INTERNATIONAL ECONOMIC REVIEW
Vol. 26, No. 2, June, 1985

MAXIMUM LIKELIHOOD AND INSTRUMENTAL VARIABLE ESTIMATION IN SIMULTANEOUS EQUATION SYSTEMS WITH ERROR COMPONENTS*

BY INGMAR R. PRUCHA[1]

1. INTRODUCTION

Error component models have been widely considered in the econometric literature. Most of these studies have focused on single equation models.[2] Exceptions are the studies of Avery [1977], Baltagi [1980], Magnus [1982] and Prucha [1984] who considered a seemingly unrelated regression (SUR) model with error components. Recently Baltagi [1981] extended the literature to the case of a simultaneous equation model with error components. He introduced, respectively, a specific generalization of the two stage least squares (2SLS) and three stage least squares (3SLS) estimators in the error component context, and derived the asymptotic distribution of those estimators.[3] To the best of my knowledge, Baltagi's estimators are the only estimators for the simultaneous equation model with error components. He did not consider classes of asymptotically equivalent estimators, nor questions of asymptotic efficiency.

In this paper we derive, assuming normality, the full information maximum likelihood (NFIML) estimator for the regression parameters and the error covariances of a linear simultaneous equation model with error components. This generalizes Amemiya's [1971] result for the single equation case, and makes it possible to discuss questions of asymptotic efficiency. We show that the NFIML estimator has an instrumental variable (IV) representation which generalizes a similar result of Hausman [1974, 1975] for the standard simultaneous equation model. The IV form of the normal equations of the NFIML estimator is used to generate a wide class of IV estimators. These estimators are typically computationally simpler and can, by construction, be viewed as numerical approximation to the NFIML estimator. We, henceforth, refer to these estimators as NFIML_A estimators. This approach is similar to that of Hendry [1976] for the

* Manuscript received April, 1983; revised May, 1984.

[1] An earlier version of this paper was presented at the European Econometric Society Meetings in Pisa, 1983. I would like to thank Roger Betancourt, Manfred Deistler, Harry Kelejian, Benedikt Poetscher, and this journal's editor, Robert A. Pollak, and referees for helpful comments and suggestions. I assume responsibility, however, for any errors.

[2] See e.g. Balestra and Nerlove [1966], Wallace and Hussain [1969], Amemiya [1971], Maddala [1971], Nerlove [1971a, b], Swamy and Arora [1972], Fuller and Battese [1974], Mundlak [1978] and Anderson and Hsiao [1981, 1982].

[3] Baltagi's generalization of the 2SLS estimator was originally suggested by Maddala [1971].

standard simultaneous equation model.[4]

All $\mathrm{NFIML_A}$ estimators have the same basic structure and differ only in the choice of the estimators of the reduced form parameters and the covariance components used in the construction of the instrument matrix. Based on easy to determine characteristics of the latter estimators we identify, respectively, a wide subclass of limited and full information estimators. We prove, for general disturbance distributions, the asymptotic equivalence of all members within each subclass; we also demonstrate their consistency and derive their asymptotic distribution. All members of the full information subclass are asymptotically equivalent with the NFIML estimator and in this sense efficient.[5] The asymptotic equivalence holds not only for the normal but also for general disturbance distributions given consistency of the NFIML estimator.

Virtually all $\mathrm{NFIML_A}$ estimators of practical interest are contained in the two subclasses considered. In particular they contain, respectively, limited and full information generalizations of the least squares dummy variable (LSDV) estimator as well as generalizations of virtually all estimators known for the standard simultaneous equation model. Typically various asymptotically equivalent generalizations are possible. Baltagi's [1981] specific generalizations of the 2SLS and 3SLS estimator turn out to be members of, respectively, the limited and full information subclasses. Several other examples of $\mathrm{NFIML_A}$ estimators, including an alternative generalization of the 3SLS estimator, are given. Furthermore, we present results that make it very easy to check if an estimator belongs to one of the two equivalence classes.

The plan of the paper is as follows: Section 2 gives the specification of the model. The NFIML estimator and its IV from are derived in Section 3. In Section 4, the class of $\mathrm{NFIML_A}$ estimators is defined and the asymptotic properties of wide subclasses of limited and full information estimators are analyzed. Section 5 contains various examples. Concluding remarks are given in Section 6. Technical issues are relegated to the appendices.

[4] There exist known transformations which reduce the simultaneous equation model with error components to a standard simultaneous equation model. It may seem appealing to try to analyze the simultaneous equation model with error components by applying the results obtained by Hausman [1974, 1975] and Hendry [1976] to one of the transformed models satisfying standard assumptions. Unfortunately, all known transformations lose information; more precisely, in all cases the transformed model no longer depends on all the parameters characterizing the original model. Hence, by applying Hausman's and Hendry's results to the transformed models no conclusions can be made about the true NFIML estimator (for instance, if it will be of IV form) or about the properties of classes of approximating estimators. Consequently the analysis must be performed in terms of the original model.

[5] In light of part of the literature on systems of equations with error components, it may seem surprising that Hendry's finding of large asymptotic equivalence classes for the standard case carries over to the error component case. In particular, for the special case of a seemingly unrelated regression model with error components, Baltagi [1980] gives results that suggest the contrary. However, those results turn out to be incorrect as is shown in Prucha [1984].

2. THE MODEL

In specifying the model, we follow Baltagi [1981]. Let there be N cross-sectional units observed over T periods; consider the following system of linear simultaneous relationships:

$$Y = e_{NT}a + YB + ZC + U \tag{1}$$

where $Y=[y_1,\ldots, y_M]$ is the $NT\times M$ matrix of observations on the M endogenous variables of the system, $Z=[z_1,\ldots, z_K]$ is the $NT\times K$ matrix of observations on the K non-stochastic exogenous slope variables and $U=[u_1,\ldots, u_n]$ is the $NT\times M$ matrix of disturbances;[6] e_{NT} is a $NT\times 1$ vector of unit elements; a, B and C are $1\times M$, $M\times M$ and $K\times M$ matrices of parameters. As a normalization rule, the diagonal elements of B are taken to be zero. Further, $(I-B)$ is assumed to be nonsingular, so that (1) has the reduced form representation

$$Y = e_{NT}\pi + Z\Pi + W, \quad [\pi, \Pi, W] = [a, C, U](I-B)^{-1}. \tag{2}$$

The disturbance vector of the j-th equation of (1) is assumed to be composed of three stochastic components, in particular

$$u_j = (I_N\otimes e_T)\mu_j + (e_N\otimes I_T)\lambda_j + v_j \qquad j = 1,\ldots, M \tag{3}$$

where the i-th element of the $N\times 1$ vector μ_j and the t-th element of the $T\times 1$ vector λ_j represent the error components specific to the i-th unit and the t-th period respectively; the $NT\times 1$ vector v_j contains the error components specific to both; e_T and e_N are $T\times 1$ and $N\times 1$ vectors of unit elements. The error components are assumed to have zero mean and the covariance structure

$$E\begin{bmatrix}\mu_j\\ \lambda_j\\ v_j\end{bmatrix}[\mu_l', \lambda_l', v_l'] = \begin{bmatrix}\sigma_{\mu jl}I_N & 0 & 0\\ 0 & \sigma_{\lambda jl}I_T & 0\\ 0 & 0 & \sigma_{0jl}I_{NT}\end{bmatrix} \qquad j, l=1,\ldots, M \tag{4}$$

with $\Sigma_\mu=(\sigma_{\mu jl})$, $\Sigma_\lambda=(\sigma_{\lambda jl})$ and $\Sigma_0=(\sigma_{0jl})$ positive definite. Define $\Sigma_1=(\sigma_{1jl})=\Sigma_0+T\Sigma_\mu$, $\Sigma_2=(\sigma_{2jl})=\Sigma_0+N\Sigma_\lambda$ and $\Sigma_3=(\sigma_{3jl})=\Sigma_0+T\Sigma_\mu+N\Sigma_\lambda$; then the covariance matrix of the disturbance vector of the j-th and l-th equation can for j, $l=1,\ldots, M$ be expressed as

$$E(u_ju_l') = \sigma_{\mu jl}A_{NT} + \sigma_{\lambda jl}B_{NT} + \sigma_{0jl}I_{NT} = \sum_{h=0}^{3} \sigma_{hjl}Q_h \tag{5}$$

where $Q_0=I_{NT}-A_{NT}/T-B_{NT}/N+J_{NT}/NT$, $Q_1=A_{NT}/T-J_{NT}/NT$, $Q_2=B_{NT}/N-J_{NT}/NT$, $Q_3=J_{NT}/NT$ with $A_{NT}=I_N\otimes e_Te_T'$, $B_{NT}=e_Ne_N'\otimes I_T$ and $J_{NT}=e_{NT}e_{NT}'$.

