working paper
department
of economics

Efficient Estimation and Identification of
Simultaneous Equation Models
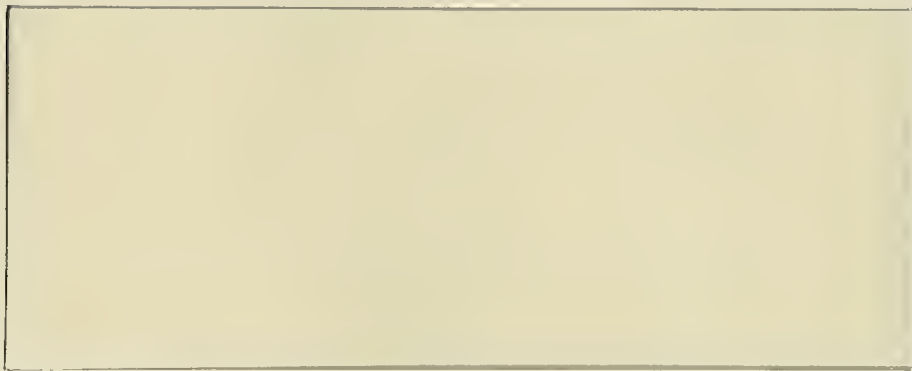with Covariance Restrictions

Jerry Hausman

Whitney Newey

William Taylor

Number 331   Revised October 1983

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

Efficient Estimation and Identification of
Simultaneous Equation Models
with Covariance Restrictions

Jerry Hausman

Whitney Newey

William Taylor

Number 331   Revised October 1983

Efficient Estimation and Identification of Simultaneous Equation
Models with Covariance Restrictions

Jerry Hausman
MIT

Whitney Newey
Princeton

William Taylor
ATT

Revised October 1983

This paper is offered in recognition of the fiftieth birthday of the Cowles Foundation.

1. Introduction

In the pioneering research in econometrics done at the Cowles Foundation, estimation technqiues 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 restriction, the problem of identification with covariance restrictions remained. Further research by Fisher (1966), Rothenberg (1971), and Wegge (1965) advanced our knowledge in this field, with Fisher's examination of certain special cases especially insightful. In a companion paper, Hausman and Taylor (1983), we 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 our solution is that covariance restrictions, if they are to be effective in aiding identification, must take one of two forms. First, covariance restrictions can cause an endogenous variable to be predetermined in a particular equation, e.g., a "relatively recursive" specification. Or, covariance restrictions can lead to an estimated residual from an identified equation being predetermined in another equation. Both of these forms of identification have ready interpretations in estimation as instrumental variable procedures, 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, pp. 154-211). The case of a diagonal

covariance matrix is also analyzed by Malinvaud (1970, pp. 678-682) and

by Rothenberg (1973, pp. 77-79 and pp. 94-115), who also does FIML

estimation in two small simultaneous equations models to assess the value

of the covariance restrictions. 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. The Cramer-Rao asymptotic lower

bound is reduced by covariance restrictions, but 3SLS does not adequately

account for these restrictions. The reason for this finding is that

simply imposing the restrictions on the covariance matrix is not adequate

---

[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).

when endogenous variables are present because of the lack of block-diagonality of the information matrix between the slope coefficients and the unknown covariance parameters. In fact, imposing the covariance restrictions on the 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 our identification results have an instrumental variable interpretation, it is natural to think of using them in a 3SLS-like instrumental variables estimator. Madansky (1964) gave an instrumental variable interpretation to 3SLS and here we augment the 3SLS estimator by the additional instruments which the covariance restrictions imply. That

---

[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).

is, in an equation where the covariance restrictions cause a previously endogeneous variable to be predetermined, we use the variable as an instrument for itself - if it is included in the equation - or just as an instrument if it is not included. In the alternative case, we use the appropriate estimated residuals from other identified equations to form instruments for a given equation. This estimator which we call the augmented three stage least squares estimator (A3SLS) is shown to be more efficient than the 3SLS estimator when effective covariance restrictions are present.

To see how efficient the A3SLS estimator is, we need to compare it to FIML which takes account of the covariance restrictions. Hausman (1975) gave an instrumental variable 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 the predicted value of the endogenous variables based only on the predetermined variables from the reduced form as instruments, when covariance restrictions are present, FIML also uses that part of the estimated residual from the appropriate reduced form equation which is uncorrelated with the residual in the equation where the endogenous variables are included. Thus more information is used in forming the instruments than in the case where

covariance restrictions are absent. More importantly, the instrumental variable interpretation of FIML leads to a straightforward proof that the A3SLS estimator is asymptotically efficient with respect to the FIML estimator. The A3SLS estimator provides a computationally convenient estimator which is also asymptotically efficient. Thus we are left with an attractive solution to both identification and estimation of the traditional simultaneous equations model. Identification and estimation both are closely related to the notion of instrumental variables which provides an extremely useful concept upon which to base our understanding of simultaneous equations model specifications.

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. The rank condition provides a stronger necessary condition than the condition of Fisher (1966). These necessary conditions apply equation by equation. We also provide a sufficient condition for identification in terms of the structural parameters of the entire system. Lastly, we provide straightforward specification tests for the covariance restrictions which can be used to test non-zero covariances.

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. Consider an industry supply curve which in the short run exhibits decreasing returns to scale. Quantity demanded is thus an appropriate included variable in the supply equation which determines price, y_1, as a function of quantity demanded, y_2. Also included in the specification of the supply equation are the quantities of fixed factors and prices of variable factors, both of which are assumed to be exogenous. The demand equation has price as a jointly endogenous explanatory variable together with an income variable assumed to be exogenous. We assume the covariance matrix of the residuals to be diagonal, since shocks from the demand side of the market are assumed to be fully captured by the inclusion of y_2 in the supply equation. The model specification in this simple case is

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

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

where we have further simplified by including only one exogenous variable in equation (2.1). 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 $E(\varepsilon_i | z_1, z_2) = 0$ for i=1,2, $\text{var}(\varepsilon_i | z_1, z_2) =$

σ_{ii}, $\sigma_{12} = \text{cov}(\varepsilon_1 \varepsilon_2 | z_1, z_2) = 0$.

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 $W_i = (Z \hat{\Pi}_j \ z_i)$, $i \neq j$, where $Z = (z_1 z_2)$ and $\hat{\Pi}_j$ is the vector of reduced form coefficients for the (other) included endogenous variables.[3] To see how FIML differs from the instrumental variables (IV) estimator, we solve for the first order conditions of the likelihood function under the assumption that the ε_i's are normally distributed. Of course, as is the case for linear simultaneous equation estimation with only coefficient restrictions, failure of the normality assumption does not lead to inconsistency of the estimates. For the two equation example, the likelihood function takes the form

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

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

where c is a constant and the X_i's and δ_i's contains the right hand side

---

[3]. 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.

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 (2.3); results for the second equation are identical. The three first order conditions are

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

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

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

Rearranging equation (2.4c) yields the familiar solution for the variance, $\sigma_{11} = (1/T)(y_1 - X_1\delta)'(y_1 - X_1\delta)$. Equation (2.4b) has the usual OLS form which is to be expected since z_1 is an exogenous variable. It is equation (2.4a) where the simultaneous equations nature of the model appears with the presence of $-T\beta_{12}/(1 - \beta_{12}\beta_{21})$ which arises from the Jacobian term in the likelihood function: see Hausman(1975).

Now the first order conditions for equations (2.4a) - (2.4c) can be solved by numerical methods which maximize the likelihood function. Koopmans et al., (1950) have a lengthy discussion of various numerical technques for maximization which must be one of the earliest treatments of this problem in the econometrics literature. But we can solve

equation (2.4a) in a particular way, using the reduced form specification to see the precise role of the covariance restrictions in the model. We first multiply equation (2.4a) by σ_{11} to get

$$(2.5) \qquad \frac{-\Pi\beta_{12}\sigma_{11}}{1-\beta_{12}\beta_{21}} + (y_1-X_1\delta_1)'y_2 = \frac{-\beta_{21}}{1-\beta_{12}\beta_{21}}(y_1-X_1\delta_1)'(y_1-X_1\delta_1)$$

$$+y_2'(y_1-X_1\delta_1) = 0.$$

We now do the substitutions from the reduced form equations $y_i = Z\Pi_i + v_i$ using the fact that $v_2 = \beta_{21}\epsilon_1/(1-\beta_{12}\beta_{21}) + \epsilon_2/(1-\beta_{12}\beta_{21})$. We transform equation (2.5) to

$$(2.6)\left(\frac{1}{1-\beta_{12}\beta_{21}}(y_2-X_2\delta_2)-v_2\right)'(y_1-X_1\delta_1)$$

$$+ (Z\Pi_2 + v_2)'(y_1-X_1\delta_1) = 0.$$

Canceling terms, we find the key result

$$(2.7) \qquad (Z\Pi_2 + \frac{\epsilon_2}{1-\beta_{12}\beta_{21}})'\epsilon_1 = 0.$$

Without the covariance restriction, we would have the result

$$(2.8) \quad (Z\Pi_2)'\epsilon_1 = 0, \qquad \Pi_2 = \left(\frac{-\beta_1\gamma_{11}}{1-\beta_{12}\beta_{21}} \quad \frac{\gamma_{22}}{1-\beta_{12}\beta_{21}}\right)'$$

which is the instrumental variable interpretation of FIML given by Hausman (1975), equation (12). But in equation (2.7), we have the additional term $\varepsilon_2/(1-\beta_{12}\beta_{21})$. What has happened is that FIML has used the covariance restrictions to form a better instrument. Remember that y_2 forms the best instrument for itself if it is predetermined in equation (2.1). But here, y_2 is jointly endogenous since it is correlated with ε_1 : from the reduced form equation

$$y_2 = Z\,\Pi_2 + \frac{\varepsilon_2}{1-\beta_{12}\beta_{21}} + \frac{\beta_{21}\varepsilon_1}{1-\beta_{12}\beta_{21}}.$$

FIML cannot use the last term in forming the instrument for y_2 since $\beta_{21}\varepsilon_1/(1-\beta_{12}\beta_{21})$ is correlated with the residual ε_1 in equation (2.1). It is this last term which makes y_2 endogenous in the first equation. However, FIML can use $\varepsilon_2/(1-\beta_{12}\beta_{21})$ because ε_2 is uncorrelated with ε_1 by the covariance restriction $\sigma_{12}= 0$. By using this term, FIML creates a better instrument than ordinary 2SLS which ignores $\varepsilon_2/(1-\beta_{12}\beta_{21})$. Our two equation example makes it clear why 3SLS is not asymptotically efficient when covariance restrictions are present. FIML uses better instruments than 3SLS and produces a better estimate of the included endogenous variables.

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 is $Z\Pi_2 + \varepsilon_2 = y_2$ 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

$$(2.9) \qquad W_1 = \left(Z \Pi_1 + \frac{\varepsilon_2}{1-\beta_{12}\beta_{21}}, z_1 \right).$$

Because of the addition of the residual term in W_1, the instrument matrix has full rank and the coefficients can be estimated. Both of these estimation results arise because we have restrictions on the matrix $B^{-1}\Sigma$, where B is the matrix of all coefficients of endogenous variables and Σ is the disturbance covariance matrix. The estimation results are closely connected with the identification results when covariance restrictions are present, see Lemma 3, Proposition 6 of Hausman and Taylor (1982), and Section 5.

