economic behavior, (ii) equations' structural disturbances are not independently, identically distributed and (iii) equations contain multi-period forecasts. The existence of a simple, limited-information alternative to full-system rational-expectations estimation á la Hansen-Sargent (1980) and Wallis (1980) is important, because the full-information method, when computationally feasible, requires careful and correct specification of all equations of the system, including those which may be of secondary interest. In addition, the limited-information technique can be used to obtain an easy, diagnostic test for misspecification in rational expectations systems.

As an application, we employed the two-step technique to estimate the price and output equations of Taylor's (1979) rational-expectations macro-model of the United States. The instrumental-variable estimates squared quite well with the full-system results reported by Taylor.

REFERENCES

[1] Amemiya, T.: "The Nonlinear Two-Stage Least-Squares Estimator," Journal of Econometrics, 2 (1974), 105-110.

[2] Amemiya, T.: "Regression Analysis When the Dependent Variable is Truncated Normal," Econometrica, 41 (1973), 997-1016.

[3] Bernanke, B.: "Estimation of a Macroeconomic Model Under the Hypothesis of Rational Expectations," mimeo, M.I.T., 1977.

[4] Dhrymes, P.: Econometrics: Statistical Foundation and Applications. New York: Springer-Verlag, 1974.

[5] Dhrymes, P., R. Berner, and D. Cummins: "A Comparison of Some Limited Information Estimators for Dynamic Simultaneous Equations Models with Autocorrelated Errors," Econometrica, 42 (1974), 311-332.

[6] Fair, R. C.: "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First-Order Serially Correlated Errors," Econometrica, 38 (1970), 507-516.

[7] Fischer, S.: "Long Term Contracts, Rational Expectations, and the Optimal Money Supply Rule," Journal of Political Economy, 85 (1977), 191-205.

[8] Flood, R. and P. Garber: "The Compatibility of Fair's Method and Errors-in-Variables," Journal of Monetary Economics, (1980)

[9] Hansen, L.: "The Asymptotic Distribution of Least-Squares Estimators with Endogenous Regressors and Dependent Residuals," mimeo, Carnegie-Mellon University, 1979.

[10] ________: "Large Sample Properties of Generalized Method of Moments Estimators", mimeo, March 1980.

[11] Hansen, L. and R. Hodrick: "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," Journal of Political Economy.

[12] Hansen, L. and T. Sargent: "Formulating and Estimating Dynamic Linear Rational Expectations Models," Journal of Economic Dynamics and Control, 1 (1979).

[13] Hatanaka, M.: "On the Global Identification of the Dynamic Simultaneous Equation Model with Stationary Disturbances," International Economic Review, 16 (1975).

[14] Hausman, J.: "Specification Tests in Econometrics," Econometrica, 46 (1978).

[15] Hayashi, F.: "Estimation of Simultaneous Equation Models under Rational Expectations: A Survey and Some New Results," mimeo, Harvard University 1980.

[16] Hayashi, F. and C. Sims: "Efficient Estimation of Time Series Models with Predetermined, but not Exogenous, Instruments," mimeo, July 1980.

[17] Huizinga, J.: "Testing the Rationality of Inflation Forecast: Theory and Evidence from the Survey of Consumer Attitudes," mimeo, University of Chicago, 1981.

[18] Kennan, J.: "The Estimation of Partial Adjustment Models with Rational Expectations," Econometrica 47 (1979), 1441-1456.

[19] McCallum, B.: "Rational Expectations and the Natural Rate Hypothesis Some Consistent Estimates," Econometrica, 46 (1976), 43-52.

[20] ________: "Topics Concerning the Formulation, Estimation and Use of Macroeconomic Models with Rational Expectations," American Statistical Association Proceedings of the Business and Economic Statistics Section, 1979, 65-72.

[21] Mishkin, F.: "The Real Interest Rate: An Empirical Investigation," forthcoming in the Carnegie-Rochester Series on Public Policy.

