EFFICIENT ESTIMATION AND IDENTIFICATION OF SIMULTANEOUS EQUATIONS MODELS WITH COVARIANCE RESTRICTIONS

Jerry A. Hausman, MIT
Whitney K. Newey, Princeton
William E. Taylor, AT&T

Econometric Research Program
Research Memorandum No. 329

Revised March 1986

This paper is offered in recognition of the fiftieth birthday of the Cowles Foundation. Hausman and Newey thank the NSF for research support. This paper is a revision of an earlier version presented at the European Econometric Meetings, 1982. G. Chamberlain, F. Fisher, A. Deaton, T. Rothenberg, P. Ruud, and three anonymous referees have provided useful comments.

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, NJ 08544

EFFICIENT ESTIMATION AND IDENTIFICATION OF SIMULTANEOUS EQUATION MODELS WITH COVARIANCE RESTRICTIONS

HEADNOTE

In this paper we consider estimation of simultaneous equations models with covariance restrictions. We first consider FIML estimation and extend Hausman's (1975) instrumental variables interpretation of the FIML estimator to the covariance restrictions case. We show that, in addition to the predetermined variables from the reduced form, FIML also uses estimated residuals as instruments for the equations with which they are uncorrelated.

A slight variation on the instrumental variables theme yields a simple, efficient alternative to FIML. Here we augment the original equation system by additional equations that are implied by the covariance restrictions. We show that when these additional equations are linearized around an initial consistent estimator and three-stage least squares is performed on the original equation system together with the linearized equations implied by the covariance restrictions, an asymptotically efficient estimator is obtained.

We also present a relatively simple method of obtaining an initial consistent estimator when the covariance restrictions are needed for identification. This estimator also makes use of additional equations that are implied by the covariance restrictions.

In the final section of the paper we consider identification from the point of view of the moment restrictions that are implied by instrument-residual orthogonality and the covariance restrictions. We show that the assignment condition of Hausman and Taylor (1983) provides necessary conditions for the identification of the structural parameters.

1. Introduction

In the pioneering research in econometrics done at the Cowles Foundation, estimation techniques for simultaneous equations models were studied extensively. Maximum likelihood estimation methods were applied to both the single equation case (LIML) and to the complete simultaneous equations models (FIML). It is interesting to note that while questions of identification were completely solved for the case of coefficient restrictions, the problem of identification with covariance restrictions remained. Further research by Fisher (1966), Rothenberg (1971), and Wegge (1965) advanced our knowledge in this field. In a companion paper, Hausman and Taylor (1983) give conditions in terms of the interaction of restrictions on the disturbance covariance matrix and restrictions on the coefficients of the endogenous variables for the identification problem. What is especially interesting about their conditions is that covariance restrictions have the interpretation that they provide additional instrumental variables for use in estimation, which links them to the situation where only coefficient restrictions are present.

For full information maximum likelihood (FIML), the Cowles Foundation research considered the case of covariance restrictions when the covariance matrix of the residuals is specified to be diagonal (Koopmans, Rubin, and Leipnik (1950)). The case of a diagonal covariance matrix is also analyzed by Malinvaud (1970) and by Rothenberg (1973). But covariance restrictions are a largely unexplored topic in simultaneous equations estimation, perhaps because of a reluctance to specify a priori restrictions on the disturbance covariances.[1] However, an important contributory cause of this situation may have been the lack of a simple, asymptotically efficient, estimation procedure for the case of covariance restrictions. Rothenberg and Leenders (1964), in their proof of the efficiency of the Zellner-Theil (1962) three stage least squares (3SLS) estimator, showed that the presence of covariance restrictions would make FIML asymptotically more efficient than 3SLS. In fact, imposing the covariance restrictions on the conventional 3SLS estimator does not improve its asymptotic efficiency. Thus efficient estimation seemed to require FIML.[2] The role of covariance restrictions in establishing identification in the simultaneous equations model was not fully understood, nor did imposing such restrictions improve the asymptotic efficiency of the most popular full information estimator. Perhaps these two reasons, more than the lack of a priori disturbance covariance restrictions, may have led to their infrequent use.

Since the identification results of Hausman and Taylor (1983) have an instrumental variable interpretation, it is natural to think of using instrumental variables as an approach to estimation when covariance

---

1. Of course, at a more fundamental level covariance restrictions are required for any structural estimation in terms of the specification of variables as exogenous or predetermined, c.f. Fisher (1966, Ch. 4).

2. Rothenberg and Leenders (1964) do propose a linearized maximum likelihood estimator which corresponds to one Newton step beginning from a consistent estimate. As usual, this estimator is asymptotically equivalent to FIML. Also, an important case in which covariance restrictions have been widely used is that of a recursive specification in which FIML coincides with ordinary least squares (OLS).

restrictions are present. Hausman (1975) gave an instrumental variables interpretation of FIML when no covariance restrictions were present, which we extend to the case with covariance restrictions. The interpretation seems especially attractive because we see that instead of using only the predetermined variables from the reduced form as instrumental variables, FIML also uses estimated residuals as instrumental variables for equations with which they are uncorrelated. Thus more instrumental variables are used to form the FIML estimator than in the case where covariance restrictions are absent.

A slight variation on the instrumental variables theme yields a useful alternative to FIML. Here we augment the 3SLS estimator by additional equations which the covariance restrictions imply. That is, a zero covariance restriction means that a pair of disturbances is uncorrelated, and therefore that the product of the corresponding residuals can itself be used in estimation as the residual of an additional equation. These additional equations are nonlinear in the parameters but can be linearized at an initial consistent estimator, and then 3SLS performed on the augmented equation system. This estimator, which we call augmented three stage least squares (A3SLS), is shown to be more efficient than the 3SLS estimator when effective covariance restrictions are present and to be at least as efficient as FIML. The A3SLS estimator thus provides a computationally convenient estimator which is also asymptotically efficient. We also consider convenient methods of using the extra equations which are implied by the covariance restrictions to form an initial consistent estimator when the covariance restrictions are necessary for identification.

In addition to the development of the A3SLS estimator, we also reconsider the assignment condition for identification defined by Hausman and Taylor (1983). We prove that the assignment condition which assigns covariance restrictions to one of the two equations from which the restriction arises provides a necessary condition for identification. A corresponding rank condition provides a stronger necessary condition than the generalized rank condition of Fisher (1966). These necessary conditions apply equation by equation. We also give a necessary and sufficient condition for local identification in terms of the structural parameters of the entire system.

2. Estimation in a Two Equation Model

We begin with a simple two equation simultaneous equation model with a diagonal covariance matrix, since many of the key results are straightforward to derive in this context.[3] The model specification we consider is

$$(2.1) \qquad y_1 = \beta_{12} y_2 + \gamma_{11} z_1 + \varepsilon_1,$$

$$(2.2) \qquad y_2 = \beta_{21} y_1 + \gamma_{22} z_2 + \varepsilon_2.$$

We assume that we have T observations so that each variable in equations (2.1) and (2.2) represents a T x 1 vector. The stochastic assumptions are $E(\varepsilon_{it}|z_1,z_2) = 0$ for i=1,2, $var(\varepsilon_{it}|z_1,z_2) = \sigma_{ii}$, $cov(\varepsilon_{1t}\varepsilon_{2t}| z_1,z_2) = \sigma_{12} = 0$.

Inspection of equations (2.1) and (2.2) shows that the order condition is satisfied so that each equation is identified by coefficient restrictions alone, so long as the rank condition does not fail. If the covariance restriction is neglected, each equation is just-identified so that 3SLS is identical to 2SLS on each equation. Note that for each equation, 2SLS uses the instruments $W_i = (Z\Pi_j, z_i)$, $i \neq j$, where $Z = (z_1, z_2)$ and Π_j is the vector of reduced form coefficients for the (other) included endogenous variables.[4]

To see how FIML differs from the instrumental variables (IV) estimator, we solve for the first order conditions of the likelihood function

---

3. Such a model might arise from a supply and a demand equation, where the specification of a diagonal disturbance covariance matrix corresponds to the assumption that the supply effect of demand shocks is fully captured by the inclusion of price in the supply equation.

4. Of course, because of the condition of just identification, a numerically identical result would be obtained if instruments $W_i = (z_1, z_2)$ were used.

under the assumption that the ε_i's are normally distributed.[5] For the two equation example, the likelihood function takes the form

$$L = c - \frac{T}{2} \log (\sigma_{11}\sigma_{22}) + T \log |1-\beta_{12}\beta_{21}| \tag{2.3}$$

$$- \frac{1}{2} [\frac{1}{\sigma_{11}} (y_1-X_1\delta_1)'(y_1-X_1\delta_1) + \frac{1}{\sigma_{22}}(y_2-X_2\delta_2)'(y_2-X_2\delta_2)]$$

where c is a constant and the X_i's and δ_i's contains the right hand side variables and unknown coefficients respectively, e.g., $X_1 = (y_2, z_1)$ and $\delta_1 = (\beta_{12}, \gamma_{11})'$.

To solve for the FIML estimator, we find the first order conditions for equation (2.1); results for the second equation are identical. The three first order conditions are

$$- \frac{T\beta_{21}}{1-\beta_{12}\beta_{21}} + \frac{1}{\sigma_{11}} (y_1-X_1\delta_1)'y_2 = 0, \tag{2.4a}$$

$$\frac{1}{\sigma_{11}}(y_1- X_1\delta_1)'z_1 = 0, \tag{2.4b}$$

---

5. Throughout the paper FIML will refer to the estimator which is obtained by performing maximum likelihood under the assumption that the disturbances are normally distributed. If they do not have a normal distribution, then FIML will be a quasi (or pseudo) maximum likelihood estimator. Of course, as is the case for linear simultaneous equation estimation with only coefficient restrictions, the FIML estimator remains consistent when the disturbances are not normally distributed.

$$(2.4c) \qquad -\frac{T}{\sigma_{11}} + \frac{1}{\sigma_{11}^2}(y_1 - X_1\delta_1)'(y_1 - X_1\delta_1) = 0.$$

Rearranging equation (2.4c) yields the familiar solution for the variance, $\sigma_{11} = (1/T)(y_1 - X_1\delta_1)'(y_1 - X_1\delta_1)$. Equation (2.4b) has the usual OLS form which is to be expected since z_1 is an exogenous variable. We can solve equation (2.4a) in a particular way, using the reduced form, to see the precise role of the covariance restrictions in the model. After multiplying equation (2.4a) by σ_{11} we obtain

$$(2.5) \qquad 0 = y_2'(y_1 - X_1\delta) - \frac{T\beta_{21}\sigma_{11}}{1-\beta_{12}\beta_{21}} = \left(Z\Pi_2 + v_2 - \frac{\beta_{21}\varepsilon_1}{1-\beta_{12}\beta_{21}}\right)'\varepsilon_1$$

$$= \left(Z\Pi_2 + \frac{\varepsilon_2}{1-\beta_{12}\beta_{21}}\right)'\varepsilon_1,$$

where we substitute $y_2 = Z\Pi + v_2$ to obtain the second equality and $v_2 = (\beta_{21}\varepsilon_1 + \varepsilon_2)/(1-\beta_{12}\beta_{21})$ to obtain the third. Without the covariance restrictions, we would have the result

$$(2.6) \qquad (Z\Pi_2)'\varepsilon_1 = 0, \quad \Pi_2 = \left(\frac{\beta_{21}\gamma_{11}}{1-\beta_{12}\beta_{21}}, \frac{\gamma_{22}}{1-\beta_{12}\beta_{21}}\right),$$

which is the instrumental variables interpretation of FIML obtained by Hausman (1975). In equation (2.5) we have the extra term $\varepsilon_2/(1-\beta_{12}\beta_{21})$, so that the instrument in equation (2.5) is a linear combination of three variables, z_1, z_2, and ε_2, each of which is uncorrelated with ε_1, rather than just z_1 and z_2. What has happened is that FIML has used the covariance restriction to add to the set of instrumental variables that are used to estimate the parameters of equation (2.1). In addition to the predetermined variables z_1 and z_2, FIML also uses ε_2 as an instrumental variable. This example makes it clear why 3SLS is not asymptotically efficient when covariance restrictions are present. The additional instrumental variable ε_2 is not utilized by the conventional 3SLS estimator.