FIML needs to be iterated to solve the first order conditions; in our two equation case, we see that the original first order condition (2.4a) or its tranformed version, equation (2.7), is nonlinear. Computational considerations when estimating FIML in this form are discussed in Hausman (1974). But we know if the covariance restrictions were not present, 3SLS (or here 2SLS) gives asymptotically efficient instruments. Since 3SLS is a linear IV estimator, it is straightforward to compute and is included in many econometric computer packages. Yet we also know that if $\sigma_{12} = 0$ then 3SLS is not asymptotically efficient. Furthermore, if $\beta_{21} \neq 0$ and $\gamma_{22}=0$, it would not be clear how to do 2SLS on

the first equation, since it is not identified by the coefficient restrictions alone. If we had to use the 2SLS instruments $y_2 (= Z \hat{\Pi}_2)$ and z_1, the instrument matrix $(W_1 ' X_1)$ would be singular as expected since the rank condition fails. The FIML solution which accounts for the covariance restriction $\sigma_{12} = 0$ is very suggestive. Suppose for the first equation in the specification of equations (2.1) and (2.2) $\hat{\epsilon}_2 = y_2 - X_2 \hat{\delta}$ is used as an instrument in addition to z_1 and z_2, as is the case for FIML. It follows immediately that the optimal estimator which use z_1, z_2, and $\hat{\epsilon}_2$ as instruments for the first equation and z_1 and z_2 as instruments for the second equation, which we call augmented 3SLS or A3SLS, is asymptotically more efficient than ordinary 3SLS since it uses more instruments. But the important question is whether A3SLS is asymptotically equivalent to FIML, as it is without covariance restrictions. We have accounted for all the restrictions in the model by adding ϵ_2 as an instrument for the first equation. A3SLS differs from FIML because it replaces efficient estimates of parameters in the instrument matrix with inefficient estimates. Unlike the case of only coefficient restrictions, this replacement does affect the asymptotic distribution of the coefficient estimator. However, this replacement is corrected for by the optimal IV estimator in such a way that A3SLS is asymptotically equivalent to FIML.

To investigate the asymptotic properties of the estimators in this case, we first calculate the information matrix for the example of equations (2.1) and (2.2). We denoted $d = (1 - \beta_{12} \beta_{21})$, let $\delta = (\beta_{12}, \gamma_{11}, \beta_{21}, \gamma_{22})'$, and $\sigma = (\sigma_{11}, \sigma_{22})$. Let N denote $\lim(1/T) Z' Z$, $e_1 = (1,0)'$, $e_2 = (0,1)'$, and $D_i = (\Pi_j, e_i)$, $i \neq j$. The lower triangle of the information

matrix is

$$
(2.10) \quad J = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} = \lim_{T \to \infty} - E \frac{1}{T} \begin{bmatrix} \dfrac{\partial^2 L}{\partial \delta \partial \delta'} & - \\ \dfrac{\partial^2 L}{\partial \delta \partial \sigma'} & \dfrac{\partial^2 L}{\partial \sigma \partial \sigma'} \end{bmatrix}
$$

where

$$
\begin{bmatrix} J_{11} \\ J_{21} \end{bmatrix} = \begin{bmatrix} \dfrac{2\beta_{21}^2 + \sigma_{22}/\sigma_{11}}{d^2} e_1 e_1' + D_1'MD_1/\sigma_{11} & \dfrac{2\beta_{12}^2 + \sigma_{11}/\sigma_{22}}{d^2} e_1 e_1' + D_2'MD_2/\sigma_{22} \\ \dfrac{1}{d^2} e_1 e_1' & \\ \dfrac{\beta_{21}}{\sigma_{11}d} e_1 e_1' & \dfrac{\beta_{12}}{\sigma_{22}d} e_2 e_1' \end{bmatrix}
$$

and

$$
J_{22} = \begin{bmatrix} -1/(2\sigma_{11}^2) & 0 \\ 0 & -1/(2\sigma_{22}^2) \end{bmatrix}
$$

We now compute the covariance matrix for $\hat{\delta}$ using the formula $J^{11} = (J_{11} - J_{12} J_{22}^{-1} J_{21})^{-1}$ to find

$$
(2.11) \quad \text{Var}(\hat{\delta}) = J^{11} = \begin{bmatrix} \dfrac{\sigma_{22}/\sigma_{11}}{d^2} e_1 e_1' + D_1'MD_1/\sigma_{11} & \dfrac{1}{d^2} e_1 e_1' \\ \dfrac{1}{d^2} e_1 e_1' & \dfrac{\sigma_{11}/\sigma_{22}}{d^2} e_1 e_1' + D_2'MD_2/\sigma_{22} \end{bmatrix}
$$

It is interesting to note that without covariance restrictions we would have

$$(2.12) \quad \text{Var}(\delta) = \begin{vmatrix} D_1'M\,D_1/\sigma_{11} & 0 \\ 0 & D_2'MD_2/\sigma_{22} \end{vmatrix}$$

so that the covariance restrictions have produced a more efficient estimator via the addition of the positive semi-definite matrix $(1/d^2)F'F$ to $(J^{11})^{-1}$, where $F = \left[\sqrt{\sigma_{22}/\sigma_{11}}\,e_1'\ ,\ \sqrt{\sigma_{11}/\sigma_{22}}\,e_1'\right]$.

We now compare these results to the A3SLS estimator, which we will denote by δ. Stack equations (2.1) and (2.2) as

$$(2.13) \quad y = X\delta + \varepsilon \ , \ \varepsilon = \begin{vmatrix} \varepsilon_1 \\ \varepsilon_2 \end{vmatrix} \ , \ X = \begin{vmatrix} X_1 & 0 \\ 0 & X_2 \end{vmatrix} \ , \ y = \begin{vmatrix} y_1 \\ y_2 \end{vmatrix}.$$

Let the $2T \times 5$ matrix of instrumental variables for the system given in equation (2.13) be $\widetilde{W}$, where

$$\widetilde{W} = \begin{vmatrix} Z, \widetilde{\varepsilon}_2, & 0 \\ \\ 0, 0, & Z \end{vmatrix} \ , \ \widetilde{\varepsilon}_2 = y_2 - X_2\widetilde{\delta}_2, \ \widetilde{\delta}_2 = (Z'X_2)^{-1}Z'y_2.$$

The A3SLS estimator is an instrumental variables estimator which satisfies

$$(2.14) \quad \widehat{\delta} = (A_T\widetilde{W}'X)^{-1}A_T\widetilde{W}'y$$

where the 5 x 4 linear combination matrix A_T is optimally chosen. Any sequence $\{A_T\}$ which satisfies plim $A_T = A$, where

$$(2.15) \quad A = \tilde{V}^{-1} \text{plim} \, (\tilde{W}'X/T)$$

and $\tilde{V}$ is the asymptotic covariance matrix of $\tilde{W}'\varepsilon/\sqrt{T}$, is optimal in the sense of obtaining the linear combination of instrumental variables with the smallest asymptotic covariance matrix for $\tilde{\delta}$. The asymptotic covariance matrix of $\hat{\delta}$ will then be

$$(2.16) \quad \text{Var} \, (\hat{\delta}) = \left[\text{plim}(X'\tilde{W}/T)\tilde{V}^{-1} \, \text{plim}(\tilde{W}'X/T) \right]^{-1}.$$

Some care must be exercised when applying a central limit theorem to calculate $\tilde{V}$, due to the instrument $\tilde{\varepsilon}_2$ depending on an estimated parameter $\hat{\delta}_2$. Let W be the instrument matrix obtained from $\tilde{W}$ by replacing $\tilde{\varepsilon}_2$ by the true disturbance ε_2. Then by $E(\varepsilon_1 \varepsilon_2) = E(\varepsilon_1 \varepsilon_2^2) = 0$ and $E(\varepsilon_1^2 \varepsilon_2^2) = \sigma_{11}\sigma_{22}$, we can use a central limit theorem to obtain

$$(2.17) \quad W'\varepsilon/\sqrt{T} = \left[Z'\varepsilon_1, \, \varepsilon_2'\varepsilon_1, Z'\varepsilon_2 \right]'/\sqrt{T} \overset{d}{\to} N(0,V)$$

where

$$V = \begin{bmatrix} \sigma_{11} N & 0 & 0 \\ 0 & \sigma_{11}\sigma_{22} & 0 \\ 0 & 0 & \sigma_{22} N \end{bmatrix}.$$

Also, since $\tilde{\varepsilon}_2 = \varepsilon_2 - X_2 \, (\hat{\delta}_2 - \delta_2) = \varepsilon_2 - X_2(Z'X_2)^{-1}Z'\varepsilon_2$ we can write

$$\widetilde{W}'\varepsilon/\!\!\sqrt{T} = W'\varepsilon/\!\!\sqrt{T} - \left[0, \varepsilon_1'X_2(Z'X_2)^{-1}Z'\varepsilon_2/\!\!\sqrt{T}, \ 0\right]'$$

$$= P_T W'\varepsilon/\!\!\sqrt{T} \ ,$$

where, for I_2 denoting a two dimensional identity matrix,

$$P_T = \begin{bmatrix} I_2 & 0 & 0 \\ 0 & 1 & -\varepsilon_1'X_2(Z'X_2)^{-1} \\ 0 & 0 & I_2 \end{bmatrix}.$$

Note that

$$\text{plim } \varepsilon_1'X_2(Z'X_2)^{-1} = \left(\text{plim } \varepsilon_1'X_2/T\right)\left(\text{plim } Z'X_2/T\right)^{-1} = (\sigma_{11}e_1'/d)(MD_2)^{-1}.$$

Then for $P = \text{plim } P_T$, equation (2.17) implies

$$(2.18) \quad \widetilde{W}'\varepsilon/\!\!\sqrt{T} \ \overset{d}{\to} \ N(0, PVP'),$$

so that $\widetilde{V} = PVP'$. We also calculate that

$$(2.19) \quad \text{plim } \widetilde{W}'X/T = \text{plim } \frac{1}{T} \begin{bmatrix} Z'X_1 & 0 \\ \widetilde{\varepsilon}_2'X_1 & 0 \\ 0 & Z'X_2 \end{bmatrix} = \begin{bmatrix} MD_1 & 0 \\ \sigma_{22}e_1'/d & 0 \\ 0 & MD_2 \end{bmatrix},$$

$$P^{-1}\text{plim}(\widetilde{W}'X/T) = \begin{bmatrix} MD_1 & 0 \\ \sigma_{22}e_1'/d & \sigma_{11}e_1'/d \\ 0 & MD_2 \end{bmatrix}.$$

From equations (2.16), (2.18), and (2.19) we can obtain the asymptotic covariance matrix for the A3SLS estimator

$$(2.20) \quad \mathrm{Var}(\widehat{\delta}) = \left[\mathrm{plim}(X'\widetilde{W}/T)(P')^{-1}V^{-1}P^{-1}\mathrm{plim}(\widetilde{W}'X/T) \right]^{-1}$$

$$= \begin{bmatrix} \dfrac{\sigma_{22}/\sigma_{11}}{d^2}\, e_1 e_1' + D_1'MD_1/\sigma_{11} & \dfrac{1}{d^2}\, e_1 e_1' \\[2ex] \dfrac{1}{d^2}\, e_1 e_1' & \dfrac{\sigma_{11}/\sigma_{22}}{d^2}\, e_1 e_1' + D_2'MD_2/\sigma_{22} \end{bmatrix}$$

which is also the FIML asymptotic covariance matrix obtained as equation

(2.11). Consequently, A3SLS is asymptotically efficient.