[22] Obstfeld, M.: "Instrumental Variables Estimation with Non-Spherical Errors," mimeo, M.I.T., 1978.

[23] Phelps, E., and J. Taylor: "Stabilizing Powers of Monetary Policy Under Rational Expectations," Journal of Political Economy, 85 (1977), 163-190.

[24] Revankar, N.: "Analysis of Regressions Containing Serially Correlated and Serially Uncorrelated Error Components," International Economic Review, 21 (1980), 185-199

[25] Salemi, M. and T. Sargent: "The Demand for Money During Hyperinflation Under Rational Expectations: II," International Economic Review 20 (1979), 741-758.

[26] Scmidt, P.: Econometrics, New York: Marcel Dekker, Inc., (1976)

[27] Sims, C.: "Macroeconomics and Reality," Econometrica, 48 (1980), 1-48.

[28] Taylor, J.: "Estimation and Control of a Macroeconomic Model with Rational Expectations," Econometrica, 47 (1979), 1267-1286.

[29] Theil, H.: Economic Forecasts and Policy. Amsterdam: North-Holland, 1961.

[30] Wallis, K.: "Econometric Implications of the Rational Expectations Hypothesis," Econometrica, 48 (1980), 49-72.

[31] White, H.: "Instrumental Variables Regression on Cross-Section Data," mimeo, May 1980.

Appendix

This appendix provides proofs of Propositions 1 through 4 of Section II. The formal model throughout the appendix is,

$$y_t = Q_t f(\delta) + \varepsilon_t \qquad t = 1,2,\ldots T \tag{33}$$

where y_t and ε_t are scalar random variables, Q_t is a k x 1 random vector, and $f(\delta)$ is a one-to-one function taking the h x 1 parameter vector δ into a space of higher dimension. In addition, it is assumed that,

(A1) The parameter δ is identified and is an interior point in a compact space,

(A2) There exists an n x 1 random vector W_t such that $E(\varepsilon_t | N_t) = 0$ where N_t is any arbitrary subset of the set $\{ W_t, W_{t-1}, W_{t-2}, \ldots \varepsilon_{t-m}, \varepsilon_{t-m-1}, \ldots \}$,

(A3) y_t, Q_t and W_t are jointly stationary and ergodic,

(A4) $(1/T)(\partial f/\partial \delta')Q'W$ converges in probability to a matrix of full column rank uniformly in δ,

(A5) $(1/T)(\partial^2 f/\partial \delta_i \partial \delta)$ converges in probability to a constant matrix uniformly in δ for i = 1,2, . . h where δ_i is the i^{th} element of δ,

(A6) y_t, Q_t and W_t have finite second and fourth moments,

(A7) $E(W_t W_t')$ is nonsingular, and

(A8) $E(\varepsilon_t^2 W_t W_t')$ exists and is finite.

We first prove consistency of the estimators in Propositions 1,2, and 4. The proof is essentially the same in all three cases, a trivial extension of the consistency proof found in Amemiya (1974). We therefore present the proof only for two-step, two-stage least squares with a known Ω and describe how to modify the proof for NL2SLS and 2S2SLS with an estimated Ω.

(*) Given the model above, the two-step, two-stage least squares estimator with known Ω, d, is a consistent estimator for the true value of δ, δ^*.