[6] The assumption that Z is non-stochastic is made to simplify the exposition. Without complications, we could relax this assumption and only maintain that Z is strongly exogenous in the sense of Engle, Hendry and Richard [1983]; we would then interpret the subsequent results as conditional on Z.

We do not require that the disturbances are normally distributed except when specifically stated. However, we assume that all fourth moments exist and that zero covariance between disturbances also indicates stochastic independence. In particular, we assume $v=[v_1',\dots, v_M']'=(P\otimes I_{NT})\xi$ where ξ is an $NMT\times 1$ vector of i.i.d. random variables with zero mean and unit variance and $\Sigma_0=PP'$.

Every equation is identified subject to zero-type parameter restrictions. Let Y_j and Z_j denote, respectively, the $NT\times M_j$ and $NT\times K_j$ matrices of observations on the endogenous and exogenous variables that appear as regressors in the j-th equation and let β_j and γ_j be the corresponding $M_j\times 1$ and $K_j\times 1$ vectors of unrestricted (non-zero) parameters; let further $a=[\alpha_1,\dots,\alpha_M]$, then the j-th equation of (1) can be written as

$$y_j = e_{NT}\alpha_j + X_j\delta_j + u_j, \quad X_j = [Y_j, Z_j], \quad \delta_j = [\beta_j', \gamma_j']'. \tag{6}$$

Define $y=\mathrm{vec}(Y)$, $X=\mathrm{diag}_M(X_j)$, $u=\mathrm{vec}(U)$, and $\delta=[\delta_1',\dots,\delta_M']'$, then the stacked form of (1) can be written as

$$y = (I_M\otimes e_{NT})a' + X\delta + u. \tag{7}$$

Equation (5) implies for the variance covariance matrix of the stacked disturbance vector

$$\Sigma = E(uu') = \Sigma_{h=0}^{3} \Sigma_h \otimes Q_h. \tag{8}$$

It will be convenient for our later discussion to introduce selector matrices L_j such that $X_j=[Y, Z]L_j$.[7] The elements of B and C can then be related to those of δ as $\mathrm{vec}\,[(B', C')']=L\delta$ with $L=\mathrm{diag}_M(L_j)$.

We further assume that the elements of $Z=(z_{tj})$ are uniformly bounded, i.e., $\sup_{t,j} z_{tj}<\infty$, and that $Z'Z/NT$ and $Z'Q_0Z/NT$ tend to finite positive definite matrices as both N and T tend to infinity in every possible way.

3. THE FULL INFORMATION MAXIMUM LIKELIHOOD ESTIMATOR

The following theorem gives the full information maximum likelihood (NFIML) estimator of the model parameters for normally distributed disturbances. The proof of the theorem is given in Appendix A.

THEOREM 1. *Consider the model specified in Section 2 and assume* $u\sim N(0, \Sigma)$. *Suppose only the first* S $(\leq M)$ *equations contain an intercept term, i.e.* $a=[\alpha', 0]$ *with* $\alpha=[\alpha_1,\dots,\alpha_S]'$. *The full information maximum likelihood estimators* $\hat{\alpha}$, $\hat{\delta}$, $\hat{\Sigma}_0$, $\hat{\Sigma}_\mu$ *and* $\hat{\Sigma}_\lambda$ *then satisfy the following system of normal equations*:

$$\hat{\delta} = [\hat{X}'\hat{\Phi}^{-1}X]^{-1}\hat{X}'\hat{\Phi}^{-1}y, \quad \hat{X} = \mathrm{diag}_M(\hat{X}_j), \tag{9a}$$

$$\hat{X}_j = (\hat{Y}, Z)L_j, \quad \hat{Y} = e_{NT}\hat{\pi} + Z\hat{\Pi},$$

$$\hat{\alpha} = \{[I_S, (\hat{\Sigma}_3^{11})^{-1}\hat{\Sigma}_3^{12}]\otimes e_{NT}'/NT\}\{y-X\hat{\delta}\}, \tag{9b}$$

[7] See Dhrymes [1978, pp. 276–7] for a further discussion of selector matrices.

(9c) $\hat{\Phi}^{-1} = \hat{\Sigma}_0^{-1} \otimes Q_0 + \hat{\Sigma}_1^{-1} \otimes Q_1 + \hat{\Sigma}_2^{-1} \otimes Q_2 + \operatorname{diag}(O_{S\times S}, \hat{\Sigma}_{3,22}^{-1}) \otimes Q_3,$

(9d) $$\begin{aligned}(N-1)(T-1)\hat{\Sigma}_0^{-1} - \hat{\Sigma}_0^{-1}\hat{U}'Q_0\hat{U}\hat{\Sigma}_0^{-1} &= \hat{\Sigma}_3^{-1} - \hat{\Sigma}_3^{-1}\hat{U}'Q_3\hat{U}\hat{\Sigma}_3^{-1}\\ &= -(N-1)\hat{\Sigma}_1^{-1} + \hat{\Sigma}_1^{-1}\hat{U}'Q_1\hat{U}\hat{\Sigma}_1^{-1}\\ &= -(T-1)\hat{\Sigma}_2^{-1} + \hat{\Sigma}_2^{-1}\hat{U}'Q_2\hat{U}\hat{\Sigma}_2^{-1},\end{aligned}$$

$$\hat{\Sigma}_0 = \hat{\Sigma}_1 + \hat{\Sigma}_2 - \hat{\Sigma}_3,\ \hat{\Sigma}_\mu = (\hat{\Sigma}_1 - \hat{\Sigma}_0)/T,\ \hat{\Sigma}_\lambda = (\hat{\Sigma}_2 - \hat{\Sigma}_0)/N,$$

(9e) $$\hat{\pi} = \hat{a}(I-\hat{B})^{-1},\ \hat{\Pi} = \hat{C}(I-\hat{B})^{-1},$$
$$\hat{a} = (\hat{\alpha}', 0),\ \operatorname{vec}[(\hat{B}', \hat{C}')'] = L\hat{\delta},$$

with $\hat{U} = Y - e_{NT}\hat{a} - Y\hat{B} - Z\hat{C}$ *and where* $\hat{\Sigma}_3^{ij}$ *and* $\hat{\Sigma}_{3ij}$ *denote for* $i, j = 1, 2$ *the* (i, j)*-th block of the matrices* $\hat{\Sigma}_3^{-1}$ *and* $\hat{\Sigma}_3$, *respectively. The underlying partition is* $(S, M-S)\times(S, M-S)$.

REMARK 1. In case there is an intercept in each equation, i.e. $S = M$, the expression for $\hat{\Phi}^{-1}$ in (9c) reduces to $\hat{\Phi}^{-1} = \Sigma_{h=0}^{2}\hat{\Sigma}_h^{-1}\otimes Q_h$. Since $Q_h e_{NT} = 0$ for $h = 0, 1, 2$ we can then simplify formula (9a) for $\hat{\delta}$ further by calculating $\hat{Y}$ as $\hat{Y} = Z\hat{\Pi}$.[8] Formula (9b) for $\hat{\alpha}$ simplifies to $\hat{\alpha} = [I_M \otimes e'_{NT}/NT][y - X\hat{\delta}]$.

For expositional simplicity, we assume for the remainder of the paper that there is an intercept term in each equation, i.e. $S = M$. All of the subsequent results can readily be generalized to the case $S < M$. We further concentrate in the following on the estimation of the slope parameters.

REMARK 2. The arrangement of the normal equations as given in (9) is of particular interest since it shows that the NFIML estimator has an instrumental variable representation. To see this, define the instrument matrix $\hat{P} = \hat{\Phi}^{-1}\hat{X}$ and note that $\hat{\delta} = (\hat{P}'X)^{-1}\hat{P}'y$. This generalizes a similar result of Hausman [1974, 1975] for the standard simultaneous equation model.[9]

REMARK 3. Amemiya [1971] derived the full information maximum likelihood estimator of the error component model in the single equation case. He paid particular attention to the estimation of the covariance components. We note that the normal equations for the NFIML estimators of the covariance components as given in (9) are direct matrix generalizations of the corresponding normal equations in the single equation case — compare Amemiya [1971, p. 7].

