BASEMENT

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LIBRARIES

working paper
department
of economics

PANEL DATA AND UNOBSERVABLE INDIVIDUAL EFFECTS

J.A. Hausman (MIT)
W.E. Taylor (Bell Labs)

Number 255 February 1980

massachusetts
institute of
technology
50 memorial drive
cambridge, mass. 02139

FIRST DRAFT
COMMENTS WELCOME

January 1980

PANEL DATA AND UNOBSERVABLE INDIVIDUAL EFFECTS

J.A. Hausman (MIT)

W.E. Taylor (Bell Labs)

Number 255 February 1980

Hausman thanks the NSF for research support. Helpful comments have been made by Rob Engle and Paul Ruud. An earlier version of the paper was presented at the Summer Econometric Meetings in Montreal, June 1979.

PANEL DATA AND UNOBSERVABLE INDIVIDUAL EFFECTS

by

Jerry A. Hausman
M.I.T.

and

William E. Taylor
Bell Laboratories

Abstract

An important purpose in pooling time series and cross section data is to control for individual-specific unobservable effects which may be correlated with other explanatory variables: e.g., latent ability in measuring returns to schooling in earnings equations or managerial ability in measuring returns to scale in firm cost functions. Using instrumental variables and the time-invariant characteristic of the latent variable, we derive

1) a test for the presence of this effect and for the over-identifying restrictions we use;

2) necessary and sufficient conditions for identification of all the parameters in the model; and

3) 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 and Balestra-Nerlove estimates - particularly a significantly higher estimate of the returns to schooling.

Digitized by the Internet Archive
in 2011 with funding from
Boston Library Consortium Member Libraries

http://www.archive.org/details/paneldataunobser00haus

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. A specification test proposed by Hausman (1978) and subsequently used in a number of applied contexts has indicated that correlated individual effects may be present in many econometric applications to individual or firm data.

The traditional technique to overcome this problem has been to eliminate the individual effects in the sample by transforming the data into deviations from individual means.[1] However, the least squares coefficient estimates from the transformed data, (which are known as "within-groups" or "fixed effects" estimates), have two important shortcomings: (1) all time invariant variables are eliminated by the transformation so that their coefficients cannot be estimated, and (2) under certain circumstances, the within-groups estimator is not fully efficient since it ignores variation across individuals in the sample. The first problem is

---

[1]This technique corresponds to Model I of the analysis of variance, e.g., Scheffé (1959). When used in analysis of covariance, errors in measured variables can create a serious problem since they are exacerbated by the data transformation.

usually the more serious, since in many applications, primary interest is attached to the unknown coefficients of these variables, e.g., to the coefficient of schooling in a wage equation specification.

To consider a specific model, let

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

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

The primary focus of this paper involves the potential correlation of α_i with the columns of X and Z. If such correlations are present, least squares (OLS) or generalized least squares (GLS) will yield biased and inconsistent estimates of both β and γ. Transforming the data into deviations from individual means eliminates the correlation problem by eliminating the time-invariant α_i; unfortunately, at the same time, it eliminates the Z_i, precluding estimation of γ. Another possible approach is to find instruments for those columns of X and Z which are considered potentially correlated with α_i and perform instrumental variables estimation on equation

(1.1) or on a single cross-section. But it may be difficult or impossible to find appropriate instruments, excluded from equation (1.1), which are not correlated with α_i. For instance, use of family background variables as instruments for schooling in a wage equation seems unlikely to eliminate bias, since the unobserved individual effect is likely to be correlated with measures of family background.

Specifications similar to equation (1.1) have been used in at least two empirical contexts. If equation (1.1) represents a cost or production function and α_i denotes the unobservable managerial efficiency of the i'th firm, Mundlak (1961) has suggested the use of the within-groups estimator to produce unbiased estimates of the remaining parameters. If Y_{it} denotes the wage of the i'th individual in the t'th time period, one of the Z_i's measures his schooling, and α_i denotes the unmeasureable component of his initial ability or ambition, then equation (1.1) represents a specification for measuring the returns to education. To the extent that unmeasureable ability and schooling are correlated, the OLS estimates are biased and inconsistant. Griliches (1977) has relied on an instrumental variables approach, using family background variables excluded from equation (1.1) as instruments. Another approach is the factor analysis model, pioneered in this context by Jöreskog (1973) and applied to the schooling problem by Chamberlain and Griliches (1975) and Chamberlain