It is clear that A3SLS must be more efficient, asymptotically, than 3SLS because 3SLS ignores the instrumental variable $\widetilde{\varepsilon}_2$ when forming the instruments. It is somewhat surprising that A3SLS is efficient, since A3SLS uses an inefficient estimator when forming $\widetilde{\varepsilon}_2$. A3SLS corrects for the use of $\widetilde{\delta}_2$ by using a different linear combination of the instruments than FIML. Rather than using $y_i = Z\Pi_i + \varepsilon_i/(1-\beta_{12}\beta_{21})$ as an instrument for y_i, i=1,2, A3SLS is a system IV estimator, with linear combination matrix

$$(2.21) \quad A = (P')^{-1}V^{-1}P^{-1}\mathrm{plim}(\widetilde{W}'X/T)$$

The reason that this correction for the use of the inefficient estimator $\tilde{\delta}_2$ in forming $\tilde{\varepsilon}_2$ results in fully efficient estimates is that the FIML estimate δ is also an IV estimator of the form given in equation (2.14). To see why this is so, note that asymptotically $\tilde{W}'\varepsilon/\sqrt{\overline{T}}$ is a nonsingular linear transformation of $W'\varepsilon \mid \sqrt{\overline{T}}$. If two sets of instrumental variables differ only by a nonsingular linear transformation, they span the same column space and therefore lead to equivalent estimators when used in an optimal manner.

Similarily, we can show that for the FIML instruments W^*, it is the case that

$$(2.22) \quad W^{*'}\varepsilon/\sqrt{\overline{T}} = S(W'\varepsilon/\sqrt{\overline{T}}) + o_p(1),$$

$$S = \begin{bmatrix} I_2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & I_2 \\ 0 & 1 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{11}e_1'/d \\ 0 & 0 \\ \sigma_{22}e_1'/d & 0 \end{bmatrix} J^{11} \begin{bmatrix} D_1'/\sigma_{11} & (e_1/\sigma_{11}d)'0 \\ 0 & e_1 & (/\sigma_{22}d) & D_2'/\sigma_{22} \end{bmatrix}$$

From equation (2.22) it follows that

$$(2.23) \quad W^{*'}\varepsilon/\sqrt{\overline{T}} = SP^{-1}\,\tilde{W}'\varepsilon/\sqrt{\overline{T}} + o_p(1).$$

We can also compute

(2.24) $\operatorname{plim} W^{*\prime}X/T = SP^{-1} \operatorname{plim} \widetilde{W}'X/T.$

Since the FIML estimator δ solves $B'_T W^{*\prime}\varepsilon = 0$, where $\operatorname{plim}B_T = B$ and

$$B = \begin{bmatrix} D_1 & 0 \\ 1/d & 0 \\ 0 & D_2 \\ 0 & 1/d \end{bmatrix},$$

it follows from equations (2.24) and (2.25) that

(2.25) $\sqrt{T}(\delta_F - \delta) = (B'W^{*\prime}X/T)^{-1} B'W^{*\prime}\varepsilon / \sqrt{T} + o_p(1)$

$= (B'S\ P^{-1}\ \widetilde{W}'X/T)\ B'S\ P^{-1}\ \widetilde{W}'\varepsilon / \sqrt{T} + o_p(1)$ so that FIML

is an instrumental variables estimator with instruments $\widetilde{W}$.

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.[4] 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:

$$(3.2) \qquad U \sim N(0, \Sigma \otimes I_T)$$

---

[4]. This set-up is fairly general since all linear restrictions on the covariance matrix can be put into this form by appropriate transformations of model specification. However, an important case which our approach does not treat occurs when unknown slope parameteres are present in the covariance matrix. This sometimes will occur when errors in variables are present in a simultaneous equations equations model, e.g., Hausman (1977).

where Σ is positive definite.[5] 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 from the likelihood functions, we will only solve for the unknown elements of Σ rather than the complete matrix as is the usual case. Using the results of Hausman and Taylor (1982) and Section 5, we assume that each equation in the model is identified by use of 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,

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

As we saw in the last section, it is the components of the (row) vector $v_i = (UB^{-1})_i$ that give the extra instruments that arise because of the covariance restrictions. 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:

$$(3.4) \qquad X_i = [\,Y_i \; Z_i\,], \; \delta_i{}' = [\,\beta_i{}' \; \gamma_i{}'\,].$$

---

[5]. If U is not normal but has the same first two moments as in equation (3.2), the FIML estimator will be consistent and asymptotically normal. However, unlike the case of no covariance restrictions, standard errors which are calculated using the normal disturbance information matrix may be inconsistent, due to third and fourth order moments of the non-normal disturbances. For this same reason, FIML assuming normal disturbances may not lead to the optimal IV estimator when the disturbances are non-normal.

It will prove convenient to stack these M equations into a system

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

where

$$y = \begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{vmatrix} \quad X = \begin{vmatrix} X_1 & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & X_M \end{vmatrix} \quad ,\delta = \begin{vmatrix} \delta_1 \\ \cdot \\ \cdot \\ \delta_M \end{vmatrix} , \quad u = \begin{vmatrix} u_1 \\ \cdot \\ \cdot \\ u_M \end{vmatrix} .$$

Likewise, we stack the reduced form equations

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

where $\widetilde{Z} = [I \times Z]$ and $\widetilde{\Pi} = [\Pi_1' \ldots \Pi_M']'$ is the vector of reduced form coefficients.

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

$$(3.7) \qquad L(B,\Gamma,\Sigma) = c + \frac{T}{2} \log \det(\Sigma)^{-1} + T \log |\det (B)|$$

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

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 $\underset{=}{u}$ to remind the reader of this fact.[6] 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 are

(3.8a) $\qquad \dfrac{\partial L}{\partial B} : T(B')^{-1} - Y'(YB + ZT)\Sigma^{-1} \underset{=}{u} 0,$

(3.8b) $\qquad \dfrac{\partial L}{\partial \Gamma} : - Z'(YB + ZT)\Sigma^{-1} \underset{=}{u} 0,$

(3.8c) $\qquad \dfrac{\partial L}{\partial \Sigma^{-1}} : T\Sigma - (YB + ZT)'(YB + ZT) \underset{=}{u} 0.$

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.

Let us consider the first order conditions in reverse order. We already
know some elements of Σ because of the covariance restrictions. The
unknown elements are then estimated by $\sigma_{ij} = (1/T)(y_i - X_i\delta_i)'(y_j - X_j\delta_j)$
where the δ's contain the estimates of the unknown elements of the B and

---

[6]. An alternative procedure is to us 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.

Γ matrices. Equation (3.8b) causes no special problems because it has the form of the first order conditions of the multivariate regression model. But it is equation (3.8a) which we must transform to put the solution into instrumental variable form. The presence of the term $T(B')^{-1}$ which arises from the Jacobian term in the log likelihood function of equation (2.7) distinguishes the problem from the non-simultaneous equations case. We now transform equation (3.8a) to eliminate $T(B')^{-1}$ and to replace the matrix Y' by the appropriate matrix of predicted Y's which are orthogonal to the U matrix.

First we transform equation (3.8a) by using the identity $\Sigma\Sigma^{-1} = I$:

$$(3.9) \qquad [T(B')^{-1}\Sigma - Y'(YB + Z\Gamma)]\Sigma^{-1} \overset{u}{=} 0.$$

Note the presence in equation (3.9) of the term $(B')^{-1}\Sigma$ which is the key term for the identification results in Hausman and Taylor (1982). We now do the substitution similar to the one we used for equations (2.5) to (2.7) in the two equation case. For equation (3.8c), we know that the elements of Σ take one of two forms. Either they equal the inner product of the residuals from the appropriate equations divided by T or they equal zero. To establish some notation, define the set N_j as the indices m which denote for the j'th row of Σ that $\sigma_{jm} = 0$. Now we return to the first part of equation (3.9) and consider the ij'th element of the matrix product

$$(3.10) \qquad [T(B')^{-1}\Sigma]_{ij} = T\sum_{k=1}^{M}\beta^{ki}\sigma_{kj} = (v'_i - \sum_{k \,\varepsilon\, N_j}\beta^{ki}u'_k)u_j$$

where β^{ik} is the ik'th element of the inverse matrix B^{-1}. Note that if no zero elements existed in column i of Σ we would have $v'_i u_j$ on the right hand side of (3.10), as in Hausman (1975, equation 11). We now use the expression from equation (3.10) and combine it with the other terms from the bracketed term in equation (3.9):

$$(3.11) \quad \left[T(B')^{-1}\Sigma - Y'(YB+Z T) \right]_{ij} = \left[v_i - \sum_{k \,\epsilon\, N_j} \beta^{ki} u_k - (Z\Pi_i) - v_i \right]' u_j$$

$$= - \left[Z\Pi_i + \sum_{k \,\epsilon\, N_j} \beta^{ki} u_k \right]' u_j$$

$$= - \left[Z\Pi_i + b'_{ji}\tilde{u}_i \right]' u_j ,$$

where $b'_{ji}\tilde{u}'_i$ corresponds to the sum of the elements β^{ij} multiplied by the residuals u_k in the set N_j.

As with equation (2.7) we see that FIML replaces the jointly endogenous variable $y_i = Z\Pi_i + v_i$ with the prediction from the predetermined variables $Z\Pi_i$ and those elements of $v_i = (UB^{-1})_i$ which are uncorrelated with u_j because of zero restrictions on the σ_{ij}'s. Thus we rewrite equation (3.9) as

$$(3.12) \quad - \left[(Z B^{-1} + \tilde{V})'(YB + Z T) \right] \Sigma^{-1} \overset{u}{=} 0$$

$$- \left[(Y + \tilde{V})'(YB + Z T) \right] \Sigma^{-1} \overset{u}{=} 0 .$$

Equation (3.12) 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 term $\hat{Y}$, we have the extra elements $\tilde{V}$ which are uncorrelated structural residuals multiplied by the appropriate elements of B^{-1}. Thus FIML uses the

covariance restrictions to form a better predictor of Y than the usual
$\hat{Y}$.

Note that if $(B^{-1}\Sigma)_{ij} = 0$ in equation (3.10), equations i and j are
relatively recursive. Then y_i is predetermined in the j'th equation
rather than jointly endogenous, [7] and equation (3.11) reduces to the same
form as equation (3.8b): columns of Y are treated like columns of Z. In
general, however, y_i is replaced by a predicted value which is composed of
two terms: a prediction of the mean from the predetermined variables and
an estimate of part of the reduced form disturbance from the uncorrelated
structural residuals. For future reference, we gather together our
transformed first order conditions which arise from equations (3.8a) -
(3.8c):

$$(3.13) \qquad - \begin{bmatrix} (B^{-1})'\Gamma'Z' + \tilde{V}' \\ Z' \end{bmatrix}(YB + Z\Gamma)\Sigma^{-1} \overset{u}{=} 0.$$

$$T\Sigma - (YB + Z\Gamma)'(YB + Z\Gamma) \overset{u}{=} 0.$$

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., Hood and Koopmans (1953) or
Rothenberg (1973).[8] Let

---

[7]. See Hausman and Taylor (1983)
[8]. 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.

$$\tilde{B} = \mathrm{diag}(\tilde{B}_1, \tilde{B}_2, \ldots, \tilde{B}_M), \quad \tilde{B}_i = [\, (\overline{B}^{\,-1})_i' \;\; 0_i'\,]\,',$$

where 0_i is an $s_i \times M$ null matrix which corresponds to the s_i included predetermined variables and $(\overline{B}')^{-1}_i$ is the matrix of rows of $(\overline{B}')^{-1}$ which correspond to the r_i included explanatory endogenous variables. $\overline{B}$ is the matrix B with normalization and exclusion restrictions imposed. Let E be the $M^2 \times M^2$ matrix whose ij'th block $(i,j = 1, \ldots M)$ is given by E_{ji}, a matrix with one in (j,i) and zeros elsewhere. With no restrictions on the disturbance covariance matrix, the information matrix for the unknown parameters is given by

$$(3.14) \quad J(\delta, \sigma^*) = \begin{bmatrix} \tilde{B}E\tilde{B}' + \mathrm{plim}\, \frac{1}{T}\, X'(\Sigma^{-1} \otimes I_T)X & \tilde{B}(\Sigma^{-1} \otimes I_M)R' \\[2ex] J_{\delta\sigma'}\cdot & \frac{1}{2}\,R(\Sigma^{-1} \otimes \Sigma^{-1})R' \end{bmatrix}$$

where R is the $M^2 \times 1/2M(M+1)$ matrix of ones and zeros that maps $\sigma^* = (\sigma_{11}, \ldots, \sigma_{M1}, \sigma_{22}, \ldots, \sigma_{M2}, \ldots, \sigma_{MM})$ into the full vector of σ_{ij}'s which ignores the symmetry restrictions: see Richard (1975). If L covariances are restricted to be zero, let S be the $(1/2)M(M+1) \times (1/2)M(M+1) - L$ selection matrix which selects the non-zero elements of σ^*. The information matrix with covariance restrictions is then identical to that in equation (3.14) with S'R substituted for R: see Appendix A, equation (A.15).