[8] For the same reason, it is readily seen that $\hat{\delta}$ is invariant against centering the data around overall sample means.

[9] See Appendix A for details of the derivation of the instrumental variable arrangement of the normal equations from the original form of the first order conditions for a maximum of the likelihood function. A crucial step in that derivation was to show that $\Sigma_{h=0}^{3}\hat{U}'Q_h\hat{U}\hat{\Sigma}_h^{-1}/NT = I_M$ despite the fact that $\hat{\Sigma}_h$ is not proportional to $\hat{U}'Q_h\hat{U}$ and that there exists no explicit solution of $\hat{\Sigma}_h$ in terms of $\hat{U}$. We note that this contrasts with the case of the standard simultaneous equation model where the NFIML estimator of the disturbance variance covariance matrix can be expressed explicitly as a function of the NFIML estimator of the disturbance matrix.

4. A GENERAL CLASS OF APPROXIMATING INSTRUMENTAL VARIABLE ESTIMATORS

In the following, we use the instrumental variable form of the normal equations of the NFIML estimator as estimator generating equations.[10] The NFIML estimators solve system (9a)–(9e) simultaneously. Such a solution may be obtained iteratively. Note that in replacing the NFIML estimators for the reduced form parameters and the covariance components on the R.H.S. of (9a) by other estimators we obtain a new estimator for δ. This estimator can be interpreted as an approximation to the NFIML estimator. Using the normal equations (9a) in this sense as estimator generating equations leads to the definition of the following general class of estimators.

DEFINITION 1. For $h=0, 1, 2$ let $\tilde{\Pi}_h$ be some estimators for Π and $\tilde{\Sigma}_h^{-1}$ some estimators for Σ_h^{-1}; further, let $\tilde{X}_{(h)}=\mathrm{diag}_M(\tilde{X}_{j(h)})$ with $\tilde{X}_{j(h)}=[\tilde{Y}_{(h)}, Z]L_j$ and $\tilde{Y}_{(h)}=Z\tilde{\Pi}_h$. Then any estimator of the form

$$(10)\qquad \tilde{\delta} = [\Sigma_{h=0}^2 \tilde{X}'_{(h)}(\tilde{\Sigma}_h^{-1}\otimes Q_h)X]^{-1}\Sigma_{h=0}^2 \tilde{X}'_{(h)}(\tilde{\Sigma}_h^{-1}\otimes Q_h)y = G(\tilde{\Pi}_0, \tilde{\Sigma}_0^{-1}, \tilde{\Pi}_1, \tilde{\Sigma}_1^{-1}, \tilde{\Pi}_2, \tilde{\Sigma}_2^{-1})$$

is called an $\mathrm{NFIML_A}$ estimator (where the subscript A is used to indicate that the estimator can be viewed as an approximation to the NFIML estimator). If any of the $\tilde{\Sigma}_h^{-1}$ are nondiagonal, the resulting estimators are referred to as full information $\mathrm{NFIML_A}$ estimators. If all $\tilde{\Sigma}_h^{-1}$ are diagonal, i.e., $\tilde{\Sigma}_h^{-1}=\mathrm{diag}_M(\tilde{\sigma}_{hjj}^{-1})$, equation (10) reduces to

$$(11)\qquad \tilde{\delta}_j = [\Sigma_{h=0}^2 \tilde{\sigma}_{jjh}^{-1}\tilde{X}'_{j(h)}Q_hX_j]^{-1}\, \Sigma_{h=0}^2 \tilde{\sigma}_{jjh}^{-1}\tilde{X}'_{j(h)}Q_hy_j$$

and we refer to those estimators as limited information $\mathrm{NFIML_A}$ estimators.

REMARK 4. Clearly the class of $\mathrm{NFIML_A}$ estimators defined above contains the NFIML estimator of the structural parameters as the special case in which $\tilde{\Pi}_h$ and $\tilde{\Sigma}_h^{-1}$ are taken to be the respective NFIML estimators of the reduced form parameters and the covariance components. All $\mathrm{NFIML_A}$ estimators are instrumental variable estimators. Let the instrument matrix be defined as $\tilde{P}=\Sigma_{h=0}^2(\tilde{\Sigma}_h^{-1}\otimes Q_h)\tilde{X}_{(h)}$, then $\tilde{\delta}=(\tilde{P}'X)^{-1}\tilde{P}'y$.

REMARK 5. The standard simultaneous equation model is a special case of our model with $\Sigma_\mu=\Sigma_\lambda=0$. Hendry [1976] showed for the standard model that virtually all known standard estimators can be viewed as numerical approximations to the standard NFIML estimator. It is, hence, evident that our class of estimators will contain generalizations of virtually all known standard estimators.[11]

[10] See Hausman [1974, 1975] and Hendry [1976] for an analogous approach applied in the standard simultaneous equation model case.

[11] Hendry's class of estimators is readily seen to constitute a special subclass of the above NFIML_A class with $\tilde{\Pi}_h=\tilde{\Pi}_0$ and $\tilde{\Sigma}_h^{-1}=\tilde{\Sigma}_0^{-1}$ for $h=1, 2$; in this case, (10) reduces to $\tilde{\delta}=\{\tilde{X}'_{(0)}[\tilde{\Sigma}_0^{-1}\otimes(I_{NT}-J_{NT}/NT)]X\}^{-1}\tilde{X}'_{(0)}[\tilde{\Sigma}_0^{-1}\otimes(I_{NT}-J_{NT}/NT)]y$.

Examples of such generalizations will be given in Section 5.

DEFINITION 2. Consider the following estimators:

(12) $\tilde{\delta}_{\text{FIDV}} = [\bar{X}'(\Sigma_0^{-1}\otimes Q_0)X]^{-1}\bar{X}'(\Sigma_0^{-1}\otimes Q_0)y = G(\Pi, \Sigma_0^{-1}, ., 0, ., 0)$,

(13) $\tilde{\delta}_{j,\text{LIDV}} = [\bar{X}'_j Q_0 X_j]^{-1}\bar{X}'_j Q_0 y_j \qquad j = 1,\dots, M$

where $\bar{X} = \text{diag}_M(\bar{X}_j)$ and $\bar{X}_j = Z[\Pi, I_K]L_j$. We, henceforth, refer to these estimators as, respectively, the FIDV and LIDV estimators; they represent (non-feasible) full and limited information NFIML_A generalizations of the single equation and seemingly unrelated regression dummy variable estimators to the simultaneous equation case.[12]

Next, we analyze the asymptotic properties of NFIML_A estimators. We note that in the subsequent analysis convergence in distribution, probability limits or limits are always understood in the sense that both N and T tend to infinity.[13] The following theorem establishes the asymptotic equivalence of the members of a wide subclass of full information NFIML_A estimators. The proof of the theorem is given in Appendix B.

THEOREM 2. *Let the model be specified as stated in Section* 2. *Consider the class of full information $NFIML_A$ estimators $\tilde{\delta}$ defined by* (10) *with* plim $\tilde{\Pi}_h = \Pi$ *for* $h = 0, 1, 2$, plim $\tilde{\Sigma}_0 = \Sigma_0$ *and* plim $T^{1-\varepsilon}\tilde{\Sigma}_1^{-1} = \text{plim } N^{1-\varepsilon}\tilde{\Sigma}_2^{-1} = 0$ *for some* $0 < \varepsilon < 1/2$.[14] *Then all members of this class are asymptotically equivalent to the FIDV estimator* (*and consequently to each other*) *in the sense that* plim $\sqrt{NT}(\tilde{\delta} - \tilde{\delta}_{\text{FIDV}}) = 0$. *In particular, any member of this class is consistent and* $\sqrt{NT}(\tilde{\delta} - \delta)$ *converges in distribution to a normal random vector with zero mean and variance covariance matrix* $\Omega = \text{plim } [\bar{X}'(\Sigma_0^{-1}\otimes Q_0)\bar{X}/NT]^{-1}$.