Proof:

By the Mean Value Theorem,

$$f(d) - f(\delta^*) = \frac{\partial f}{\partial \delta'}\Big|_{\delta^+} (d - \delta^*) \tag{34}$$

for some δ^+ on a line segment joining d and δ^*. Therefore,

$$RW'Qf(d) - RW'Qf(\delta^*) = RW'Q \frac{\partial f}{\partial \delta'}\Big|_{\delta^+} (d - \delta^*) \tag{35}$$

$\Omega^{-1} = R'R$. Such a nonsingular R will exist because by (A8) and (A2) Ω exists and is a symmetric, positive definite matrix. Upon substituting $Qf(\delta^*) = y - \varepsilon$ into equation (35) and dividing by T, one gets

$$A - B = RW'Q \frac{\partial f}{\partial \delta'}\Big|_{\delta^+} (d - \delta^*)(1/T) \tag{36}$$

where $A = R(W'\varepsilon)/T$ and $B = RW'(y-Qf(d))/T$. Now $\operatorname{plim}_{T\to\infty} A = 0$ because R is nonstochastic, $E(W'\varepsilon) = 0$ by (A2), and by (A3) W and ε are ergodic. In addition, $\operatorname{plim}_{T\to\infty} B = 0$ because $0 \le B'B = (1/T^2)\ \Lambda(d) \le (1/T^2)\ \Lambda(\delta^*) = A'A$.

Thus the right-hand side of equation (35) converges in probability to 0. However, by (A4) and the fact that R is nonsingular, this can only be true if $\operatorname{plim}_{T\to\infty} d = \delta^*$.

QED

In order to modify this proof for NL2SLS all that is necessary is to replace $\Lambda(\ )$ with $\phi(\ )$ and redefine R so that $(W'W/T)^{-1} = R'R$. That $(W'W/T)^{-1}$

exists is guaranteed asymptotically by (A7) and (A3). This leaves only the additional assumption of $\underset{T\to\infty}{\text{plim}}\, R$ existing, which is assumed to hold. To modify the proof for 2S2SLS with an estimated Ω, all that is necessary is that R be defined by $\hat{\Omega}^{-1} = R'R$ and the additional assumption that $\underset{T\to\infty}{\text{plim}}\, R$ exists.

Before leaving the issue of consistency, one additional point is in order. The two-step, two-stage least squares estimator with estimated Ω is consistent provided only that $\underset{T\to\infty}{\text{plim}}\, (W'\varepsilon/T) = 0$ and does not require $\hat{\Omega}$ to be consistent. Of course misspecification of the true Ω may lead to W being picked so that $\underset{T\to\infty}{\text{plim}}\, (W'\varepsilon/T) \neq 0$, however.

We now turn to proving the asymptotic distribution of NL2SLS is as stated in Proposition 1. We again follow the proof of Amemiya (1974) and rely on the Central Limit Theorem used by Hansen (1979).

(**) Given the model above, the NL2SLS estimator defined by that $\hat{\delta}$ which minimizes the value of equation (17), has the property that if δ^* is the true value of δ,

$$\sqrt{T}\,(\hat{\delta} - \delta^*) \xrightarrow{d} N(0, \underset{T\to\infty}{\text{plim}}\, T^2 (V'W(W'W)^{-1}W'V)^{-1} D (V'W(W'W)^{-1}W'V)^{-1})$$

where $D = V'W(W'W)^{-1}\Omega(W'W)^{-1}W'V$ and

$$V = Q \left.\frac{\partial f}{\partial \delta'}\right|_{\delta^*} .$$

Proof:

By expanding the first order conditions for minimizing equation (17) around the true value δ^* we obtain,

$$\sqrt{T}\,(\hat{\delta} - \delta^*) = -\left[\frac{1}{T}\frac{\partial^2\phi}{\partial\delta\partial\delta'}\bigg|_{\delta^+}\right]^{-1} (1/\sqrt{T})\,\frac{\partial\phi}{\partial\delta}\bigg|_{\delta^*} \tag{37}$$

where again δ^+ is on the line segment joining $\hat{\delta}$ and δ^*. In addition,

$$\frac{\partial\phi}{\partial\delta}\bigg|_{\delta^*} = -\,2\,\frac{\partial f}{\partial\delta'}\bigg|_{\delta^*} Q'W(W'W)^{-1}W'\varepsilon \tag{38}$$

and

$$\frac{\partial^2\phi}{\partial\delta\partial\delta'}\bigg|_{\delta^*} = 2\,\frac{\partial f}{\partial\delta'}\bigg|_{\delta^*} Q'W(W'W)^{-1}W'Q\,\frac{\partial f}{\partial\delta'}\bigg|_{\delta^*} - 2\,H \tag{39}$$

where H is the matrix whose i^{th} row is,

$$\varepsilon'W(W'W)^{-1}W'\,\frac{\partial^2 f}{\partial\delta_i\partial\delta'}\bigg|_{\delta^*}. \tag{40}$$