Two other important cases can be examined with our simple two equation model. First, suppose that $\beta_{21} = 0$. The specification is then triangular, and given the diagonal covariance matrix, the model is recursive. Here, the FIML instrument in equation (2.5) is $Z\Pi_2 + v_2 = y_2$ so that y_2 is predetermined and FIML becomes OLS as expected. The second case returns to $\beta_{12} \neq 0$ but sets $\gamma_{22} = 0$. The first equation is no longer identified by coefficient restrictions alone, but it is identified by the covariance restrictions because the FIML instruments are

$$(2.7) \qquad W_1 = \left(Z\,\Pi_2 + \frac{\varepsilon_2}{1-\beta_{12}\beta_{21}}, z_1\right).$$

Because of the addition of the residual term in W_1, the instrument matrix has full column rank and the coefficients can be estimated.

In general, FIML needs to be iterated to solve the first order conditions. To obtain a convenient and efficient alternative to FIML we make use of the moment restrictions which are implied by the covariance restrictions. The variables z_1 and z_2 can be used as instruments because they are both uncorrelated with ε_1 and ε_2. In addition FIML can use ε_1 as an instrument for ε_2 and ε_2 as an instrument for ε_1 because $E(\varepsilon_{1t}\varepsilon_{2t}) = 0$. We can account for this extra moment restriction by augmenting equations

(2.1) and (2.2) with the additional equation

$$(2.8) \qquad (y_{1t} - X_{1t}\delta_1)(y_{2t} - X_{2t}\delta_2) = e_t,$$

where e_t has mean zero.

This additional equation is nonlinear in the parameters, but when an initial consistent estimator $\hat{\delta}$ is available it can be linearized around the initial estimate. A first order Taylor's expansion of equation (2.8) around $\hat{\delta}$ gives

$$\hat{\varepsilon}_{1t}\hat{\varepsilon}_{2t} - \hat{\varepsilon}_{2t}X_{1t}(\delta_1 - \hat{\delta}_1) - \hat{\varepsilon}_{1t}X_{2t}(\delta_2 - \hat{\delta}_2) = r_t,$$

where $\hat{\varepsilon}_{it} = y_{it} - X_{it}\hat{\delta}_i$ and r_t is equal to e_t plus a second order term. Collecting terms with unknown values of the parameters on the right-hand side gives

$$(2.9) \qquad \hat{\varepsilon}_{1t}\hat{\varepsilon}_{2t} + \hat{\varepsilon}_{2t}X_{1t}\hat{\delta}_1 + \hat{\varepsilon}_{1t}X_{2t}\hat{\delta}_2 = \hat{\varepsilon}_{2t}X_{1t}\delta_1 + \hat{\varepsilon}_{1t}X_{2t}\delta_2 + r_t.$$

The parameter vector δ can now be estimated, while accounting for the presence of covariance restrictions, by joint 3SLS estimation of equations (2.1), (2.2), and (2.9), imposing the cross equation restrictions and using a vector of ones as the instrument for the last equation.[6] We will refer to this estimator as augmented 3SLS (A3SLS), since the original equation has

[6]. A consistent estimator of the joint covariance matrix of the disturbances of the augmented equations system can be obtained in the usual way from the residuals for equations (2.1), (2.2), and (2.9) with $\delta = \hat{\delta}$. Also, z_1 and z_2 can also be used as instruments for equation (2.9) without affecting the efficiency of the estimator. See Section 4.

been augmented by an equation which is generated by the covariance restriction. It will be shown below that in general the A3SLS estimator is asymptotically equivalent to FIML when the disturbances of the original equation system are normally distributed and is also efficient relative to FIML when the disturbances are nonnormal.

Direct use of the extra moment restrictions also yields a fruitful approach to estimation when covariance restrictions are needed for identification. For example, suppose that $\gamma_{22}=0$ in equation (2.2). Then z_2 is no longer useful as an instrument because it does not appear in the reduced form. However, we can still obtain an estimator of the unknown parameter vector $(\beta_{12}, \gamma_{11}, \beta_{21})$ by utilizing the covariance restriction. Consider an estimator $\hat{\delta}$ which is obtained as the solution to the three equations

$$(2.10a)\quad z_1'(y_1 - \hat{\beta}_{12}y_2 - \gamma_{11}z_1) = 0,$$

$$(2.10b)\quad z_1'(y_2 - \hat{\beta}_{21}y_1) = 0,$$

$$(2.10c)\quad (y_2 - \hat{\beta}_{21}y_1)'(y_1 - \hat{\beta}_{12}y_2 - \hat{\gamma}_{11}z_1) = 0.$$

The first two equations use the instrumental variable moment conditions that z_{1t} is orthogonal to ε_{1t} and ε_{2t} while the third equation uses the moment condition $E[\varepsilon_{1t}\varepsilon_{2t}]=0$. This estimator is a generalized method of moments (GMM, see Hansen (1982)) estimator which uses the moment condition implied by the covariance restriction in addition to the usual instrumental variable orthogonality conditions. Generally the solution to such an equation will require iteration because the product of two residuals is quadratic in the unknown parameters, although in this simple example, which has a recursive structure and is thus covered by the results of Section 4.2 of Hausman and Taylor (1983), iteration is not required. Note that the solution to equation (2.10) can be obtained by first solving (2.10b) for $\hat{\beta}_{21}$ and then solving equations (2.10a) and (2.10c) for $\hat{\beta}_{12}$ and $\hat{\gamma}_{11}$, which amounts to first doing 2SLS on equation (2.2) using z_1 as an instrument, and then doing 2SLS on equation (2.1) using z_1 and $\hat{\varepsilon}_2 = y_2 - \hat{\beta}_{21} y_1$ as instruments.

3. FIML Estimation in the M-equation Case

We now turn to the general case where zero restrictions are present on some elements of the covariance matrix, but the covariance matrix is not necessarily assumed to be diagonal. We consider the standard linear simultaneous equations model where all identities are assumed to have been substituted out of the system of equations:

$$(3.1) \qquad YB + Z\Gamma = U$$

where Y is the TxM matrix of jointly endogenous variables, Z is the TxK matrix of predetermined variables, and U is the TxM matrix of the structural disturbances of the system. The model has M equations and T observations. It is assumed that B is nonsingular and that Z is of full rank. We assume that plim (1/T) (Z'U) = 0, and that the second order moment matrices of the current predetermined and endogenous variables have nonsingular probability limits. Lastly, if lagged endogenous variables are included as predetermined variables, the system is assumed to be stable.

The structural disturbances are assumed to be mutually independent and identically distributed as a nonsingular M-variate normal distribution:

$$(3.2) \qquad \text{vec}U \sim N(0, \Sigma \otimes I_T)$$

where Σ is positive definite.[7] However, we allow for restrictions on the elements of Σ of the form $\sigma_{ij} = 0$ for $i \neq j$, which distinguishes this from the case that Hausman (1975) examined. In deriving the first order conditions for the likelihood function, we will only solve for the unknown

7. Here and elsewhere vec(.) denotes the usual column vectorization.

elements of Σ rather than the complete matrix as is the usual case. Using the results of Hausman and Taylor (1983) and Section 5, we assume that each equation in the model is identified by use of zero coefficient restrictions on the elements of B and Γ and covariance restrictions on elements of Σ.

We will make use of the reduced form specification,

$$(3.3) \qquad Y = -Z\Gamma B^{-1} + UB^{-1} = Z\,\Pi + V.$$

The other form of the original system of equations which will be useful is the so-called "stacked" form. We use the normalization rule $B_{ii} = 1$ for all i and then rewrite each equation in regression form where only unknown parameters appear on the right-hand side:

$$(3.4) \qquad y_i = X_i\delta_i + u_i$$

where $X_i = [Y_i, Z_i]$, $\delta_i' = [\beta_i', \gamma_i']$, Y_i is the Txr_i matrix of included endogenous variables, Z_i is a Txs_i matrix of included predetermined variables, and δ_i is the $q_i = r_i + s_i$ dimensional vector of structural coefficients for the ith equation. It will prove convenient to stack these M equations into a system

$$(3.5) \qquad y = X\delta + u$$

where

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix}, \quad X = \text{diag}[X_1, \ldots, X_M] = \begin{bmatrix} X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_M \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_M \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \vdots \\ u_M \end{bmatrix}.$$

Note that δ is the $q = \Sigma_i q_i$ dimensional vector of structural coefficients. Likewise, we stack the reduced form equations

$$(3.6) \qquad y = \tilde{Z}\tilde{\Pi} + v$$

where $\tilde{Z} = [I_M \otimes Z]$ and $\tilde{\Pi} = \text{vec}(\Pi)$ is the vector of reduced form coefficients.

The log likelihood function arises from the model specification in equation (3.1) and the distribution assumption of equation (3.2):

$$(3.7) \qquad L(B, \Gamma, \Sigma) = c - \frac{T}{2} \log \det(\Sigma) + T \log |\det(B)|$$

$$- \frac{T}{2} \text{tr}\left[\frac{1}{T} \Sigma^{-1}(YB + Z\Gamma)'(YB + Z\Gamma)\right]$$

where the constant c is disregarded in maximization procedures. We now calculate the first order necessary conditions for a maximum by matrix differentiation. The procedures used and the conditions derived are the same as in Hausman (1975, p. 730). To reduce confusion, we emphasize that we only differentiate with respect to unknown parameters, and we use the symbol $\overset{u}{=}$ to remind the reader of this fact.[8] Thus the number of equations in each block of the first order conditions equals the number of unknown parameters; e.g., the number of equations in (3.8a) below equals the number of unknown parameters in B rather than M^2. The first order conditions for δ are

$$\text{(3.8a)} \qquad \frac{\partial L}{\partial B} : T(B')^{-1} - Y'(YB + Z\Gamma)\Sigma^{-1} \overset{u}{\equiv} 0,$$

$$\text{(3.8b)} \qquad \frac{\partial L}{\partial \Gamma} : - Z'(YB + Z\Gamma)\Sigma^{-1} \overset{u}{\equiv} 0,$$

In particular, note that we cannot postmultiply equation (3.8b), or later, the transformed versions of equation (3.8a), to eliminate Σ^{-1}, as a simple two equation example will easily convince the reader.

To state the first order condition for the unrestricted Σ parameters, let σ^* be a $M(M+1)/2$ dimensional column vector of the distinct elements of Σ, $\sigma^* = (\sigma_{11}, \ldots, \sigma_{M1}, \ldots, \sigma_{22}, \ldots, \sigma_{M2}, \ldots, \sigma_{MM})'$, and let R' be the $M^2 x(M+1)M/2$ matrix such that $\text{vec}(\Sigma) = R'\sigma^*$ (see Richard (1975)). Noting that for a symmetric, positive definite matrix A, $\partial \ln\det(A)/\partial \text{vec}A = \text{vec}(A^{-1})$, $\partial \text{tr}(U'UA)/\partial \text{vec}A = \text{vec}(U'U)$, and $\partial \text{vec}(A^{-1})/\partial \text{vec}A = - A^{-1} \otimes A^{-1}$, we obtain the vector first order condition

8. An alternative procedure is to use Lagrange Multiplier relationships of the type $0 = \sigma_{ij} = (y_i - X_i\delta_i)'(y_j - X_j\delta_j)$ for known elements of Σ, but the approach adopted in the paper seems more straightforward. We are grateful to an anonymous referee and Paul Ruud for pointing out an error in the FIML derivation in a previous draft.

(3.9) $-(\text{vec}[T\Sigma - (YB+Z\Gamma)'(YB+Z\Gamma)])' \; \Sigma^{-1} \otimes \Sigma^{-1} R' \overset{u}{=} 0$

Consider equation (3.8a). Using the identities $\Sigma\Sigma^{-1}=I$, $Y=Z\Pi+V$, and $V=(YB+Z\Gamma)B^{-1}$ we obtain

(3.10) $[(B')^{-1}(T\Sigma-(YB+Z\Gamma)'(YB+Z\Gamma))-(Z\Pi)'(YB+Z\Gamma)]\Sigma^{-1} \overset{u}{=} 0.$

Equation (3.9) characterizes $T\Sigma-(YB+Z\Gamma)'(YB+Z\Gamma)$ in the zero covariance restrictions case, while from equation (3.10) we see that it is precisely the presence of this term which causes the instrumental variables for Y to differ from $Z\Pi$. We can use equations (3.9) and (3.10) to obtain an instrumental variable interpretation for FIML with covariance restrictions.