The following lemma proves helpful in analyzing the class of estimators considered in Theorem 2. The proof of the lemma is not difficult and hence omitted.

LEMMA 1. *Let* $\tilde{\Sigma}_0$, $\tilde{\Sigma}_\mu$ *and* $\tilde{\Sigma}_\lambda$ *be any consistent estimators of* Σ_0, Σ_μ *and* Σ_λ. *Define* $\tilde{\Sigma}_1^{-1} = [\tilde{\Sigma}_0 + T\tilde{\Sigma}_\mu]^{-1}$ *and* $\tilde{\Sigma}_2^{-1} = [\tilde{\Sigma}_0 + N\tilde{\Sigma}_\lambda]^{-1}$, *then* plim $T^{1-\varepsilon}\tilde{\Sigma}_1^{-1} =$ plim $N^{1-\varepsilon}\tilde{\Sigma}_1^{-1} = 0$ *for all* $0 < \varepsilon < 1/2$.

REMARK 6. We can now identify two important subclasses of estimators that satisfy the assumptions of Theorem 2:

(i) The class of full information NFIML_A estimators based on consistent esti-

[12] For a definition of the dummy variable estimator in the single and seemingly unrelated regression case see, respectively, Wallace and Hussain [1969] and Baltagi [1980].

[13] Compare e.g. Wallace and Hussain [1969], Amemiya [1971] and Baltagi [1981]. We also note that the model considered in this paper is static. Results by Anderson and Hsiao [1981, 1982] for the single equation model suggest that in case of a dynamic model, particular attention will have to be paid to initial conditions.

[14] Note that lim $T^{1-\varepsilon}\Sigma_1^{-1} = \text{lim } N^{1-\varepsilon}\Sigma_2^{-1} = 0$ for any $0 < \varepsilon < 1/2$.

mators of the reduced form parameters Π and the covariance components Σ_0, Σ_μ and Σ_λ (in the sense of Lemma 1),

(ii) The class of feasible FIDV estimators $\tilde{\delta}_{\mathrm{FFIDV}}=G(\tilde{\Pi}_0, \tilde{\Sigma}_0^{-1}, ., 0, ., 0)$ based on consistent estimators for Π and Σ_0.

Most full information $\mathrm{NFIML_A}$ estimators of practical interest will fall into one of these two categories. Theorem 2 should hence make case by case considerations of the asymptotic properties of full information $\mathrm{NFIML_A}$ estimators essentially unnecessary.

REMARK 7. We have the following result concerning asymptotic efficiency: Given the consistency of the NFIML estimators it follows immediately from Remark 6 that the NFIML estimator $\hat{\delta}$ and the members of the class of full information $\mathrm{NFIML_A}$ estimators considered in Theorem 2 are asymptotically equivalent. We note that the latter class contains various members that are computationally considerably simpler than the NFIML estimator.

The following theorem identifies a wide subclass of asymptotically equivalent limited information $\mathrm{NFIML_A}$ estimators. Since the proof and the discussion of the theorem would be analogous to that of Theorem 2 we omit both.

THEOREM 3. *Let the model be specified as stated in Section* 2. *Consider the class of limited information $NFIML_A$ estimators $\tilde{\delta}_j$ defined by* (11) *with* $\operatorname{plim} \tilde{\Pi}_h = \Pi$ *for* $h=0,1,2$, $\operatorname{plim} \tilde{\sigma}_{0jj}=\sigma_{0jj}$ *and* $\operatorname{plim} T^{1-\varepsilon}\tilde{\sigma}_{1jj}^{-1}=\operatorname{plim} N^{1-\varepsilon}\tilde{\sigma}_{2jj}^{-1}=0$ *for some* $0<\varepsilon<1/2$. *Then all members of this class are asymptotically equivalent with the LIDV estimator (and consequently to each other) in the sense that* $\operatorname{plim}\sqrt{NT}(\tilde{\delta}_j-\tilde{\delta}_{j,\mathrm{LIDV}})=0$. *In particular, any member of this class is consistent and $\sqrt{NT}(\tilde{\delta}_j-\delta_j)$ converges in distribution to a normal random vector with zero mean and covariance matrix* $\Omega_j=\sigma_{0jj} \operatorname{plim} [\bar{X}_j' Q_0 \bar{X}_j/NT]^{-1}$.

5. EXAMPLES OF ESTIMATORS

In the following, we give some specific examples of $\mathrm{NFIML_A}$ estimators. As remarked above, virtually all known estimators for the standard simultaneous equation system have generalizations within the $\mathrm{NFIML_A}$ class. For expositional reasons, we concentrate on generalizations of full information estimators. We start out with a brief discussion of consistent estimators for Σ_0, Σ_μ, Σ_λ and Π.

5.1. *On Consistent Estimators of the Covariance Components.* The following estimators of the covariance components are generalizations of the single equation analysis of variance (AOV) estimators:[15] $\tilde{\Sigma}_0=\tilde{U}'Q_0\tilde{U}/[(N-1)(T-1)]$, $\tilde{\Sigma}_\mu=\tilde{U}'[(T-1)Q_1-Q_0]\tilde{U}/[T(N-1)(T-1)]$ and $\tilde{\Sigma}_\lambda=\tilde{U}'[(N-1)Q_2-Q_0]\tilde{U}/[N(N-1)(T-1)]$ where $\tilde{U}$ is some estimator of the disturbance matrix U. It then follows that $\tilde{\Sigma}_1=\tilde{\Sigma}_0+T\tilde{\Sigma}_\mu=\tilde{U}'Q_1\tilde{U}/(N-1)$ and

[15] Those generalizations have been introduced by Baltagi [1980] in the context of a SUR model with error components.

$\tilde{\Sigma}_2=\tilde{\Sigma}_0+N\tilde{\Sigma}_\lambda=\tilde{U}'Q_2\tilde{U}/(T-1)$.[16]

LEMMA 2. *Consider the model of Section 2. Let* $\tilde{U}=Y-e_{NT}\tilde{a}-Y\tilde{B}-Z\tilde{C}$ *be an estimator of the disturbance matrix based on consistent parameter estimators. Then the AOV estimators for* $\tilde{\Sigma}_0$, $\tilde{\Sigma}_\mu$ *and* $\tilde{\Sigma}_\lambda$ *defined above are consistent estimators for* Σ_0, Σ_μ *and* Σ_λ.

The proof of the lemma is not difficult and hence omitted here. Lemma 1 and Lemma 2 combined imply that the assumptions of Theorem 2 with respect to the covariance component estimators are satisfied for AOV estimators based on consistent residual estimators.

5.2. *On Consistent Estimators of the Reduced Form Parameters.* The reduced form (2) can be written in stacked notation as $y=(I_M\otimes e_{NT})\text{vec}(\pi)+(I_M\otimes Z)\text{vec}(\Pi)+w$ where $w=[(I-B')^{-1}\otimes I_{NT}]u$. Clearly $E(w)=0$ and $E(ww')=\Sigma_{h=0}^{3}\psi_h\otimes Q_h$ with $\psi_h=(I-B')^{-1}\Sigma_h(I-B)^{-1}$. The reduced form model can hence be viewed as a SUR model with error components and identical regressors in each equation. Each reduced form equation by itself represents a single equation error component model. The consistency of the following estimators of the reduced form parameters is obvious from the existing literature on SUR and single equation error component models.[17]

The simplest consistent estimator is the ordinary least squares (OLS) estimator, $\tilde{\Pi}_{\text{OLS}}=[Z'(I-J_{NT}/NT)Z]^{-1}Z'(I-J_{NT}/NT)Y$. Other consistent estimators that do not depend on estimators of the covariance components are: $\tilde{\Pi}_h=[Z'Q_hZ]^{-1}Z'Q_hY$ for $h=0, 1, 2$. The estimator corresponding to $h=0$ is the least squares dummy variable (LSDV) estimator, that corresponding to $h=1$ is the between group least squares (LSBG) estimator and that corresponding to $h=2$ is the between time least squares (LSBT) estimator.[18] Note that the LSDV estimator and the seemingly unrelated regression dummy variable (SURDV) estimator, $\text{vec}(\tilde{\Pi}_{\text{SURDV}})=[\psi_0^{-1}\otimes Z'Q_0Z]^{-1}[\psi_0^{-1}\otimes Z'Q_0]y$, are identical. The LSDV estimator is, therefore, for normally distributed disturbances, also asymptotically efficient.[19]

[16] Amemiya [1971, pp. 1 and 7–8] has shown for the single equation case that the AOV estimators can be interpreted as first stage estimators in the interation process of obtaining maximum likelihood estimators. Note that by replacing the NFIML estimator $\tilde{\Sigma}_0$ in (9d) by the AOV estimator for Σ_0 we obtain the AOV estimators for Σ_1, Σ_2 and Σ_μ, Σ_λ. Hence, Amemiya's interpretation of the AOV estimators carries over to the systems case.

