Econometrica, Vol. 49, No. 6 (November, 1981)

PANEL DATA AND UNOBSERVABLE INDIVIDUAL EFFECTS[1]

BY JERRY A. HAUSMAN AND WILLIAM E. TAYLOR[2]

An important purpose in combining time-series and cross-section data is to control for individual-specific unobservable effects which may be correlated with other explanatory variables. Using exogeneity restrictions and the time-invariant characteristic of the latent variable, we derive (i) a test for the presence of this effect and for the over-identifying restrictions we use, (ii) necessary and sufficient conditions for identification, and (iii) the asymptotically efficient instrumental variables estimator and conditions under which it differs from the within-groups estimator. We calculate efficient estimates of a wage equation from the Michigan income dynamics data which indicate substantial differences from within-groups or Balestra-Nerlove estimates—particularly, a significantly higher estimate of the returns to schooling.

1. INTRODUCTION

AN IMPORTANT BENEFIT from pooling time-series and cross-section data is the ability to control for individual-specific effects—possibly unobservable—which may be correlated with other included variables in the specification of an economic relationship. Analysis of cross-section data alone can neither identify nor control for such individual effects. To consider a specific model, let

$$Y_{it} = X_{it}\beta + Z_i\gamma + \alpha_i + \eta_{it} \qquad (i = 1, \dots, N;\ t = 1, \dots, T), \tag{1.1}$$

where β and γ are k and g vectors of coefficients associated with time-varying and time-invariant observable variables respectively. The disturbance η_{it} is assumed uncorrelated with the columns of (X, Z, α) and has zero mean and constant variance σ_η^2 conditional on X_{it} and Z_i. The latent individual effect α_i is assumed to be a time-invariant random variable, distributed independently across individuals, with variance σ_α^2.

Our primary focus is the potential correlation of α_i with the columns of X and Z. In the presence of such correlations, least squares (OLS) and generalized least squares (GLS) yield biased and inconsistent estimates of the parameters $(\beta, \gamma, \sigma_\eta^2, \sigma_\alpha^2)$. The traditional technique to overcome this problem is to eliminate the individual effects in the sample by transforming the data into deviations from individual means.[3] Unfortunately, the OLS coefficient estimates from the transformed data (known as "within-groups" or "fixed effects" estimators) have two important defects: (1) *all* time-invariant variables are eliminated by the transformation, so that γ cannot be estimated, and (2) under certain circumstances, the within-groups estimator is not fully efficient since it ignores variation across

[1] Early versions of this paper were presented at the Econometric Society Summer Meetings in Montreal and at the Fourth World Congress in Aix en Provence.

[2] Hausman thanks the NSF for financial support. Rob Engle and Paul Ruud have given useful comments.

[3] This technique corresponds to Model I of the analysis of variance, e.g., Scheffé [**19**]. When used in analysis of covariance, errors in measured variables can create a serious problem since they are exacerbated by the data transformation.

individuals in the sample. The first problem is generally the more serious since in some applications, primary interest is attached to the unknown coefficients of the time-invariant variables, e.g., to the coefficients of schooling in the wage equation below.

Another possible approach in the simultaneous equations model is to find instruments for those columns of X and Z which are potentially correlated with α_i. However, it may be difficult to find appropriate instruments, excluded from equation (1.1), which are uncorrelated with α_i, and, in any case, such procedures ignore the time-invariant characteristic of the latent effect.

Specifications similar to equation (1.1) have been used in at least two empirical contexts. In the cost and production function literature where α_i denotes the managerial efficiency of the ith firm, Hoch [**9**] and Mundlak [**14**] have suggested the use of the within-groups estimator to produce unbiased estimates of the remaining parameters. If Y_{it} denotes the wages of the ith individual in the tth time period, one of the Z_i's measures his schooling, and α_i denotes his unmeasurable ability or ambition, then equation (1.1) represents a standard specification for measuring the returns to education. To the extent that ability and schooling are correlated, the OLS estimates are biased and inconsistent. Griliches [**4**] has relied on an instrumental variables approach, using family background variables excluded from equation (1.1) as instruments.[4] Another approach is the factor analytic model, pioneered in this context by Jöreskog [**10**] and applied to the schooling problem by Chamberlain and Griliches [**2**] and Chamberlain [**1**]. This model relies for identification upon orthogonality assumptions imposed upon observable and unobservable components of α_i. The method we propose does not assume a specification of the components of α_i and may be less sensitive to our lack of knowledge about the unobservable individual-specific effect.

Instead, our method uses assumptions about the correlations between the columns of (X, Z) and α_i. If we are willing to assume that certain variables among the X and Z are uncorrelated with α_i, then conditions may hold such that all of the β's and γ's may be consistently and efficiently estimated. Intuitively, the columns of X_{it} which are uncorrelated with α_i can serve two functions because of their variation across both individuals and time: (i) using deviations from individual means, they produce unbiased estimates of the β's, and (ii) using the individual means, they provide valid instruments for the columns of Z_i that are correlated with α_i.

One needs to be quite careful in choosing among the columns of X for those variables which are uncorrelated with α_i. For instance, in the returns to schooling example, it may be safe to assume that ability is uncorrelated with health or age but less so with unemployment. An important feature of our method is that, in certain circumstances, the non-correlation assumptions can be tested, so that the method need not rely on totally a priori assumptions.

[4] A specification test developed below suggests that, in our sample, family background is likely to be correlated with the latent individual effect.

The plan of the paper is as follows. In Section 2, we formally set up the model and consider the standard estimation procedures. Using these procedures, we develop and compare three specification tests for correlation between α_i and the columns of X and Z, generalizing results of Hausman [**6**]. In the event of such correlation, we propose a consistent but inefficient estimator of all the parameters of the model. In Section 3, we find conditions under which the parameters are identified and develop an efficient instrumental variables estimator that accounts for the variance components structure of the disturbance. We also derive a test of the correlation assumptions necessary for identification and estimation, applying results from Hausman and Taylor [**8**]. Finally, in Section 4, we apply the procedure to an earnings function, focusing on the returns to schooling. The results indicate that when the correlation of α_i with the independent variables is taken into account, traditional estimates of the return to schooling are revised markedly.

2. PRELIMINARIES

2.1 *Conventional Estimation*

We begin by developing the model in equation (1.1) slightly and examining properties of conventional estimators in the absence and presence of specification errors of the form $E(\alpha_i \mid X_{it}, Z_i) \neq 0$. Thus

$$\begin{aligned} Y_{it} &= X_{it}\beta + Z_i\gamma + \epsilon_{it}, \\ \epsilon_{it} &= \alpha_i + \eta_{it}, \end{aligned} \tag{2.1}$$

where we have reason to believe that $E(\epsilon_{it} \mid X_{it}, Z_i) = E(\alpha_i \mid X_{it}, Z_i) \neq 0$. Note that, somewhat unconventionally, X_{it} and Z_i denote $TN \times k$ and $TN \times g$ matrices respectively, whose subscripts indicate variation over individuals $(i = 1, \ldots, N)$ and time $(t = 1, \ldots, T)$. Observations are ordered first by individual and then by time, so that α_i and each column of Z_i are TN vectors having blocks of T identical entries within each $i = 1, \ldots, N$.

The prior information our procedure uses is the ability to distinguish columns of X and Z which are asymptotically uncorrelated with α_i from those which are not. For fixed T, let

$$\begin{aligned} &\operatorname*{plim}_{N\to\infty} \frac{1}{N} X'_{1it}\alpha_i = \mathbf{0}, \qquad \operatorname*{plim}_{N\to\infty} \frac{1}{N} Z'_{1i}\alpha_i = \mathbf{0}, \\ &\operatorname*{plim}_{N\to\infty} \frac{1}{N} X'_{2it}\alpha_i = h_x, \qquad \operatorname*{plim}_{N\to\infty} \frac{1}{N} Z'_{2i}\alpha_i = h_z, \end{aligned} \tag{2.2}$$

where $X_{it} = [X_{1it} \vdots X_{2it}]$ of dimensions $[TN \times k_1 \vdots TN \times k_2]$, $Z_i = [Z_{1i} \vdots Z_{2i}]$ of dimensions $[TN \times g_1 \vdots TN \times g_2]$, and the k_2, g_2 vectors h_x, h_z are assumed unequal to zero.

It will prove helpful to recall the menu of conventional estimators for (β, γ) in equation (2.1), ignoring the misspecification in equations (2.2). Letting ι_T denote a T vector of ones, two convenient orthogonal projection operators can be defined as

$$P_V = \left[I_N \otimes \frac{1}{T}\iota_T\iota_T' \right], \qquad Q_V = I_{NT} - P_V,$$

which are idempotent matrices of rank N and $TN - N$ respectively. With data grouped by individuals, P_V transforms a vector of observations into a vector of group means: i.e., $P_V Y_{it} = (1/T)\sum Y_{it} \equiv Y_{i\cdot}$. Similarly Q_V produces a vector of deviations from group means: i.e., $Q_V Y_{it} = \tilde{Y}_{it} = Y_{it} - Y_{i\cdot}$. Moreover, Q_V is orthogonal by construction to any time-invariant vector of observations: $Q_V Z_i = Z_i - (1/T)\sum_{t=1}^{T} Z_i = \mathbf{0}$.