Suppose there are a total of L distinct covariance restrictions and let S be a $M^2 \times L$ selection matrix of rank L such that the covariance restrictions are given by $S'\text{vec}(\Sigma) = 0$.[9] The FIML residuals can be added to the list of instrumental variables by allowing $W=[\tilde{Z},(I_M \otimes U)S]$ to be the T x MK+L matrix of instrumental variables, where $U = YB+Z\Gamma$. Note that each column of S corresponds to exactly one covariance restriction $\sigma_{ij} = 0$ for $i \neq j$. The column of S which corresponds to $\sigma_{ij} = 0$ will either select u_i as an instrument for equation j or u_j as an instrument for equation i. In the two equation example of Section 2 with $\sigma_{12}= 0$, S is either (0,1,0,0)' or (0,0,1,0)'. Then, for example, if
S = (0,1,0,0)',

$$W = \begin{bmatrix} Z & 0 & u_2 \\ 0 & Z & 0 \end{bmatrix},$$

9. S is not unique because of symmetry of Σ.

and u_2 is selected as an instrument for equation 1. Also let $\tilde{B}_i = [(B^{-1})_i, 0_i]$, where 0_i is a M x s_i null matrix which corresponds to the s_i included exogenous variables and $(B^{-1})_i$ is the matrix of columns of B^{-1} corresponding to the r_i included right-hand side endogenous variables, and let $\tilde{B} = \text{diag}(\tilde{B}_1', \ldots, \tilde{B}_M')$. Let P be the M^2 dimensional permutation matrix such that PvecA = vecA' for any M dimensional square matrix A. Let $D_i = [\Pi_i, I_i]$, where Π_i is the columns of Π corresponding to the included right hand side endogenous variables and I_i is the selection matrix such that $Z_i = ZI_i$, and let $\tilde{D} = \text{diag}(D_1', \ldots, D_M')$.

Theorem 3.1: For $H' = [\tilde{D}\Sigma^{-1} \otimes I_K, \tilde{B}(I_M \otimes \Sigma)(I+P)S[S'\Sigma \otimes \Sigma(I+P)S]^{-1}]$, the FIML estimator δ satisfies

$$\delta = (H'W'X)^{-1}H'W'Y . \tag{3.11}$$

Equation (3.11) demonstrates the essential difference for FIML estimation which arises between the case of no covariance constraints and the present situation. We see that in addition to the usual instrumental variables $\tilde{Z}$, we have the extra instrumental variables $(I_M \otimes U)S$ which are uncorrelated structural residuals. Thus FIML uses the covariance restrictions to form a better instrumental variables estimator.[10]

We now calculate the asymptotic Cramer-Rao bound for the estimator. Under our assumptions, we have a linear structural model for an i.n.i.d. specification. We do not verify regularity conditions here since they have been given for this model before, e.g., Koopmans and Hood (1953) or

10. When Σ is restricted to be diagonal, FIML can be given the IV interpretation discussed in Hausman and Taylor (1983).

Rothenberg (1973).[11] Let $\tilde{\sigma}$ be the $[M(M+1)/2]-L$ dimensional vector of unrestricted elements of σ^* and let $\tilde{S}$ be the $([M(M+1)/2]-L) \times M(M+1)/2$ matrix such that $\tilde{\sigma} = \tilde{S}\sigma^*$. Also, let $Q = \text{plim}(Z'Z/T)$.

Theorem 3.2: If the structure disturbances are normally distributed then the information matrix for the unknown parameter vectors δ and $\tilde{\sigma}$ is given by

$$(3.12)\quad J(\delta, \tilde{\sigma}) = \begin{bmatrix} \tilde{B}P\tilde{B}' + \text{plim}\, \frac{1}{T} X'(\Sigma^{-1} \otimes I_T)X & \tilde{B}(\Sigma^{-1} \otimes I_M)R'\tilde{S}' \\ \tilde{S}R(\Sigma^{-1} \otimes I_M)\tilde{B}' & \frac{1}{2}\tilde{S}R(\Sigma^{-1} \otimes \Sigma^{-1})R'\tilde{S}' \end{bmatrix}.$$

Further, the inverse of the Cramer-Rao bound for δ is, for $F = R(\Sigma^{-1} \otimes \Sigma^{-1})R'$,

$$(3.13)\quad (J^{\delta\delta})^{-1} = \tilde{D}(\Sigma^{-1} \otimes Q)\tilde{D}' + 2\tilde{B}(\Sigma^{-1} \otimes I_M)R'[F^{-1} - \tilde{S}'(\tilde{S}F\tilde{S}')^{-1}\tilde{S}]R(\Sigma^{-1} \otimes I_M)\tilde{B}'.$$

The first term in equation (3.13) is the inverse of the 3SLS asymptotic covariance matrix. Since the second term is positive semi-definite, 3SLS is asymptotically inefficient relative to FIML, in the presence of covariance restrictions.

Further insight into the efficiency gain from imposing the covariance restrictions can be obtained by examining the diagonal covariance matrix case. Let P_{ij} be the ijth M-dimensional square block of P, i,j=1,...,M, and let P* be the $M^2 \times M$ matrix $P^* = [P_{11},\ldots,P_{MM}]'$.

11. The most straightforward approach to regularity conditions is to use the reduced form. The reduced form has a classical multivariate least squares specification subject to nonlinear parameter restrictions. Since the likelihood function is identical for either the structural or reduced form specification, the more convenient form can be used for the specific problem being considered.

Corollary 3.3: If the structural disturbances are normally distributed and the covariance matrix is diagonal the information matrix for δ and $\tilde{\sigma}$ is given by

$$(3.14)\quad J(\delta, \sigma_{11}, \ldots, \sigma_{MM}) = \begin{bmatrix} \tilde{B}P\tilde{B}' + \text{plim}\,\frac{1}{T} X^{-1}(\Sigma^{-1} \otimes I_T)X & \tilde{B}P^*\Sigma^{-1} \\ \Sigma^{-1} P^{*\prime}\, \tilde{B}' & \frac{1}{2}\Sigma^{-1}\Sigma^{-1} \end{bmatrix} .$$

Further, the inverse of the Cramer-Rao lower bound for δ is

$$(3.15)\quad (J^{\delta\delta})^{-1} = \tilde{D}(\Sigma^{-1} \otimes Q)\tilde{D}' + \tilde{B}(P + \Sigma^{-1} \otimes \Sigma - 2P^*P^{*\prime})\tilde{B}' .$$

The first term in equation (3.15) is the inverse of the 3SLS asymptotic covariance matrix. By comparing the first term with the second term we can easily see that the larger is the covariance matrix Σ of the disturbance vector U_t relative to the second moment matrix Q of the predetermined variables from the reduced form, the larger is the efficiency gain which can be obtained by imposing the covariance restrictions. For example, if Σ is multiplied by a scalar which is larger than one then the second term is unaffected while the first term decreases. In other words, the efficiency gain from imposing covariance restrictions will be relatively large where the population r-squared for the reduced form equations is relatively small, as might be the case in cross-section data.

4. Instrumental Variables in the M-Equation Case

A convenient and efficient alternative to FIML can be obtained by utilizing the extra moment conditions which follow from the covariance restrictions. In the absence of covariance restrictions conventional instrumental variable estimators, such as 2SLS and 3SLS, make use of the fact that the instrumental variables are orthogonal to the disturbances. The covariance restrictions add moment conditions which can also be used in estimation. Let S be a selection matrix with $S'\mathrm{vec}\,\Sigma=0$, as introduced in Section 3, and let $e_t = S'\mathrm{vec}(U_t'U_t)$, where U_t is the tth row of U. Then the covariance restrictions imply that the Lx1 vector e_t has mean zero conditional on Z_t. We can account for these additional moment restrictions by augmenting the original M equations with the L additional equations

$$(4.1) \qquad S'[(y_t-X_t\delta)\otimes(y_t-X_t\delta)] = e_t, \quad (t=1,\ldots,T),$$

where $y_t = (y_{1t}, \ldots, y_{Mt})'$ and $X_t = \mathrm{diag}[X_{1t}, \ldots, X_{Mt}]$.

These additional equations are nonlinear (quadratic) in the parameters. When an initial consistent estimator $\hat{\delta}$ of the parameter vector δ is available this nonlinearity can be eliminated by linearizing the extra equations around $\hat{\delta}$. Using the fact that for an M-dimensional square matrix A it is the case that $\mathrm{vec}(A') = P\mathrm{vec}(A)$, and using $\partial\mathrm{vec}(U_t'(y_t-X_t\delta)')/\partial\delta = \partial((y_t-X_t\delta)\otimes U_t')/\partial\delta = -X_t\otimes U_t'$ we can calculate the first-order Taylor's expansion of equation k of (4.1) around $\hat{\delta}$,

$$(4.2) \qquad \hat{e}_{tk} + S_k'(I+P)(X_t\otimes\hat{U}_t')\hat{\delta} = S_k'(I+P)(X_t\otimes\hat{U}_t')\delta + r_{kt}.$$

where I is an M^2 dimensional identity matrix, S_k is the kth column of S, $\hat{U}_t'$

$= y_t - X_t\hat{\delta}$, $\hat{e}_{tk} = S_k'(\hat{U}_t' \otimes \hat{U}_t')$, r_{kt} is equal to e_{tk} plus a second order term, and terms with δ are collected on the right-hand side. Let

$$y_{ak} = (\hat{e}_{1k} + S_k'(I+P)(X_1 \otimes \hat{U}_1')\hat{\delta}, \ldots, \hat{e}_{Tk} + S_k'(I+P)(X_T \otimes \hat{U}_T')\hat{\delta})'$$

be the Tx1 vector of observations on the left-hand side variable of equation (4.2) and let

$$X_{ak} = [(X_1' \otimes \hat{U}_1)(I+P)S_k, \ldots, (X_T' \otimes \hat{U}_T)(I+P)S_k]'$$

be the Txq matrix of right-hand side variables of equation (4.2). We can then write the observations for the linearized equation (4.2) as

$$(4.3) \qquad y_{ak} = X_{ak}\delta + r_k, \quad (k=1,\ldots,L),$$

where $r_k = (r_{k1},\ldots,r_{kT})'$. An estimator of δ which accounts for the presence of covariance restrictions can now be obtained by joint 3SLS estimation of the L equations (4.3) and the original equation system (3.5), using a vector of ones as the instrumental variable for equation (4.3), which estimator we will refer to as augmented 3SLS (A3SLS).

To obtain the A3SLS estimator it is convenient to stack the additional equations (4.3) which follow from the covariance restrictions with the observations for the original equation system to form an augmented equation system. Let $y_A = (y', y_{a1}', \ldots, y_{aL}')'$ be the (M+L)T dimensional vector of observations on the left-hand side variables and $X_A = [X', X_{a1}', \ldots, X_{aL}']'$ the (M+L)Txq matrix of observations on the right-hand side variables of both equations (3.5) and (4.3). Then we can write the augmented equation system which adds the L equations (4.3) to the original equation system (3.5) as

$$(4.4) \qquad y_A = X_A \delta + r,$$

where r is equal to $(u', r_{a1}', \ldots, r_{aL}')'$.

To form the A3SLS estimator we use Z as instrumental variables for each of the original equations of the augmented system and a Tx1 vector of ones, which we will denote by α, for each of the additional equations. The corresponding (M+L)Tx(MK+L) matrix of instrumental variables for the augmented equation system will be $Z_A = \text{diag}[I_M \otimes Z, I_L \otimes \alpha]$. We also estimate the covariance matrix of the augmented system $\Omega = E[(U_t,e_t')'(U_t,e_t')]$ using the estimator $\hat{\Omega} = (1/T)\Sigma_{t=1}^{T}(\hat{U}_t,\hat{e}_t')'(\hat{U}_t,\hat{e}_t')$ which can be formed from the residuals of the linearized system evaluated at $\delta=\hat{\delta}$. The A3SLS estimator $\hat{\delta}_A$ can then be obtained as

$$(4.5) \qquad \hat{\delta}_A = (\hat{X}_A'(\hat{\Omega}^{-1} \otimes I_T)X_A)^{-1} \hat{X}_A'(\hat{\Omega}^{-1} \otimes I_T)y_A,$$

where $\hat{X}_A = Z_A(Z_A'Z_A)^{-1} Z_A'X_A$.