(1978). The factor analysis approach relies for identification upon orthogonality assumptions which must be imposed on observable and unobservable components of α_i. The method presented in this paper 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 specify which variables among the included right hand side variables of equation (1.1) are uncorrelated with the individual effects, conditions may hold such that all of the β's and γ's may be consistently estimated. By combining the unbiased within-groups estimates of the β's with the biased between-groups estimates of the β's and γ's, adjustments can be made which produce consistent estimates of γ and more efficient estimates of β. An alternative approach which uses these assumptions observes that 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_{it} for those variables which are uncorrelated with α_i. For instance, in our returns to schooling example, it may be safe to assume that health status and age are uncorrelated with α_i, but one might be reluctant to assume that unemployment and α_i were uncorrelated. 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 totally on *a priori* assumptions.

The plan of the paper is as follows. In Section 2, we formally set up the model and consider estimates proposed in the literature for cases in which α_i is uncorrelated or correlated with some of the independent variables. In the latter case, we propose a consistent but inefficient estimator of *all* the parameters in the model. In Section 3, we discuss a variety of tests which determine when such correlations may be present, generalizing results of Hausman (1978). In Section 4, 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 model. We derive a test of the correlation assumptions necessary for identification and estimation, applying results from Hausman and Taylor (1980). Section 5 connects our results with Mundlak's (1978) paper and derives Gauss-Markov properties of our estimator in special cases. Finally, in Section 6, we apply the procedure to an earnings function, focusing on the returns to schooling. These 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 its properties in the absence and presence of specification errors of the form $E(\alpha_i|X_{it},Z_i) \neq 0$. Let

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

where we have reason to believe that $E(\varepsilon_{it}|X_{it},Z_i) = E(\alpha_i|X_{it},Z_i) \neq \underset{\sim}{0}$. That is, some of the measured variables among the X_{it} and the Z_i are correlated with the unobservable individual-specific effects α_i. It will prove convenient to distinguish columns of X and Z which are asymptotically uncorrelated with α_i from those which are not. For fixed T, let

$$\begin{aligned} &\underset{N\to\infty}{\text{plim}} \frac{1}{N} X_{1i.}'\alpha_i = \underset{\sim}{0}, \qquad \underset{N\to\infty}{\text{plim}} \frac{1}{N} Z_{1i}'\alpha_i = \underset{\sim}{0}, \\ &\underset{N\to\infty}{\text{plim}} \frac{1}{N} X_{2i.}'\alpha_i = h_x \neq \underset{\sim}{0}, \text{ and } \underset{N\to\infty}{\text{plim}} \frac{1}{N} Z_{2i}'\alpha_i = h_z \neq \underset{\sim}{0} \end{aligned} \tag{2.2}$$

where $X_{it} = [X_{1it} \vdots X_{2it}]$, $Z_i = [Z_{1i} \vdots Z_{2i}]$, and the dimensions of X and Z are $TN\times k = [TN\times k_1 \vdots TN\times k_2]$ and $TN\times g = [TN\times g_1 \vdots TN\times g_2]$

respectively. Note that, somewhat unconventionally, X_{it} and Z_i denote matrices whose subscripts indicate variation over individuals ($i=1,\dots,N$) and time ($t=1,\dots,T$). Observations are ordered first by individual; α_i and each column of Z_i are thus TN vectors having T identical entries for each $i=1,\dots,N$.

We are thus assuming that k_2 columns of X_{it} and g_2 columns of Z_i are correlated (asymptotically) with the time-invariant unobservable α_i: $E(\alpha_i|X_{it},Z_i) \neq 0$. Implicitly, we are also assuming that there are no other observable exogenous variables which - along with the X_{it} and Z_i - could enable us to write $E(\alpha_i|X_{it},Z_i)$ as a linear function of observables plus an orthogonal error. In addition, we assume no knowledge of other relationships in which the unobservable α_i enters in a similar or known fashion. In sum, we are thinking of α_i as an inherently unmeasurable individual-specific effect about which we have only the prior information embodied in equations (2.1) and (2.2). Operationally, this means that we cannot obtain a consistent estimate of the conditional mean of Y_{it} from available observable data without further assumptions regarding the relative magnitudes of (k_1,k_2,g_1,g_2).