Transform model (2.1) by Q_V, obtaining

$$Q_V Y_{it} = Q_V X_{it}\beta + Q_V Z_i\gamma + Q_V\alpha_i + Q_V\eta_{it},$$

which simplifies to

$$\tilde{Y}_{it} = \tilde{X}_{it}\beta + \tilde{\eta}_{it}. \tag{2.3}$$

Least squares estimates of β in equation (2.3) are Gauss-Markov (for the transformed equation) and define the within-groups estimator

$$\hat{\beta}_W = (X_{it}'Q_V X_{it})^{-1} X_{it}'Q_V Y_{it} \equiv (\tilde{X}_{it}'\tilde{X}_{it})^{-1}\tilde{X}_{it}'\tilde{Y}_{it}.$$

Since the columns of $\tilde{X}_{it}$ are uncorrelated with $\tilde{\eta}_{it}$, $\hat{\beta}_W$ is unbiased and consistent for β regardless of possible correlation between α_i and the columns of X_{it} or Z_i. The sum of squared residuals from this equation can be used to obtain an unbiased and consistent estimate of σ_η^2, as we shall see shortly.

To make use of between-group variation, transform model (2.1) by P_V, obtaining

$$Y_{i\cdot} = X_{i\cdot}\beta + Z_i\gamma + \alpha_i + \eta_{i\cdot}. \tag{2.4}$$

Least squares estimates of β and γ in equation (2.4) are known as between-groups estimators (denoted $\hat{\beta}_B$ and $\hat{\gamma}_B$) and because of the presence of α_i, both $\hat{\beta}_B$ and $\hat{\gamma}_B$ are biased and inconsistent if $E(\alpha_i \mid X_{it}, Z_i) \neq 0$. Similarly, the sum of squared residuals from equation (2.4) provides a biased and inconsistent estimator for $\mathrm{var}(\alpha_i + \eta_{i\cdot}) = \sigma_\alpha^2 + (1/T)\sigma_\eta^2$ when $E(\alpha_i \mid X_{it}, Z_i) \neq 0$.

In the absence of misspecification, the optimal use of within and between groups information is a straightforward calculation. From equation (2.1), $\mathrm{cov}(\epsilon_{it} \mid X_{it}, Z_i) \equiv \Omega = \sigma_\eta^2 I_{TN} + \sigma_\alpha^2[I_N \otimes \iota_T\iota_T'] = \sigma_\eta^2 I_{TN} + T\sigma_\alpha^2 P_V$, a familiar block-diagonal matrix. Observe that the problem is merely a linear regression equation with a non-scalar disturbance covariance matrix. Assuming α_i, η_{it} to be normally distributed, the within-groups and between-groups coefficient estimators and the sums of squared residuals from equations (2.3) and (2.4) are jointly sufficient statistics for $(\beta, \gamma, \sigma_\alpha^2, \sigma_\eta^2)$. The Gauss-Markov estimator, then, is the minimum variance matrix-weighted average of the within and between groups estimators, where the weights depend upon the variance components. Following Maddala [**13**], the solution can be written as

$$\begin{pmatrix} \hat{\beta}_{GLS} \\ \hat{\gamma}_{GLS} \end{pmatrix} = \tilde{\Delta} \begin{pmatrix} \hat{\beta}_B \\ \hat{\gamma}_B \end{pmatrix} + (I - \tilde{\Delta}) \begin{pmatrix} \hat{\beta}_W \\ \mathbf{0} \end{pmatrix}$$

where $\tilde{\Delta} = (V_B + V_W)^{-1} V_W$ and V_B, V_W denote the covariance matrices of the between and within groups estimators of β and γ. This is frequently known as the Balestra-Nerlove estimator. It requires knowledge of the variance components σ_α^2 and σ_η^2 but consistent estimates may be substituted for the variance components without loss of asymptotic efficiency.[5] Of course, if $E(\alpha_i \mid X_{it}, Z_i) \neq 0$, this GLS estimator will be biased and if h_x or h_z is unequal to $\mathbf{0}$, it will be inconsistent, since it is a matrix-weighted average of the consistent within-groups estimator and the inconsistent between-groups estimator.

For both numerical and analytical convenience, we can express the Gauss-Markov estimator in a slightly different form. Let $\theta = [\sigma_\eta^2/(\sigma_\eta^2 + T\sigma_\alpha^2)]^{1/2}$. Then:

PROPOSITION 2.1: *The $TN \times TN$ non-singular matrix*

$$\Omega^{-1/2} = \theta P_V + Q_V = I_{TN} - (1 - \theta) P_V$$

transforms the disturbance covariance matrix Ω into a scalar matrix.

PROOF:

$$\Omega^{-1/2} \Omega \Omega^{-1/2} = [\theta P_V + Q_V][\sigma_\eta^2 I_{TN} + T\sigma_\alpha^2 P_V][\theta P_V + Q_V]$$
$$= \theta^2 (\sigma_\eta^2 + T\sigma_\alpha^2) P_V + \sigma_\eta^2 Q_V = \sigma_\eta^2 I_{TN}.$$

This works because the N and $TN - N$ basis vectors of the column spaces of P_V and Q_V respectively span the eigenspaces of Ω corresponding to the two distinct eigenvalues $\sigma_\eta^2 + T\sigma_\alpha^2$ and σ_η^2; see Nerlove [**17**].

Transforming equation (2.1) by $\Omega^{-1/2}$ is thus equivalent to a simple differencing of the observations, as pointed out by Hausman [**6**].

$$\Omega^{-1/2} Y_{it} = \Omega^{-1/2} X_{it} \beta + \Omega^{-1/2} Z_i \gamma + \Omega^{-1/2} \alpha_i + \Omega^{-1/2} \eta_{it},$$

or

$$(2.5) \qquad Y_{it} - (1 - \theta) Y_{i\cdot} = [X_{it} - (1 - \theta) X_{i\cdot}] \beta + \theta Z_i \gamma + \theta \alpha_i + [\eta_{it} - (1 - \theta) \eta_{i\cdot}].$$

[5] The small sample implications of this substitution are explored in Taylor [**20**].

OLS estimates of (β, γ) in equation (2.5) are Gauss-Markov, provided $E(\alpha_i \mid X_{it}, Z_i) = 0$. If misspecification is present, the fact that α_i appears in equation (2.5) means that the GLS estimates will be inconsistent.

2.2 *Specification Tests Using Panel Data*

A crucial assumption of the cross-section regression specification is that the conditional expectation of the disturbances given knowledge of the explanatory variables is zero. An extremely useful property of panel data is that by following the cross-section panel over time, we can test this assumption. Moreover, the fact that the within, between, and GLS estimators are affected differently by the failure of this assumption, suggests that we may base a test of it on functions of these statistics.

The specification tests which we consider test the null hypothesis H_0: $E(\alpha_i \mid X_{it}, Z_i) = \mathbf{0}$ against the alternative $H_1: E(\alpha_i \mid X_{it}, Z_i) \neq \mathbf{0}$. If H_0 is rejected, we might try to reformulate the cross-section specification in the hope of finding a model in which the orthogonality property holds. Alternatively, we might well be satisfied with using an estimator which permits consistent estimation of (β, γ) by controlling for the correlation between the explanatory variables and the latent α_i. An asymptotically efficient procedure for doing this is outlined below.

From the three classical estimators for β in equation (2.1), we are naturally led to construct three different specification tests. Let Δ denote the upper-left $k \times k$ block of $\tilde{\Delta}$.

(i) *GLS vs. within*: Form the vector $\hat{q}_1 = \hat{\beta}_{GLS} - \hat{\beta}_W$. Under $H_0, \operatorname{plim}_{N\to\infty} \hat{q}_1 = \mathbf{0}$, while under $H_1, \operatorname{plim}_{N\to\infty} \hat{q}_1 = \operatorname{plim}_{N\to\infty} \hat{\beta}_{GLS} - \beta \neq \mathbf{0}$. Hausman [6] showed that $\operatorname{cov}(\hat{q}_1) = \operatorname{cov}(\hat{\beta}_W) - \operatorname{cov}(\hat{\beta}_{GLS})$, so that a χ^2 test is easily constructed, based on the length of $\hat{q}_1$. This test has been used fairly frequently and has appeared to be quite powerful.

(ii) *GLS vs. between*: Form the vector $\hat{q}_2 = \hat{\beta}_{GLS} - \hat{\beta}_B$. Again, under H_0, $\operatorname{plim}_{N\to\infty} \hat{q}_2 = \mathbf{0}$, while under $H_1, \operatorname{plim} \hat{q}_2 = (I - \Delta)\operatorname{plim}_{N\to\infty}(\hat{\beta}_B - \beta)$, so that deviations of $\hat{q}_2$ from the zero vector cast doubt upon H_0. Using the asymptotic Rao-Blackwell argument in Hausman [6], $\operatorname{cov}(\hat{q}_2) = \operatorname{cov}(\hat{\beta}_B) - \operatorname{cov}(\hat{\beta}_{GLS})$, which gives rise to another χ^2 statistic.

(iii) *Within vs. between*: Form the vector $\hat{q}_3 = \hat{\beta}_W - \hat{\beta}_B$. Under $H_0, \operatorname{plim}_{N\to\infty} \hat{q}_3 = \mathbf{0}$ and under $H_1, \operatorname{plim}_{N\to\infty} \hat{q}_3 = \beta - \operatorname{plim}_{N\to\infty} \hat{\beta}_B \neq \mathbf{0}$. Also, $\operatorname{cov}(\hat{q}_3) = \operatorname{cov}(\hat{\beta}_W) + \operatorname{cov}(\hat{\beta}_B)$ since the within and between estimators lie in orthogonal subspaces. Thus the length of $\hat{q}_3$ yields a third χ^2 statistic.