[17] References to that literature are given in the introduction.

[18] See e.g. Swamy and Arora [1972] for a general discussion of the LSDV, LSBG and LSBT estimators.

[19] This follows from the asymptotic equivalence of the SURDV estimator and the seemingly unrelated regression generalized least squares (GLS) estimator. While this result is well known, it also follows immediately from Theorem 2. The above discussion implies further that all feasible single equation and seemingly unrelated regression GLS estimators for the reduced form parameters are asymptotically efficient, given they are e.g. based on AOV estimators of the reduced form covariance components constructed from consistent residuals.

5.3. *Generalizations of the 3SLS and FIVE Estimator.* Baltagi [1981] introduced an error components 3SLS estimator, henceforth referred to as the $EC3SLS_1$ estimator, as a weighted average of three standard 3SLS estimators applied to different transformations of the model (1). The following discussion interprets the $EC3SLS_1$ estimator in relation to the NFIML estimator. Denoting the dependence of $\tilde{X}$ and $\tilde{X}_j$ on (say) $\tilde{\Pi}$ more explicitly as $\tilde{X}(\tilde{\Pi})$ and $\tilde{X}_j(\tilde{\Pi})$ the estimator is given by

$$(5.1) \qquad \tilde{\delta}_{EC3SLS_1} = \{\sum_{h=0}^{2} \tilde{X}'(\tilde{\Pi}_h)[\tilde{\Sigma}_h^{-1}\otimes Q_h]X\}^{-1} \sum_{h=0}^{2} \tilde{X}'(\tilde{\Pi}_h)[\tilde{\Sigma}_h^{-1}\otimes Q_h]y$$

where the $\tilde{\Sigma}_h$ are AOV estimators. It is readily seen that the $EC3SLS_1$ estimator is a member of the $NFIML_A$ class. The residuals used in constructing these AOV estimators $\tilde{\Sigma}_h$ for $h=0$, 1, 2 are based on, respectively, the estimators $\tilde{\delta}_{j(h)}=[\tilde{X}'_j(\tilde{\Pi}_h)Q_hX_j]^{-1}\tilde{X}'_j(\tilde{\Pi}_h)Q_hy_j$ obtained by applying the 2SLS formula to transformations of the model. It is not difficult to show that the latter estimators are consistent. As a consequence of Lemmata 1 and 2 and the discussion of the last subsection it is then readily seen that the $EC3SLS_1$ estimator satisfies the assumptions of Theorem 2. Next, consider the following feasible FIDV estimator $\tilde{\delta}_{(0)}=\{\tilde{X}'(\tilde{\Pi}_0)[\tilde{\Sigma}_0^{-1}\otimes Q_h]X\}^{-1}\tilde{X}'(\tilde{\Pi}_0)[\tilde{\Sigma}_0^{-1}\otimes Q_h]y$ used in constructing the $EC3SLS_1$ estimator. It is readily seen that this estimator also satisfies the assumptions of Theorem 2. Both estimators are hence asymptotically equivalent to each other as well as to any other member of the class of full information $NFIML_A$ estimators defined by Theorem 2.

The instruments of the $EC3SLS_1$ estimator are formed from three different estimators of the reduced form parameters. In analogy to the NFIML estimator, it seems appealing to form the instruments from only one (efficient) estimator for Π. We thus, introduce the following alternative generalization of the 3SLS estimator, which we shall refer to as the $EC3SLS_2$ estimator:

$$(5.2) \qquad \tilde{\delta}_{EC3SLS_2} = \{\tilde{X}'(\tilde{\Pi}_0)[\sum_{h=0}^{2} \tilde{\Sigma}_h^{-1}\otimes Q_h]X\}^{-1}\tilde{X}'(\tilde{\Pi}_0)[\sum_{h=0}^{2} \tilde{\Sigma}_h^{-1}\otimes Q_h]y.$$

Clearly the advantage of this formulation could only be in terms of the small sample properties of the estimator. According to Theorem 2, the $EC3SLS_1$ and $EC3SLS_2$ estimators are asymptotically equivalent.[20]

Both of the above EC3SLS estimators make use of OLS type estimators of the reduced form parameters. Hence, those estimators cannot be computed in case there are more exogenous variables in the system than observations. The number or regressors in each equation is typically small as compared to the total number of exogenous variables in the system. Following the approach of Brundy and Jorgenson [1971, 1974] for the standard simultaneous equation model we may

[20] Obviously, various other (reasonable) generalization of the 3SLS estimator are possible. For instance, we may base all AOV covariance component estimators on the feasible LIDV estimator $\tilde{\delta}_{J(0)}$.

get initial consistent estimators of the structural parameters, say $\check{B}_{IV}$ and $\check{C}_{IV}$, from some single equation IV procedure. Those estimators can then be used to form a consistent estimator of the reduced form parameters as $\check{\Pi}_{IV}=\check{C}_{IV}(I-\check{B}_{IV})^{-1}$ and to form a consistent estimator of the disturbance matrix, say $\check{U}$. We introduce the following generalization of the FIVE estimator, say the ECFIVE estimator:

$$(5.3) \qquad \tilde{\delta}_{\mathrm{ECFIVE}} = \{\tilde{X}'(\check{\Pi}_{IV})[\sum_{h=0}^{2} \check{\Sigma}_h^{-1}\otimes Q_h]X\}^{-1}\tilde{X}'(\check{\Pi}_{IV})[\sum_{h=0}^{2} \check{\Sigma}_h^{-1}\otimes Q_h]y$$

where $\check{\Sigma}_h$ denotes the AOV covariance component estimators based on $\check{U}$. Again, this estimator is seen to be a member of the asymptotic equivalence class defined by Theorem 2.

6. CONCLUSION

In this paper, we considered the estimation of a system of linear simultaneous relationships from N cross sectional units observed over T periods. The model considered in particular is the linear simultaneous equation model with error components of Baltagi [1981]. We derived the full information maximum likelihood estimator for that model and referred to that estimator as the NFIML estimator. We showed that the normal equations of the NFIML estimator can be put into an instrumental variable form. The normal equations in instrumental variable form were then used as estimator generating equations to define a wide class of instrumental variable estimators that can be viewed as numerical approximations to the NFIML estimator. We referred to those estimators as $\mathrm{NFIML_A}$ estimators. The class of $\mathrm{NFIML_A}$ estimators was shown to contain generalizations of virtually all known estimators for the standard simultaneous equation model. We give theorems concerning the asymptotic properties of the $\mathrm{NFIML_A}$ estimator as both N and T tend to infinity. In particular, we established the existence of wide asymptotic equivalence classes of, respectively, full and limited information estimators. Baltagi's [1981] generalizations of the 2SLS and 3SLS estimator are found to be members of those equivalence classes.

The results of this paper simplify and clarify the asymptotic appraisal of estimators for the simultaneous equations model with error components. However, the results also suggest that to discriminate between estimators further research is needed concerning the properties of those estimators when both N and T are small or when only T or N is large. Furthermore, an extension of the present analysis to the case of dynamic models seems desirable. Results of Anderson and Hsiao [1981, 1982] for the single equation model suggest that such an extension will have to pay particular attention to initial conditions. Also, a generalization of the present model to the nonlinear case seems of interest.

University of Maryland, U. S. A.