The inverse of the corresponding Cramer-Rao bound for the slope coefficients is given by

$$(3.15) \qquad (J^{11})^{-1} = \operatorname*{plim}_{T \to \infty} \tilde{D}'(I_M \otimes Z')(\Sigma^{-1} \otimes I_T)(I_M \otimes Z)\tilde{D}$$

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

$$- S(S'R(\Sigma^{-1} \otimes \Sigma^{-1})R'S)^{-1}S']R(\Sigma^{-1} \otimes I_M)\tilde{B}'$$

where $\tilde{D} = \operatorname{diag}(D_1, \ldots, D_M)$, and $D_i = [\Pi_i I_i]$, where I_i is a selection matrix which choses the explanatory variables for the i'th structural equation: $ZI_i = Z_i$: see Appendix A, equation (A.14). The first term in equation (3.15) is the inverse of the covariance matrix for the 3SLS estimates of the slope coefficients. Since the second term can be shown to be positive semi-definite, 3SLS is asymptotically inefficient relative to FIML, in the presence of covariance restrictions.

For a diagonal covariance matrix, these expressions simply further. If E'^* is the $M \times M^2$ matrix given by $[E_{11} E_{22} \cdots E_{MM}]$ then the information matrix reduces to

$$J(\delta, \sigma_{11} \cdots \sigma_{MM}) = \begin{bmatrix} \tilde{B}E\tilde{B}' + \operatorname*{plim}_{T \to \infty} X'(\Sigma^{-1} \otimes I_T)X & \tilde{B}E^*\Sigma^{-1} \\ \Sigma^{-1}E'^*\tilde{B}' & \frac{1}{2}\Sigma^{-1}\Sigma^{-1} \end{bmatrix}.$$

Similarly, the Cramer-Rao bound for the slope coefficients is given by

$$(J^{11})^{-1} = \tilde{B}(E + \Sigma^{-1} \otimes \Sigma - 2E^*E'^*)\tilde{B}'$$

$$+ \operatorname*{plim}_{T \to \infty} \frac{1}{T}\tilde{D}'(\Sigma^{-1} \otimes Z'Z)\tilde{D}.$$

We now wish to compare these results with the limiting covariance matrix of our A3SLS instrumental variables estimator.

4. Instrumental Variables in the M-Equation Case

To analyze the limiting distribution of the A3SLS estimator in the general case, it will be helpful to rewrite the FIML estimator from equations (3.13) in an instrumental variables form, following the derivation in Hausman (1975) for the unrestricted Σ case. Since we have one equation in equation (3.13) for each unknown δ, we take each equation from the first block of the gradient (3.13) corresponding to the system of equations (3.1), impose the normalization $\beta_{ii} = 1$ and the exclusion restrictions on the slope parameters, and stack the results in the form of equation (3.5). We then have

$$(4.1) \qquad \tilde{X}'(\tilde{\Sigma} \times I)^{-1}(y - \tilde{X}\tilde{\delta}) = 0$$

so

$$(4.2) \qquad \tilde{\delta} = (\tilde{X}'(\tilde{\Sigma} \times I)^{-1}X)^{-1}(\tilde{X}'(\tilde{\Sigma} \times I)^{-1})y$$

$$= (W_F'X)^{-1}W_F'y$$

where $\tilde{X} = \mathrm{diag}\,(\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_M)$, $\tilde{X}_i = \left[Z(\widetilde{\Gamma B})_i^{-1} + (\tilde{V})_i, Z_i \right]$, $\tilde{\Sigma}$ has its known elements set to zero and its unknown elements calculated by $\sigma_{ij} = (1/T)(y_i - X_i\tilde{\delta}_i)'(y_j - X_j\tilde{\delta}_j)$. The instruments are given by

$$(4.3) \qquad W_F' = \tilde{X}'(\tilde{\Sigma} \otimes I)^{-1}.$$

Since the instruments depend on elements of δ, the resulting problem is nonlinear and needs to be solved by iterative methods. While we could iterate on equation (4.2) and see if covergence occured, better methods exist: see Hausman (1974).

The ordinary 3SLS estimator

$$(4.4) \qquad \tilde{\delta}_3 = (X'[\tilde{\Sigma} \otimes P_z]^{-1} X)^{-1} X'[\tilde{\Sigma} \otimes P_z]^{-1} y$$

can be written in the form of equation (4.2), where the instruments are

$$(4.5) \qquad W_3' = X'[I \otimes P_z][\tilde{\Sigma} \otimes I]^{-1}$$

and $P_z = Z(Z'Z)^{-1} Z'$. Compare equations (4.5) and (4.3) and note that FIML and 3SLS differ in their prediction of the explanatory variables $(\tilde{X}')$ in forming their instruments. Ordinary 3SLS projects X onto the space spanned by the exogenous variables Z, whereas FIML replaces Y_i in equation j by $Z(\Gamma B)_i^{-1} + \tilde{V}_{ij}$.

The system A3SLS estimator differs from the ordinary 3SLS estimator in the set of variables which it takes to be uncorrelated with the structural disturbance in each equation. For 3SLS, these are simply the exogenous variables Z and they are the same for all M structural equations. For A3SLS, these instruments are augmented for each equation by the residuals which are assumed uncorrelated with the given disturbance by the covariance restrictions. The set of predetermined

variables thus differs across structural equations for the A3SLS estimator.

Residuals can be added to the list of instruments by allowing $\hat{W} = [\; \tilde{Z}, (I_M \; x \; \hat{U})\tilde{S}]$ to be the matrix of instrumental variables, where U is a T x M matrix of estimated residuals and $\tilde{S}$ is a selection matrix. For example, in the two equation example of Section 2, in order to use $\hat{u}_2$ as an instrument for equation (2.1), let $\tilde{S} = (0,1,0,0)'$ so that

$$W = \begin{bmatrix} Z & 0 & \hat{u}_2 \\ 0 & Z & 0 \end{bmatrix}.$$

We will assume that $\tilde{S}$ is an M^2 x L matrix, where each column correspond to a distinct covariance restriction $\sigma_{ij} = 0$ for some $i \neq j$. This column of $\tilde{S}$ will either select $\hat{u}_i$ as an instrument for equation j or $\hat{u}_j$ as an instrument for equation i.

Instrumental variables estimators which use residuals can be obtained from

(4.6) $\qquad \hat{\delta}_A = (W_A'X)^{-1} W_A'y,$

where $W_A = WA_T$ is the Txq matrix of instruments, $W_A'X$ is nonsingular, A_T is a (MK + L) x q linear combination matrix, and q is the dimension of δ. Depending on the choice of A_T, many different system instrumental variables estimators can be obtained from equation (4.6). For example, if $A_T' = [X(I_M \otimes Z(Z'Z)^{-1}),0]$ where 0 is a qxL matrix of zeros, then $\hat{\delta}_A$ is the vector of 2SLS estimates.

The A3SLS estimator will be $\hat{\delta}_A$ with A_T chosen so as to minimize the asymptotic covariance matrix of $\hat{\delta}_A$.

The asymptotic distribution theory for $\hat{\delta}_A$ is complicated by the fact that W contains residuals. Following the usual instrumental variables analysis, we substitute $y = X\delta + u$ into equation (4.7) to obtain

(4.7)
$$\sqrt{T}(\hat{\delta}_A - \delta) = (W_A'X/T)^{-1} W_A'u/\sqrt{T}$$
$$= (A_T'\hat{W}'X/T)^{-1} A_T'\hat{W}'u/\sqrt{T}$$

To use equation (4.7) to obtain the asymptotic distribution of δ_A, it is useful to be able to apply a central limit theorem to $W'u/\sqrt{T}$, which is the vector of $\sqrt{T}$ normalized sums of cross-products of instrumental variables and disturbances . To use a central limit theorem the presence of residuals needs to be accounted for, which requires us to be specific concerning the way in which U was obtained. We will assume that each equation is identified by coefficient restrictions alone, so that $\text{rank}(D_i) = \text{rank}([\Pi_i, I_i]) = q_i$, i=1,...,M, where $q_i = r_i + s_i$ is the dimension of δ_i .[8a]

Since $\text{plim}(Z'X_i/T) = ND_i$, (i=1,...,M), for $N = \text{plim}(Z'Z/T)$ nonsingular, we can obtain

(4.8)
$$\text{plim}[\tilde{Z}'X/T] = (I_M \otimes N)\tilde{D}.$$

which has rank q. Let A_{1T} be a MK x q matrix satisfying $\text{plim } A_{1T} = A_1$, where $A_1' \text{plim}(\tilde{Z}'X/T)$ is nonsingular.

[8a]. Estimation when covariance restrictions are necessary for identification is treated in Newey (1983). Included in this treatment is a relatively simple means of obtaining an initial estimator, and the appropriate A3SLS estimator.