In considering these three tests, Hausman [6] conjectured that the first test might be better than the third, since $\operatorname{cov}(\hat{q}_2) \geqq \operatorname{cov}(\hat{q}_1)$; while Pudney [18] conjectured that the second might be better than the third because $\hat{\beta}_{GLS}$ is efficient.[6] (Actually, $\operatorname{cov}(\hat{q}_3) \geqq \operatorname{cov}(\hat{q}_1)$.) However, since $\hat{\beta}_{GLS} = \Delta\hat{\beta}_B + (I - \Delta)\hat{\beta}_W$

[6] Pudney actually considered using estimates of ϵ_{it} from the three estimators and basing tests upon the sample covariance $X'\hat{\epsilon}$, using either the within or GLS estimate of β to form $\hat{\epsilon}$. However, the tests are considerably simpler to apply by directly comparing the estimates of β; Pudney was unaware that using $\hat{\beta}_{GLS}$ to form $\hat{\epsilon}$ is equivalent to the second test.

where the matrix Δ is non-singular,[7] the vectors $(\hat{q}_1, \hat{q}_2, \hat{q}_3)$ are all nonsingular transformations of one another. Thus $\hat{q}_1 = -\Delta\hat{q}_3$ and $\hat{q}_2 = (I - \Delta)\hat{q}_3$. If we denote $\text{cov}(\hat{q}_k)$ by $V_k (k = 1, 2, 3)$,

$$\hat{q}_1' V_1^{-1} \hat{q}_1 = \hat{q}_3' \Delta' [\Delta V_3 \Delta']^{-1} \Delta \hat{q}_3 = \hat{q}_3' V_3^{-1} \hat{q}_3,$$

and

$$\hat{q}_2' V_2^{-1} \hat{q}_2 = \hat{q}_3' (I - \Delta)' [(I - \Delta) V_3 (I - \Delta)']^{-1} (I - \Delta) \hat{q}_3 = \hat{q}_3' V_3^{-1} \hat{q}_3,$$

so that the following proposition holds:

PROPOSITION 2.2: *The chi-square statistics for tests* $(1 - 3)$ *are numerically exactly identical.*

2.3 *Consistent but Inefficient Estimation*

If the specification test of the previous section rejects H_0, it may still be possible to obtain consistent estimates of both β and γ from the within-groups regression. The estimates of the variance components derived from these estimators will be used below in our efficient instrumental variables procedure.

Let

$$\hat{d}_i = Y_{i\cdot} - X_{i\cdot}\hat{\beta}_W = \left[P_V - X_{i\cdot}(X_{it}' Q_V X_{it})^{-1} X_{it}' Q_V \right] Y_{it}$$

be the TN vector of group means estimated from the within-groups residuals. Expanding this expression,

$$\hat{d}_i = Z_i \gamma + \alpha_i + \left[P_V - X_{i\cdot}(\tilde{X}_{it}' \tilde{X}_{it})^{-1} \tilde{X}_{it}' \right] \eta_{it}. \tag{2.6}$$

Treating the last two terms as an unobservable mean zero disturbance, consider estimating γ from equation (2.6). Since α_i is correlated with the columns of Z_{2i}, both OLS and GLS will be inconsistent for γ. Consistent estimation is possible, however, if the columns of X_{1it}—uncorrelated with α_i by assumption—provide sufficient instruments for the columns of Z_{2i} in equation (2.6). A necessary condition for this is clearly that $k_1 \geqq g_2$: that there be at least as many exogenous time-varying variables as there are endogenous time-invariant variables. Note that this condition figures prominently in the discussion of identification in Section 3.

The two stage least squares (2SLS) estimator for γ in equation (2.6) is

$$\hat{\gamma}_W = (Z_i' P_A Z_i)^{-1} Z_i' P_A \hat{d}_i \tag{2.7}$$

where $A = [X_{1it} \vdots Z_{1i}]$ and P_A is the orthogonal projection operator onto its

[7] If σ_α^2 and σ_η^2 are unknown and must be estimated, the above identities hold exactly in finite samples, since the estimated weight matrix $\hat{\Delta}$ is non-singular.

column space. The sampling error is given by

$$\hat{\gamma}_W - \gamma = (Z'P_A Z)^{-1} Z'P_A \Big[\alpha_i + \big\{ P_V - X_{i.}(\tilde{X}_{it}'\tilde{X}_{it})^{-1}\tilde{X}_{it}' \big\} \eta_{it} \Big],$$

and under the usual assumptions governing X and Z, the 2SLS estimator is consistent for γ, since for fixed T, $\text{plim}_{N\to\infty}(1/N)A'\alpha_i = \mathbf{0}$ and $\text{plim}_{N\to\infty}(1/N) X_{it}'\eta_{it} = \mathbf{0}$. The fact that the $\hat{d}_i$ are calculated from the within-groups residuals suggests that if $\hat{\beta}_W$ is not fully efficient, then $\hat{\gamma}_W$ may not be fully efficient either.

Having consistent estimates of β and—under certain circumstances—γ, we can construct consistent estimators for the variance components. First, a consistent estimate of σ_η^2 can always be derived from the within-groups residuals. If $Q_{\tilde{X}}$ denotes $I_{TN} - \tilde{X}_{it}(\tilde{X}_{it}'\tilde{X}_{it})^{-1}\tilde{X}_{it}'$, we can write the sum of squares of within-group residuals as

$$\tilde{Y}_{it}' Q_{\tilde{X}} \tilde{Y}_{it} = \tilde{\eta}_{it}' Q_{\tilde{X}} \tilde{\eta}_{it} = \tilde{\eta}_{it}'\tilde{\eta}_{it} - \tilde{\eta}_{it}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{\eta}_{it}.$$

Thus

$$\begin{aligned}\underset{N\to\infty}{\text{plim}}\, s_\eta^2 &= \underset{N\to\infty}{\text{plim}} \frac{1}{N(T-1)} \tilde{Y}_{it}' Q_{\tilde{X}} \tilde{Y}_{it} = \underset{N\to\infty}{\text{plim}} \frac{1}{N(T-1)} \tilde{\eta}_{it}'\tilde{\eta}_{it} - 0 \\ &= \underset{N\to\infty}{\text{plim}} \frac{1}{N(T-1)} \eta_{it}' Q_V \eta_{it} = \sigma_\eta^2,\end{aligned}$$

since rank $(Q_V) = N(T-1)$.

Finally, whenever we have consistent estimators for both β and γ, a consistent estimator for σ_α^2 can be obtained. Let $s^2 = (1/N)(Y_{i.} - X_{i.}\hat{\beta}_W - Z_i\hat{\gamma}_W)'(Y_{i.} - X_{i.}\hat{\beta}_W - Z_i\hat{\gamma}_W)$; then

$$\underset{N\to\infty}{\text{plim}}\, s^2 = \underset{N\to\infty}{\text{plim}} \frac{1}{N}(\alpha_i + \eta_{i.})'(\alpha_i + \eta_{i.}) = \sigma_\alpha^2 + \frac{1}{T}\sigma_\eta^2$$

so that $s_\alpha^2 = s^2 - (1/T)s_\eta^2$ is consistent for σ_α^2.

3. INSTRUMENTAL VARIABLES ESTIMATORS

3.1 *Identification*

In this section, we discuss identification of some or all of the elements of (β, γ), using only the prior information in equations (2.2) and the time-invariant nature of the latent variable. Because the only component of the disturbance which is correlated with an explanatory variable is time-invariant, any vector orthogonal to a time-invariant vector can be used as an instrument. In particular, $Q_V\alpha_i = \mathbf{0}$ by construction, so that time-invariance provides at least $TN - N$ linearly independent instruments in equation (2.1): namely, $TN - N$ basis vectors of the column space of Q_V.

Unfortunately, as noted in the Introduction, Q_V is also orthogonal to Z_i, which violates the requirement that instruments be correlated with all of the explanatory variables.[8] The familiar results on identification in linear models are easily extended to cover this case: consider the canonical linear simultaneous equations model

$$Y = X\beta + e \tag{3.1}$$

where some of the k columns of X are endogenous and the matrix Z contains T observations on all variables for which $\operatorname{plim}_{T\to\infty}(1/T)Z'e = \mathbf{0}$. Project equation (3.1) onto the column space of Z:

$$P_Z Y = P_Z X\beta + P_Z e. \tag{3.2}$$

Now, if λ is a k vector of known constants, we have the following lemma.

LEMMA: *A necessary and sufficient condition for a particular (set of) linear function(s) $\lambda'\beta$ to be identifiable in equation* (3.1) *is that $\lambda'\beta$ be estimable in equation* (3.2).

The proof follows readily from a result of F. Fisher [**3**, Theorem 2.7.2, p. 56]; details are available in Hausman-Taylor [**7**].

Using this result, it is easy to see that without any prior information in equations (2.2), all of the elements of β in equation (2.1) are identifiable. Simply project equation (2.1) onto the column space of all the exogenous variables—such a projection operator is Q_V—and observe that all linear functions of β are estimable, since $X_{it}'Q_V X_{it}$ is non-singular. No linear functions of γ are estimable.[9]

If prior information is provided by equations (2.2) (i.e., $k_1 > 0$, $g_1 > 0$), then the columns of X_{1it} and Z_{1i} must be added to the list of instruments. Let A denote the matrix $[Q_V \vdots X_{1it} \vdots Z_{1i}]$ and let P_A be the orthogonal projection operator onto its column space. Then, corresponding to the familiar rank condition, we have the following proposition:

PROPOSITION 3.1: *A necessary and sufficient condition that the entire vector of parameters (β, γ) be identified in equation* (2.1) *is that the matrix*

$$\begin{bmatrix} X_{it}' \\ \cdots \\ Z_i' \end{bmatrix} P_A \left(X_{it} \vdots Z_i \right)$$

be non-singular.