To derive consistent and efficient estimators for (β,γ) in equation (2.1), it will be helpful to recall the menu of appropriate estimators in the absence of misspecification. If we let ι_T denote a T vector of ones, two convenient orthogonal projection operators can be defined as

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

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., $P_V Y_{it} = \frac{1}{T}\sum Y_{it} \equiv Y_{i.}$. Similarly Q_V produces a vector of deviations from group means: i.e., $Q_V Y_{it} = \tilde{Y}_{it} = Y_{it} - Y_{i.}$. Moreover, Q_V is orthogonal by construction to any time-invariant vector of observations: $Q_V Z_i = Z_i - \frac{1}{T}\sum_{t=1}^{T} Z_i = \underset{\sim}{0}$.

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. As pointed out in the Introduction, this within-groups estimator has two serious defects: (i) it ignores between group variation in the data, and (ii) the transformation Q_V eliminates time-invariant observables such as Z_i.

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

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

or

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

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 $E(\alpha_i | X_{it}, Z_i) \neq 0$. Similarly, the sum of squared residuals from equation (2.4) provides a biased and inconsistent estimator for $\mathrm{Var}(\alpha_i + \eta_{i.}) = \sigma_\alpha^2 + \frac{1}{T}\sigma_\eta^2$ when $E(\alpha_i | X_{it}, Z_i) \neq 0$.

In the absence of misspecification, the optimal use of within and between groups information is a straightforward calculation. Let

$$Y_{it} = X_{it}\beta + Z_i\gamma + \varepsilon_{it}$$

where $E(\varepsilon_{it}|X_{it},Z_i) = 0$ and $\mathrm{cov}(\varepsilon_{it}) = \Omega = \sigma_\eta^2 I_{TN} + \sigma_\alpha^2[I_N \otimes \iota_T\iota_T'] = \sigma_\eta^2 I_{TN} + T\sigma_\alpha^2 P_V$, 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 and η_{it} to be normally distributed, it is easy to show that the within 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 optimal matrix-weighted average of the between and within groups estimators, where the weights depend upon the variance components σ_α^2 and σ_η^2, and are chosen to $\min_\Delta \mathrm{var}(\hat\beta_{GLS}'\hat\gamma_{GLS}')' = \Delta V_B\Delta' + (I-\Delta)V_W(I-\Delta)'$, where V_B, V_W denote the covariance matrices of the between and within groups coefficient estimator. The solution can be written as

$$\begin{pmatrix}\hat\beta_{GLS}\\ \hat\gamma_{GLS}\end{pmatrix} = \Delta\begin{pmatrix}\hat\beta_B\\ \hat\gamma_B\end{pmatrix} + (I-\Delta)\begin{pmatrix}\hat\beta_W\\ 0\end{pmatrix}$$

(see, e.g., Maddala (1971)), where

$$\Delta = \left\{(X:Z)'P_V(X:Z) + \frac{\sigma_\eta^2 + T\sigma_\alpha^2}{\sigma_\eta^2}(X:Z)'Q_V(X:Z)\right\}^{-1}(X:Z)'P_V(X:Z)$$

$$= (V_B+V_W)^{-1}V_B.$$

This is frequently known as the Balestra-Nerlove estimator; it requires knowledge of the variance components σ_α^2 and σ_η^2 but one can substitute consistent estimates for the variance components without loss of asymptotic efficiency.[2] Observe that if $E(\alpha_i|X_{it},Z_i) \neq 0$, these Gauss-Markov estimators will be biased and if $h_x \neq 0$ and $h_z \neq 0$, they will be inconsistent, since they are matrix-weighted averages 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. Nerlove (1971) shows that Ω has two distinct eigenvalues, $\sigma_\eta^2 + T\sigma_\alpha^2$ of multiplicity N and σ_η^2 of multiplicity TN-N; from equations (2.3) and (2.4), it follows that the N and TN-N basis vectors spanning the column spaces of P_V and Q_V span the eigenspaces of Ω corresponding to the eigenvalues $\sigma_\eta^2 + T\sigma_\alpha^2$ and σ_η^2 respectively. Thus if we weight these basis

[2] The finite sample implications of this substitution are explored in Taylor (1979).

vectors by $\theta = \left[\sigma_\eta^2/(\sigma_\eta^2 + T\sigma_\alpha^2)\right]^{1/2}$, we obtain the following.