APPENDIX A

PROOF OF THEOREM 1. By assumption, the disturbance vector u is distributed multivariate normal with zero mean and variance covariance matrix Σ. The log-likelihood function of the sample is hence given by

$$\text{(A.1)}\qquad \pounds(\alpha, \delta, \Sigma_\mu, \Sigma_\lambda, \Sigma_0 | Y, Z) = \text{const} + NT \ln\{|I-B|\} + \frac{1}{2}\ln\{|\Sigma^{-1}|\} - \frac{1}{2}u'\Sigma^{-1}u.$$

We use in the following the conventions on matrix differentiation of Dhrymes [1978, pp. 523–540]. The first order derivatives of $\pounds(\cdot)$ with respect to the unrestricted parameters are given by

$$\text{(A.2)}\qquad \begin{aligned} &\frac{\partial \pounds}{\partial \alpha'} = -\frac{1}{2}\frac{\partial u'\Sigma^{-1}u}{\partial \alpha'}, \quad \frac{\partial \pounds}{\partial \delta'} = NT\cdot\frac{\partial \ln\{|I-B|\}}{\partial \delta'} - \frac{1}{2}\frac{\partial u'\Sigma^{-1}u}{\partial \delta'}, \\ &\frac{\partial \pounds}{\partial \Sigma_s} = +\frac{1}{2}\frac{\partial \ln\{|\Sigma^{-1}|\}}{\partial \Sigma_s} - \frac{1}{2}\frac{\partial u'\Sigma^{-1}u}{\partial \Sigma_s} \qquad s=\mu, \lambda, 0.^{1} \end{aligned}$$

The first order conditions for a maximum of the log-likelihood function are obtained by equating the above derivatives to zero. We need to find explicit expressions for them. Proposition 90 in Dhrymes [1978, p. 523] implies that $\partial \operatorname{vec}(I-B)/\partial\delta = -\operatorname{diag}_M[I_M, O_{M\times K})]L$. Applying Proposition 102 in Dhrymes [1978, p. 533] and recognizing that $\operatorname{diag}_M[(I_M, 0_{M\times K})'] \operatorname{vec}[(I-B')^{-1}] = \operatorname{vec}\{[(I-B)^{-1}, 0_{M\times K}]'\}$ then yields

$$\text{(A.3)}\qquad \partial \ln\{|I-B|\}/\partial\delta' = -L' \operatorname{vec}\{[(I-B)^{-1}, 0_{M\times K}]'\}.$$

We adopt the following notation: σ_s^{jl} denotes the (j, l)-th element of Σ_s^{-1} and $\sigma^{\cdot j}$ and $\sigma^{l\cdot}$, respectively, the j-th column and the l-th row of Σ_s^{-1} for $s=0,\ldots, 3$. Baltagi [1980] showed that $\Sigma^{-1} = \Sigma_{h=0}^{3}\Sigma_h^{-1}\otimes Q_h$. By using the results given in Remark 45 of Dhrymes [1978, p. 522] we get $u'\Sigma^{-1}u = \operatorname{tr}\{\Sigma_{h=0}^{3}\Sigma_h^{-1}U'Q_hU\}$. Corollary 38 in Dhrymes [1978, p. 540] implies that

$$\partial\Sigma_1^{-1}/\partial\sigma_{\mu jl} = -\Sigma_1^{-1}\{\partial(\Sigma_0 + T\Sigma_\mu)/\partial\sigma_{\mu jl}\}\Sigma_1^{-1} = -T\sigma_1^{\cdot j}\sigma_1^{l\cdot}.^{2}$$

Similarly it is seen that $\partial\Sigma_3^{-1}/\partial\sigma_{\mu jl} = -T\sigma_3^{\cdot j}\sigma_3^{l\cdot}$. From these results and since $\partial\Sigma_0^{-1}/\partial\sigma_{\mu jl} = \partial\Sigma_2^{-1}/\partial\sigma_{\mu jl} = 0$ it follows that

$$\text{(A.4a)}\qquad \partial[u'\Sigma^{-1}u]/\partial\sigma_{\mu jl} = -T\operatorname{tr}\{\sigma_1^{\cdot j}\sigma_1^{l\cdot}U'Q_1U + \sigma_3^{\cdot j}\sigma_3^{l\cdot}U'Q_3U\}$$

[1] Note that $\sigma_{sjl} = \sigma_{slj}$ for $s=\mu, \lambda, 0$. It turns out that differentiating $\pounds(\cdot)$ with respect to the elements of Σ_μ, Σ_λ and Σ_0 without taking those symmetry restrictions into account leads to the same first order conditions as when those restrictions are taken into account. Since it simplifies the presentation, we neglect those restrictions in the subsequent computation of the derivatives.

[2] Note that $\partial\Sigma_\mu/\partial\sigma_{\mu jl} = e_{\cdot j}e_{l\cdot}$ where $e_{\cdot j}$ and $e_{l\cdot}$ are the j-th column and the l-th row of the $M\times M$ identity matrix.

$$= - T\sigma_1^{l\cdot} U' Q_1 U \sigma_1^{\cdot j} - T\sigma_3^{l\cdot} U' Q_3 U \sigma_3^{\cdot j}.$$

Recall that $\Sigma = \Sigma_{h=0}^{3} \Sigma_h \otimes Q_h$. Noting that the idempotent matrices Q_0, Q_1, Q_2 and Q_3 are orthogonal to each other with $\mathrm{tr}(Q_0) = NT - T - N + 1$, $\mathrm{tr}(Q_1) = N - 1$, $\mathrm{tr}(Q_2) = T - 1$ and $\mathrm{tr}(Q_3) = 1$ and applying Corollary 30 in Dhrymes [1978, p. 534] it is further seen that

(A.4b) $$\partial[\ln|\Sigma^{-1}|]/\partial\sigma_{\mu jl} = \mathrm{tr}\{\Sigma(\partial\Sigma^{-1}/\partial\sigma_{\mu jl})\} = - T\mathrm{tr}\{\Sigma_1 \sigma_1^{\cdot j}\sigma_1^{l\cdot} \otimes Q_1 + \Sigma_3 \sigma_3^{\cdot j}\sigma_3^{l\cdot} \otimes Q_3\} = - \sigma_1^{lj} T(N-1) - \sigma_3^{lj} T.^3$$

Analogously we obtain

(A.4c) $$\partial[u'\Sigma^{-1}u]/\partial\sigma_{\lambda jl} = - N\sigma_2^{l\cdot} U' Q_2 U \sigma_2^{\cdot j} - N\sigma_3^{l\cdot} U' Q_3 U \sigma_3^{\cdot j},$$

(A.4d) $$\partial[\ln|\Sigma^{-1}|]/\partial\sigma_{\lambda jl} = - \sigma_2^{lj} N(T-1) - \sigma_3^{lj} N,$$

(A.4e) $$\partial[u'\Sigma^{-1}u]/\partial\sigma_{0jl} = - \sigma_0^{l\cdot} U' Q_0 U \sigma_0^{\cdot j} - \sigma_1^{l\cdot} U' Q_1 U \sigma_1^{\cdot j} - \sigma_2^{l\cdot} U' Q_2 U \sigma_2^{\cdot j} - \sigma_3^{l\cdot} U' Q_3 U \sigma_3^{\cdot j},$$

(A.4f) $$\partial[\ln|\Sigma^{-1}|]/\partial\sigma_{0jl} = - \sigma_0^{lj}(NT - T - N + 1) - \sigma_1^{lj}(N-1) - \sigma_2^{lj}(T-1) - \sigma_3^{lj}.$$

By assumption only the first S equations contain an intercept so that $u = y - [(I_S, 0_{S\times M-S})' \otimes e_{NT}]\alpha - X\delta$. Using the results (A.3), (A.4) and Propositions 94 and 95 in Dhrymes [1978, p. 526] we obtain the following set of first order conditions by equating the derivatives (A.2) equal to zero:

(A.5a) $$\frac{\partial £}{\partial\alpha'} = [(I_S, 0_{S\times M-S}) \otimes e'_{NT}]\hat{\Sigma}^{-1}\hat{u} = 0,$$

(A.5b) $$\frac{\partial £}{\partial\delta'} = - NT \cdot L' \operatorname{vec}\begin{bmatrix} (I - \hat{B}')^{-1} \\ 0_{K\times M} \end{bmatrix} + X'\hat{\Sigma}^{-1}\hat{u} = 0,$$

(A.5c) $$2\frac{\partial £}{\partial\Sigma_\mu} = - T(N-1)\hat{\Sigma}_1^{-1} - T\hat{\Sigma}_3^{-1} + T\hat{\Sigma}_1^{-1}\hat{U}'Q_1\hat{U}\hat{\Sigma}_1^{-1} + T\hat{\Sigma}_3^{-1}\hat{U}'Q_3\hat{U}\hat{\Sigma}_3^{-1} = 0,$$