Corresponding to the order condition, we have the following proposition:

PROPOSITION 3.2: *A necessary condition for the identification of (β, γ) in equation* (2.1) *is that $k_1 \geqq g_2$.*

[8] In this case, the order condition is grossly overfulfilled and the rank condition fails.

[9] This fact underscores the importance of the observation that the disturbance in equation (2.6) is correlated with Z_i, so that prior information from equations (2.2) is necessary to estimate γ.

PROOF: The rank condition follows from the Lemma. For the order condition, $\text{rank}[P_A(X_{it} \vdots Z_i)] \leqq \text{rank}[P_A X_{it}] + \text{rank}[P_A Z_i] = k + \text{rank}[P_A Z_i]$, so that a necessary condition for the matrix in Proposition 3.1 to be non-singular is that $\text{rank}[P_A Z_i] \geqq g$. Since Z_i is orthogonal to $Q_V, k_1 \geqq g_2$ is necessary for $\text{rank}[P_A Z_i]$ to equal g, which completes the proof.

Identification in structural equations with panel data thus has a number of noteworthy features. First, given the assumption that α_i is correlated with the columns of X and Z, it is remarkable to find that the coefficients of the time-varying variables are identified while those of the time-invariant variables are not. Similarly, one variance component (σ_η^2) is identifiable while the other (σ_α^2) is not. The parameters $(\gamma, \sigma_\alpha^2)$ can be identified if the prior information (2.2) provides enough instruments—at least one for every endogenous column of Z_i. Curiously, the k_1 exogenous columns of X_{it} which are *included* in the structural equation (2.1) are the only candidates for these identifying instruments. This contrasts with the conventional simultaneous equations model in which *excluded* exogenous variables—such as family background in the traditional measurement of the return to education—are required to identify and estimate the parameters of a structural equation. Intuitively, this works because only the time-invariant component of the disturbance is correlated with (X_2, Z_2). Since $X_{1it} = \tilde{X}_{1it} + X_{1i\cdot}$, $\tilde{X}_{1it}$ can be used as an instrument for X_1 and $X_{1i\cdot}$ can be an instrument for Z_{2i}.

3.2 *Estimation*

When the parameters of equation (2.1) are identified by means of a specified set of variables which can be used as instruments, a consistent and asymptotically efficient estimator for (β, γ) can be constructed. Except for the fact that the disturbance covariance matrix Ω is non-scalar, equations (2.1) and (2.2) represent an ordinary structural equation and a list of exogenous and endogenous variables from which the reduced form can be calculated. Thus if Ω were known, 2SLS estimates of (β, γ) in

$$\Omega^{-1/2} Y_{it} = \Omega^{-1/2} X_{it} \beta + \Omega^{-1/2} Z_i \gamma + \Omega^{-1/2} \epsilon_{it}, \tag{3.3}$$

taking X_1, Z_1 as exogenous, would be asymptotically efficient in the sense of converging in distribution to the limited information maximum likelihood estimator.

More precisely, the information in equations (2.1) and (2.2) can be expressed as a system consisting of a single structural equation and two multivariate reduced form equations:

$$\begin{aligned} &\Omega^{-1/2} Y_{it} = \Omega^{-1/2} X_{it} \beta + \Omega^{-1/2} Z_i \gamma + \Omega^{-1/2} \epsilon_{it}, \\ (3.4) \qquad &X_{2it} = X_{1it} \pi_{11} + Z_{1i} \pi_{12} + Q_V \pi_{13} + v_{1it}, \\ &Z_{2i} = X_{1it} \pi_{21} + Z_{1i} \pi_{22} + Q_V \pi_{23} + v_{2it}. \end{aligned}$$

This system is a convenient form for discussing the efficiency of 2SLS estimators in equations (2.1) and (2.2). Equations (3.4) are triangular (because the bottom two equations are reduced forms) but not recursive (because v_1 and v_2 are correlated with α_i). In addition, the reduced form equations are, of course, just-identified. Since the disturbance covariance matrix in equations (3.4) is unknown, OLS is inconsistent while 3SLS is fully efficient; see Lahiri-Schmidt **[11]**. Finally, since the reduced forms are just-identified, 3SLS estimates of (β, γ) in the entire system are identical to 3SLS estimates of (β, γ) in the first equation alone (see Narayanan **[16]**), and these are, of course, just the 2SLS estimates. Thus 2SLS estimates of (β, γ) in equation (3.3) are fully efficient in the sense that they coincide asymptotically with FIML estimators from the system (3.4).

Finally, 2SLS estimates of (β, γ) in equation (3.3) are equivalent to OLS estimates of (β, γ) in

$$(3.5) \qquad P_A \Omega^{-1/2} Y_{it} = P_A \Omega^{-1/2} X_{it} \beta + P_A \Omega^{-1/2} Z_i \gamma + P_A \Omega^{-1/2} \epsilon_{it},$$

where P_A is the orthogonal projection operator onto the column space of the instruments $A = [Q_V \vdots X_{1it} \vdots Z_{1i}]$. Least squares applied to this equation is computationally convenient: (i) the transformation $\Omega^{-1/2}$ can be done by $(1-\theta)$-differencing the data as in equation (2.5), (ii) the projection of the exogenous variables onto the column space of A yields the variables themselves, and (iii) the projection of the endogenous variables onto the column space of A can be calculated using only time averages, rather than the entire TN vectors of observations (see Appendix B).

For Ω known, then, the calculation of asymptotically efficient estimators is straightforward; but the only case of practical interest is where Ω is unknown and must be estimated. The question that immediately arises is how Ω should be estimated and whether an iterative procedure is necessary for efficient estimation of (β, γ).

Consider the feasible analog of equation (3.5), where $\hat{\Omega}^{-1/2}$ is any consistent estimator for Ω, e.g., from Section 2.3:

$$(3.6) \qquad P_A \hat{\Omega}^{-1/2} Y_{it} = P_A \hat{\Omega}^{-1/2} X_{it} \beta + P_A \hat{\Omega}^{-1/2} Z_i \gamma + P_A \hat{\Omega}^{-1/2} \epsilon_{it}.$$

PROPOSITION 3.3: *For any consistent estimator $\hat{\Omega}$ of Ω, OLS estimates of (β, γ) in equation* (3.6) *have the same limiting distribution as OLS estimates in equation* (3.5), *based upon a known Ω.*

PROOF: For fixed T, it is straightforward to verify that $\sqrt{N}[\hat{\beta}(\hat{\Omega}) - \hat{\beta}(\Omega)] \xrightarrow{p} \mathbf{0}$; details are available in Hausman-Taylor [7].

We have thus shown that the 2SLS estimates of the parameters in equation (3.3)—using any consistent estimate of Ω—are asymptotically efficient. These estimators are identical to OLS estimates of (β, γ) in equation (3.6); for future reference, we denote them by $(\hat{\beta}^*, \hat{\gamma}^*)$.

In Appendix A, we derive the following characteristics of $(\hat{\beta}^*, \hat{\gamma}^*)$, depending upon the degree of over-identification. If the model is under-identified ($k_1 < g_2$), $\hat{\beta}^* = \hat{\beta}_W$ and $\hat{\gamma}^*$ does not exist. If the model is just-identified ($k_1 = g_2$), $\hat{\beta}^* = \hat{\beta}_W$ and $\hat{\gamma}^* = \hat{\gamma}_W$, as defined in equation (2.7). If the model is overidentified ($k_1 > g_2$), $(\hat{\beta}^*, \hat{\gamma}^*)$ differ from and are more efficient than $(\hat{\beta}_W, \hat{\gamma}_W)$. Finally, following Mundlak [**15**], we explore the relationship between $\hat{\beta}^*$ and the Gauss-Markov estimator, when the latter exists.

3.3 *Testing the Identifying Restrictions*

More efficient estimates of β and consistent estimates of γ require prior knowledge that certain columns of X and Z are uncorrelated with the latent α_i. An important feature of our procedure is that when the parameters are over-identified, all of these prior restrictions can be tested. This is an extremely useful and unusual characteristic: unusual in that it provides a test for the identification of γ, and useful since the maintained hypothesis need contain only the relatively innocuous structure of equation (2.1). It works, basically, because β is always identified and $\hat{\beta}_W$ provides a consistent benchmark against which all (or some) of the restrictions in equation (2.2) can be tested by comparing $\hat{\beta}^*$ with $\hat{\beta}_W$. The principles of such tests are outlined in Hausman [**6**], and extended in Hausman-Taylor [**8**].

The null hypothesis is of the form

$$H_0: \operatorname*{plim}_{N\to\infty} \frac{1}{N} X'_{1i\cdot}\alpha_i = \mathbf{0} \qquad \text{and} \qquad \operatorname*{plim}_{N\to\infty} \frac{1}{N} Z'_{1i}\alpha_i = \mathbf{0}.$$

Under H_0, both $\hat{\beta}_W$ and $\hat{\beta}^*$ are consistent, while under the alternative, plim $\hat{\beta}^* \neq$ plim $\hat{\beta}_W = \beta$; thus deviations of $\hat{q} = \hat{\beta}^* - \hat{\beta}_W$ from the zero vector cast doubt on the null hypothesis.