Let $\tilde{\delta}$ be an instrumental variables estimator which satisfies

(4.9) $$\tilde{\delta} = (A_{1T}'\tilde{Z}'X)^{-1} A_{1T}'\tilde{Z}'y.$$

We now will assume that the matrix of residuals U is obtained from $\hat{U}_{ti} = y_{it} - X_{it}\tilde{\delta}_i$, (t=1,...,T, i=1,...,M). That is, the residuals used to form the instrument matrix W are obtained from an estimator $\hat{\delta}$ which uses the predetermined variables $\tilde{Z}$ as instruments. For example, U could be obtained from the 2SLS or the 3SLS estimator. This assumption allows us to purge $\tilde{\delta}$ from $\hat{W}'u/\surd T$ as follows. Let

$$M_{2T} = -[\sum_{t=1}^{T} U_t' \otimes (\partial U_t'/\partial \delta)]/T,$$

where U_t is the tth row of U and $\partial U_t'/\partial \delta = -\operatorname{diag}[X_{1t},...,X_{Mt}]$. Then

(4.10) $$\tilde{S}'(I_M \otimes \hat{U})'u = \tilde{S}' \sum_{t=1}^{T} \hat{U}'_t \otimes \hat{U}'_t$$

$$= \tilde{S}'[\sum_{t=1}^{T} U_t' \otimes U_t' - TM_{2T} (\tilde{\delta} - \delta)]$$

$$= \tilde{S}'[(I_M \otimes U)'u - TM_{2T}(A_{1T}'\tilde{Z}'X)^{-1}A_{1T}'\tilde{Z}'u]$$

where the last equality is obtained by substituting $y = X\delta + u$ into equation (4.9). Let P_T and W be denoted by

$$P_T = \left| \begin{array}{cc} I & 0 \\ MK & \\ -T\tilde{S}'M_{2T}(A_{1T}'\tilde{Z}'X)^{-1}A_{1T}' & I_L \end{array} \right| , \quad W = [\tilde{Z},(I_M \otimes U)\tilde{S}].$$

Note that W is obtained from $\hat{W}$ by replacing $\hat{U}$ by the true disturbance matrix U. It follows from equation (4.10) that

(4.11) $\qquad \hat{W}'u/\sqrt{T} = P_T W'u/\sqrt{T},$

so that $\hat{W}'u/\sqrt{T}$ is a nonsingular linear combination of $W'u/\sqrt{T}$, which consists of cross-products of predetermined variables and disturbances. It is now straightforward to use a central limit theorem. Let $e_t = (U_t \otimes U_t)\tilde{S}$ be the 1 x L vector of cross-products of true disturbances corresponding to the zero covariance restrictions. Suppose that, conditional on the 1 x K vector Z_t of contemporaneous predetermined variables , U_t has constant moments up to the fourth order. Then by the orthogonality of disturbances and predetermined variables, the covariance restrictions, and the absence of autocorrelation an appropriate central limit theorem gives

(4.12) $W'u/\sqrt{T} \overset{d}{\to} N(0,V)$

$$V = \text{plim} \frac{1}{T} \sum_{t=1}^{T} E([U_t \otimes Z_t, e_t]'[U_t \otimes Z_t, e_t] \mid Z_t)$$

$$= \left[\begin{array}{cc} \Sigma \otimes N & V_{12} \\ E(e_t'U_t) \otimes \text{plim}(\sum_{t=1}^{T} Z_t/T) & E(e_t'e_t) \end{array} \right]$$

where the last equality follows by $e_t'(U_t \otimes Z_t) = (e_t'U_t) \otimes Z_t$.

If the disturbances are normally distributed then the off diagonal blocks of V are zero, since all third order moments of a joint normal distribution are zero. Let $M_2 = \text{plim } M_{2T}$, so that

$$(4.13) \qquad M_2 = \text{plim } \frac{1}{T} \begin{bmatrix} \text{diag}(u_1'X_1,\ldots,u_1'X_M) \\ \\ \text{diag}(u_M'X_1,\ldots,u_M'X_M) \end{bmatrix} = E(I_M \otimes \Sigma)\tilde{B}'$$

where Σ_i is the ith row of Σ ($i=1,\ldots,M$), $\text{plim}(u_i X_j/T) = \Sigma_i \tilde{B}_j$, and the last equality in equation (4.14) follows from Lemma A1 of Appendix A. Also, let $\text{plim } P_T = P$, so that equations (4.8) and (4.13) we can compute

$$(4.14) \qquad P = \begin{bmatrix} I_{MK} & 0 \\ -\tilde{S}'M_2(A_1'(I_M \otimes N)\tilde{D})^{-1}A_1' & I_L \end{bmatrix}$$

Then by equation (3.7) we have

$$(4.15) \qquad \hat{W}'u/\sqrt{T} \xrightarrow{d} N(0,PVP'),$$

so that PVP' is the asymptotic covariance matrix of $\hat{W}'u/\sqrt{T}$.

The rest of the derivation of the asymptotic distribution of $\hat{\delta}_A$ is straightforward. Our assumptions are sufficient to guarantee that $\text{plim } \tilde{\delta} = \delta$. Then by $\text{plim}(\hat{U}'X_i/T) = \Sigma \tilde{B}_i$ ($i=1,\ldots,M$), and by equation

(4.8),

$$(4.16) \qquad \text{plim } \hat{W}'X/T = \text{plim } \begin{bmatrix} (I_M \otimes N) \tilde{D} \\ \tilde{S}'(I_M \otimes \Sigma) \tilde{B}' \end{bmatrix}.$$

Let $\overline{G}$ denote $\overline{G} = \text{plim } \hat{W}'X/T$. We will assume that plim $A_T = A$ and that $A'\overline{G}$ is nonsingular. By equations (4.7), (4.15), and (4.16) we obtain

$$(4.17) \qquad \sqrt{T}(\hat{\delta}_A - \delta) \xrightarrow{d} N(0, (A'\overline{G})^{-1} A'PVP'A(\overline{G}'A)^{-1}).$$

The A3SLS estimator is obtained by choosing an optimal linear combination matrix A. We will assume that the covariance matrix V is nonsingular. In the normal disturbances case, the nonsingularity of V follows from previous assumptions: see Lemma 5.1 below . Note also that P is nonsingular. Since the asymptotic covariance matrix of $W'u/\sqrt{T}$ is PVP', the linear combination matrix A* which minimize the asymptotic covariance of the instrumental variables estimator δ_A (e.g., see White (1982) satisfies

$$(4.18) \qquad A^* = (PVP')^{-1}\overline{G} = (P^{-1})'V^{-1}P^{-1}\overline{G}.$$

The asymptotic covariance matrix of δ_A will be

$$(4.19) \qquad \text{Var}(\delta_{A^*}) = (\overline{G}'(P^{-1})'V^{-1}P^{-1}\overline{G})^{-1}.$$

Since $W'X/T$ is a natural estimator of its probability limit $\overline{G}$, we

only need to consider consistent estimators of P and V to obtain an estimate A_T^* satisfying plim $A_T^* = A^*$, which is required for implementation of A3SLS. We will assume that sample averages of up to fourth order cross-products of elements of $X_t = [Y_t, Z_t](t=1,\ldots,T)$ have finite probability limits. Then by plim $\tilde{\delta} = \delta$, a consistent estimator of $\hat{V}_T$ of V can be obtained as follows. Let $\hat{e}_t = (\hat{U}_t \otimes \hat{U}_t)\tilde{S}$ and

$$\hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} [\hat{U}_t, \hat{e}_t]'[\hat{U}_t, \hat{e}_t] = \begin{bmatrix} \Sigma & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}.$$

Then we can take

$$(4.20) \qquad V_T = \begin{bmatrix} \hat{\Sigma} \ \times \ (Z'Z/T) & \hat{V}_{12T} \\ \hat{\Omega}_{21} \ \times \ (\sum_{t=1}^{T} Z_t/T) & \hat{\Omega}_{22} \end{bmatrix}.$$

A consistent estimator $\hat{P}_T$ of P can be obtained by replacing M_{2T} in the definition of P_T by an estimate $\hat{M}_{2T}$ satisfying plim $\hat{M}_{2T} = M_2$. By plim $\tilde{\delta} = \delta$ such an estimator is given by

$$\hat{M}_{2T} = E(I_M \ \times \ \hat{U})'X/T$$

so that we can let

$$(4.21) \qquad P_T = \begin{bmatrix} I_{MK} & 0 \\ -\tilde{S}' \hat{M}_{2T}(A_{1T}'\tilde{Z}'X/T)^{-1} A_{1T}' & I_L \end{bmatrix}$$

Then we can take $A_T^* = (P_T^{-1})'V_T^{-1} P_{\bar{T}}1W'X/T$, so that the A3SLS estimator can be obtained from equation (4.6),

$$(4.22) \qquad \hat{\delta}_{A*} = (X'\hat{W}(\hat{P}_T^{-1})'\hat{V}_T^{-1}\hat{P}_T^{-1}\hat{W}'X)^{-1} X'\hat{W}(\hat{P}_T^{-1})'\hat{V}_T^{-1}\hat{P}_T^{-1}\hat{W}'y.$$

A consistent estimator of the asymptotic covariance matrix of δ_{A*} is also given by

$$(4.23) \qquad \widehat{Var(\hat{\delta}_{A*})} = [(X'\hat{W}/T)(\hat{P}_T^{-1})'\hat{V}_T^{-1}\hat{P}_T^{-1}(\hat{W}'X/T)]^{-1}.$$

Computation of the A3SLS estimator (and estimating its covariance matrix) is only a little more laborious than computation of 3SLS except for the presence of $\hat{P}_T^{-1}$ in equation (4.22). Note though that because P_T is block triangular,

$$(4.24) \qquad \hat{P}_T^{-1} = \begin{bmatrix} I_{MK} & 0 \\ \tilde{S}'M_{2T}(A_{1T}'\tilde{Z}'X/T)^{-1} A_{1T}' & I_L \end{bmatrix}$$

so that $\hat{P}_T$ need not be inverted numerically. Also, M_{2T} consists mostly of zeros, and if the estimator $\hat{\delta}$ used to form the instrumental variable residuals U is 2SLS, then both $(A_{1T}'\tilde{Z}X/T)^{-1}$ and A_{1T}' will be block diagonal.[9]

---

[9]. It may appear that FIML is more easy to compute than A3SLS. However, A3SLS is a linear estimator (once U is chosen) so that no iterative process is needed to obtain A3SLS. FIML is also complicated by the fact with covariance restrictions, it is more difficult to concentrate out the covariance matrix parameters.

It remains for us to demonstrate that for the general M-equation case, if the disturbances are normally distributed then A3SLS is asymptotically efficient. To obtain this result, it is convenient to simplify the asymptotic covariance matrix for A3SLS. Let the $(MK + L)$ x q matrix G be denoted by $G = P^{-1}\overline{G}$, so that the asymptotic covariance matrix of A3SLS is $(G'V^{-1}G)^{-1}$. From equations (4.8) and (4.10) it follows that,

$$(4.25) \qquad G = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} = \begin{bmatrix} (I_M \otimes N)\widetilde{D} \\ \widetilde{S}'(E + I_{M^2})(I_M \otimes \Sigma)\widetilde{B}' \end{bmatrix}$$

To prove efficiency of A3SLS we can compute $G'V^{-1}G$ for the normal disturbance case and show that this matrix is equal to

$J_{\delta\delta'} - J_{\delta\widetilde{\sigma}'}(J_{\widetilde{\sigma}\widetilde{\sigma}'})^{-1}J_{\widetilde{\sigma}\delta'}$, where $\widetilde{\sigma} = \sigma^*S$ is the $(1/2)M(M + 1)-L$ vector of

distinct, unrestricted elements of Σ. Since the derivation of this equality is tedious, we relegate it to Appendix B.

Lemma 4.1: If $N = \text{plim}(Z'Z/T)$ and $\Sigma = E(U_t'U_t)$ are nonsingular and U has a joint normal distribution then V and $J_{\widetilde{\sigma}\widetilde{\sigma}'}$ are nonsingular and

$$G'V^{-1}G = J_{\delta\delta'} - J_{\delta\widetilde{\sigma}'}(J_{\widetilde{\sigma}\widetilde{\sigma}'})^{-1}J_{\widetilde{\sigma}\delta'}.$$

The reason that A3SLS is efficient in the general case is the same as for the example of Section 2. Using computations similar to those of Section

2 (see equations (2.24) and (2.25)) we can show that there is a nonsingular matrix H such that

(4.26) $\qquad W'_{A*}u/\sqrt{T} = HW'_F u/\sqrt{T} + o_p(1), \quad W'_{A*}X/T = HW'_F X/T + o_p(1)$

so that, asymptotically, the A3SLS instruments are a nonsingular linear transformation of the FIML instruments.

Lemma 4.1 also has implications for the question of identification of the system of equations using covariance restriction. When the hypothesis of Lemma 4.1 are satisfied, $J_{\tilde\sigma\tilde\sigma}$, is nonsingular so that the information matrix J is nonsingular if and only if the q dimensional square matrix $J_{\delta\delta'} - J_{\delta\tilde\sigma'}(J_{\tilde\sigma\tilde\sigma'})^{-1}J_{\tilde\sigma\delta'}$ is nonsingular. Since V^{-1} is also nonsingular, it follows that J is nonsingular if and only if G has rank q. Therefore, local identification of the parameters of a system of simultaneous equations subject to covariance restrictions is related to the matrix G . In the next section we will derive necessary and sufficient conditions for local identification by studying the properties of G.