(A.5d) $$2\frac{\partial £}{\partial\Sigma_\lambda} = - N(T-1)\hat{\Sigma}_2^{-1} - N\hat{\Sigma}_3^{-1} + N\hat{\Sigma}_2^{-1}\hat{U}'Q_2\hat{U}\hat{\Sigma}_2^{-1} + N\hat{\Sigma}_3^{-1}\hat{U}'Q_3\hat{U}\hat{\Sigma}_3^{-1} = 0,$$

(A.5e) $$2\frac{\partial £}{\partial\Sigma_0} = - (N-1)(T-1)\hat{\Sigma}_0^{-1} - (N-1)\hat{\Sigma}_1^{-1} - (T-1)\hat{\Sigma}_2^{-1} - \hat{\Sigma}_3^{-1} + \hat{\Sigma}_0^{-1}\hat{U}'Q_0\hat{U}\hat{\Sigma}_0^{-1} + \hat{\Sigma}_1^{-1}\hat{U}'Q_1\hat{U}\hat{\Sigma}_1^{-1} + \hat{\Sigma}_2^{-1}\hat{U}'Q_2\hat{U}\hat{\Sigma}_2^{-1} + \hat{\Sigma}_3^{-1}\hat{U}'Q_3\hat{U}\hat{\Sigma}_3^{-1} = 0,$$

[3] Note that $\mathrm{tr}\{\Sigma_s \sigma_s^{\cdot j}\sigma_s^{l\cdot} \otimes Q_s\} = \mathrm{tr}\{\Sigma_s \sigma_s^{\cdot j}\sigma_s^{l\cdot}\}\cdot\mathrm{tr}(Q_s)$ and $\mathrm{tr}\{\Sigma_s \sigma_s^{\cdot j}\sigma_s^{l\cdot}\} = \sigma_s^{lj}$ for $s = 1, 3$.

with

(A.5f) $$\hat{u} = \mathrm{vec}(\hat{U}) = y - [(I_S, 0_{S\times M-S})'\otimes e_{NT}]\hat{\alpha} - X\hat{\delta}$$

(A.5g) $$\hat{\Sigma}^{-1} = \sum_{h=0}^{3} \hat{\Sigma}_h^{-1}\otimes Q_h, \quad \hat{\Sigma}_0 = \hat{\Sigma}_1 + \hat{\Sigma}_2 - \hat{\Sigma}_3$$

and where $\hat{\Sigma}_\mu = (\hat{\Sigma}_1 - \hat{\Sigma}_0)/T$ and $\hat{\Sigma}_\lambda = (\hat{\Sigma}_2 - \hat{\Sigma}_0)/N$. Dividing (A.5c) by T and (A.5d) by N and subtracting the resulting equations from (A.5e) yields

(A.5h) $$-(N-1)(T-1)\hat{\Sigma}_0^{-1} + \hat{\Sigma}_3^{-1} + \hat{\Sigma}_0^{-1}\hat{U}'Q_0\hat{U}\hat{\Sigma}_0^{-1} - \hat{\Sigma}_3^{-1}\hat{U}'Q_3\hat{U}\hat{\Sigma}_3^{-1} = 0.$$

Equation (A.5h) corresponds to the first of the normal equations (9d) of Theorem 1. The other two equations of (9d) are obtained by substituting (A.5h) into (A.5c) and (A.5d). Note that e_{NT} is othogonal to Q_0, Q_1 and Q_2; consequently,

(A.6) $$[(I_S, 0_{S\times M-S})\otimes e'_{NT}]\hat{\Sigma}^{-1} = (I_S, 0_{S\times M-S})\hat{\Sigma}_3^{-1}\otimes e'_{NT}.$$

Making use of this result and (A.5f) it is readily seen that (A.5a) implies (9b) of the normal equations as given by Theorem 1. To obtain the instrumental variable equations (9a) note from (A.5c, d, h) that

(A.7) $$NT.I_M = \sum_{h=0}^{3} \hat{U}'Q_h\hat{U}\hat{\Sigma}_h^{-1}.$$

Further let $\underline{X} = [Y, Z]$ so that $X = (I_M\otimes\underline{X})L$; then by applying Corollary 33 in Dhrymes [1978, p. 520] to (A.5b) it is readily seen that

(A.8) $$\frac{\partial \pounds}{\partial\delta'} = L'\,\mathrm{vec}\,\{[NT(I-\hat{B})^{-1}, 0_{M\times K}]' - \underline{X}'[\sum_{h=0}^{3} Q_h\hat{U}\hat{\Sigma}_h^{-1}]\} = 0.$$

Now let $\hat{\underline{X}} = [\hat{Y}, Z]$ with $\hat{Y} = [e_{NT}, Z][\hat{a}', \hat{C}']'(I-\hat{B})^{-1}$, and $\hat{X} = (I_M\otimes\hat{\underline{X}})L = \mathrm{diag}_M(\hat{X}_i)$. Since $(I-\hat{B}')^{-1}\hat{U}' = Y' - \hat{Y}'$ it follows that by premultiplying (A.7) with $(I-\hat{B}')^{-1}$ and substituting the result into (A.8) we get

(A.9) $$L'\,\mathrm{vec}\,\{\hat{\underline{X}}'[\sum_{h=0}^{3} Q_h\hat{U}\Sigma_h^{-1}]\} = 0.$$

Applying Corollary 23 in Dhrymes [1978, p. 520] to (A.9) and making use of (A.5f) and (A.6) yields

(A.10) $$\hat{X}'\hat{\Sigma}^{-1}\hat{u} = \hat{X}'\{[\hat{\Sigma}_3^{11}, \hat{\Sigma}_3^{12}]'\otimes e_{NT}\}\hat{\alpha} + \hat{X}'[\sum_{h=0}^{3} \hat{\Sigma}_h^{-1}\otimes Q_h][X\hat{\delta} - y] = 0.$$

Premultiplying (9b) with $\hat{X}'\{[\hat{\Sigma}_3^{11}, \hat{\Sigma}_3^{12}]'\otimes e_{NT}\}$, subtracting the result from (A.10) and observing that $\hat{\Sigma}_{3,22}^{-1} = \hat{\Sigma}_3^{22} - \hat{\Sigma}_3^{21}(\hat{\Sigma}_3^{11})^{-1}\hat{\Sigma}_3^{12}$ yields (9a). Q.E.D.

APPENDIX B

PROOF OF THEOREM 2. To prove the theorem we first show that

(B.1) $$\mathrm{plim}\frac{1}{NT}\{\sum_{h=0}^{3} \tilde{X}'_{(h)}(\tilde{\Sigma}_h^{-1}\otimes Q_h)X - \bar{X}'(\Sigma_0^{-1}\otimes Q_0)X\} = 0,$$

(B.2) $$\text{plim}\, \tilde{X}'_{(h)}(\tilde{\Sigma}_h^{-1}\otimes Q_h)u/\sqrt{NT} = 0 \qquad h = 1, 2,$$

(B.3) $$\text{plim}\, \{\tilde{X}'_{(0)}(\tilde{\Sigma}_0^{-1}\otimes Q_0)u/\sqrt{NT} - \bar{X}(\Sigma_0^{-1}\otimes Q_0)u/\sqrt{NT}\} = 0,$$

(B.4) $$\bar{X}'(\Sigma_0^{-1}\otimes Q_0)u/\sqrt{NT} \xrightarrow{i.d.} N(0, \Omega).$$

By assumption, the matrices $Z'Z/NT$ and $Z'Q_0Z/NT$ converge to finite nonsingular limiting matrices as both N and T tend to infinity; we define $M_{ZQZ} = \lim Z'Q_0Z/NT$. This implies that also the matrices $Z'Q_hZ/NT$ converge for $h=1, 2$ to finite limiting matrices. Since convergence in probability to zero can be proven by showing that both the sequences of means and variances converge to zero and since the matrices Q_h are orthogonal to each other, it is not difficult to see from (5) that