To form a χ^2 text based on $\hat{q}$, premultiply equation (2.1) by $Q_Z\Omega^{-1/2} = [I_{TN} - Z_i(Z_i'Z_i)^{-1}Z_i']\Omega^{-1/2}$ and consider the within-groups and efficient estimators for β in the transformed equation. Letting $X^* = Q_Z\Omega^{-1/2}X_{it}$,

$$\text{(3.8)} \qquad \begin{aligned} \hat{q} &= \left[(X^{*\prime}P_AX^*)^{-1}X^{*\prime}P_AQ_Z - (X^{*\prime}Q_VX^*)^{-1}X^{*\prime}Q_VQ_Z\right]\Omega^{-1/2}Y_{it} \\ &= DY^* \end{aligned}$$

where $Q_VQ_Z\Omega^{-1/2} = Q_V$ and $Y^* = \Omega^{-1/2}Y$ has a scalar covariance matrix. The specification test statistic is given by

$$\begin{aligned} \hat{m} &= \hat{q}'\left[\operatorname{cov}(\hat{q})\right]^+\hat{q} = \hat{q}'\left[\operatorname{cov}(\hat{\beta}_W) - \operatorname{cov}(\hat{\beta}^*)\right]^+\hat{q} \\ &= \hat{q}'\left[\sigma_\eta^2 DD'\right]^+\hat{q} \end{aligned}$$

where the second equality follows from the asymptotic Rao-Blackwell argument in Hausman [**6**] and $(\ )^+$ denotes any generalized inverse. Following Hausman-Taylor [**8**], we have the next proposition.

PROPOSITION 3.4: *Under the null hypothesis, $\hat{\sigma}_\eta^2 \hat{m}$ converges in distribution to a χ_d^2 random variable, where $\hat{\sigma}_\eta^2$ is any consistent estimate of σ_η^2, and $d = \text{rank}(D)$ $= \min[k_1 - g_2, TN - k]$.*

PROOF: From Propositions 4.1 and 4.2 of Hausman-Taylor [**8**], we obtain the limiting distribution of m and the fact that

$$\text{rank}(D) = \min\left[\text{rank}(X^{*\prime}P_H), \text{rank}\left(I - X^*(X^{*\prime}Q_V X^*)^{-1}X^{*\prime}Q_V\right)\right]$$

where P_H projects onto the orthocomplement of the column space of Q_V in the column space of A: i.e., onto the column space of $[X_{1i\cdot} \vdots Z_{1i}]$. Under the usual linear independence assumptions, the second term in brackets equals $TN - k$. For the first term, $\text{rank}(X^{*\prime}P_H) = \min[k, \text{rank}(Q_Z P_H)]$, and since $P_H = P_Z P_H + Q_Z P_H$, $\text{rank}(Q_Z P_H) = (k_1 + g_1) - (g_1 + g_2) = (k_1 - g_2)$.

This specification test of the identifying restrictions in equation (2.2) has some noteworthy features. The number of restrictions nominally being tested is $k_1 + g_1$, in the sense that if any of the restrictions in equation (2.2) is false, $\hat{q}$ should differ from $\mathbf{0}$. Yet the degrees of freedom for the χ^2 test depend upon the number of *overidentifying* restrictions $(k_1 - g_2)$. Indeed, when the model is just-identified, $\hat{\beta}^* = \hat{\beta}_W$ (see Section 3.2), so that $\hat{q}$ is identically zero and the degrees of freedom are zero. Finally, note that the alternative hypothesis does not require that any of the columns of X or Z be uncorrelated with α_i. Hence *all* of the exogeneity information about X and Z is subject to test by this procedure.[10]

4. ESTIMATING THE RETURNS TO SCHOOLING

Measuring the returns to schooling has received extensive attention lately, and much discussion has focused on the potential correlation between individual-specific latent ability and schooling. (See Griliches [**4**] for an excellent survey.) Since the sample we use does not contain an *IQ* measure, it would seem likely on *a priori* grounds that the schooling variable and α_i are correlated. Yet as Griliches [**4**] points out, it is not clear in which direction the schooling coefficient will be biased. While a simple story of positive correlation between ability and schooling leads to an upward bias in the OLS estimate, a model in which the choice of the amount of schooling is made *endogenous* can lead to a negative correlation between the chosen amount of schooling and ability. In fact, both Griliches [**4**] and Griliches, Hall, and Hausman [**5**] find that treating schooling as endogenous and using family background variables as instruments leads to a *rise* in the estimated schooling coefficient of about 50 percent.[11] Since our method does not

[10] This test compares instrumental variables estimators under nested subsets of instruments: $\hat{\beta}^*$ uses $[Q_V \vdots X_1 \vdots Z_1]$ and $\hat{\beta}_W$ uses $[Q_V]$. If one wishes to test particular columns of X_1 or Z_1 for exogeneity, while maintaining a just-identifying set of instruments, a similar test can be constructed by comparing $(\hat{\beta}^*, \hat{\gamma}^*)$ with $(\hat{\beta}_W, \hat{\gamma}_W)$. If the model is overidentified under the maintained hypothesis, compare $(\hat{\beta}^*, \hat{\gamma}^*)$ using the different instruments. See Hausman-Taylor [**8**] for details.

[11] Using a specification test of the type Wu [**21**] and Hausman [**6**] propose, we find a statistically significant difference between the IV and OLS estimates. Chamberlain [**1**] also finds a significant increase in the schooling coefficient, comparing OLS with estimates from his two factor model.

require excluded exogenous instruments, it is of some interest to see if the estimated return to schooling remains higher than the OLS estimate.

Our sample consists of 750 randomly chosen prime age males, age 25–55, from the non-Survey of Economic Opportunity portion of the Panel Study on Income Dynamics (PSID) sample. We consider two years, 1968 and 1972, to minimize problems of serial correlation apart from the permanent individual component.[12] The sample contains 70 non-whites for which we use a 0-1 variable, a union variable also treated as 0-1, a bad health binary variable, and a previous year unemployment binary variable. The two continuous explanatory variables are schooling and either experience or age.[13] The PSID data do not include IQ. The National Longitudinal Survey (NLS) sample for young men would provide an IQ measure, but problems of sample selection would need to be treated (as in Griliches, Hall, and Hausman [**5**]) which would cause further econometric complications. Perhaps of more importance is the fact that for the NLS sample, IQ has an extremely small coefficient in a log wage specification (e.g., between .0006 and .002 in Griliches, Hall, and Hausman [**5**]); and if it is included in the specification, it has only a small effect on the schooling coefficient. Thus we use the PSID sample without an *IQ* measure, and our results should be interpreted with this exclusion in mind.

Table I gives the results of traditional estimators for our sample. The first two columns show the OLS and GLS estimates respectively, which assume no correlation between the explanatory variables and α_i. The estimates are reasonably similar, especially the schooling coefficient, which in both cases equals .067. The effects of experience and race stay the same, while the remaining three coefficients change somewhat, though they are not estimated very precisely. Note that the correlation coefficient across the four year period ($\rho = .623$) indicates that the latent individual effect is important. The finding that an additional year of schooling leads to a 6.7 percent higher wage is very similar to other OLS results from both PSID and other data sets.

In the third column of Table I, we present the within-groups estimate of the wage equation specification. All the time-invariant variables are eliminated by the data transformation, leaving only experience, bad health, and unemployed last year. As we have seen, the estimates of these coefficients are unbiased even if the variables are correlated with the latent individual effect. The point estimates change markedly from the first two columns: bad health by 26 percent, unemployment by 33 percent, and experience by 59 percent.[14] Comparing the within-

[12] Lillard and Willis [**12**] demonstrate within a random coefficients framework that a first order autoregressive process remains even after the permanent individual effect is accounted for. Our estimation technique can easily be extended to account for an autoregressive process, but for expository purposes we use a simpler case. Note that we are not investigating the dynamics of wages or earnings here.

[13] Experience was used as either experience with present employer or as (age − schooling − 5). Qualitatively, the results are similar, so we report results using the latter definition. As the results show, use of age yields very similar results for the schooling coefficient. Unlike Griliches [**4**], we are not attempting to separate the influence of age from that of experience.

[14] Percentage changes are calculated as differences in natural logarithms.

TABLE I

DEPENDENT VARIABLE: LOG WAGE

		OLS[a]	GLS	Within	IV/GLS
Experience	(*it*)	.0132	.0133	.0241	.0175
		(.0011)	(.0017)	(.0042)	(.0026)
Bad Health	(*it*)	−.0843	−.0300	−.0388	−.0249
		(.0412)	(.0363)	(.0460)	(.0399)
Unemployed	(*it*)	−.0015	−.0402	−.0560	−.0636
Last Year		(.0267)	(.0207)	(.0295)	(.0245)
Race	(*i*)	−.0853	−.0878	—	−.0542
		(.0328)	(.0518)	—	(.0588)
Union	(*i*)	.0450	.0374	—	.0733
		(.0191)	(.0296)	—	(.0434)
Years of	(*i*)	.0669	.0676	—	.0927
Schooling		(.0033)	(.0052)	—	(.0191)
Other Variables		constant	constant	constant	constant
		time	time	time	time
NOBS		1500	1500	1500	1500
S.E.R.		.321	.192	.160	.193
RHO		—	.623	—	—
Specification Test			$\chi_3^2 = 20.2$		$\chi_3^2 = 8.70$
Instruments					father's ed.
					mother's ed.
					poor

[a] Reported standard errors are inconsistent since they do not account for the variance components.

groups and GLS estimates, using results in Hausman [6], we test the hypothesis that some of the explanatory variables in our log wage specification are correlated with the latent α_i. Under the null hypothesis, the statistic is distributed as χ_3^2, and since we compute $\hat{m} = 20.2$, we can reject the null hypothesis with any reasonable size test. This confirms Hausman's [6] earlier finding that misspecification was present in a similar log wage equation.