Proposition 2.1: The TN×TN non-singular matrix

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

transforms the disturbance $\alpha_i + \eta_{it}$ into a sequence of independent and identically distributed random variables.

Proof: Basis vectors of the column spaces of P_V and Q_V can be chosen to diagonalize Ω. To make the resulting matrix scalar, it is necessary to multiply P_V by the square root of the ratio of the two distinct eigenvalues:

$$\Omega^{-1/2}\,\Omega\,\Omega^{-1/2} = [\theta P_V + Q_V]\left[\sigma_\eta^2 I_{TN} + T\sigma_\alpha^2 P_V\right][\theta P_V + Q_V]$$

$$= \theta^2(\sigma_\eta^2 + T\sigma_\alpha^2)P_V + \sigma_\eta^2 Q_V = \sigma_\eta^2 I_{TN}.$$

Alternatively, note that

$$\Omega^{-1/2}\varepsilon_{it} = [I_{TN} - (1-\theta)P_V](\alpha_i + \eta_{it})$$

$$= \alpha_i - (1-\theta)\alpha_i + \eta_{it} - (1-\theta)\eta_{i.} = \theta(\alpha_i + \eta_{i.}) + \tilde{\eta}_{it};$$

since the last two terms are orthogonal,

$$\operatorname{cov}(\Omega^{-1/2}\varepsilon_{is}, \Omega^{-1/2}\varepsilon_{jt}) \begin{cases} = 0 & s \neq t \text{ or } i \neq j \\ = \sigma_\eta^2 & s = t \text{ and } i = j \end{cases}.$$

We can then premultiply equation (2.1) by $\Omega^{-1/2}$, or - equivalently - transform the data so that

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

(2.5)

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

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

2.2 Consistent But Inefficient Estimation

Despite correlation between the unobservables and the observable explanatory variables, we saw in Section 2.1 that $\hat{\beta}_W$ is unbiased and consistent for β but makes no use of between-group variation in the data. Furthermore, $Q_V Z_i\gamma = 0$, so that it appears impossible to obtain an estimate of γ from the within-group data. Under appropriate assumptions about k_1 and g_2, (the number of exogenous X's and endogeneous Z's), it is possible to obtain consistent estimates of γ, using the residuals from the within-groups regression. Let

$$\hat{d}_{it} = Y_{it} - X_{it}\hat{\beta}_W = \{I_{NT}-X_{it}(X_{it}'Q_V X_{it})^{-1}X_{it}'Q_V\}\, Y_{it}$$

be the TN vector of group means estimated from the within-group residuals. This simplifies to

$$(2.6)\quad \hat{d}_{it} = Z_i\gamma + \alpha_i + \{I_{NT} - X_{it}(\tilde{X}_{it}'\tilde{X}_{it})^{-1}\tilde{X}_{it}'\}\ \eta_{it}$$

Treating the last two terms as an unobservable disturbance, consider estimating γ in 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} which are correlated with α_i. A necessary condition for this - and thus for the identification of γ in equation (2.6) - is clearly that $k_1 \geq g_2$: that there be at least as many exogenous time-varying variables as there are endogenous time-invariant variables. We shall return to the question of identification of (β,γ) from equation (2.1) in the next section.

The 2SLS estimator for γ in equation (2.6) is

$$(2.7)\qquad \hat{\gamma}_W = (Z_i'P_WZ_i)^{-1}Z_i'P_W\hat{d}_{it}$$

where W denotes the instruments X_1 and Z_1, and P_W is the orthogonal projection operator onto the column space of W. The sampling error is given by

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

and under the usual assumptions governing the X and Z processes, the 2SLS estimator is consistent for γ, since for fixed T, $\text{plim}\ \frac{1}{N} W'\alpha_i = \underset{\sim}{0}$ and $\text{plim}\ \frac{1}{N} X_{it}'\eta_{it} = \underset{\sim}{0}$. The fact that the $\hat{d}_{it}$ are calculated from the within-group residuals suggests that if $\hat{\beta}_W$ is not fully efficient, then $\hat{\gamma}_W$ in equation (2.7) may not be fully efficient.