The form of the A3SLS estimator differs from that of a standard 3SLS estimator in two respects. The first is that there are cross-equation restrictions in the augmented equation system. The parameters which enter the original equation system also enter the additional equations which arise from the covariance restrictions. The second way A3SLS is different is that different instruments are used for different equations, so that Z_A does not have a Kronecker product form. This second difference can be eliminated when the matrix of instrumental variables Z includes a column of ones, as will be assumed below and as is usually the case in applications, by using all the columns of Z as instrumental variables for the last L equations of the augmented equation system. The resulting estimator, say $\tilde{\delta}_A$, then has the more familiar 3SLS form

$$(4.5a) \qquad \tilde{\delta}_A = (X_A'(\hat{\Omega}^{-1} \otimes Z(Z'Z)^{-1}Z')X_A)^{-1} X_A'(\hat{\Omega}^{-1} \otimes Z(Z'Z)^{-1}Z')y_A .$$

This alternative estimator may be easier to compute, since some standard 3SLS computer programs can be used. Although $\hat{\delta}_A$ and $\tilde{\delta}_A$ will generally not be numerically equal, they will be asymptotically equivalent under our conditions, as we show below.

We have derived the A3SLS estimator by linearizing the extra equations which arise from the covariance restrictions. We could have proceeded by linearizing the FIML first order conditions (i.e. by obtaining the Rothenberg and Leenders (1964) linearized MLE) but the resulting estimator would be more complicated than the A3SLS estimator because of the presence of the parameters of the disturbance covariance matrix for the original system. The A3SLS estimator has the advantage that it has a familiar form and, at least when the entire matrix Z is used as the instrumental variables for the additional equations, can be implemented by using existing software (e.g. TSP). Furthermore, as will be shown below, the A3SLS estimator will often be efficient relative to FIML when U_t is nonnormal.

An initial consistent estimator of δ is required to form the A3SLS estimator. When the covariance restrictions are not needed for identification, then an IV estimator such as 2SLS or 3SLS will suffice. When covariance restrictions are necessary for identification, we must take a different approach. Direct use of the moment restrictions also provides a useful approach to obtaining an initial estimator of δ. Let

$$g_{1T}(\delta) = (I_M \otimes Z)'u/T, \quad g_{2T}(\delta) = S'\mathrm{vec}(U'U)/T,$$

$$g_T(\delta) = (g_{1T}(\delta)', g_{2T}(\delta)')'.$$

Note that $g_T(\delta)$ is a MK+L dimensional vector of sample moments which has expectation zero when δ is equal to the true parameter value. It is formed of the usual MK dimensional vector of products of instrumental variables and residuals plus an additional L dimensional vector formed as the sample average of the vector of residuals e_t for the additional equations which arise from the covariance restriction. A GMM estimator $\hat{\delta}_\Psi$ which utilizes the covariance restrictions can now be obtained by minimizing a quadratic form in the moment functions $g_T(\delta)$, i.e. by solving

$$\text{(4.6)} \qquad \min_D g_T(\delta)'\Psi_T g_T(\delta),$$

where D is a subset of R^q and Ψ_T is a positive semi-definite matrix. A similar estimation method has recently been suggested by Rothenberg (1983), who motivates the GMM method as a modified minimum distance approach.

The minimization problem (4.6) is quartic in the parameters of δ and may therefore not be very difficult to solve in many circumstances. To minimize the computational complexity of this estimator we have not included products of e_t with components of Z_t, other than that with $Z_{t1} \equiv 1$, in the vector of moment functions.[12] The computational complexity of $\hat{\delta}_\Psi$ can be further reduced by choosing $\Psi_T = \text{diag}[I_q, 0]$. This choice is one which minimizes the role of the vector of quadratic functions $g_{2T}(\delta)$ in the minimization problem (4.6) and thus may simplify the computation of $\hat{\delta}_\Psi$. For this choice of Ψ_T the minimization problem (4.6) is using just enough covariance restrictions to give just-identification.

Some insight concerning the nature of the A3SLS estimator can be obtained by considering whether A3SLS, like FIML, has an interpretation as an instrumental variables estimator for the original equation system, where residuals are used as instruments in addition to predetermined variables. Residuals can be added to the list of instrumental variables by allowing $\hat{W} = [\mathcal{Z}, (I_M \otimes \hat{U})S]$ to be the T x (MK+L) matrix of instrumental variables for the original equation system. A3SLS then has an instrumental variables interpretation if and only if there is a (MK+L) x q linear combination matrix H^* such that $\hat{\delta}_A$ satisfies

$$(4.7) \qquad \hat{\delta}_A = (H^{*\prime}\hat{W}'X)^{-1}H^{*\prime}\hat{W}'y.$$

Such an H^* will generally not exist if $\hat{\Omega}_{12} \neq 0$, where we partition $\hat{\Omega}$ and Ω

12. When $E[e_t U_t] \neq 0$, these other cross-products of e_t with components of Z_t would need to be included in the moment vector if the optimal GMM estimator (see Hansen (1982)) is to be as efficient as A3SLS. Although $Z_{t1} \equiv 1$ is the only instrumental variable used for each of the additional equations, the linear combination of the instrumental variables of all the equations which multiplies each component of e_t in the A3SLS estimator may include all of the components of Z_t when $E(e_t U_t) \neq 0$.

comformably with (U_t, e_t'). When $\hat{\Omega}_{12} \neq 0$, the transformed instrumental variables $(\hat{\Omega}^{-1} \otimes I_T)Z_A$ will lead to all the columns of Z, and not just α, being used as instruments for the additional equations. When all the columns of Z are used as instruments for the additional equations $\hat{\delta}_A$ is implicitly using the information that all the predetermined variables are uncorrelated with products of disturbances that correspond to covariance restrictions, and not only the information that such disturbance products have expectation zero. If $\hat{\Omega}_{12}$ is constrained to be equal to zero, which would be the actual value of Ω_{12} if ε_t had a distribution that is symmetric around zero (e.g. normal), then an instrumental variables interpretation of $\hat{\delta}_A$ can be obtained.

To obtain an instrumental variables interpretation of A3SLS we need to be specific about what estimator of δ is used to form the residual matrix $\hat{U}$. We will assume that the initial estimator $\hat{\delta}$ satisfies

$$(4.8) \qquad \tilde{H}' g_T(\hat{\delta}) = 0 \quad .$$

for some (MK+L)xq linear combination matrix $\tilde{H}$. For example if $\tilde{H}'$ is equal to $[X'\tilde{Z}(I_M \otimes (Z'Z)^{-1}), 0]$, then equation (4.8) gives the normal equations for the vector of 2SLS estimators. More generally, equation (4.8) can be thought of as the first order condition for the GMM estimator of equation (4.6), with $\tilde{H} = \Psi_T \partial g_T(\hat{\delta})/\partial\delta$.

Theorem 4.1: If $\hat{\delta}$ satisifies equation (4.8) and $\tilde{H}'W'X$ is nonsingular, and $\hat{\Omega}_{12} = 0$, then there exists a (MK+L)xq linear combination matrix H* such that equation (4.7) is satisfied. The matrix H^* is given in the appendix.[13]

To obtain the asymptotic properties of the A3SLS estimator it is convenient to assume that certain regularity conditions are satisfied. Let $|x| = \max_i |x_i|$ for $x = (x_1, \ldots, x_n)$.

Assumption 4.1: The observations (U_t, Z_t) are independently not identically distributed such that there exists γ, $M > 0$ such that

$$E[|U_t|^{4+\gamma}] < M, \quad E[|Z_t|^{4+\gamma}] < M, \quad (t=1,2,\ldots)$$

This assumption could be relaxed to allow dependent observations as long as the disturbance vectors for different observations are independent.

Assumption 4.2: For all t, U_t has constant conditional raw moments up to the fourth order which do not depend on t. Also, $Z_{t1} \equiv 1$, $E[U_t] = 0$, $E[e_t] = 0$, $\text{plim}\ Z'Z/T = Q$, and the probability limits of averages of all products up to the fourth order of elements of Z_t and U_t exist.

Besides specifying that the instrumental variables include a vector of ones, this assumption rules out heteroskedasticity in either the original equation system or the additional equations which arise from the covariance

13. It can be shown that H^* is the optimal choice of a linear combination matrix to use when using residuals as instrumental variables for the original equation system when $\Omega_{12} = 0$. The formula in equation (4.7) was analyzed in an earlier version of this paper.

restrictions, as well as specifying that the disturbance vector U_t has mean zero and that the covariance restrictions are satisfied. Let

$$G_2 = S'(I+P)(I_M \otimes \Sigma)\tilde{B}',$$

$$G = [\tilde{D}, G_2']'.$$

Assumption 4.3: The matrices Ω and Q are nonsingular, and the matrix G has rank q.

As will be discussed in Section 5, rank(G) = q is a condition for local identification of δ under covariance restrictions. Let S_A be the (M+L)K x (MK+L) selection matrix such that $Z_A = (I_{M+L} \otimes Z)S_A$, and let

$$V^* = (G'S_A'(\Omega^{-1} \otimes Q)S_A G)^{-1}.$$

The following result gives the asymptotic distribution of the A3SLS estimator.

Theorem 4.2: If Assumptions 4.1-4.3 are satisfied and the initial estimator $\hat{\delta}$ is such that $\sqrt{T}(\hat{\delta}-\delta)$ is bounded in probability then

$$\sqrt{T}(\hat{\delta}_A - \delta) \xrightarrow{d} N(0,V^*)$$

and $\text{plim}[\sqrt{T}(\hat{\delta}_A - \tilde{\delta}_A)] = 0$. Furthermore, $\text{plim}[T(\hat{X}_A'(\hat{\Omega}^{-1} \otimes I_T)\hat{X}_A)^{-1}] =$ $\text{plim}[T(X_A'(\tilde{\Omega}^{-1} \otimes Z(Z'Z)^{-1}Z')X_A)^{-1}] = V^*$.

This result says that V^* is the asymptotic covariance matrix of $\hat{\delta}_A$, that the other version $\tilde{\delta}_A$ of A3SLS is asymptotically equivalent to $\hat{\delta}_A$, and that the usual estimator for the asymptotic covariance matrix of each of these estimators is consistent. The hypothesis that $\sqrt{T}(\hat{\delta}-\delta)$ is bounded in probability will be satisfied if $\hat{\delta}$ is asymptotically normal. If the covariance restrictions are not needed for identification then both the 2SLS and 3SLS estimators will satisfy this hypothesis; while if the covariance restrictions are necessary for identification then the GMM estimator discussed above can be used as an initial estimator when forming A3SLS. The verification that the GMM estimator is asymptotically normal when δ is identified and Assumptions 4.1-4.3 are satisfied is routine, so we omit this verification.

The important question concerning the asymptotic properties of the A3SLS estimator is its asymptotic efficiency. The following result says that the A3SLS estimator is asymptotically efficient when the disturbance vector U_t is normally distributed.

Theorem 4.3: If Assumptions 4.1 - 4.3 are satisfied and the distribution of U_t conditional on Z_t is normal then $\sqrt{T}(\hat{\delta}_A - \hat{\delta}_{FIML})$ converges in probability to zero.

This result has a straightforward explanation. From the instrumental variables interpretation of FIML given in Section 3 we can see that FIML uses the fact that Z is orthogonal to the disturbances and that ε_i and ε_j are orthogonal for $\sigma_{ij} = 0$. This information is exactly that which is used to form the augmented equation system and the A3SLS estimator. Furthermore, as stated in the following result, the A3SLS estimator is asymptotically equivalent to the best nonlinear 3SLS (BNL3SLS) estimator for the augmented equation system and is therefore efficient in the class of nonlinear instrumental variable estimators for the augmented equation system. Let $\hat{\delta}_B$ be a BNL3SLS estimator (Amemiya (1977)) for the augmented equation system consisting of equations (3.5) and (4.1).[14]

Theorem 4.4: If Assumptions 4.1 - 4.3 are satisfied then $\sqrt{T}(\hat{\delta}_A - \hat{\delta}_B)$ converges in probability to zero.