In the last column of Table I, we present traditional instrumental variables estimates of the wage equation, treating schooling as endogenous. Family background variables are used as excluded instruments: father's education, mother's education, and a binary variable for a poor household. The estimated schooling coefficient rises to .0927, which echoes previous results of Griliches [4] and Griliches, Hall, and Hausman [5]. Under the null hypothesis that the instruments are uncorrelated with α_i, the point estimates should be close to the consistent within-groups estimates. Note that these estimates are somewhat closer to the within-groups estimates than the original OLS estimates. We might conclude that the instruments have lessened the correlation of schooling with α_i by replacing schooling with a linear combination of background variables; yet the result of the specification test is $\hat{m} = 8.70$, which again indicates the presence of correlation between the instruments and α_i. We conclude that family background variables are inappropriate instruments in this specification, perhaps because unmeasured individual effects may be transmitted from parents to children.

In the first two columns of Table II, we present the results of our estimation

TABLE II

DEPENDENT VARIABLE: LOG WAGE

	HT/IV[a]	HT/GLS	HT/GLS	HT/GLS	HT/GLS
Experience	.0217	.0217	–	.0268	.0241[b]
(*it*)	(.0027)	(.0031)	—	(.0037)	(.0045)
Experience Squared	—	—	—	−.00012	—
(*it*)	—	—	—	(.00015)	—
Age	—	—	.0147	—	—
(*it*)	—	—	(.0028)	—	—
Bad Health	−.0535	−.0278	−.0228	−.0243	−.0388
(*it*)	(.0468)	(.0307)	(.0318)	(.0318)	(.0348)
Unemployment Last Year	−.0556	−.0559	−.0634	−.0634	−.0560[b]
(*it*)	(.0311)	(.0246)	(.0265)	(.0236)	(.0279)
Race	−.0257	−.0278	−.0046	−.0014	−.0175
(*i*)	(.0531)	(.07518)	(.0824)	(.0662)	(.0764)
Union	.1245	.1227	.1648	.1449	.2240
(*i*)	(.0560)	(.0473)	(.0721)	(.0598)	(.2863)
Years of Schooling	.1247[b]	.1246[b]	.1311[b]	.1315[b]	.2169[b]
(*i*)	(.0380)	(.0434)	(.0490)	(.0319)	(.0979)
Other Variables	constant time	constant time	constant time	constant time	constant time
NOBX	1500	1500	1500	1500	1500
S.E.R.	.352	.190	.196	.189	.629
RHO	—	.661	.678	.674	.817
Specification Test	—	$\chi_2^2 = 2.24$	—	—	$\chi_0^2 = 0.0$

[a] Reported standard errors are inconsistent since they do not account for the variance components.
[b] Treated as if correlated with α_i.

method. We assume that X_1 contains experience, bad health, and unemployment last year, all initially assumed to be uncorrelated with α_i. Z_1 is assumed to contain race and union status, while Z_2 contains schooling, which is allowed to be correlated with α_i. The estimated schooling coefficient rises to .125, which is 62 percent above the original OLS estimate and 30 percent above the traditional instrumental variables estimate. The effect of race has now almost disappeared: its coefficient has fallen from −.085 in the OLS regression to −.028. The effects of experience and union status have risen substantially, while that of bad health has fallen.

Using the test from Section 3.3, we compare the within-groups and efficient estimates of the X_1 coefficients. Observe that the unemployment coefficient is now very close to the within estimate, while bad health and experience have moved considerably closer to the within-groups estimates from either the OLS or instrumental variables estimates. The specification test statistic is $\hat{m} = 2.24$ which is distributed as χ_2^2 under the null hypothesis of no correlation between the instruments and α_i. While m is somewhat higher than its expected value of 2.0 under H_0, we would not reject the hypothesis that the columns of X_1 and Z_1 are uncorrelated with the latent individual effect.

We next examine how robust these estimates are to small changes in specification. Column 3 of Table II replaces experience with age. While experience is arguably correlated with α_i through its schooling component, age can be taken as uncorrelated unless important cohort effects are present. The results are quite similar to our previous findings. The effect of schooling is .131, only slightly higher than the .125 found previously. Race again has little or no effect, while the coefficients of health and unemployment are similar to those in the specification with experience. In the next column of Table II, we include experience and experience squared as explanatory variables.[15] Again, the results are in general agreement with the original specification. The schooling coefficient increases from .125 to .132, and race still has little effect. We conclude that our main results are reinforced by these alternative specifications.

Our final specification relaxes some of the noncorrelation assumptions between α_i and the explanatory variables. We remove experience and unemployment from the X_1 category, allowing them to be correlated with α_i. Now X_1 contains only bad health. The model is just-identified, so that the efficient estimates of the coefficients of the X_{it} variables are identical to the within-groups estimates. The specification test of Section 3.3 has zero degrees of freedom, and no specification test can be performed. The asymptotic standard errors have now risen to the point where coefficient estimates are quite imprecise, especially the schooling coefficient. Nevertheless, it is interesting to note that the point estimate of the schooling coefficient has risen to .217.

Thus all of our different estimation methods have led to a rise in the size of the schooling coefficient. Removing potentially correlated instruments has had a substantial effect: the point estimates change and their standard errors increase. All methods which control for correlation with the latent individual effects increase the schooling coefficient over those which do not; and this is certainly not the direction that many people concerned about ability bias would have expected.

5. SUMMARY

In this paper, we have developed a method for use with panel data which treats the problem of correlation between explanatory variables and latent individual effects. Making use of time-varying variables in two ways—to estimate their own coefficients and to serve as instruments for endogenous time-invariant variables—allows identification and efficient estimation of both β and γ. The method is a two-fold improvement over the within-groups estimator: it is more efficient and it produces estimates of the coefficients of time-invariant variables. In the wage equation example, it performs better than traditional instrumental variables methods, which rely on excluded exogenous variables for instruments.

[15] Neither of these alternative specifications pass the specification test if estimated by OLS and compared with the appropriate within-groups estimator. In both specifications, the latent individual effects are correlated with the explanatory variables.

Perhaps most important, we derive a specification test of the appropriateness of the identifying exogeneity restrictions in equation (2.2). Since the within-groups estimates of β always exist and are consistent, they provide a benchmark against which further results—using the information in equations (2.2)—can be compared. If this specification test is satisfied, we can be confident in the consistency of our final results, since the maintained hypothesis embodied in the within-groups estimator is so weak.

Massachusetts Institute of Technology
and
Bell Telephone Laboratories, Inc.

Manuscript received February, 1980; revision received October, 1980.

APPENDIX A

1. *Special Cases*

Depending upon $(k_1 - g_2)$, the degree of over-identification, the consistent and asymptotically efficient estimator $(\hat{\beta}^*, \hat{\gamma}^*)$ exhibits some interesting peculiarities. Since we shall be interested in estimating β and γ separately from equation (3.6), two generic formulae will prove convenient. Let $Y = X_1\beta_1 + X_2\beta_2 + e$.

LEMMA: *The following two expressions for the OLS estimator of β_1 are identical: (i) "parse out" X_2 by premultiplying the model by $Q_2 = I - X_2(X_2'X_2)^{-1}X_2'$, so that*

$$\hat{\beta}_1 = (X_1'Q_2X_1)^{-1}X_1'Q_2Y;$$

(ii) remove the OLS estimates of $X_2\beta_2$ from Y and regress that on X_1, so that

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'\left[I - X_2(X_2'Q_1X_2)^{-1}X_2'Q_1\right]Y.$$

Now, suppose the parameters in equation (2.1) are under-identified.

(a) $k_1 = g_1 = 0$. Here, the set of instruments is only $A = [Q_V]$. Since $\hat{\Omega}^{-1/2} = I_{NT} - (1-\hat{\theta})P_V$, $P_A\hat{\Omega}^{-1/2} = Q_V$, so that $\hat{\beta}^* = \hat{\beta}_W$.

(b) $k_1 < g_2$. Here, the instruments are $[Q_V \vdots Z_{1i} \vdots X_{1i.}]$ which we write as $A = [Q_V \vdots H]$. Consider equation (3.6). $P_A\hat{\Omega}^{-1/2}Z_i = P_H\hat{\Omega}^{-1/2}Z_i$, since $Q_V\hat{\Omega}^{-1/2}Z_i = \mathbf{0}$. When $k_1 < g_2$, $P_H\hat{\Omega}^{-1/2}Z_i$ is not of full column rank since the dimension of the column space of H is $g_1 + k_1$ and Z_i has $g_1 + g_2$ linearly independent columns. Thus there exists a g vector ξ such that $P_A\hat{\Omega}^{-1/2}Z_i\xi = \mathbf{0}$, so that γ is not identifiable since γ and $(\gamma + \xi)$ are observationally equivalent in equation (3.6). To calculate $\hat{\beta}^*$, "parse out" $P_A\hat{\Omega}^{-1/2}Z_i$ in equation (3.6) and run OLS. The column space of $P_A\hat{\Omega}^{-1/2}Z_i$ equals the column space of H, so projecting $P_A\hat{\Omega}^{-1/2}X_{it}$ onto the orthocomplement of $P_A\hat{\Omega}^{-1/2}Z_i$ yields Q_VX_{it}. Thus $\hat{\beta}^* = \hat{\beta}_W$ in the generic underidentified case, and there is no consistent estimator for γ.

Consider the just-identified case.