Having consistent estimates of β and - under certain circumstances - γ, we can construct consistent estimators for the variance components σ_α^2 and σ_η^2. First, a consistent estimate of σ_η^2 can always be derived from the within-group residuals; i.e., from the least squares residuals from equation (2.3). 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

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

so that if $s_\eta^2 = \frac{1}{N(T-1)} \tilde{Y}'_{it} Q_{\tilde{X}} \tilde{Y}_{it}$,

$$\underset{N\to\infty}{\text{plim}}\ s_\eta^2 = \underset{N\to\infty}{\text{plim}} \frac{1}{N(T-1)} \tilde{Y}'_{it} Q_{\tilde{X}} \tilde{Y}_{it}$$

$$= \underset{N\to\infty}{\text{plim}} \frac{1}{N(T-1)} \tilde{\eta}'_{it} \tilde{\eta}_{it} - 0 = \underset{N\to\infty}{\text{plim}} \frac{1}{N(T-1)} \eta'_{it} Q_V \eta_{it}$$

$$= \sigma_\eta^2$$

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

$$\hat{\sigma}^2 = \frac{1}{N} (Y_{i.} - X_{i.}\hat{\beta}_W - Z_i\hat{\gamma}_W)^1 (Y_{i.} - X_{i.}\hat{\beta}_W - Z_i\hat{\gamma}_W);$$

then

$$\underset{N\to\infty}{\text{plim}}\ \hat{\sigma}^2 = \underset{N\to\infty}{\text{plim}} \frac{1}{N} (Y_{i.} - X_{i.}\beta - Z_i\gamma)'(Y_{i.} - X_{i.}\beta - Z_i\gamma)$$

$$= \underset{N\to\infty}{\text{plim}} \frac{1}{N} (\alpha_i + \eta_{i.})'(\alpha_i + \eta_{i.})$$

$$= \sigma_\alpha^2 + \frac{1}{T}\sigma_\eta^2$$

so that $s_\alpha^2 = \hat{\sigma}^2 - \frac{1}{T} s_\eta^2$ is consistent for σ_α^2.

3. Specification Tests Using Panel Data

A crucial assumption of the cross-section regression specification $Y_i = X_i\beta + \varepsilon_i$ $(i = 1,\ldots,N)$ is that the conditional expectation of the disturbances given knowledge of the right hand side variables is zero: $E(\alpha_i|X_i) = \underset{\sim}{0}$. A great advantage of panel data is that following the cross-section panel over time allows a test of this hypothesis. To derive such a test, we consider the random effects specification of equation (1.1), including the time-invariant Z_i among the X_{it} for notational convenience:

$$Y_{it} = X_{it}\beta + \alpha_i + \eta_{it} \quad (i = 1,\ldots,N; t = 1,\ldots,T). \tag{3.1}$$

The unobservable disturbance has been broken into two terms, the first of which reflects unobservable individual characteristics unchanging over time which are not represented in $X_{it}\beta$. The η_{it} are random shocks which we assume to be orthogonal to α_i and the X_{it}.

The specification tests which we consider test the null hypothesis

$$H_0: \; E(\alpha_i|X_{it}) = \underset{\sim}{0},$$

against the alternative that $E(\alpha_i|X_{it}) \neq \underset{\sim}{0}$. 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 the slope parameters by controlling for the correlation between α_i and X_{it}. An asymptotically efficient procedure for doing this is outlined in the latter half of this paper.

Recall the three estimators for β in equation (3.1) - $\hat{\beta}_W$, $\hat{\beta}_B$, $\hat{\beta}_{GLS}$ - which we discussed in the previous section. Since these estimators have different properties under the null and alternative hypotheses, we are led naturally to form three different specification tests.

(1) GLS vs. within. Under the null hypothesis, $\hat{\beta}_{GLS}$ is efficient, while under the alternative, it is inconsistent. $\hat{\beta}_W$ is consistent under both, but inefficient. Consider the vector

$$\hat{q}_1 = \hat{\beta}_{GLS} - \hat{\beta}_W.$$

Under H_0, $\underset{N\to\infty}{\text{plim}}\ \hat{q}_1 = \underset{\sim}{0}$, while under H_1, $\underset{N\to\infty}{\text{plim}}\ \hat{q}_1 \neq \underset{\sim}{0}$, since $\text{plim}\ \hat{\beta}_{GLS} \neq \beta = \text{plim}\ \hat{\beta}_W$. Hausman (1978) showed that $\text{var}(\hat{q}_1) = \text{var}(\hat{\beta}_W) - \text{var}(\hat{\beta}_{GLS})$ so a X^2 test is easily formed. This test has been used fairly frequently and has appeared to be quite powerful.