5.   Identification

During the discussion of A3SLS we assumed that the coefficient
vector δ of the system of equations

$$y = X \delta + u$$

is identified by coefficient restrictions alone. It is well known,
though, 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 a necessary condition for first order identification is that
there must exist an assignment of residuals as instruments which has the
property that for each equation the matrix of cross-products of
instrumental variables and right-hand side variables has rank equal to
the number of coefficients to be estimated in the equation.

Lemma 4.1 implies that a necessary and sufficient condition for
nonsingularity of the information matrix J is that the matrix G have rank
q, which is also equivalent to the condition that the Jacobian, with
respect to unknown parameters, of the equation system

(5.1) $\Pi B + \Gamma = 0, \quad \Sigma - B'\Omega B = 0$

has full rank, when Π and Ω are taken to be fixed constants and the parameter restrictions are imposed. The matrix $G_1 = \text{plim}(\tilde{Z}'X/T)$ is familiar from the analysis of identification via coefficient restrictions. The matrix G_2 has an interesting structure. The ith 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 $\Sigma_j(B^{-1})_i = \text{plim}(u_j'X_i/T)$ appears, and the elements corresponding to δ_j, where $\Sigma_i(B^{-1})_j = \text{plim}(u_i'X_j/T)$ appears. For example, in a three equation simultaneous equation system, where the ℓth row of G_2, which we will denote by $(G_2)_\ell$, corresponds to $\sigma_{13} = 0$, we have

$$(G_2)_\ell = [\text{plim}(u_3'X_1/T), \; 0, \; \text{plim}(u_1'X_3/T)].$$

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 to <u>either</u> assign u_i as an instrument for equation

j $\underline{or}$ assign u_j as a residual for equation i, but $\underline{not}$ both. We can think of an assignment of residuals as instruments to equations as an L-tuple $a_p = (a_{p1},\ldots,a_{pL})$, where p indexes the assignment and each element $a_{p\ell}$ of a_p corresponds to a unique restriction σ_{ij}, with $a_{p\ell} = i$ if u_j is assigned to equation i and $a_{p\ell} = j$ if u_i is assigned to equation j.

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, p, of residuals as instruments 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 and $C_{pi} = \text{plim}(W'_{pi}X_i/T)$ be the population cross-product matrix of instrumental variables and right-hand side variables for equation i, i=1,...,M. The following necessary condition for identification is proved in Appendix B.

Theorem 5.1: $\underline{\text{If } N = \text{plim}(Z'Z/T) \text{ and } \Sigma \text{ are nonsingular, then if the}}$ $\underline{\text{information matrix is nonsingular there exists an assignment p* such}}$ $\underline{\text{that}}$

(5.2) $\qquad \text{rank}(C_{p*i}) = q_i, \quad i=1,\ldots,M.$

Theorem 5.1 says that a necessary condition for first-order local identification is that after assigning residuals as instruments the cross-product matrix of instrumental variables and right-hand side variables has rank equal to the number of coefficients in the 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 the following order

condition from Theorem 5.1. Let $a_i = \max(0, q_i - K)$ $(i=1,\ldots,M)$.

Corollary 5.2: <u>If Σ and N are nonsingular then J nonsingular implies</u> <u>that there exists an assignment of residuals as instruments such that at</u> <u>least a_i residuals are assigned to equation i, for i=1,...,M.</u>

For the ith equation, a_i is the number of instrumental variables which are required for estimation, in addition to the predetermined variables Z. Therefore, Corollary 5.2 says that first-order local identification implies the existence of an assignment of residuals such that there are enough instruments for each equation. Following Geraci's (1976) analysis of identification of a simultaneous equation system with measurement error we can obtain an algorithm for determining whether or not such an assignment exists. Let R^i be the set of indices ℓ such that $\sigma_{i\ell} = 0$ for some ℓ, for all i=1,...,M. For $a_i > 0$, let $R^i_1,\ldots,R^i_{a_i}$ be a_i copies of R^i.

Let R be the $\sum\limits_{i=1}^{M} a_i$ tuple with components equal to R^i_j for (j=1,...a_i, i=1,...M).

Theorem 5.3: <u>There exists an assignment of residuals as instruments such</u> <u>that for each i=1,...,M at least a_i residuals are assigned as instruments</u> <u>to each equation if and only if for each n=1,..., $\sum\limits_{i=1}^{M} a_i$ the union of any</u> <u>n components of R contains at least n distinct indexes ℓ.</u>

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 that when only coefficient restrictions are present the condition that $\mathrm{plim}(Z'X_i/T)$ have rank $q_i, i=1,\ldots,M$ is easily translated into a more transparent condition on the structural parameters $A = [B',\Gamma']'$. We can also state an equivalent rank condition to $\mathrm{plim}(W'_{pi}X_iT)$ having rank q, when covariance restrictions are present. For an assignment p, let $\widetilde{\Sigma}_{pi}$ be the rows of Σ corresponding for residuals which are assigned as instruments to the ith equation, $i=1,\ldots,M$. Let ϕ_i be the $(M-1-q_i) \times MK$ 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 5.4: <u>For a particular assignment p and an equation i, the rank of equation C_{pi} equals q_i if and only if rank $[A'\phi_i', \widetilde{\Sigma}_{pi}'] = M-1$.</u>

We prove this result in Appendix B. Together with Lemma 5.1, Lemma 5.4 implies the following necessary rank condition for identification of a linear simultaneous equations system subject to covariance restrictions.

Theorem 5.5: <u>If Σ and N are nonsingular, then nonsingularity of J implies that there exists an assignment, p*, of residuals such that</u>

(5.3) $\operatorname{rank}\left[A'\phi_i', \widetilde{\Sigma}_{p*i}'\right] = M - 1, \quad (i=1,\ldots,M).$

This rank condition is a strengthening of Fisher (1966). Fisher shows that if a system of equations is first-order locally identified then

(5.4) $\operatorname{rank}\left[A'\phi_i', \widetilde{\Sigma}_i'\right] = M - 1,$

where $\widetilde{\Sigma}_i$ is the matrix of <u>all</u> rows $(\Sigma)_k$ of Σ such that $\sigma_{ik} = 0$. Theorem 5.5 strengthens this condition by requiring that equation (5.3) only hold for those rows of $\widetilde{\Sigma}$ corresponding to residuals which are assigned to equation i.

It is sufficient for first-order local identification that the rank of G equals q. It would be useful to have other sufficient conditions for local identification which are more readily interpretable in terms of the structural parameters. We can obtain a sufficiency result which is the system analog of Lemma 5.4

Theorem 5.6: <u>The rank of G equals q if and only if</u>

(5.5) $\operatorname{rank}\left(\left[\operatorname{diag}(\phi_1, \ldots, \phi_M, \widetilde{S}')\right]\bullet\left[I_M \times A', (I_M \otimes \Sigma)(E+I_{M^2})\right]'\right) = M^2 - M.$

6. Testing the Overidentifying Covariance Restrictions

Since covariance restrictions specify that the distribution of unobservables has certain properties, such restrictions may have weaker a priori justification than coefficient restrictions, so that it is useful to have available a test for the validity of overidentifying covariance restrictions. The case we considered for the A3SLS estimator the simultaneous system was identified in the absence of covariance restrictions, so that these restrictions only help in obtaining a more efficient A3SLS estimator of the structural parameters δ. If the restrictions are false then the A3SLS estimator will not be consistent. We can use these facts to form a Hausman test based on the difference of the 3SLS and A3SLS estimators. Consider the test statistic

$$(6.1) \quad m = T(\tilde{\delta}_{A*} - \tilde{\delta}_{3SLS})'[Var(\tilde{\delta}_{3SLS}) - Var(\tilde{\delta}_{A*})]^{-}(\tilde{\delta}_{A*} - \tilde{\delta}_{3SLS}),$$

where A^{-} denotes a generalized inverse of a matrix A. Under the null hypothesis that the covariance restrictions are true this test statistic will have an asymptotic chi-squared distribution. Except in an exceptional case, where adding an additional covariance restriction $\sigma_{ij} = 0$ does not improve the efficiency of enough components of $\tilde{\delta}$, the degrees of freedom of this test will be $\min(q,L)$. The case $L < q$ is of most practical interest, since it seems unlikely that more covariance restrictions than structural parameters will be available (e.g., see the example of Section 2). When m has degrees of freedom L and the disturbances are distributed normally, then because 3SLS and A3SLS are

asymptotically equivalent to FIML without and with covariance restrictions, respectively, m will be asymptotically equivalent to the classical tests; see Holly (1982). When the disturbances are not normally distributed this test will have the optimality properties discussed in Newey (1983).

The statistic m can be computed by forming a Hausman (1978) test on a subset of L coefficients. This method corresponds to a particular choice of generalized inverse of the difference of variance matrices in equation (6.1), but the test statistic will be numerically invariant to such a choice of g-inverse as long as the same estimator $\hat{\Sigma}$ of Σ is used throughout; see Newey (1983). Thus, the specification test proposed here is asymptotically equivalent under the null hypothesis and local alternatives to the Wald and LM tests which seem more difficult to compute. Since it is likely that both the A3SLS and 3SLS estimators would both be computed in an applied situation, comparison of the estimates provides a convenient test of the underlying covariance restriction assumptions.

APPENDIX A

We will first enumerate some properties of the matrix E which will be useful when obtaining the information matrix. Let A and B be M-dimensional square matrices. Let A_i be the ith row of A, (i=1,...,M). Let Q be a matrix obtained from R by $Q_i = R_i$ for each row of R corresponding to $\sigma_{kk} = 0$ and $Q_i = (1/2) R_i$ for each row corresponding to $\sigma_{k\ell}$, $k \neq \ell$

Lemma A1: <u>The matrix E satisfies</u>

(i) $\qquad\qquad E' = E,$

(ii) $\qquad\qquad E^2 = I_{M^2},$

(iii) $\qquad\qquad E(A \otimes B)E = B \otimes A,$

(iv) $\qquad\qquad E(I_M \otimes A) = \begin{bmatrix} I_M \otimes A_1 \\ \vdots \\ I_M \otimes A_M \end{bmatrix},$

(v) $\qquad\qquad (1/2)(E + I_{M^2}) = R'Q,$

(vi) $\qquad\qquad ER' = R', \quad EQ' = Q',$

(vii) $\qquad$ <u>For A nonsingular,</u> $\partial^2 \ln|det A| / \partial vec A\, \partial(vec A)' =$
$$- (I_M \otimes A^{-1'}) E (I \otimes A^{-1})$$

Proof: (i): Follows from $E_{ij}' = E_{ji}$. (ii): The ijth block of E^2 equals

$\sum_m E_{mi}E_{jm}$. Also, $E_{mi}E_{jm} = 0$ for $i \neq j$ and $E_{mi}E_{im} = E_{mm}$ so that $\sum_m E_{mi}E_{jm} = 0$ for

$i \neq j$ and $\sum_m E_{mi}E_{im} = \sum_m E_{mm} = I_m$. (iii): The ijth block of $E(A \times B)E$ is

$\sum_{mn} a_{mn} E_{mi}BE_{jn} = b_{ij}\sum_{mn} a_{mn} E_{mn} = b_{ij}A$. (iv): The ijth block of $E(I_M \times A)$ is

$E_{ji}A$, which is a matrix of zeros, except for the jth row, where A_i

appears. (v): Number the rows of R' by $h(M-1) + i$, h, i=1,...M. Also,

number the columns of Q by $j(M-1) + k, j, k = 1, ...M$. Note that the h(M-1)

+ i and i (M-1) + h rows of R' are identical, because these rows each

give σ_{ih} in the equality $\sigma = R'\sigma_*$. Similiarly the j(M-1) +k and k(M-1)

+j columns of Q are identical. Also note that the h(M-1)+i row of R' is

all zeros except in ecxect for a one in the place which selects

σ_{ih} from σ_*. The j(M-1)+k column of Q is all zeros except for a one

(one-half) in the place which selects σ_{jk} from σ_*, for $j = k(j \neq k)$. It

follows that the i(M-1)+i row of R'Q has a one in the i(M-1)+i place and

zeros elsewhere, and that the h(M-1)+i row of R'Q, for $h \neq i$ has one-half

in the h(M-1)+i element and the i(M-1)+h elements and zeros elsewhere.