(c) $k_1 = g_2$. Here, again, the rank of $P_H\hat{\Omega}^{-1/2}Z_i$ equals the rank of H, so that $\hat{\beta}^* = \hat{\beta}_W$. The OLS estimate of γ in equation (3.6) is identical to the OLS estimate of γ in

$$P_A\hat{\Omega}^{-1/2}Y_{it} - P_A\hat{\Omega}^{-1/2}X_{it}\hat{\beta}_W = P_A\hat{\Omega}^{-1/2}Z_i\gamma + P_A\hat{\Omega}^{-1/2}\epsilon_{it}$$

by the previous Lemma, since $\hat{\beta}_W$ is the OLS estimate of β in equation (3.6). Thus $\hat{\gamma}^*$ can be written as

$$\hat{\gamma}^* = (Z_i'\hat{\Omega}^{-1/2}P_A\hat{\Omega}^{-1/2}Z_i)^{-1}Z_i'\hat{\Omega}^{-1/2}P_A[P_A\hat{\Omega}^{-1/2}Y_{it} - P_A\hat{\Omega}^{-1/2}X_{it}\hat{\beta}_W]$$

$$= (Z_i'\hat{\Omega}^{-1/2}P_A\hat{\Omega}^{-1/2}Z_i)^{-1}Z_i'\hat{\Omega}^{-1/2}P_A[Y_{i\cdot} - X_{i\cdot}\hat{\beta}_W]$$

$$= (Z_i'P_AZ_i)^{-1}Z_i'P_A\hat{d}_i = \hat{\gamma}_W$$

since $\hat{\Omega}^{-1/2}Z_i = \hat{\theta}Z_i$. For the just-identified case, then, our instrumental variables estimator coincides with the within-groups estimator of both β and γ.

Suppose the parameters in equation (2.1) are over-identified.

(d) $k_2 = g_2 = 0$. Here, A coincides with the explanatory variables in equation (3.6), so that the 2SLS and OLS estimators are identical. For $\hat{\Omega}^{-1/2}$, it is identical to the feasible GLS estimator in Section 2.1; for known $\Omega^{-1/2}$, it is Gauss-Markov.

(e) $k_1 > g_2$. The column rank of $P_A\hat{\Omega}^{-1/2}Z_i$ is now g and the column space of $P_A\hat{\Omega}^{-1/2}Z_i$ differs from that of H. Thus $\hat{\beta}^*$ will differ from $\hat{\beta}_W$ in the over-identified case. By the Lemma, $\hat{\gamma}^*$ is calculated from the regression of $P_A\hat{\Omega}^{-1/2}Y - P_A\hat{\Omega}^{-1/2}X\hat{\beta}^*$ on $P_A\hat{\Omega}^{-1/2}Z_i$, so that $\hat{\gamma}^*$ will differ from $\hat{\gamma}_W$, which we derived from the regression of $P_A\hat{\Omega}^{-1/2}Y - P_A\hat{\Omega}^{-1/2}X\hat{\beta}_W$ on $P_A\hat{\Omega}^{-1/2}Z_i$.

Since $(\hat{\beta}^*, \hat{\gamma}^*)$ are asymptotically efficient, $(\hat{\beta}_W, \hat{\gamma}_W)$ are inefficient in the over-identified case. Intuitively, this can be explained by regarding the within-groups estimators as 2SLS estimators which ignore the instruments $X_{1i\cdot}$ and $Z_{1\cdot}$. It is a peculiar feature of this model that ignoring these instruments only matters when the parameters are over-identified.

2. *Mundlak's Model*

A final special case is the model discussed at length by Mundlak [**15**] in which no time-invariant observables are present and all explanatory variables are correlated with α_i:

$$\text{(A.1)} \qquad Y_{it} = X_{it}\beta + \alpha_i + \eta_{it}.$$

The relationship between α and X is expressed by Mundlak through the "auxillary" regression $\alpha_i = X_{i\cdot}\pi + \omega_i$ where no prior information is assumed about π. Mundlak shows that (i) if α_i is correlated with every column of $X_{i\cdot}$ (π is unconstrained), the Gauss-Markov estimator for β is the within-groups estimator $\hat{\beta}_W$, and (ii) if α_i is uncorrelated with every column of $X_{i\cdot}$ ($\pi = \mathbf{0}$), the G-M estimator for β is the GLS estimator $\hat{\beta}_{GLS}$ in equation (2.5), assuming Ω to be known.

Recognizing that case (i) is just-identified ($k_1 = g_2 = 0$) and case (ii) is overidentified ($k_2 = g_2 = 0$), the discussion in (c) and (d) above shows that the 2SLS estimator $\hat{\beta}^*$ is identical to the G-M estimator in both cases. More to the point, if α_i is uncorrelated with some columns of $X_{i\cdot}$ and correlated with others (π obeys some linear restrictions), the model is overidentified ($k_1 > g_2 = 0$) and case (e) above shows that $\hat{\beta}^*$ is asymptotically efficient relative to $\hat{\beta}_W$. Thus it is only in the two extremes (i) and (ii) that $\hat{\beta}_W$ or $\hat{\beta}_{GLS}$ is appropriate.

We can use this characterization of the G-M estimator, however, to examine the relationship between $\hat{\beta}^*$ and the G-M estimator, should the latter exist. Suppose Ω is known, and we premultiply Mundlak's model (A.1) by $\Omega^{-1/2}$ and reparameterize for convenience:

$$\text{(A.2)} \qquad \Omega^{-1/2}Y_{it} = \Omega^{-1/2}X_{it}SS^{-1}\beta + \Omega^{-1/2}\alpha_i + \Omega^{-1/2}\eta_{it}$$

$$= \Omega^{-1/2}M_{it}\xi + \epsilon_{it}^*$$

where $M_{it} = X_{it}S$, $\xi = S^{-1}\beta$, $\epsilon_{it}^* = \theta\alpha_i + \eta_{it} - (1-\theta)\eta_{i\cdot} = \alpha_i^* + \eta_{it}^*$ and the non-singular transformation S is chosen so that

$$S'(X_{it}'X_{it})S = I_k.$$

Since the X_{it} are random variables in the analysis, the matrix S, being a function of the X_{it}, will be random also; since some X_{it} are endogenous S will also be endogenous.

Let us specify prior information about the correlation between X_{it} and α in a somewhat more flexible manner than Mundlak's. Let h_x denote the k vector of probability limits (for fixed T)

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} X_{it}'\alpha_i \equiv h_x = \operatorname{plim} \frac{1}{N} S^{-1}M_{it}'\alpha_i \equiv S^{-1}h_M$$

where h_M denotes the corresponding vector of (asymptotic) correlations between α_i and M_{it}. We can express prior information on h_x as $r(r < k)$ homogeneous linear restrictions

$$Rh_x = \mathbf{0} = RS^{-1}Sh_x \equiv R^*h_M$$

which yield r homogeneous restrictions on h_M. Note that (i) the exogeneity information in equations (2.2) can be expressed as $Rh_x = \mathbf{0}$ where each row of R has a single 1 and the rest zeroes; (ii) the previous results on identification and estimation go through, taking the columns of $X_{it}R_i'$ as exogeneous where $R_i(i = 1, \ldots, r)$ is a row of R; (iii) homogeneous restrictions on h_x correspond uniquely to homogeneous restrictions on $\tilde{\pi}$ in Mundlak's specification; i.e., $Rh_x = \mathbf{0} \Rightarrow \operatorname{plim}_{N\to\infty}(1/N)R(X_{i.}'X_{i.})(X_{i.}'X_{i.})^{-1}X_{i.}'\alpha_i \Rightarrow \tilde{R}\pi = 0$ where $\tilde{R} = R(X_{i.}'X_{i.})$.

In the model (A.2), then, certain linear combinations of the colums of M_{it} are assumed uncorrelated with α_i^* and all of the columns of M_{it} are orthogonal.

PROPOSITION A.1: *The 2SLS estimator $\hat{\xi}^*$ in equation (A.2) is Gauss-Markov for ξ.*

PROOF: Let F denote the $k \times k$ non-singular matrix

$$F = \left[R' \;\vdots\; B' \right]$$

where the columns of $B'(k \times k - r)$ are $k - r$ basis vectors for the column space of $I_k - R'(RR')^{-1}R$. Now, reparameterize equation (A.2) as

$$\Omega^{-1/2}Y_{it} = \Omega^{-1/2}M_{it}FF^{-1}\xi + \epsilon_{it}^*$$

$$= \Omega^{-1/2}\left[M_{it}R' \;\vdots\; M_{it}B' \right]F^{-1}\xi + \epsilon_{it}^*$$

which we write as

$$\Omega^{-1/2}Y_{it} = \Omega^{-1/2}L_{1it}\delta_1 + \Omega^{-1/2}L_{2it}\delta_2 + \epsilon_{it}^* \tag{A.3}$$

where $\delta = [\delta_1' \vdots \delta_2'] = F^{-1}\xi$. Consider 2SLS estimates of δ in equation (A.3), using as instruments $A = [Q_V \vdots L_{1it}]$ since $\operatorname{plim}_{N\to\infty}(1/N)L_{1it}'\alpha_i^* = \operatorname{plim}(1/N)RM_{it}'\Omega^{-1}\alpha_i = \mathbf{0}$ by assumption. By construction, $\Omega^{-1/2}L_1$ and $\Omega^{-1/2}L_2$ are orthogonal, and $P_A L_1 = L_1$, so the 2SLS estimator

$$\hat{\delta}_1^* = \left(L_{1it}'\Omega^{-1}L_{1it}\right)^{-1}L_{1it}'\Omega^{-1}Y_{it}$$

coincides with the GLS estimator (for known Ω). It is Gauss-Markov for δ_1 in this model since all columns of L_1 are uncorrelated with ϵ_{it}^* and L_2 is orthogonal to L_1. Similarly, the 2SLS estimator for δ_2 is

$$\hat{\delta}_2^* = \left(L_2'\Omega^{-1/2}Q_V\Omega^{-1/2}L_2\right)^{-1}L_2'\Omega^{-1/2}Q_V\Omega^{-1/2}Y_{it}$$

since $P_A L_2 = Q_V L_2$. Since $Q_V\Omega^{-1/2} = Q_V$, this simplifies to $\hat{\delta}_2^* = (L_2'Q_VL_2)^{-1}L_2'Q_VY$ which is the within-groups estimator. Using Mundlak's result (i) above, $\hat{\delta}_2^*$ is G-M for δ_2 since every column of L_2 is correlated with α_i, and L_2 is orthogonal to L_1. Hence $\hat{\delta}^* = [\hat{\delta}_1^{*\prime} : \hat{\delta}_2^{*\prime}]'$ is G-M for δ, and since F is a non-singular, non-stochastic matrix, $\hat{\xi}^* = F\hat{\delta}^*$ is Gauss-Markov for $F\delta = \xi$. This completes the proof.