The asymptotic efficiency of the A3SLS can now be seen to result from the fact that both FIML and A3SLS use the same information, i.e. are members of the same class of estimators, and A3SLS is asymptotically efficient in this class.

Theorem 4.4 also sheds some light on the comparison of FIML and A3SLS estimators when the disturbances of the original equation system do not have a multivariate normal distribution. The efficiency of A3SLS in the class of instrumental variable estimators for the augmented equation system does not depend in any way on normality because we have not imposed any particular form for the covariance matrix of the augmented equation system. If the disturbance vector of the original equation system is normally distributed then third raw moments of the disturbances are zero and fourth raw moments consist of products of second raw moments, and the covariance matrix of the

14. A BNL3SLS estimator is derived in the appendix.

disturbance vector of the augmented equation system is given by[15]

$$\Omega_N = \begin{bmatrix} \Sigma & 0 \\ 0 & S'[(\Sigma \otimes \Sigma)(I+P)]S \end{bmatrix}.$$

FIML imposes this special form for Ω and as a result may be asymptotically inefficient relative to A3SLS if U_t is nonnormal. FIML also imposes the special form of Ω when forming standard error estimates from the inverse information matrix formula so that the usual standard error formulae for $\hat{\delta}_{FIML}$ may be wrong if U_t is nonnormal.[16]

To conclude our discussion of estimation, it is useful to note that the results of this section do not apply only to the case of zero covariance restrictions. All of the above results, including asymptotic efficiency of A3SLS apply without modification to the case of linear homogenous restrictions $S'\text{vec}(\Sigma) = 0$, where S need not be a selection matrix.

---

15. See Henderson and Searle (1979) for the expression for Ω_{22}.

16. Similar points about the consequences of nonnormality for the properties of FIML when covariance restrictions are present have been made by others in the context of panel data, e.g., Chamberlain (1982).

5. Identification

It is well known that covariance restrictions can help to identify the parameters of a simultaneous equations system (see the references in Hsiao (1983)). Hausman and Taylor(1983) have recently provided necessary and sufficient conditions for identification of a single equation of a simultaneous system using covariance restrictions, and have suggested a possible interpretation of identification of a simultaneous system which is stated in terms of an assignment of residuals as instruments. In this section we show that such an assignment condition provides useful conditions for identification.

The asymptotic covariance matrix V^* of A3SLS is well defined only if $\mathrm{rank}(G) = g$, so that this condition is a natural one to consider when analyzing the identification of δ from the conditional moment restrictions $E(U_t|Z_t) = 0$, $E(S'\mathrm{vec}(U_t'U_t)|Z_t) = 0$. We can relate this condition to familiar identification conditions for the structural parameters of a simultaneous equations system. As discussed by Rothenberg (1971) nonsingularity of the information matrix is a necessary and sufficient condition for first order local identification at any regular point of the information matrix.[17] Nonsingularity of the information matrix for the normal disturbance case is equivalent to the matrix G having full column rank, as stated in the following result.

Lemma 5.1: If Q and Σ are nonsingular then the information matrix J is nonsingular if and only if $\mathrm{rank}(G) = q$.

---

17. A regular point of the information matrix is one where the information matrix has constant rank in a neighborhood of the point. The set of such points has measure one, Rothenberg (1971). Also, see Sargan (1983) for a discussion of the relationship between identification and first order identification.

Thus the rank of G plays a crucial role in determining the identification of the structural parameters of a simultaneous equations system.

The matrix G has an interesting structure. The first MK rows form a matrix $G_1' = \tilde{D}'$ which is familiar from the analysis of identification via coefficient restrictions. The covariance restrictions play a role in determining the rank of G through the matrix of the last L rows $G_2 = S'(I+P)(I_M \otimes \Sigma)\tilde{B}' = S'(I+P)\text{plim}(I_M \otimes U)'X/T$. Let Σ_j denote the jth row of Σ. The kth row of G_2, corresponding to the covariance restriction $\sigma_{ij} = 0$, has zero for each element except for the elements corresponding to δ_i, where $\Sigma_j \tilde{B}_i = \text{plim}(u_j'X_i/T)$ appears, and the elements corresponding to δ_j, where $\Sigma_i \tilde{B}_j = \text{plim}(u_i'X_j/T)$ appears. Thus the kth row of G_2 contains both the covariance of u_j with the right-hand side variables for equation i and the covariance of u_i with the right-hand side variables for equation j. We can exploit this structure to obtain necessary conditions for identification which are stated in terms of using residuals as instruments.

An assignment of residuals as instruments is a choice for each covariance restriction $\sigma_{ij} = 0$ to <u>either</u> assign u_i as an instrument for equation j <u>or</u> u_j as an instrument for equation i, but <u>not</u> both. Since for each covariance restriction there are two distinct ways of assigning a residual as an instrument there are 2^L possible distinct assignments. For each assignment of residuals as instruments, which we will index by $p = 1,\ldots,2^L$, let U_{pi} be the (possibly nonexistent) matrix of observations on the disturbances assigned to equation i. Let $W_{pi} = [Z, U_{pi}]$ be the resulting matrix of instrumental variables for equation i and $C_{pi} = \text{plim}(W_{pi}'X_i/T)$ the population cross-product matrix of instrumental variables and right-hand side variables for equation i.

Theorem 5.2 If rank(G) = q then there exists an assignment p such that

$$\text{rank}(C_{pi}) = q_i, \quad (i=1,\ldots,M). \tag{5.1}$$

This result means that for a regular point a necessary condition for first order identification is that there is an assignment of residuals as instruments such that population cross-product matrix of instrumental variables and right hand side variables has full column rank for each equation. Note that if $\text{rank}(C_{pi}) = q_i$ there must be at least q_i instrumental variables for the ith equation. We can thus obtain an order condition for residuals assigned as instruments. Let $a_i = \max(0, q_i-K)$ be the number of instrumental variables which are required for the ith equation in addition to Z to have the same number of instrumental variables as right-hand side variables.

Corollary 5.3: If rank(G) = q then there exists an assignment p such that at least a_i residuals are assigned as instruments to equation i, i = 1,...,M.

This order condition says that there must exist an assignment of residuals as instruments so that there are enough instruments to estimate each equation. We can use Hall's Theorem on the existence of a system of distinct representatives, in a similiar fashion to the use of this theorem by Geraci (1976), to obtain an algorithm for determining whether or not such an assignment of residuals as instruments exists. Let R_i be the set of indices k of distinct covariance restrictions such that the kth covariance restriction is $\sigma_{ij}=0$ for some $j \neq i$. Note that R_i is just the set of indices of distinct covariance restrictions which involve equation i.

Theorem 5.4: There exists an assignment of residuals as instruments such that for each i=1,...,M at least a_i residuals are assigned as instruments to equation i if and only if for each subset J of $\{1,\ldots,M\}$, $\bigcup_{i \in J} R_i$ contains at least $\sum_{i \in J} a_i$ elements.

So far, each of the identification results of this section have been stated in terms of the number and variety of instruments for each equation; see Koopmans et. al. (1950). It is well known (see Fisher (1966) pp. 52-56) that when only coefficient restrictions are present the condition that $\text{plim}(Z'X_i/T)$ have rank q_i can be translated into a more transparent condition on the structural parameters $A = [B', \Gamma']'$. When covariance restrictions are present we can also state a rank condition which is equivalent to $\text{plim}(W_{pi}'X_i/T)$ having rank q_i. For an assignment p, let Σ_{pi} be the rows of Σ corresponding to residuals which are assigned as instrumental variables to the ith equation. Let ϕ_i be the $(M-1-q_i)x(M+K)$ selection matrix such that the exclusion restrictions on the ith equation can be written as $\phi_i A_i = 0$, where A_i is the ith column of A.

Lemma 5.5: For a particular assignment p and an equation i, the rank of C_{pi} equals q_i if and only if

(5.2) $\text{rank}[A'\phi_i', \Sigma_{pi}'] = M-1$.

When this result is combined with Theorem 5.2 we can see that Theorem 5.2 is a stronger necessary condition than Fisher's (1966, Theorem 4.6.2) Generalized Rank Condition, which says that a necessary condition for identification of the ith equation is that the rank of $[A'\phi_i', \tilde{\Sigma}_i']$ is M-1, where $\tilde{\Sigma}_i$ is all the rows Σ_j of Σ such that $\sigma_{ij}=0$. Theorem 5.2 strengthens this necessary condition by requiring that the rank condition only hold for those rows of Σ corresponding to residuals which are assigned to equation i.

So far we have only presented necessary conditions for identification. We can also give a sufficient condition for local identification which includes the recursive case of Proposition 9 of Hausman and Taylor (1983).

Theorem 5.6: If for a subset of covariance restrictions there is exactly one assignment p of residuals as instruments such that $\text{rank}(C_{pi}) = q_i$, $(i=1,\ldots,M)$, then $\text{rank}(G) = q$.

We do not know whether the existence of an assignment p such that the rank condition (5.1) is satisfied is sufficient for local identification when there is more than one such assignment condition. We have not been able to prove that the assignment condition is sufficient in general or to construct a counter example.

The previous results have the virtue that they can be checked on an equation by equation basis. It is possible to give a necessary and sufficient condition for local identification in terms of the structural parameters, although this result involves the restrictions on all the equations and is not readily interpretable in terms of instrumental variables.

Theorem 5.7: The matrix G has full column rank if and only if

$$(5.3) \qquad \mathrm{rank}(\mathrm{diag}[\phi_1,\ldots,\phi_M,S'].[I_M \otimes A',(I_M \otimes \Sigma)(I+P)]') = M(M-1).$$

This result is a special case of the necessary and sufficient conditions for local identification given by Rothenberg (1971) and Wegge (1965), since G having full column rank is equivalent to the information matrix being nonsingular.

The identification results of this section are local in nature. The question of global identification with general zero covariance restrictions is more difficult because the moment functions $g_T(\delta)$ are nonlinear (quadratic) in the parameters. In fact Bekker and Pollock (1984) have recently given an example of a system of simultaneous equations subject to exclusion and zero covariance restrictions which is locally but not globally identified. Thus the problem of global identification remains somewhat problematical in the general exclusion and zero covariance restriction case without further restrictions on the parameter space.

MATHEMATICAL APPENDIX

Some properties of the permutation matrix P are useful for deriving the information matrix. From Henderson and Searle (1979) we know that P is symmetric, $P^{-1} = P$, and for any M-dimensional square matrices A and B $P(A \otimes B)=(B \otimes A)P$. Let $R^{-'}$ be the $M^2 \times (M\cdot(M+1)/2)$ matrix obtained from R' by replacing the rows of R' corresponding to σ_{ij}, $i \neq j$, by 1/2 times the corresponding row of R' and by leaving the rows of R' corresponding to σ_{ii} unchanged (see Richard (1975)). Then $PR' = R'$, $PR^{-'} = R^{-'}$ $2R'R^{-} = I + P$, $(R(\Sigma^{-1} \otimes \Sigma^{-1})R')^{-1} = R^{-}(\Sigma \otimes \Sigma)R^{-'}$; and $R^{-}R' = I$.