Then the h(M-1)+i row of $R'Q-(1/2)I_{M^2}$ has a 1/2 in the i(M-1)+h place

and zeros elsewhere. Consider the i(M-1)+h of (1/2)E, which will be

the hth row of $(1/2)[E_{1i}, E_{2i}, \ldots, E_{Mi}]$, and will thus have zeros

elsewhere but in the h(M-1) +i position, where a 1/2 will appear. (vi)

It is known that $QR' = I_{M(M+1)/2}$: see Richard (1975). Then ER' = ER'QR' =

$E(E + I)R'/2 = (I + E)R'/2$ which implies ER' = R'. The proof of EQ' = Q'

is similar. (vii): Suppose detA < 0. By Theil (1971), $\partial \ln |\det A| / \partial$

$a_{ij} = \partial \ln(-\det A)/\partial a_{ij} = (-1/\det A)[\partial(-\det A)/\partial a_{ij}] = a^{ji}$, where $A^{-1} = [a^{ij}]$,

80

that $\partial^2 \ln |\det A|,/\partial a_{ij}\partial a_{hk} = a^{jh}a^{ki}$.

Consider the j, kth block of $-(I_M \otimes A^{-1})'E(I_M \otimes A^{-1})$, which is $-A^{-1'}E_{kj}A^{-1}$. The i,h th element of this matrix is $-(A_i^{-1})'E_{kj}A_h^{-1} = -a^{ki}a^{jh}$, where A_ℓ^{-1} is the ℓth column of A^{-1}. Since this order is the same used to form $\text{vec}A=(a_{11}, \ldots, a_{M1}, a_{12}, \ldots, a_{M2} \ldots, a_{MM})'$ (vii) follows.

The information matrix is given by

$$(A.1) \quad J(\delta,\sigma) = \text{plim}-\frac{1}{T} \begin{bmatrix} \partial^2 L_T/\partial\delta\partial\delta' & \partial^2 L_T\partial\delta\partial\sigma' \\ \partial^2 L_T'\partial\sigma\partial\delta' & \partial^2 L_T\partial\sigma\partial\sigma' \end{bmatrix}$$

where H is the Hession matrix of the log likelihood function L_T given in equation (3.7) and σ is a vector of elements of Σ. We derive the information matrix by ignoring symmetry of Σ when taking derivatives of L_T, followed by accounting for symmetry of Σ by transforming $J(\delta,\sigma)$, as in Richard (1975).

By matrix differentiation

$$(A.2) \quad \frac{\partial^2 (y - X\delta)'(\Sigma^{-1} \otimes I_T)(y - X\delta)}{\partial\delta\partial\delta'} = X'(\Sigma^{-1} \otimes I_T) \; X.$$

Using the exclusion restrictions and Lemma A1(vii), we compute

$$(A.3) \quad \partial^2 \ln |\det \bar{B}|/\partial\delta\partial\delta' = \tilde{B} E \tilde{B}',$$

so that

$$(A.4) \quad J_{\delta\delta'} = \tilde{B} E \tilde{B}' + \text{plim} \left[X'(\Sigma^{-1} \otimes I_T) X \right]/T.$$

Lemma A1(vii) is also useful in obtaining $J\sigma\sigma'$, for $\sigma = \text{vec } \Sigma$, since

$$(A.5) \quad \partial^2 \ln \det \Sigma / \partial\sigma\partial\sigma' = -(I_M \otimes \Sigma^{-1}) E (I_M \otimes \Sigma^{-1}).$$

We can use the result that for a non-singular matrix A,

$$(A.6) \quad \partial A^{-1} / \partial a_{ij} = - A^{-1} E_{ij} A^{-1}.$$

and matrix differentiation to obtain

$$(A.7) \quad \partial^2 u' (\Sigma^{-1} \otimes I_T) u/\partial\sigma_{\ell m}\partial\sigma_{hk} = u'(\Sigma^{-1} E_{\ell m}\Sigma^{-1} E_{hk}\Sigma^{-1} \otimes I_T) u$$

$$+ u' (\Sigma^{-1} E_{hk} \Sigma^{-1} E_{\ell m} \Sigma^{-1} \otimes I_T) u$$

$$= \text{tr}(\Sigma^{-1} (E_{\ell m} \Sigma^{-1} E_{hk} + E_{hk} \Sigma^{-1} E_{\ell m})\Sigma^{-1} U'U)$$

where the last equality follows by

$$u'(A \otimes I_T) u = \text{tr}(A'U'U').$$

for any M dimensional square matrix A. This equation and $\text{plim}(U'U/T) =$ imply

(A.8) $\quad \text{plim} - \dfrac{1}{T} \dfrac{\partial^2}{\partial\sigma_{1m}\partial\sigma_{hk}} \left[\dfrac{1}{2} u'(\Sigma^{-1} \otimes I_T) u \right]$

$$(1/2) \left[\text{tr}(\Sigma^{-1}E_{\ell m} \Sigma^{-1} E_{hk}) + \text{tr}(\Sigma^{-1}E_{hk} \Sigma^{-1}E_{\ell m}) \right] = \sigma^{k\ell}\sigma^{hm}, \text{ so that}$$

(A.9) $\quad \text{plim} - \dfrac{1}{T} \dfrac{\partial^2}{\partial\tilde\sigma\partial\tilde\sigma'} \left[\dfrac{1}{2} u'(\Sigma^{-1} \otimes I_T) u \right] = (I_M \otimes \Sigma^{-1}) E (I_M \otimes \Sigma^{-1})$

Equations (A.9) and (A.5) imply

(A.10) $J_{\tilde\sigma \tilde\sigma'} = (1/2) (I_M \otimes \Sigma^{-1}) E (I_M \otimes \Sigma^{-1}).$

To obtain $J_{\delta\tilde\sigma'}$, we again use equation (A.6) to compute

(A.11) $\partial^2 L_T / \partial\delta\partial\sigma_{hk} = - (1/2) X' \left[(\Sigma^{-1}E_{hk} \Sigma^{-1} + \Sigma^{-1}E_{kh} \Sigma^{-1} \otimes I_T \right] u$

from which we compute

(A.12) $J_{\delta\sigma'} = (1/2) \tilde B(\Sigma^{-1} \otimes I_T) (E + I_{M^2})$

To complete the derivation of the information matrix, we now incorporate the symmetry restrictions $\sigma_{ij} = \sigma_{ji}$, $i \neq j$. Let $\sigma^* = (\sigma_{11} \cdots, \sigma_{M1}, \sigma_{22}, \cdots \sigma_{M2}, \cdots \sigma_{MM})$. By Lemma A1 and $ER' = R'$ for the matrix which

has zeros and ones and satisfies vec Σ = $R'\sigma^{*'}$, the information matrix is

$$(A.13) \quad J(\delta,\sigma^*) = \begin{bmatrix} \tilde{B} \ E \ \tilde{B}' + \text{plim} \ X'(\Sigma^{-1} \otimes I_T) \ X/T & \dot{\tilde{B}}(\Sigma^{-1} \otimes I_M)R' \\ J_{21} & (1/2)R(\Sigma^{-1} \otimes \Sigma^{-1})R' \end{bmatrix}$$

For the case of covariance restrictions, the information matrix is obtained from equation (A.13) by deleting the rows and columns of $J(\delta,\sigma^*)$ corresponding to those covariances restricted to be zero. If L covariances are restricted to be zero, let S be the (1/2) M (M+1) x (1/2) M (M+1) - L selection matrix for which $\tilde{\sigma}$ S is the vector of unrestricted elements of $\tilde{\sigma}$. For example, for M=2 with the restriction σ_{12}= 0 imposed we have

$$S = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

The information matrix with covariance restrictions is then

$$(A.14) \quad J(\delta, \sigma^{**}) = \begin{bmatrix} \tilde{B} \ E \ \tilde{B} + \text{plim} \ X'(\Sigma^{-1} \otimes I_T) \ X/T \ \tilde{B}(\Sigma^{-1} \otimes I_M)R'S \\ (1/2)S'R(\Sigma^{-1} \otimes \Sigma^{-1})R'S \end{bmatrix}$$

To compute the Cramer-Rao lower bound, we use the fact that

(A.15) $\left(R(\Sigma^{-1} \otimes \Sigma^{-1})R' \right)^{-1} = Q(\Sigma \otimes \Sigma) \, Q'$

Lemma A1(iii), (v), (vi) yields

(A.16) $\quad \tilde{B} \, (\Sigma^{-1} \otimes I_M) \, 2R'Q \, (\Sigma \otimes \Sigma) \, Q'R(\Sigma^{-1} \otimes I_M) \, \tilde{B}'$

$\qquad\qquad = \tilde{B} \, E \, \tilde{B}' + \tilde{B} \, (\Sigma^{-1} \otimes \Sigma) \, \tilde{B}'$

We can also compute

(A.17) $\quad \text{plim } X'(\Sigma^{-1} \otimes I_T) \, X = \tilde{D}' \, (\Sigma^{-1} \otimes N) \, \tilde{D} + \tilde{B}' \, (\Sigma^{-1} \otimes \Sigma) \, \tilde{B}.$

where $N = \text{plim } Z'Z/T$. Subtracting and adding equation (A.16), we can obtain the inverse of the Cramer-Rao lower bound

(A.18) $(J^{11})^{-1} = \tilde{B} \, E \, \tilde{B}' + \text{plim } X'(\Sigma^{-1} \otimes I_T) \, X/T$

$\qquad\qquad -2\tilde{B}(\Sigma^{-1} \otimes I_M)R'S[\, S'R(\Sigma^{-1} \otimes \Sigma^{-1})R'S\,]^{-1}S'R(\Sigma^{-1} \otimes I_M)\tilde{B},$

$\qquad\qquad = \tilde{D}'(\Sigma^{-1} \otimes D)\tilde{D}$

$\qquad\qquad +2\tilde{B}(\Sigma^{-1} \otimes I_M)R'[\, F^{-1} - S(S'FS)^{-1}S'\,]R(\Sigma^{-1} \otimes I_M)\tilde{B}$

where $\qquad\qquad F = R(\Sigma^{-1} \otimes \Sigma^{-1})R'.$

APPENDIX B

Proof of Lemma 4.1: For normally distributed disturbances V_{12} and V_{21} are each zero matrices. By Σ and N nonsingular, $V_{11} = \Sigma \otimes N$ is also nonsingular. Let R and Q be the M^2 x $(1/2)M(M+1)$ matrices defined in Appendix A. Let $\tilde{U}_t = (U_{t1}^2, U_{t1} \bullet U_{t2}, \cdots, U_{tM} \bullet U_{t1}, U_{t2}^2, \cdots, U_{tM}^2)'$ be the $(1/2)M(M+1)$ x 1 vector of distinct products of disturbances, so that

$$U_t' \text{ x } U_t' = R' \tilde{U}_t.$$

Let $\bar{S}' = \tilde{S}'R'$, so that $\tilde{S}'(U_t' \text{ x } U_t') = \tilde{S}' \tilde{U}_t$ and $\bar{S}'$ is a L x $(1/2)M(M+1)$ selection matrix. It then follows from Richard (1975)and $E\left[S'\tilde{U}_t\right] = 0$ that

$$(B.1) \quad V_{22} = E\left[\bar{S}'\tilde{U}_t\tilde{U}_t'\bar{S}\right] = Var(S'\tilde{U}_t) = \bar{S}'Var(\tilde{U}_t)\bar{S} = 2\bar{S}'Q(\Sigma \text{ x } \Sigma)Q'\bar{S}.$$

For Σ nonsingular, $\Sigma \otimes \Sigma$ is also nonsingular, and since Q has full row rank and $\tilde{S}$ has full column rank, it follows that V_{22} is nonsingular, and consequently that V is nonsingular. Also, from Appendix A, it follows that if Σ is nonsingular, so is $J_{\tilde{\sigma}\tilde{\sigma}}$.