Two related questions immediately emerge. First, is $\hat{\beta}^* = S\hat{\xi}^*$ Gauss-Markov for β, since S is non-singular? Secondly, what became of the intuition that 2SLS estimates were biased and thus not Gauss-Markov?

PROPOSITION A.2: *The 2SLS estimator $\hat{\beta}^*$ coincides with $S\hat{\xi}^*$ but $\hat{\beta}^*$ is biased for β and not Gauss-Markov.*

PROOF: Calculate $\hat{\beta}^*$ directly using 2SLS in the model

$$P_A\Omega^{-1/2}Y_{it} = P_A\Omega^{-1/2}X_{it}\beta + P_A\Omega^{-1/2}\epsilon_{it}$$

where $A = [Q_V \vdots L_{1it}]$ is the appropriate set of instruments here, as well as in equation (A.3). Then

$$\begin{aligned}\hat{\beta}^* &= [X_{it}'\Omega^{-1/2}P_A\Omega^{-1/2}X_{it}]^{-1}X_{it}'\Omega^{-1/2}P_A\Omega^{-1/2}Y_{it}\\ &= S[SX'\Omega^{-1/2}P_A\Omega^{-1/2}XS]^{-1}SX'\Omega^{-1/2}P_A\Omega^{-1/2}Y\\ &= S\hat{\xi}^*.\end{aligned}$$

Thus $\hat{\beta}^*$ is a non-singular transformation of the G-M estimator $\hat{\xi}^*$: i.e.,

$$\hat{\beta}^* = S\hat{\xi}^* \quad \text{and} \quad \beta \equiv S\xi$$

so that $\hat{\beta}^* - \beta = S(\hat{\xi}^* - \xi)$. However, recall that S is a function of the matrix X_{it}; it is endogenous and in calculating moments of $\hat{\beta}^* - \beta$, we cannot condition on it. Hence, in general, $E(\hat{\beta}^* - \beta) = ES(\hat{\xi}^* - \xi) \neq SE(\hat{\xi}^* - \xi) = 0$, and $\operatorname{cov}[\hat{\beta}^* - \beta] = \operatorname{cov}[S(\hat{\xi}^* - \xi)] \neq S[\operatorname{cov}(\hat{\xi}^* - \xi)]S'$ where $\operatorname{cov}(\hat{\xi}^* - \xi)$ attains the Cramer-Rao bound.

A final anomalous property of $\hat{\beta}^*$ follows from these propositions. Suppose the original design matrix X_{it} were orthogonal, so that $X_{it}'X_{it} = I_k$. Then the 2SLS estimator $\hat{\beta}^*$ using $[Q_V \vdots X_{it}R']$ as instruments would be both unbiased and Gauss-Markov. One rarely finds a G-M estimator in a simultaneous equations problem; one does in this model because 2SLS estimates when *all* the explanatory variables are correlated with α_i are identical to the within-groups estimators, and these are unbiased in finite samples. This is because the set of instruments in this model is just the columns of Q_V, which are orthogonal to α_i in the sample, not just in expectation or as a probability limit.

APPENDIX B

Computational Details

We now consider the estimation method proposed in Section 3.2 from the standpoint of computational convenience. Equation (3.6) and Proposition 3.3 state the basic theoretical results. Given initial consistent instrumental variables estimates of (β, γ), we can estimate Ω and transform the variables by θ-differencing the data. The model now is of the form of equation (3.6) and OLS estimates will be asymptotically efficient.

The main difficulty that arises is computational: how to do instrumental variables when the data matrix (of order $T \times N$) may exceed the computational capacity of much econometric software. If this occurs, using equation (3.4), calculate predicted values of X_2 and Z_2 from their reduced forms. The predicted $\hat{Z}_{2i}$'s are formed from a sample size N regression of Z_{2i} on the columns of $X_{1i.}$ and Z_{1i}. For the $\hat{X}_{2it}$'s, rather than doing a sample size $T \times N$ regression, an equivalent procedure is to form $\hat{X}_{2it} = X_{2it} - X_{2i.} + \hat{X}_{2i.}$. The last term, $\hat{X}_{2i.}$, is calculated from the sample size N regression of $X_{2i.}$ on $X_{1i.}$ and Z_{1i}. Then the calculated $\hat{X}_{2it}$ and $\hat{Z}_{2i}$ are used with the X_{1it} and the Z_{1i} in an OLS regression to obtain consistent estimates of both β and γ.[16] A similar technique works with the transformed variables in equation (3.6) which yields asymptotically efficient estimates of β and γ. The reason that calculating $\hat{X}_{2it}$ in this manner is equivalent to the more cumbersome approach of a $T \times N$ sample regression of X_{2it} on instruments as indicated in equation (3.4) is that Q_V is orthogonal

[16]One note of caution, however. The estimates of the variance from the second stage are inconsistent, for the same reason that doing 2SLS in two steps yields inconsistent variance estimates in the second stage. One must use the estimated coefficients and equation (2.1) without the estimated variables on the right hand side.

to any time-invariant variable. Thus parsing out Q_V in the second and third equations of (3.4) is equivalent to premultiplying them by P_V, and $\hat{X}_{2i\cdot}$ and $\hat{Z}_{2i}$ can be calculated from the sample size N regressions on $X_{1i\cdot}$ and $Z_{1\cdot}$. To get $\hat{X}_{2it}$, we must add $Q_V\hat{\pi}_{13}$ to $\hat{X}_{2i\cdot}$, so that $\hat{X}_{2it}$ is given by $\tilde{X}_{2it} + \hat{X}_{2i\cdot}$.

If computational capacity is not a difficulty, a standard instrumental variables package can be used, with $\tilde{X}_{1it}$, $X_{1i\cdot}$, $\tilde{X}_{2it}$, and Z_{1i} as instruments. The variables which are time invariant have T identical entries for each individual i. So long as Proposition 3.1 is satisfied, the parameters are identified and the number of columns of $X_{1i\cdot}$ is at least as great as the number of columns of Z_{2i} (i.e., $k_1 \geqq g_2$). Note again how the columns of X_{1it} serve two roles: both in estimation of their own coefficients and as instruments for the columns of Z_{2i}.

REFERENCES

[1] Chamberlain, G.: "Omitted Variable Bias in Panel Data: Estimating the Returns to Schooling," *Annales de l'INSEE*, 30/31, (1978), 49–82.

[2] Chamberlain, G., and Z. Griliches: "Unobservables with a Variance Components Structure: Ability, Schooling, and the Economic Success of Brothers," *International Economic Review*, 16(1975), 422–449.

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

[4] Griliches, Z.: "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45(1977), 1–22.

[5] Griliches, Z., B. Hall, and J. Hausman: "Missing Data and Self-Selection in Large Panels," *Annales de l'INSEE*, 30/31(1978), 137–176.

[6] Hausman, J.: "Specification Tests in Econometrics," *Econometrica*, 46(1978), 1251–1272.

[7] Hausman, J., and W. Taylor: "Panel Data and Unobservable Individual Effects," M.I.T. Discussion Paper, January, 1980.

[8] ———: "Comparing Specification Tests and Classical Tests," M.I.T. Discussion Paper, September, 1980.

[9] Hoch, I.: "Estimation of Production Function Parameters and Testing for Efficiency," *Econometrica*, 23(1955), 325.

[10] Jöreskog, K.: "A General Method for Estimating a Linear Structural Equation System," in *Structural Equation Models in the Social Sciences*, ed. by A. S. Goldberger and O. D. Duncan. New York: Seminar Press, 1973.

[11] Lahiri, K., and P. Schmidt: "On the Estimation of Triangular Structural Systems," *Econometrica*, 46(1978), 1217–1222.

[12] Lillard, L., and R. Willis: "Dynamic Aspects of Earning Mobility," *Econometrica*, 46(1978), 985–1012.

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

[14] Mundlak, Y.: "Empirical Production Function Free of Management Bias," *Journal of Farm Economics*, 43(1961), 44–56.

[15] ———: "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46(1978), 69–86.

[16] Narayanan, R.: "Computation of Zellner-Theil's Three Stage Least Squares Estimates," *Econometrica*, 37(1969), 298–306.

[17] Nerlove, M.: "A Note on Error Components Models," *Econometrica*, 39(1971), 359–382.

[18] Pudney, S.: "The Estimation and Testing of Some Error Components Models," London School of Economics Discussion Paper, 1979.

[19] Scheffé, H.: *The Analysis of Variance*. New York: John Wiley & Sons, 1959.

[20] Taylor, W.: "Small Sample Considerations in Estimation from Panel Data," *Journal of Econometrics*, 13(1980), 203–223.

[21] Wu, D.: "Alternative Tests of Independence Between Stochastic Regressors and Disturbances," *Econometrica*, 41(1973), 733–750.