(2) <u>GLS vs. between</u>. Under H_0, $\hat{\beta}_B$ is inefficient while under the alternative hypothesis it is inconsistent and $\operatorname*{plim}_{N\to\infty} \hat{\beta}_B \neq \operatorname*{plim}_{N\to\infty} \hat{\beta}_{GLS} \neq \beta$. Thus deviations of the vector

$$\hat{q}_2 = \hat{\beta}_{GLS} - \hat{\beta}_B$$

from the zero vector cast doubt upon the null hypothesis. Using Hausman's (1978) results, $\operatorname{var}(\hat{q}_2) = \operatorname{var}(\hat{\beta}_B) - \operatorname{var}(\hat{\beta}_{GLS})$, which gives rise to another chi-square statistic.

(3) <u>Within vs. between</u>. As we have seen, under H_0, $\operatorname*{plim}_{N\to\infty} \hat{\beta}_B = \operatorname*{plim}_{N\to\infty} \hat{\beta}_W = \beta$, whereas under the alternative hypothesis, $\operatorname*{plim}_{N\to\infty} \hat{\beta}_W = \beta \neq \operatorname*{plim}_{N\to\infty} \hat{\beta}_B$. Also, from the characterization in Section 2, the within and between groups estimators lie in orthogonal subspaces so that $\hat{\beta}_W$ and $\hat{\beta}_B$ are uncorrelated. Thus if

$$\hat{q}_3 = \hat{\beta}_W - \hat{\beta}_B,$$

$\operatorname{var}(\hat{q}_3) = \operatorname{var}(\hat{\beta}_W) + \operatorname{var}(\hat{\beta}_B)$, and a third chi-square statistic is available.

In considering these three tests, Hausman (1978) conjectured that the first test might be better than the third since $V(\hat{q}_2) \geq V(\hat{q}_1)$; while Pudney (1979) conjectured that the second test might be better than the third because

$\hat{\beta}_{GLS}$ is efficient.[3] (Actually, $\mathrm{var}(\hat{q}_3) \geq \mathrm{var}(\hat{q}_1)$.)

A somewhat surprising event occurs, however, if we parameterize the relationship between α_i and the X_{it}. Using the specification of Mundlak (1978) that

$$\alpha_i = X_{i\cdot}\pi + \omega_i \tag{3.3}$$

and assuming that ω_i and η_{it} are independent joint normal leads to a straightforward maximum likelihood problem. Assuming that we know Ω for simplicity, what we might call the Holy Trinity of statistical tests appears. That is, the three tests outlined above correspond to the likelihood ratio, Lagrange multiplier (Rao efficient score), and Wald test respectively.

This, however, creates a problem. Assuming Ω to be known, the within and between groups estimators are jointly sufficient for β so that no other information should be present in the data. In addition, we know the likelihood ratio, Lagrange multiplier and Wald tests to be identical for testing linear restrictions on linear models and the null hypothesis $E(\alpha_i | X_{it}) = \underset{\sim}{0}$ corresponds - in this special case - to the linear restriction $\pi = \underset{\sim}{0}$.

---

[3] Pudney actually considered using estimates of ε_{it} from the three estimators and then basing tests upon the sample covariance $X'\hat{\varepsilon}$, using either the within or GLS estimate of β to form $\hat{\varepsilon}$. However, the tests are considerably simpler to apply by directly comparing the $\hat{\beta}$'s; Pudney was not aware that using $\hat{\beta}_{GLS}$ to form $\hat{\varepsilon}$ is equivalent to the second test.

In this case, it is evident that the tests must be identical, and it is straightforward to demonstrate this identity in general. Recall that $\hat{\beta}_{GLS}$ can be written as a matrix-weighted average of $\hat{\beta}_B$ and $\hat{\beta}_W$,

$$\hat{\beta}_{GLS} = \Delta\hat{\beta}_B + (I-\Delta)\hat{\beta}_W$$

where the weight matrix Δ is non-singular.[4] While this form of the GLS estimator is computationally inconvenient, it is extremely easy to derive the relationships among tests (1-3) from it. Considering the tests in turn

$$\hat{q}_1 = \hat{\beta}_{GLS} - \hat{\beta}_W = \Delta(\hat{\beta}_B-\hat{\beta}_W) = -\Delta\hat{q}_3$$

and

$$\hat{q}_2 = \hat{\beta}_{GLS} - \hat{\beta}_B = (I-\Delta)(\hat{\beta}_W-\hat{\beta}_B) = (I-\Delta)\hat{q}_3$$

so that the three tests are all non-singular transformations of each other. Their operating characteristics must therefore be identical. Indeed,

<u>Proposition 3.1</u>: The chi-square statistics for tests (1-3) are numerically exactly identical.