(B.4) $$\text{plim}\, (Z'Q_hu_j)/NT = 0 \qquad h = 0, 1, 2,$$

(B.5) $$\text{plim}\, (Z'Q_1u_j)/(T^{1-\varepsilon}\sqrt{NT}) = \text{plim}\, (Z'Q_2u_j)/(N^{1-\varepsilon}\sqrt{NT}) = 0$$

for all $0<\varepsilon<1/2$. Let $\tilde{Y}=Z\tilde{\Pi}$ where $\tilde{\Pi}$ is some consistent estimator for Π; since $Y=e_{NT}\pi+Z\Pi+U(I-B)^{-1}$ it follows from (B.4) and $Q_he_{NT}=0$ for $h=0, 1, 2$ that

(B.6) $$\text{plim}\, \frac{1}{NT}[\tilde{Y}, Z]'Q_h[Y, Z] = [\Pi, I_K]'[\lim \frac{1}{NT}(Z'Q_hZ)][\Pi, I_K] \qquad h = 0, 1, 2.$$

Note that $\tilde{X}'_{(h)}(\tilde{\Sigma}_h^{-1}\otimes Q_h)X = L'\{\tilde{\Sigma}_h^{-1}\otimes[\tilde{Y}_{(h)}, Z]'Q_h[Y, Z]\}L$. Since $\text{plim}\tilde{\Sigma}_1^{-1} = \text{plim}\,\tilde{\Sigma}_2^{-1}=0$ result (B.1) then follows from (B.6) and the fact that the matrices $Z'Q_hZ/NT$ converge to finite limiting matrices. (B.6) implies further that

(B.7) $$\Omega = \text{plim}\, \bar{X}'(\Sigma_0^{-1}\otimes Q_0)X/NT = \text{plim}\, \bar{X}'(\Sigma_0^{-1}\otimes Q_0)\bar{X}/NT = R'(\Sigma_0^{-1}\otimes M_{ZQZ})R$$

with $R=\text{diag}_M[(\Pi, I_K)L_j]$. Since each equation is identified subject to zero type parameter restrictions it follows (along the lines of Schmidt [1976, pp. 205–207]) that R is of full column rank. Since Σ_0 and M_{ZQZ} are nonsingular, it follows also that Ω is nonsingular. Next, note that $\tilde{X}'_{(1)}(\tilde{\Sigma}_1^{-1}\otimes Q_1)u/\sqrt{NT} = L'\{T^{1-\varepsilon}\tilde{\Sigma}_1^{-1}\otimes [\tilde{\Pi}_1, I_K]\}\{(I_M\otimes Z'Q_1)u\}/(T^{1-\varepsilon}\sqrt{NT})$. Result (B.2) is then readily seen to hold for $h=1$ because of (B.5) and since $\text{plim}\, T^{1-\varepsilon}\tilde{\Sigma}_1^{-1}=0$. The other probability limit in (B.2) follows analogously.

Because of (3) and $Q_0(I_N\otimes e_T)=Q_0(e_N\otimes I_T)=0$ we have $Q_0u_j=Q_0v_j$. By assumption $v=(P\otimes I_{NT})\xi$ where ξ is a $MNT\times 1$ vector of i.i.d. random variables with zero mean and unit variance and $\Sigma_0=PP'$. Applying a standard central limit theorem (compare e.g. Theil [1971, p. 380–381]) it then follows that $(I_M\otimes Z'Q_0)u/\sqrt{NT}=(I_M\otimes Z'Q_0)v/\sqrt{NT} \xrightarrow{i.d.} N(0, \Sigma_0\otimes M_{ZQZ})$. This result and the consistency of $\tilde{\Pi}$, $\tilde{\Sigma}_0$ are then readily seen to imply (B.3) and (B.4) since $\tilde{X}'_{(0)}(\tilde{\Sigma}_0^{-1}\otimes Q_0)u/\sqrt{NT}=L'(\tilde{\Sigma}_0^{-1}\otimes[\tilde{\Pi}_0, I_K]')(I_M\otimes Z'Q_0)u/\sqrt{NT}$ and $\bar{X}'(\Sigma_0^{-1}\otimes Q_0)u/\sqrt{NT}=L'(\Sigma_0^{-1}\otimes[\Pi, I_K]')(I_M\otimes Z'Q_0)u/\sqrt{NT}$.

Clearly $\tilde{\delta}=\delta+[\Sigma_{h=0}^2 \tilde{X}'_{(h)}(\tilde{\Sigma}_h^{-1}\otimes Q_h)X]^{-1}\Sigma_{h=0}^2 \tilde{X}'_{(h)}(\tilde{\Sigma}_h^{-1}\otimes Q_h)u$ and $\bar{\delta}_{\text{FIDV}}=\delta+[\bar{X}'(\Sigma_0^{-1}\otimes Q_0)X]^{-1}\bar{X}'(\Sigma_0^{-1}\otimes Q_0)u$; combining (B.1)–(B.4) and (B.6) it is then not difficult to see that plim $\sqrt{NT}(\tilde{\delta}-\tilde{\delta}_{\text{FIDV}})=0$ and $\sqrt{NT}(\tilde{\delta}_{\text{FIDV}}-\delta) \xrightarrow{i.d.} N(0, \Omega)$.

Q. E. D.

REFERENCES

AMEMIYA, T., "The Estimation of the Variances in a Variance-Components Model," *International Economic Review*, 12 (February, 1971), 1–13.

ANDERSON, T. W. AND C. HSIAO, "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76 (September, 1981), 598–606.

——— AND ——— "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18 (January, 1982), 47–82.

AVERY, R. B., "Error Components and Seemingly Unrelated Regressions," *Econometrica*, 45 (January, 1977), 199–209.

BALESTRA, P. AND M. NERLOVE, "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," *Econometrica*, 34 (July, 1966), 585–612.

BALTAGI, B. M., "On Seemingly Unrelated Regressions with Error Components," *Econometrica*, 48 (September, 1980), 1547–1551.

———, "Simultaneous Equations with Error Components," *Journal of Econometrics*, 17 (September, 1981), 189–200.

BRUNDY, J. M. AND D. W. JORGENSON, "Efficient Estimation of Simultaneous Equations by Instrumental Variables," *Review of Economics and Statistics*, 53 (August, 1971), 207–224.

——— AND ——— "Consistent and Efficient Estimation of Systems of Simultaneous Equations by Means of Instrumental Variables," in, P. Zarembka, ed., *Frontiers in Econometrics* (New York: Academic Press, 1974), 215–244.

DHRYMES, P. J., *Introductory Econometrics* (New York: Springer Verlag, 1978).

ENGLE, R. F., D. F. HENDRY AND J. -F. RICHARD, "Exogeneity," *Econometrica*, 51 (March, 1983), 277–304.

FULLER, W. A. AND G. E. BATTESE, "Estimation of Linear Models with Crossed-Error Structure," *Journal of Econometrics*, 2 (May, 1974), 67–78.

HAUSMAN, J. A., "Full Information Instrumental Variables Estimation of Simultaneous Equation Models," *Annals of Economic and Social Measurement*, 3 (October, 1974), 641–652.

———, "An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models," *Econometrica*, 43 (July, 1975), 727–738.

HENDRY, D. F., "The Structure of Simultaneous Equation Estimators," *Journal of Econometrics*, 4 (February, 1976), 51–88.

MADDALA, G. S., "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica*, 39 (March, 1971), 341–358.

MAGNUS, J. R., "Multivariate Error Components Analysis of Linear and Nonlinear Regression Models by Maximum Likelihood," *Journal of Econometrics*, 19 (August, 1982), 239–285.

MUNDLAK, Y., "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46 (January, 1978), 69–86.

NERLOVE, M., "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections," *Econometrica*, 39 (March, 1971a), 359–382.

———, "A Note on Error Components Models," *Econometrica*, 39 (March, 1971b), 383–396.

PRUCHA, I. R., "On the Asymptotic Efficiency of Feasible Aitken Estimators for Seemingly Unrelated Regression Models with Error Components," *Econometrica*, 52 (January, 1984), 203–207.

SWAMY, P. A. V. B. AND S. S. ARORA, "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models," *Econometrica*, 40 (March, 1972), 261–275.

THEIL, H., *Principles of Econometrics* (New York: Wiley, 1971).

WALLACE, T. D. AND A. HUSSAIN, "The Use of Error Components Models in Combining Cross Section with Time Series Data," *Econometrica*, 37 (January, 1969), 55–72.