Proof of Theorem 3.1: Let $\tilde{S}$ be the matrix defined immediately preceding Theorem 3.2 in the text. Let $S^{*'}$ be a $L \times M(M+1)/2$ selection matrix such that the covariance restrictions are given by $0 = S^{*'}\sigma^{*} = S^{*'}R^{-}\text{vec}\Sigma$. Let $A' = \tilde{S}'R\Sigma^{-1} \otimes \Sigma^{-1}$ and $B = \Sigma \otimes \Sigma R^{-'}S^{*}$. Note that A' has rank $M(M+1)/2-L$ and B has rank L. Also by $\tilde{S}'S^{*} = 0$, $A'B = \tilde{S}'RR^{-'}S^{*} = 0$. Further, $I-P$ has rank $M(M-1)/2$ and $(I-P)A = \Sigma^{-1} \otimes \Sigma^{-1}(I-P)R'\tilde{S} = 0$. Similarly $(I-P)B = 0$. It follows that $C = [I-P,A]'$ has rank M^2-L, and that the columns of B are a basis for the nullspace of C. Let $y = \text{vec}(T\Sigma-U'U)$. By symmetry $(I-P)y=0$ while $A'y = 0$ follows by equation (3.9). It follows that $y = Bx$ for some $L \times 1$ vector x. By the definition of S^{*}, $S^{*'}R^{-}y = -S^{*'}R'\text{vec}(U'U)$. Substituting for y and solving gives $x = -(S^{*'}R^{-}\Sigma \otimes \Sigma R^{-'}S^{*})^{-1}S^{*'}R^{-}\text{vec}(U'U)$. Solving for y, and noting that $S'(I+P) = 2S^{*'}R^{-}$ for any S as discussed above, we obtain

$$\text{(A.1)} \quad \text{vec}(T\Sigma - U'U) = -\Sigma \otimes \Sigma(I+P)S[S'\Sigma \otimes \Sigma(I+P)S]^{-1}S'\text{vec}(U'U).$$

$$\text{(A.2)} \quad \text{vec}[(B')^{-1}(T\Sigma - U'U)\Sigma^{-1}]$$

$$= [I_M \otimes (B')^{-1}][\Sigma^{-1} \otimes I_M]\text{vec}(T\Sigma - U'U)$$

$$= [I_M \otimes (B')^{-1}][I_M \otimes \Sigma(I+P)S[S'\Sigma \otimes \Sigma(I+P)S]^{-1} S'\text{vec}(U'U).$$

Selecting from equations (3.9) and (3.8) according to the unrestricted elements of $[B, \Gamma]$ as in Hausman (1975), and using $\text{vec}(U'U) = (I \otimes U)'u$, gives

$$\text{(A.3)} \quad H'W'u = 0.$$

Proof of Theorem 3.2: An expression for the information matrix which ignores symmetry of Σ and the covariance restrictions on Σ is given in Hausman (1983), where the notation is identical to that used here except that P is there denoted by E. Equation (3.12) then follows by $\text{vec}(\Sigma) = R'\tilde{S}\tilde{\sigma}$ and the chain rule, using $R(I_M \otimes \Sigma)P(I_M \otimes \Sigma)R' = R(\Sigma \otimes \Sigma)PR' = R(\Sigma \otimes \Sigma)R'$ and $(I+P)R' = 2R'$. To obtain the expression for the Cramer-Rao bound note that

$$\text{(A.4)} \quad \tilde{B}(\Sigma^{-1} \otimes I_M)R'(\tfrac{1}{2}F)^{-1}R(\Sigma^{-1} \otimes I_M)\tilde{B}$$

$$= \tilde{B}(\Sigma^{-1} \otimes I_M)R'R^{-}(2\Sigma \otimes \Sigma)R^{-\prime}R(\Sigma^{-1} \otimes I_M)\tilde{B}'$$

$$= \frac{1}{2}\tilde{B}(\Sigma^{-1} \otimes I_M)(I+P)(\Sigma \otimes \Sigma)(I+P)(\Sigma^{-1} \otimes I_M)\tilde{B}'$$

$$= \tilde{B}P\tilde{B}' + \tilde{B}(\Sigma^{-1} \otimes \Sigma)\tilde{B}'.$$

We can also compute

$$\text{(A.5)} \qquad \text{plim}X'(\Sigma^{-1} \otimes I_T)X/T = \tilde{D}(\Sigma^{-1} \otimes Q)\tilde{D}' + \tilde{B}(\Sigma^{-1} \otimes \Sigma)\tilde{B}'.$$

Using equation (A.5) to obtain the upper-left block of the information matrix and adding and subtracting equation (A.4) to the partitional inverse formula yields equation (3.13).

Proof of Corollary 3.3: When the covariance matrix is restricted to be diagonal we compute $R'\tilde{S}' = \text{diag}(e_1,\ldots,e_M) = P^*$ where e_i is the ith unit vector. Also, $\Sigma^{-1} \otimes \Sigma^{-1}$ is a diagonal matrix, with element $(\sigma_{ii})^{-2}$ in the ith position of the ith block, so that the lower right block of the information matrix is as given in equation (3.14). To obtain the upper right block, note that $\Sigma^{-1} \otimes I_M = \text{diag}\,[(\sigma_{11})^{-1}I_M,\ldots,(\sigma_{MM})^{-1}I_M]$, so that $(\Sigma^{-1} \otimes I_M)P^* = \text{diag}\,[(\sigma_{11})^{-1}e_1,\ldots,(\sigma_{MM})^{-1}e_M] = P^*\Sigma^{-1}$. The form of the Cramer-Rao bound now follows from the partitioned inverse formula.

Proof of Theorem 4.1: Note that $\hat{\Omega}_{12} = 0$ implies

$$\hat{X}_A'(\hat{\Omega}^{-1} \otimes I_T) = X_A' \,\text{diag}\,[\hat{\Sigma}^{-1} \otimes Z(Z'Z)^{-1}Z',\ \hat{\Omega}_{22}^{-1} \otimes \alpha(\alpha'\alpha)^{-1}\alpha']$$

$$= X_A'Z_A\Psi_T^*Z_A'$$

where $\Psi_T^* = \text{diag}\,[\hat{\Sigma}^{-1} \otimes (Z'Z)^{-1},\ \hat{\Omega}_{22}^{-1}(\alpha'\alpha)^{-1}]$. Also, $\sum_{t=1}^{T} X_t \otimes \hat{U}_t' = (I_M \otimes \hat{U})'X$

so that for $\tilde{Z} = I_M \otimes Z$

$$\text{(A.6)} \qquad Z_A{}'X_A = [X'\tilde{Z},\ X'(I_M \otimes \hat{U})(I+P)S\,]'$$

$$= (X'\hat{W} + [0, X'(I_M \otimes \hat{U})PS\,])'$$

$$= \hat{W}'X + \begin{bmatrix} 0 \\ S'P(I_M \otimes \hat{U})'X(\tilde{H}'\hat{W}'X)^{-1}\tilde{H}'\hat{W}'X \end{bmatrix} = \tilde{A}\hat{W}'X$$

where

$$\tilde{A} = I_{MK+L} + \begin{bmatrix} 0 \\ S'P(I_M \otimes \hat{U})'X(\tilde{H}'\hat{W}'X)^{-1}\ \tilde{H}' \end{bmatrix}$$

Also note that $\sum_{t=1}^{T} U_t' \otimes U_t' = \text{vec}(U'U) = (I_M \otimes U)'u$, so that $g_T(\hat{\delta}) = \hat{W}'(y-X\hat{\delta})$, and equation (4.8) implies $\hat{\delta} = (\tilde{H}'\hat{W}'X)^{-1}\ \tilde{H}'\hat{W}'y$. It follows that

$$\text{(A.7)} \qquad Z_A'y_A = (y'\tilde{Z},\ [(y-X\hat{\delta})'(I_M \otimes \hat{U}) + \hat{\delta}'X'(I_M \otimes \hat{U})(I+P)\,]S\,)'$$

$$= (y'\tilde{Z},\ y'(I_M \otimes \hat{U})S + y'\hat{W}\tilde{H}(X'\hat{W}\tilde{H})^{-1}X'(I_M \otimes \hat{U})PS)'$$

$$= \hat{W}'y + \begin{bmatrix} 0 \\ S'P(I_M \otimes \hat{U})'X(\tilde{H}'\hat{W}'X)^{-1}\tilde{H}'\hat{W}'y \end{bmatrix} = \tilde{A}\hat{W}'y$$

The conclusion then follows with

$$\text{(A.8)} \qquad H^{*\prime} = X_A{}'Z_A \Psi_T^{*}\tilde{A}.$$

Proof of Theorem 4.2: By $\sqrt{T}(\hat{\delta}-\delta)$ bounded in probability we have $\operatorname{plim}\hat{\delta}=\delta$. Then Markov's weak law of large numbers and the uniform boundedness of $4+\gamma$ moments of the data imply $\operatorname{plim}[\sum_{t=1}^{T}\hat{e}_t\hat{e}_t'/T - \sum_{t-1}^{T} e_te_t'/T] = 0$.

Markov's weak law of large numbers and the uniform boundedness of fourth order moments of the disturbance also imply $\operatorname{plim}(\sum_{t=1}^{T} e_te_t'/T) = E(e_te_t') =$ Ω_{22}. It follows that $\operatorname{plim}\hat{\Omega}_{22} = \operatorname{plim}(\sum_{t=1}^{T} e_te_t'/T) = \Omega_{22}$, and similarly that $\operatorname{plim}\hat{\Omega} = \Omega$.

In an analogous fashion it follows that for $k = 1,\ldots,L$,

$$\text{(A.9)} \qquad \operatorname{plim}(Z'X_{ak}/T) = \operatorname{plim}[\sum_{t=1}^{T} Z_t'S_k'(I+P)(X_t\otimes \hat{U}_t')/T]$$

$$= \lim \sum_{t=1}^{T} E[Z_t'S_k'(I+P)E(X_t \otimes U_t'|Z_t)]/T$$

$$= [\lim \sum_{t=1}^{T} E(Z_t')/T]S_k'(I+P)(I_M \otimes \Sigma)\tilde{B}'$$

$$= Qe_1G_{2k}$$

where e_1 is the first K-dimensional unit vector and G_{2k} is the kth row of G_2. Also note that $\operatorname{plim} \tilde{Z}'X/T = (I_M \otimes Q)\tilde{D}'$. Defining $\tilde{Z}_A = I_{M+L} \otimes Z$, it follows that

$$(A.10)\qquad \text{plim}(\tilde{Z}_A'X_A/T) = \tilde{Q}_A S_A G,$$

where $\tilde{Q}_A = I_{M+L} \otimes Q$. Then by $\text{plim}(\tilde{Z}_A'\tilde{Z}_A/T) = \tilde{Q}_A$ and $Z_A = \tilde{Z}_A S_A$, we have

$$(A.11)\qquad \text{plim}(Z_A'X_A/T) = S_A'\tilde{Q}_A S_A G,\quad \text{plim}(Z_A'Z_A/T) = S_A'\tilde{Q}_A S_A,$$

and

$$(A.12)\qquad \text{plim}[Z_A'(\hat{\Omega}^{-1} \otimes I_T)X_A/T] = \text{plim}[S_A'(I_{M+L} \otimes Z)'(\hat{\Omega}^{-1} \otimes I_T)X_A/T]$$

$$= \text{plim}[S_A'(\hat{\Omega}^{-1} \otimes I_K)\tilde{Z}_A'X_A/T] = S_A'(\Omega^{-1} \otimes Q)S_A G.$$

It follows from (A.11) and (A.12) that

$$(A.13)\quad \text{plim}[T(\hat{X}_A'(\hat{\Omega}^{-1} \otimes I_T)X_A)^{-1}] = \text{plim}(X_A'Z_A(Z_A'Z_A)^{-1}Z_A'(\hat{\Omega}^{-1} \otimes I_T)X_A/T)^{-1}$$

$$= (G'(S_A'\tilde{Q}_A S_A)(S_A'\tilde{Q}S_A)^{-1}S_A'(\Omega^{-1} \otimes Q)S_A G)^{-1} = V^*$$

and that

$$(A.14)\qquad \text{plim}[X_A'Z_A(Z_A'Z_A)^{-1}] = G'(S_A'\tilde{Q}_A S_A)(S_A\tilde{Q}_A S_A)^{-1} = G'.$$