<u>Proof</u>: Recall that $\text{var}(\hat{q}_3) = \text{var}(\hat{\beta}_B) + \text{var}(\hat{\beta}_W) \equiv V_3$. Then $\text{var}(\hat{q}_1) = \Delta V_3\Delta' \equiv V_1$ and $\text{var}(\hat{q}_2) = (I-\Delta)V_3(I-\Delta)' \equiv V_2$. Thus

---

[4] Note that if σ_α^2 and σ_η^2 are unknown and must be estimated, the above identity holds exactly in finite samples, using the estimated weight matrix $\hat{\Delta}$.

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

and

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

Since the chi-square statistics which define the tests are identical, it makes no difference which test is used. Computationally, the first test might be preferable since it requires calculation of only those estimators which might be used under either the null (GLS) or alternative (within-groups) hypothesis. Note that Hausman (1978) shows that this test - and, as we have just shown, tests (2) and (3) - can be set up as an F test in an auxilliary regression so that direct calculation of the quadratic form is unnecessary.

The important result in this section, however, is that all three tests are identical. Despite intuition and folk wisdom to the contrary, it makes no difference which comparison is used, in testing for the presence of correlation between α_i and the columns of X and Z. In the next section, we extend these specification tests to determine when the prior information in equation (2.2) - upon which our identificatio and estimation results depend - is in agreement with the data.

4. Instrumental Variables Estimators

4.1 Identification

In this section, we address the question of the identification of some or all of the elements of (β,γ) using only the prior information embodied in equation (2.2) and the time-invariance characteristic of the latent variable α_i. Because the only component of ε_{it} which is correlated with the explanatory variables is time-invariant, any vector that is orthogonal to a time-invariant vector can be used as an instrument, and TN-N linearly independent vectors with this characteristic can always be constructed. Recall from Section 2 that

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

is an idempotent matrix of rank TN-N which transforms a TN vector into deviations from individual means. Thus any set of TN-N basis vectors for the column space Q_V is orthogonal to any time-invariant vector. In particular, $Q_V\alpha_i = 0$, from which we conclude that there are always at least TN-N instruments available in equation (2.1).

Unfortunately, as noted in the introduction, Q_V is also orthogonal to Z_1 which violates the requirement that instruments be correlated with all of the explanatory variables. We thus need to specialize the familiar results on identification in linear models to identification of subsets of parameters. Consider the canonical linear simultaneous equations model

$$Y = X\beta + e \tag{4.1}$$

where some columns of X are endogenous and the matrix Z contains T observations on all variables for which $\operatorname{plim}_{T\to\infty} Z'e = 0$. Consider the projection of equation (4.1) onto the column space of Z:

$$P_Z Y = P_Z X\beta + P_Z e. \tag{4.2}$$

Now, if λ is a k vector of known constants,

<u>Lemma</u>: A necessary and sufficient condition for β to be identified in equation (4.1) is that every linear function $\lambda'\beta$ be estimable in equation (4.2).

This useful result follows immediately from a theorem of F. Fisher (1966, Theorem 2.7.2, p. 56) which implies that the parameters of a structural equation are identified if and only if the two stage least squares estimator is well-defined, which in turn is equivalent to the non-singularity of the matrix $X'P_Z X$. A function $\lambda'\beta$ is estimable in equation (4.2) if and only if λ' lies in the row space of $P_Z X$ (Scheffé

(1959), Theorem 1, p. 13). For this to hold for any λ, $P_Z X$ must be of full column rank, which completes the proof.

Suppose that the conditions of this lemma are not attained, as occurs in equation (2.1), taking elements of the column space of Q_V as exogenous. Then

Corollary: A necessary and sufficient condition for a particular (set of) linear function(s) $\lambda'\beta$ to be identified in equation (4.1) is that $\lambda'\beta$ be estimable in equation (4.2).