For normally distributed disturbances, V_{12} and V_{21} are both zero matrices, so that

(B.2) $G'VG = G_1'V_{11}^{-1} G_1 + G_2'V_{22}^{-1} G_2 = \tilde{D}' (\Sigma^{-1} \otimes N) \tilde{D} + G_2'V_{22}^{-1} G_2.$

We now proceed to calculate $G_2'V_{22}^{-1} G_2$. Using equations (4.26), the

definition of $\bar{S}$, and Lemma A1(v),

(B.3) $G_2 = \tilde{S} (E + I)(I_M \otimes \Sigma)\tilde{B}' = 2\tilde{S}'R'Q(I_M \otimes \Sigma)\tilde{B}' = 2\bar{S}'Q(I_M \otimes \Sigma)\tilde{B}'.$

Let $F = R(\Sigma^{-1} \otimes \Sigma^{-1})R'$, so that by eqution (A.15) we have $F^{-1} = Q(\Sigma \otimes \Sigma)$ Q'. Then by equations (B.1) and (B.3)

(B.4) $G_2'V_{22}^{-1}G_2 = 2\tilde{B}(\Sigma^{-1} \otimes I_M) R'F^{-1}\bar{S} (\bar{S}' F^{-1}\bar{S})^{-1} \bar{S}'F^{-1}R(\Sigma^{-1} \otimes I_M)\tilde{B}'.$

Then from equations (B.5), (B.1), and (A.18) the conclusion will follow

if

(B.6) $F^{-1}\bar{S} (\bar{S}'F^{-1}\bar{S})^{-1} \bar{S}'F^{-1} = F^{-1}+ S(S'FS)^{-1}S'$

Note that S selects the unrestricted components of σ_* and $\bar{S}$ the

restricted. Therefore, rank (S) + rank $(\bar{S})$ = rank (F) and $\bar{S}'S = 0$, and

(B.7) $F^{-1/2} \bar{S}(\bar{S}'F^{-1}\bar{S})^{-1}\bar{S}'F^{-1/2} + F^{1/2}S (S'FS)^{-1}S'F^{1/2} = I$

since the matrix on the left-hand side of equation (B.7) is the sum of

two orthogonal, indempotent matrices, and the sum of their ranks equals

their dimension. The conclusion follows by substituting equation (B.7) into equation (B.6).

The following Lemma will be useful in obtaining proofs of the identification results of Section 5. Suppose for the moment that G is square. For a particular assignment of residuals as instruments, which is indexed by $p=1,\ldots 2^L$, let $\tilde{C}_p = \text{diag}(C_{p1},\ldots, C_{pM})$

Lemma B1: For some 2^L-tuple of positive integers $(\ell_1,\ldots,\ell_2 L)$

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

Proof: Let the rows of G_2 be denoted by s_k, $k=1,\ldots,L$. Each k corresponds to a restriction $\sigma_{ij} = 0$ 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

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

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

$$- G_p = \begin{bmatrix} (I_M \otimes N)\tilde{D} \\ \tilde{s}_p \end{bmatrix},$$

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

$$(3.18) \quad \det(-G) = \det(-\tilde{G}_1) + \det(-\tilde{G}_2).$$

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 $(I_M \otimes N)\tilde{D}$ 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 $\tilde{N}_p$ from $\tilde{G}_p$. That is, $\tilde{N}_p = E_p\tilde{G}_p$, 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 $\det(-G) = (-1)^q\det(G)$ and for each

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

Proof of Theorem 5.1:

If $\text{rank}(G) = q$, then there exists a q-dimensional square submatrix

of G, denoted by $\tilde{G}$, which is nonsingular. The matrix $\overline{G}$ is obtained by

deleting $MK+L-q$ rows of G. Each row of $(I \times N)\tilde{D}$ which is deleted

corresponds to ignoring an a variable in Z when considering instruments

for an equation i. For each i let Zi denote the predetermined variables

which remain as instruments for equation i after forming G. Each row of

G_2 deleted corresponds to ignoring a covariance restriction. Let $k = 1$,

..., L index the remaining covariance restrictions. For each asignment of

disturbances as instruments, indexed as before by p, from the remaining

covariance restrictions, let $W_{pi} = (Z_i, U_{pi})$ be the matrix of

observations on the instrumental variables for equation i, and let

$\overline{C}_{pi} = \text{plim}(\overline{W}_{pi}'X/T)$ and $\overline{\tilde{C}}_{pi} = \text{diag}(\overline{C}_{p1},...,\overline{C}_{pM})$. Then from Lemma B1 it

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

Then $\overline{G}$ non-singular implies $\det(\overline{\tilde{C}}_{\overline{p}}) \neq 0$ for some $\overline{p}$ and consequently

$\text{rank}(\overline{\tilde{C}}_{\overline{p}}) = q$. Since $\overline{\tilde{C}}_{\overline{p}}$ is block diagonal,

$$(B.8) \quad \sum_{i=1}^{M} \text{rank}(\overline{C}_{\overline{p}i}) = \text{rank}(\widetilde{\overline{C}}_{p}) = q.$$

Each $\overline{C}_{\overline{p}i}$ has q_i columns so that $\text{rank}(\overline{C}_{\overline{p}i}) \leqslant q_i$ for $i=1,\ldots,M$, and

consequently equation (B.8) implies rank $(C_{\overline{p}i}) = q_i$. Now let m* be an

assignment of disturbances as instruments such that the covariance

restrictions indexed by $\overline{k} = 1,\ldots,\overline{L}$ have disturbances assigned as the

assignment indexed by $\overline{p}$ and for the other covariance restrictions the

disturbances are assigned in any feasible fashion. Then for $i=1,\ldots,M$

$$q_i \geqslant \text{rank}(C_{p*i}) = \text{rank}(\text{plim}\left[Z:U_{p*i}\right]'X_i/T) \geqslant \text{rank}(\text{plim}\left[Z_i:U_{\overline{p}i}\right]'X_i/T)= q_i$$

so that $q_i = \text{rank}(C_{p*i})$.

Proof of Theorem 4.3: This follows as in the algorithm for the

assignment condition of Geraci (1977).

Proof of Lemma 5.4: 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 $\phi A_1 = 0$ 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 $\phi AB^{-1} = [E_1', (B^t)^{-1}\Gamma_1']'$ where E_1 is an $(M-1-r_1) \times M$ 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

$$(B.9) \quad \text{rank}(FB^{-1}) = \text{rank} \begin{bmatrix} \Gamma_1 & (B^{-1})_1 \\ \Sigma_1 & (B^{-1})_1 \end{bmatrix} + M-1-r_1 .$$

Now consider N_1. Note that for any $j \neq 1$, $\text{plim } u_j'X_1/T = [\text{plim}(u_j'Y_1/T), \text{plim}(u_j'Z_1/T)] = [\text{plim}(u_j'V_1/T), 0_1] = [\Sigma_j(B^{-1})_1, 0_1]$,

where 0_1 is a $1 \times s_1$ vector of zeros and Σ_j is the jth row of Σ. By C non-singular

$$(B.10) \quad \text{rank}(C_1) = \text{rank} \left(\begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \text{plim} \overset{D_1}{} (u_j'X_1/T) \end{bmatrix} \right)$$

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

By column reduction, using the columns of $[I_1' \ 0_1']'$, equation (B.10) implies

$$(B.11) \qquad \text{rank}(C_1) = \text{rank} \begin{bmatrix} \Gamma_1(B^{-1})_1 \\ \Sigma_1(B^{-1})_1 \end{bmatrix} + s_1.$$

Then equations (B.9) and (B.11) imply $M-1-\text{rank}(F) = q_1-\text{rank}(C_1)$, from which the conclusion of the proposition follows.

Proof of Theorem 5.6: This proof follows closely the proof of Lemma 5.4. Let $\widetilde{F} = \text{diag}(\phi_1,\ldots,\phi_M, \widetilde{S}') \bullet (I_M \otimes A', I_M \otimes \Sigma(E+I))'$.

Post-multiplication of F by $I_M \otimes B^{-1}$ and row reduction using E_i, $i=1,\ldots,M$ as in the proof of Lemma 5.4 gives

$$(B.12) \qquad \text{rank } \widetilde{F} (I_M \otimes B^{-1}) = \text{rank}(G) - \sum_{i=1}^{M} s_i + M^2 - M - \sum_{i=1}^{M} r_i = (\text{rank}(G)-q)$$
$$+ M^2 - M.$$

REFERENCES

[1] Berndt, E. B., B. Hall, R. Hall, and J. A. Hausman: "Estimation and Inference in Nonlinear Structural Models," Annals of Economic and Social Measurement, 3 (1974), 653-665.

[2] Brundy, J. M., and D. W. Jorgenson: "Efficient Estimation of Simultaneous Equations by Instrumental Variables," Review of Economics and Statistics, 53 (1971), 207-224.

[3] Dhrymes, P. J. : Econometrics. New York: Harper and Row, 1970.

[4] Dhrymes, P. J. : "A Simplified Structural Estimator for Large-Scale Econometric Models," The Australian Journal of Statistics, 13 (1971), 168-175.

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

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

[7] Hausman, J. A.: "Full Information Instrumental Variable Estimation of Simultaneous Equation Systems," Annals of Economic and Social Measurement, 3 (1974), 641-652.

[8] Hausman, J. A.: "An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Non-linear Econometric Models," Econometrica, 43 (1975), 727-738.

[9] Hausman, J.A. : "Errors in Variables in Simultaneous Equation Models," Journal of Econometrics, 5 (1977), 389-401.

[10] Hausman, J. A., and W. K. Newey: "The Information Matrix for the Linear Simultaneous Equations System," M.I.T. Working Paper, 1982.

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

[12] Hausman, J. A. : "Specification Tests in Econometrics," Econometrica, 46 (1978), 1251-1271.

[13] Holly, Alberto: "A Remark on Hausman's Specification Test", Econometrica, 50 (1982), 749-759.

[14] 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.

[15] 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).

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

[17] Malinvaud, E. : Statistical Methods of Econometrics, 2nd edition. Amsterdam: North Holland, 1970.

[18] Newey, W. K., Specification Testing and Estimation Using a Generalized Method of Moments, Unpublished Ph.D Thesis, M.I.T., 1983.

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

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

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

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

[23] Theil, Henri,:Principles of Econometrics, New York: John Wiley and Sons, 1971.

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

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