Next, using $(X_t \otimes \hat{U}_t')(\hat{\delta} - \delta) = (U_t - \hat{U}_t)' \otimes \hat{U}_t' = \text{vec}[\hat{U}_t'(U_t - \hat{U}_t)]$ and $P\text{vec}(A) = \text{vec}(A')$ we obtain, for $k = 1,\ldots,L$,

$$(A.15)\qquad Z'(y_{ak}-X_{ak}\delta)/\sqrt{T} = \sum_{t=1}^{T} Z_t'(\hat{e}_{tk}+S_k'(I+P)(X_t\otimes \hat{U}_t')(\hat{\delta}-\delta))/\sqrt{T}$$

$$= [\sum_{t=1}^{T} Z_t'e_{tk}/\sqrt{T}] + \sum_{t=1}^{T} Z_t'S_k'\text{vec}[\hat{U}_t'\hat{U}_t-U_t'U_t+(U_t-\hat{U}_t)'\hat{U}_t+\hat{U}_t'(U_t-\hat{U}_t)]/\sqrt{T}$$

$$= [\sum_{t=1}^{T} Z_t'e_{tk}/\sqrt{T}] + \sum_{t=1}^{T} Z_t'S_k'\text{vec}[(U_t-\hat{U}_t)'(\hat{U}_t-U_t)]/\sqrt{T}$$

$$= [\sum_{t=1}^{T} Z_t'e_{tk}/\sqrt{T}] + [\sum_{t=1}^{T} Z_t'S_k'(X_t\otimes(\hat{U}_t-U_t))/T]\sqrt{T}(\hat{\delta}-\delta),$$

where the second term after the last equality converges in probability to zero by $\sqrt{T}(\hat{\delta}-\delta)$ bounded in probability. By the Liapunov central limit theorem the first term after the last equality is bounded in probability, as is $Z'u_i/\sqrt{T}$, for $i = 1,\dots,M$. Let $\nu = (u',e_{11},\dots,e_{T1},\dots,e_{1L},\dots,e_{TL})'$. Then by (A.13), (A.14), and (A.15) we obtain

$$(A.16)\qquad \sqrt{T}(\hat{\delta}_A-\delta) = (\hat{X}_A'(\hat{\Omega}^{-1}\otimes I_T)X_A/T)^{-1}\hat{X}_A'(\hat{\Omega}^{-1}\otimes I_T)(y_A-X_A\delta)/\sqrt{T}$$

$$= V^*X_A'Z_A(Z_A'Z_A)^{-1}S_A'(I_{M+L}\otimes Z)'(\Omega^{-1}\otimes I_T)(y_A-X_A\delta)/\sqrt{T} + o_p(1)$$

$$= V^*G'S_A'(\Omega^{-1}\otimes I_K)\tilde{Z}_A'\nu/\sqrt{T} + o_p(1),$$

where $o_p(1)$ denotes a term that converges in probability to zero. The asymptotic distribution result for $\hat{\delta}_A$ now follows from

$$(A.17)\qquad \tilde{Z}_A'\nu/\sqrt{T} \xrightarrow{d} N(0,\ \Omega\otimes Q),$$

which is implied by the Liapunov central limit theorem, and from (A.16).

To show that $\text{plim}[\sqrt{T}(\hat{\delta}_A - \tilde{\delta}_A)] = 0$, note that

$$(A.18)\quad \sqrt{T}(\tilde{\delta}_A - \delta) = (X_A'\tilde{Z}_A(\hat{\Omega}^{-1} \otimes (Z'Z)^{-1})\tilde{Z}_A'X_A/T)^{-1}X_A'\tilde{Z}_A(\hat{\Omega}^{-1} \otimes (Z'Z)^{-1})\tilde{Z}_A'(y_A - X_A\delta)/\sqrt{T}$$

$$= (G'S_A'\tilde{Q}_A(\Omega^{-1} \otimes Q^{-1})\tilde{Q}_A S_A G)^{-1}G'S_A'\tilde{Q}_A(\Omega^{-1} \otimes Q^{-1})\tilde{Z}_A'\nu/\sqrt{T} + o_p(1)$$

$$= V^*G'S_A'(\Omega^{-1} \otimes I_K)\tilde{Z}_A'\nu/\sqrt{T} + o_p(1),$$

where the second equality uses (A.10). The conclusion follows by subtracting equation (A.18) from (A.16).

The remainder of the conclusions of Theorem (4.2) follow from (A.10) and (A.11), as in the proof of (A.12). We omit details.

Proof of Theorem 4.3: From the instrumental variables interpretation of $\hat{\delta}_{FIML}$ given in equation (3.11) and the fact that $W'u = g_T(\delta)$ we see that $\hat{\delta}_{FIML}$ satisfies the normal equation

$$(A.19)\qquad 0 = H'W'(y - X\delta) = H'g_T(\delta)$$

where H' is defined in Theorem 3.1 and $\text{plim}\, H' = \bar{H}' = [\bar{D}(\Sigma^{-1} \otimes I_K),$ $G_2'[S'(\Sigma \otimes \Sigma)(I+P)S]^{-1}]$. Note that

$$\text{(A.20)} \quad \text{plim}\ \partial g_T(\hat{\delta}_{FIML})/\partial\delta = \text{plim}[-X'(I_M \otimes Z)/T,\ -X'(I_M \otimes U)(I+P)S/T]'$$

$$= -[\bar{D}(I_M \otimes Q),\ \bar{B}(I_M \otimes \Sigma)(I+P)S\ ]'$$

$$= S_A'\tilde{Q}_A S_A G$$

where the last equality uses the fact that $S_A'\tilde{Q}_A S_A = \text{diag}[I_M \otimes Q,\ I_L]$. Then, using the usual mean value expansion argument, (A.19) gives

$$\text{(A.21)} \quad \sqrt{T}(\hat{\delta}_{FIML} - \delta) = (H'\partial g_T(\hat{\delta}_{FIML})/\partial\delta)^{-1}\ H'\sqrt{T}g_T(\delta) + o_p(1)$$

$$= (\bar{H}'S_A'Q_A S_A G)^{-1}\ \bar{H}'Z_A'\nu/\sqrt{T} + o_p(1)$$

$$= (G'S_A'(\Omega_N^{-1} \otimes I_K)S_A S_A'\tilde{Q}_A S_A G)^{-1}\ G'S_A'(\Omega_N^{-1} \otimes I_K)S_A S_A'\tilde{Z}_A'\nu/\sqrt{T} + o_p(1)$$

$$= (G'S_A'(\Omega_N^{-1} \otimes Q)S_A G)^{-1} G'S_A'(\Omega_N^{-1} \otimes I_K)\tilde{Z}_A'\nu/\sqrt{T} + o_p(1),$$

where we define $\Omega_N \equiv \text{diag}[\Sigma,\ S'[(\Sigma \otimes \Sigma)(I+P)]S\ ]$ and we have used the facts $\bar{H}' = G'S_A'(\Omega_N^{-1} \otimes I_K)S_A$ and $S_A'(\Omega_N^{-1} \otimes I_K)S_A S_A' = S_A'(\Omega_N^{-1} \otimes I_K)$ to obtain the third and fourth equalities. Now, when U_t has a multivariate normal distribution, $\Omega = \Omega_N$ (see Henderson and Searle (1979)). The conclusion then follows by subtracting equation (A.21) from (A.16).

Proof of Theorem 4.4: Let $G_1 = \bar{D}'$ so that $G = [G_1', G_2']'$ with

$G_2 = S'(I+P)(I_M \otimes \Sigma)\hat{B}'$. Note that

$$\text{(A.22)}\quad E[\partial U_t'/\partial\delta | Z_t] = -E(X_t | Z_t) = -(I_M \otimes Z_t)\hat{D}' = -(I_M \otimes Z_t)G_1$$

and

$$\text{(A.23)}\quad E[\partial e_t/\partial\delta | Z_t] = -E[S'(I+P)(X_t \otimes U_t') | Z_t]$$

$$= -G_2 = -(I_L \otimes Z_t)(I_L \otimes e_1)G_2,$$

where e_1 denotes the first K dimensional unit vector. Equation (A.22) gives the form of the optimal instruments for the original equations and (A.23) gives the form of the optimal instruments for the additional equations implied by the covariance restrictions; see Amemiya (1977). Stacking these two sets of instruments together it follows that $\mathcal{Z}_A S_A G$ is the matrix of optimal instrumental variables for the augmented equations system. Let $e_{tk}(\delta) = S_k'[(y_t - X_t\delta) \otimes (y_t - X_t\delta)]$ and let $\nu(\delta) = ((y-X\delta)', e_{11}(\delta), \ldots, e_{T1}(\delta), \ldots, e_{1L}(\delta), \ldots, e_{TL}(\delta))'$ be the vector of residuals for the augmented equation system. A BNL3SLS estimator $\hat{\delta}_B$ can be obtained by solving

$$\text{(A.24)}\qquad 0 = G'S_A'\mathcal{Z}_A'(\Omega^{-1} \otimes I_T)\nu(\hat{\delta}_B) = G'S_A'(\Omega^{-1} \otimes I_K)\mathcal{Z}_A'\nu(\hat{\delta}_B)$$

The calculations leading up to equation (4.2) in the text show that $\partial\nu(\delta)/\partial\delta = -X_A(\delta)$, where the δ argument indicates that the residual vector $y-X\delta$ is used in place of $y-X\hat{\delta}$ in the formation of $X_A(\delta)$. Consistency of $\hat{\delta}_B$ can be shown in the usual way, so that the usual mean value expansion and solving for $\hat{\delta}_B$ gives

$$\text{(A.25)} \qquad \sqrt{T}(\hat{\delta}_B - \delta) = (G'S_A'(\Omega^{-1} \otimes I_K)\tilde{Z}_A'X_A(\hat{\delta})/T)^{-1} G'S_A'(\Omega^{-1} \otimes I_K)\tilde{Z}_A'\nu/\sqrt{T} + o_p(1)$$

$$= V^*G'S_A'(\Omega^{-1} \otimes I_K)\tilde{Z}_A'\nu/\sqrt{T} + o_p(1),$$

where the equality follows from $\operatorname{plim}[\tilde{Z}_A'X_A(\hat{\delta})/T] = \tilde{Q}_A S_A G$ as in the proof of (A.10). The conclusion follows by subtracting (A.25) from (A.16).

Proof of Lemma 5.1: Nonsingularity of Σ and Q implies that $S_A'(\Omega_N^{-1} \otimes Q)S_A$ is nonsingular, where the terms in this expression are defined in the proofs of Theorems 4.2 and 4.3. When rank(G) = q the asymptotic equivalence of $\hat{\delta}_{FIML}$ and $\hat{\delta}_A$ implies that the information matrix is equal to $G'S'_A(\Omega_N^{-1} \otimes Q)S_A G$. It can also be shown by some matrix algebra, which is available upon request from the authors, that this equality continues to hold when rank(G) < q. The conclusion then follows by $S_A'(\Omega_N^{-1} \otimes Q)S_A$ positive definite.

Because rank(G) = q if and only if there is a nonsingular q-dimensional submatrix of G we can assume without loss of generality that G is square, which simplifies the identification proofs. The following Lemma will prove useful. Let $\tilde{G} = S_A'\tilde{Q}_A S_A G$ and, for a particular assignment, p let $\tilde{C}_p = \operatorname{diag}[C_{p1}, \ldots, C_{pM}]$.

Lemma A: For some 2^L-tuple of positive integers $(\ell_1,\ldots,\ell_{2^L})$

$$\det(\tilde{G}) = \sum_{p=1}^{2^L} (-1)^{\ell_p} \det(\tilde{C}_p).$$

Proof: Let the rows of G_2 be denoted by s_k, $k=1,\ldots,L$. Each k corresponds to a restrictrion $\sigma_{ij} = 0$ for some $i \neq j$. Further, each s_k is a sum of two 1xq vectors, $s_{ki} + s_{kj}$ where s_{ki} has $\text{plim}(u_j'X_i/T)$ for the subvector corresponding to δ_i and zeros for all other subvectors and s_{kj} has $\text{plim}(u_i'X_j/T)$ for the subvector corresponding to δ_j and zeros for all other subvectors. We can identify s_{ki} with an assignment of residual j to equation i and s_{kj} with an assignment of residual i to equation j. We have

$$\tilde{G} = \begin{bmatrix} (I_M \otimes Q)\,\tilde{D} \\ s_{1i} + s_{1j} \\ \vdots \\ s_{Li} + s_{Lj} \end{bmatrix},$$

where we drop k subscript on i and j for notational convenience. For each of the 2^L distinct assignments, indexed by p, let

$$\tilde{G}_p = \begin{bmatrix} (I_M \otimes Q)\tilde{D} \\ \tilde{s}_p \end{bmatrix},$$

where $\tilde{s}_p$ is the Lxq matrix which has its kth row s_{ki} if u_j is assigned to equation i or s_{kj} if u_i is assigned to equation j. The determinant of a matrix is a linear function of any particular row of the matrix. It follows that if L = 1

(A.26) $\det(\tilde{G}) = \det(\tilde{G}_1) + \det(\tilde{G}_2)$.