Now $(1/\sqrt{T})W'\varepsilon$ converges in distribution to $N(0,\Omega)$ because it is equal to $(1/\sqrt{T})\sum_{i=1}^{T}(q_i = \varepsilon_i W_i')$ and the q_i here satisfy the same criteria that the z_i do in Hansen (1979). This convergence in distribution, along with (A4) and (A7) - which assure the convergence in probability of

$(1/T)\,\frac{\partial f}{\partial\delta'}\Big|_{\delta^*} Q'W(W'W/T)^{-1}$ - imply that,

$$(1/\sqrt{T})\;\frac{\partial\phi}{\partial\delta}\bigg|_{\delta^*} \xrightarrow{d} N(0,\underset{T\to\infty}{\text{plim}}\;(V'W/T)(W'W/T)^{-1}\Omega(W'W/T)^{-1}(W'V/T)). \tag{41}$$

In addition,

$$\underset{T\to\infty}{\text{plim}}\,\frac{1}{T}\frac{\partial^2\phi}{\partial\delta\partial\delta'}\bigg|_{\delta^*} = \underset{T}{\text{plim}}\,\frac{1}{T}\frac{\partial f'}{\partial\delta}\bigg|_{\delta^*} Q'W(W'W)^{-1}W'Q\,\frac{\partial f}{\partial\delta'}\bigg|_{\delta^*} \tag{42}$$

because $\underset{T\to\infty}{\text{plim}}\,(1/T)H = 0$ by (A5), (A7) and the fact that $\underset{T\to\infty}{\text{plim}}\,(1/T)\,W'\varepsilon = 0$.

The existence of the probability limit on the right-hand side of equation (42) is guaranteed by (A4).

Combining equations (41) and (42) and a lemma of Amemiya (1973) which guarantees that $\operatorname*{plim}_{T\to\infty} \frac{1}{T} \frac{\partial^2 \phi}{\partial\delta\partial\delta'}\Big|_{\delta^+} = \operatorname*{plim}_{T\to\infty} \frac{1}{T} \frac{\partial^2 \phi}{\partial\delta\partial\delta'}\Big|_{\delta^*}$ yields the desired result,

$$\sqrt{T}\,(\hat{\delta} - \delta^*) \xrightarrow{d} N(0, \operatorname*{plim}_{T\to\infty} T^2 (V'W(W'W)^{-1}W'V)^{-1} D (V'W(W'W)^{-1}W'V)^{-1}) \qquad (43)$$

where $D = V'W(W'W)^{-1}\Omega(W'W)^{-1}W'V$.

QED

The proof of the asymptotic distribution of two-step, two-stage least squares with an estimated Ω is the same as the proof when Ω is known so we provide a proof only for the case where one uses a consistently estimated $\hat{\Omega}$.