Then induction on L gives $\det(G) = \sum_{p=1}^{2^L} \det(\tilde{G}_p)$

Now consider $\tilde{G}_p$ for each p. The matrix $(I_M \otimes Q)\tilde{D}'$ is block diagonal, where the column partition corresponds to δ_i for i=1,...,M, and the ith diagonal block is plim $Z'X_i/T$. Further the kth row of $\tilde{s}_p$ consists of zeros except for the subvector corresponding to δ_i where plim$(u_j'X_i/T)$ appears. Then by interchanging pairs of rows of $\tilde{G}_p$, we can obtain $\tilde{C}_p$ from $\tilde{G}_p$. That is, $\tilde{C}_p = E_p\tilde{G}_p$, where E_p is a product of matrices which interchange a pair of rows of $\tilde{G}_p$. Note that E_p satisfies $E_p'E_p = I$, so that $\det(E_p) = (-1)^{\ell_p}$ for ℓ_p equal to 1 or 2. It follows that $\det(\tilde{G}_p) = (-1)^{\ell_p} \det(\tilde{C}_p)$. Then since for each p, $\det(\tilde{G}_p) = (-1)^{\ell_p} \det(\tilde{C}_p)$, $\det(G) = \sum_{p=1}^{2^L} \det(\tilde{G}_p) = \sum_{p=1}^{2^L} (-1)^{\ell_p} \det(\tilde{C}_p)$

Proof of Theorem 5.2: If rank(G) = q then by Lemma A $\text{rank}(\tilde{C}_p) = q$ for some p. Since $\tilde{C}_p$ is block diagonal, with diagonal blocks C_{pi},(i=1,...,M), we have $\sum_{i=1}^{M} \text{rank}(C_{pi}) = q$. Each C_{pi} has q_i columns so that $\text{rank}(C_{pi}) \leq q_i$, and consequently $\text{rank}(C_{pi}) = q_i$,(i=1,...,M).

Proof of Theorem 5.4: Hall's Theorem (see Geraci (1976)) states that, for a set S, a collection of subsets $S_1,\ldots,S_N$ has an associated N-tuple $(s_1,\ldots,s_N)$ of distinct elements of S with $s_i \in S_i$, (i=1,...,N), if and only if for each $n=1,\ldots,N$ the union of every n subsets S_i contains n distinct elements of S. For each i such that $a_i > 0$ let $\tilde{R}_i$ be an a_i tuple consisting of a_i copies of R_i and let $\tilde{R}$ be the $\sum_{i=1}^{M} a_i$ tuple $(\tilde{R}_1,\ldots,\tilde{R}_M)$. Note that each distinct k contained in a copy of R_i corresponds to a residual u_j assigned as an instrumental variable to equation i, where the kth covariance restriction is $\sigma_{ij} = 0$. Therefore, an assignment of residuals such that for each i at least a_i residuals are assigned to each equation corresponds to a vector $(s_1,\ldots,s_N)$, with $N = \sum_{i=1}^{M} a_i$, of distinct elements of $\{1,\ldots,L\}$ such that each copy of R_i has at least one corresponding element s_m. It follows immediately from

Hall's Theorem that such an assignment will exist if and only for each $n = 1,\ldots, \sum_{i=1}^{M} a_i$ the union of any n components of $\tilde{R}$ contains at least n distinct elements. We will refer to this condition as condition R, and will show that it is equivalent to the condition given in the theorem.

Suppose that condition R is true. For any particular subset J, consider choosing $\sum_{i \varepsilon J} a_i$ components of $\tilde{R}$ equal to all the a_i copies of R_i for $i \varepsilon J$. Then the union of these components of $\tilde{R}$ is just the union of R_i for all $i \varepsilon J$, and by condition R must have $\sum_{i \varepsilon J} a_i$ distinct elements. Now suppose that the condition given in the statement of the theorem is true. Consider the union of n components of $\tilde{R}$. Since all of the components of $\tilde{R}$ are R_i sets, there is a subset J such that this union is equal to the union of R_i for $i \varepsilon J$, which must have at least $\sum_{i \varepsilon J} a_i$ distinct elements by the condition from the theorem. Furthermore, since for each $i \varepsilon J$ there are exactly a_i copies of R_i among the components of $\tilde{R}$, $\sum_{i \varepsilon J} a_i \geq n$.

Proof of Lemma 5.5: We drop the p subscript for notational convenience. We also assume i=1. Note that the first column of Σ_1 consists entirely of zeros, since to qualify as an instrument for the first equation a disturbance u_j must satisfy $E(u_1 u_j) = \sigma_{1j} = 0$. Let e_1 be an M dimensional unit vector with a one in the first position and zeros elsewhere. Then $\phi_1 A e_1 = 0$ and the covariance restrictions imply $Fe_1 = 0$ where $F = (A'\phi', \Sigma_1')'$. Note that rank $(FB^{-1}) = \text{rank}(F)$. Also $FB^{-1}B_1 = Fe_1 = 0$ where B_1 is the first column of B, so that the first column of F is a linear combination of the other columns of F by $B_{11} = 1$. Let Γ_1 be the rows of Γ corresponding to the excluded predetermined variables. Then $\phi_1 AB^{-1} = [E_1', (B')^{-1}\Gamma_1']'$ where E_1 is an $(M-1-r_1)xM$ matrix for which each row has a one in the position corresponding to a distinct excluded endogenous variable and zeros elsewhere. Let $(B^{-1})_1$ be the columns of B^{-1} corresponding to included right-hand side endogenous variables. Note that $FB^{-1} = \begin{bmatrix} E_1 \\ \Gamma_1 B^{-1} \\ \Sigma_1 B^{-1} \end{bmatrix}$.

Then row reduction of FB^{-1} using the rows of E_1, and the fact that the first column of FB^{-1} is a linear combination of the other columns imply

$$\text{(A.27)} \quad \text{rank}(FB^{-1}) = \text{rank}\begin{bmatrix} \Gamma_1 & (B^{-1})_1 \\ \Sigma_1 & (B^{-1})_1 \end{bmatrix} + M-1-r_1.$$

Now consider C_1. Note that for any $j \neq 1$, $\operatorname{plim} u_j' X_1/T = [\Sigma_j (B^{-1})_1, 0_1]$, where 0_1 is a $1 \times s_1$ vector of zeros and Σ_j is the jth row of Σ. By C_1 non-singular

$$\text{(A.28)} \qquad \operatorname{rank}(C_1) = \operatorname{rank}\left(\begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix} \left[\operatorname{plim} D_1 (u_j' X_1/T)\right]\right)$$

$$= \operatorname{rank} \begin{bmatrix} \Pi_1 & I_1 \\ \Sigma_1 (B^{-1})_1 & 0_1 \end{bmatrix} = \operatorname{rank} \begin{bmatrix} \Gamma (B^{-1})_1 & I_1 \\ \Sigma_1 (B^{-1})_1 & 0_1 \end{bmatrix}.$$

By column reduction, using the columns of $[I_1' \; 0_1']'$, equation (A.28) implies

$$\text{(A.29)} \qquad \operatorname{rank}(C_1) = \operatorname{rank} \begin{bmatrix} \Gamma_1 (B^{-1})_1 \\ \Sigma_1 (B^{-1})_1 \end{bmatrix} + s_1.$$

Then equations (A.27) and (A.29) imply $M-1-\operatorname{rank}(F) = q_1 - \operatorname{rank}(C_1)$, from which the conclusion of the proposition follows.

Proof of Theorem 5.6: Note that for any assignment p with $\operatorname{rank}(C_{pi}) < q_i$ for some i it follows that $\det(\tilde{C}_p) = 0$. Then the conclusion follows from Lemma A1, since the determinant of G is nonzero because exactly one of the determinants $\det(\tilde{C}_p)$ is nonzero.

Proof of Theorem 5.7: This proof follows closely the proof of Lemma 5.5. Let $\tilde{F} = \operatorname{diag}(\phi_1, \ldots, \phi_M, S') \bullet (I_M \otimes A', (I_M \otimes \Sigma)(P+I))'$.

Post-multiplication of $\tilde{F}$ by $I_M \otimes B^{-1}$ and row reduction using E_i, $i=1,\ldots,M$ as in the proof of Lemma 5.5 gives

(A.30) $$\operatorname{rank} \tilde{F}(I_M \otimes B^{-1}) = \operatorname{rank}(G) - \sum_{i=1}^{M} s_i + M^2 - M - \sum_{i=1}^{M} r_i = (\operatorname{rank}(G) - q) + M^2 - M.$$

REFERENCES

Amemiya, T.: "Tha Maximum Likelihood and Nonlinear Three Stage Least Squáres Estimator in the General Nonlinear Simultaneous Equations Model," Econometrica, 45 (1977), 955-968.

Bekker, P. A., and D. S. G. Pollock: "Identification of Linear Stochastic Models with Covariance Restrictions," mimeo, University of Tilsburg, 1984.

Chamberlain, G.: "Panel Data," in Z. Griliches and M. Intriligator, Handbook of Econometrics, Vol. 2, Amsterdam: North-Holland, 1984.

Fisher, F.: The Identification Problem in Econometrics, New York: McGraw-Hill, 1966.

Geraci, Vincent J.: "Identification of Simultaneous Equation Models with Measurement Error," Journal of Econometrics, 4 (1976), 263-283.

Hansen, L. P.: "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica, 50 (1982), 1029-1054.

Hausman, J. A.: "Simultaneous Equation Estimation," in Z. Griliches and M. Intriligator, Handbook of Econometrics, Vol. 1, Amsterdam: North-Holland, 1983.

Hausman, J. A. and W. E. Taylor: "Identification in Linear Simultaneous Equations Models with Covariance Restrictions: An Instrumental Variables Interpretation," Econometrica, 51 (1983), 1527-1549.

Henderson, H. V., and S. R. Searle: "Vec and Vech Operators for Matrices, with Some Uses in Jacobians and Multivariate Statistics," Canadian Journal of Statistics, 7 (1979), 65-81.

Hsiao, C.: "Identification," in Z. Griliches and M. Intriligator, Handbook of Econometrics, Vol. 1, Amsterdam: North-Holland, 1983.

Koopmans, T. C., and W. C. Hood: "The Estimation of Simultaneous Linear Economic Relationships," in W. C. Hood and T. C. Koopmans (eds.), Studies in Econometric Method, Cowles Commission Monograph 14, New York: Wiley, 1953.

Koopmans, T. C., H. Rubin, and R. B. Leipnik: "Measuring the Equation Systems of Dynamic Economics," in Koopmans, T. C., (ed.), Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph 10, New York: Wiley, 1950.

Malinvaud, E.: Statistical Methods of Econometrics, Amsterdam: North-Holland, 1970.

Madansky, A.: "On the Efficiency of Three-Stage Least Squares Estimation," Econometrica, 32 (1964), 51-56.

Richard, J. F.: "A Note on the Information Matrix of the Multivariate Normal Distribution," Journal of Econometrics, 3 (1975), 57-60.

Rothenberg, T. J.: "Identification in Parametric Models," Econometrica, 39 (1971), 577-592.

Rothenberg, T. J.: Efficient Estimation with a priori Information, Cowles Foundation Monograph 23, New Haven: Yale University Press, 1973.

Rothenberg, T. J.: "Estimation of Simultaneous Equations Models Under Covariance Restrictions: A Minimum Distance Approach," mimeo, Berkeley, 1983.

Rothenberg, T. J. and C. T. Leenders: "Efficient Estimation of Simultaneous Equations Systems," Econometrica, 32 (1964), 57-76.

Sargan, J. D.: "Identification and Lack of Identification," Econometrica, 51 (1983), 1605-1633.

Wegge, L.: "Identifiability Criteria for a System of Equations as a Whole," The Australian Journal of Statistics, 7 (1965), 67-77.

White, H.: "Nonlinear Regression on Cross-Section Data," Econometrica, 49 (1980), 721-746.

Zellner, A. and H. Theil: "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations," Econometrica, 30 (1962), 54-78.