(***) Given the model above, the 2S2SLS estimator defined by that d which minimizes equation (21) - replacing Ω with a consistent estimate $\hat{\Omega}$ - has the property that if δ^* is the true value of δ,

$$\sqrt{T}\,(d - \delta^*) \xrightarrow{d} N(0, \operatorname*{plim}_{T\to\infty} T^2(V'W\Omega^{-1}W'V)^{-1}) \tag{44}$$

<u>Proof</u>:

Proceding as in the proof of (**) we can obtain the analogues of equations (37) - (40) where we merely replace ϕ in equation (37) with Λ and $(W'W)^{-1}$ in equations (38) - (40) with $\hat{\Omega}^{-1}$ to get an expression for $\sqrt{T}\,(d - \delta^*)$. We then can use the fact that $\sqrt{T}\,(W'\varepsilon)$ converges in distribution to $N(0,\Omega)$ to show,

$$(1/T)^{3/2} \left.\frac{\partial\Lambda}{\partial\delta}\right|_{\delta^*} \xrightarrow{d} N(0, \operatorname*{plim}_{T\to\infty} (V'W/T)\hat{\Omega}^{-1}\Omega\,\hat{\Omega}^{-1}(W'V/T)). \tag{45}$$

However note that the asymptotic variance in equation (45) can be rewritten as $\operatorname*{plim}_{T\to\infty} (V'W/T)\Omega^{-1}(W'V/T)$ since $\hat{\Omega}$ is a consistent estimate of Ω. Combining this with the fact that ,

$$\operatorname*{plim}_{T\to\infty} (1/T^2) \left.\frac{\partial^2\Lambda}{\partial\delta\partial\delta'}\right|_{\delta^*} = \operatorname*{plim}_{T\to\infty} (V'W/T)\Omega^{-1}(W'V/T) \tag{46}$$

yields the desired result,

$$\sqrt{T}\,(d - \delta^*) \xrightarrow{d} N(0, \operatorname{plim} T^2(V'W\Omega^{-1}W'V)^{-1}). \tag{47}$$

QED

Before proceding to the last proof of this appendix, it is useful to examine more clearly why the asymptotic distribution of the two-step, two-stage estimator does not depend on whether Ω is estimated consistently or known a priori. In the standard Generalized Least Squares case one has to have the information matrix be block diagonal to obtain such a result. The reason that 2S2SLS is different is that while most GLS procedures need $(1/\sqrt{T})(W'\hat{\Omega}^{-1}\varepsilon)$ to converge in distribution, 2S2SLS needs only the asymptotic distribution of $(1/\sqrt{T})W'\varepsilon$. The only place $\hat{\Omega}^{-1}$ appears in $\sqrt{T}\,(d - \delta^*)$ is in terms which converge in probability.

We now turn to the last proof, which formalizes that 2S2SLS is more efficient than NL2SLS.

(****) Given the model above, the 2S2SLS estimator is more efficient than the NL2SLS estimator in the sense that the asymptotic variance of the NL2SLS estimator minus the asymptotic variance of the 2S2SLS estimator is a positive semi-definite matrix.

Proof:

In order to show A - B is positive semi-definite it is suffcient to show that $B^{-1} - A^{-1}$ is positive semi-definite. Hence we will show,

$$V'W\Omega^{-1}W'V - V'W(W'W)^{-1}W'V\{V'W(W'W)^{-1}\Omega(W'W)^{-1}W'V\}^{-1}V'W(W'W)^{-1}W'V \quad (48)$$

is positive definite. By factoring a V'W and W'V from equation (48) we see that it is positive semi-definite if and only if,

$$\Omega^{-1} - (W'W)^{-1}W'V\{V'W(W'W)^{-1}\Omega(W'W)^{-1}W'V\}^{-1}V'W(W'W)^{-1} \quad (49)$$

is. But equation (49) can be put in the form,

$$(R'R)^{-1} - H(H'R'RH)^{-1}H' \quad (50)$$

where $\Omega = R'R$ and $H = (W'W)^{-1}W'V$. We can now use the fact that if R is an m x k matrix of rank k and H is a k x r matrix of rank r $(r \leq k)$, then equation (50) is positive semi-definite. (See Schmidt (1976) p163). Note that as defined R is k x k and nonsingular since it is the Cholesky decompositon of a positive definite and symmetric matrix and $H = (W'W)^{-1}W'V$ has full column rank by (A4) and (A7).