Clearly, if $\lambda'\beta$ is estimable in equation (4.2), then it is identified. On the other hand, if $\lambda'\beta$ is identified in (4.1), it has a consistent estimator $a'Y$ for which

$$\operatorname*{plim}_{T\to\infty} a'Y = \operatorname*{plim}_{T\to\infty} a'X\beta + \operatorname*{plim}_{T\to\infty} a'e = \lambda'\beta$$

for all β. Thus plim $a'e = 0$ so that a lies in the column space of Z, the set of all exogenous variables. Hence $a = P_Z b$ for some T vector b and

$$\operatorname{plim} a'X\beta = \operatorname{plim} b'P_Z X\beta = \lambda'\beta$$

for all β, so that λ' lies in the (asymptotic) row space of $P_Z X$. By the previously cited theorem of Scheffe, $\lambda'\beta$ is thus (asymptotically) estimable in equation (4.2), completing the proof.

Returning to the question of identification in equations (2.1) and (2.2), we observe that even if none of the columns of X or Z is exogenous ($k_1 = g_1 = 0$), all of the elements of β are identified: 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. The two stage least squares (2SLS) estimator for β in this case is

$$\hat{\beta}_{2SLS} = (X_{it}' Q_V X_{it})^{-1} X_{it}' Q_V Y_{it} = (\tilde{X}_{it}' \tilde{X}_{it})^{-1} \tilde{X}_{it}' \tilde{Y}_{it} = \hat{\beta}_W$$

which is identical to the within-groups estimator. In this case ($k_1 = g_1 = 0$), it is easy to verify that γ is not identified.[5]

If prior information suggests that certain columns of X and Z are exogenous ($k_1 > 0$, $g_1 > 0$), then the columns of X_{1it} and Z_{1i} must be added to the list of instruments.

---

[5] This fact underscores the importance of the observation that the disturbance in equation (2.6) is correlated with Z_i, so that instruments are required to estimate γ.

Let W denote the matrix $[Q_V \vdots X_{1it} \vdots Z_{1i}]$ and let P_W be the orthogonal projection operator onto the column space of W. Then, corresponding to the familiar rank condition, we have

Proposition 4.1: A necessary and sufficient condition that the entire vector of parameters (β,γ) be identified in equation (2.1) is that the matrix

$$\begin{pmatrix} X_{it}' \\ \cdots \\ Z_i' \end{pmatrix} P_W(X_{it} \vdots Z_i)$$

be non-singular.

Corresponding to the order condition, we have

Proposition 4.2: A necessary condition for the identification of (β,γ) in equation (2.1) is that $k_1 \geq g_2$.

Proof: The first proposition is a simple restatement of the earlier Lemma. Proposition 4.2 asserts that we must have as many (or more) exogenous X's as we have endogenous Z's: a familiar enough requirement from the instrumental variables literature, but here it is used to identify an otherwise unidentified subset of the parameters. To prove it, observe that rank $[P_W(X_{it} \vdots Z_i)] \leq$ rank $[P_W X_{it}]$ + rank $[P_W Z_i]$ =

k + rank $[P_W Z_i]$ so that a necessary condition for the matrix in Proposition 4.1 to be non-singular is that rank $[P_W Z_i] = g$. Since Z_i is orthogonal to Q_V, $k_1 \geq g_2$ is necessary for rank $[P_W Z_i]$ to equal g, which completes the proof.

This discussion of identification in structural models with panel data has revealed a few noteworthy features. First, given only the assumption that individual-specific unobservable components cause some explanatory variables to be correlated with the disturbance, it is remarkable to find that the coefficients of the time-varying variables are identified while those of the time-invariant observations are not. Second, Mundlak (1978) has shown that when all the columns of X and Z are correlated with α_i, (i.e., $k_1=g_1=0$), $\hat{\beta}_W$ is Gauss-Markov for β. In this case, the 2SLS estimator coincides with the within-groups estimator for β and the components of γ are not identified.

Finally, identification of γ can be attained by finding additional instruments - at least one for every endogenous column of Z_i. Curiously, the k_1 exogenous columns of X_{1it} which are <u>included</u> in the structural equation (2.1) in question, are the only candidates for these identifying instruments. This contrasts with the conventional simultaneous equations model in which <u>excluded</u> 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 error is correlated with (X_2, Z_2). Since $X_{1it} = \tilde{X}_{1it} + X_{1i.}$, $\tilde{X}_{1it}$ can be used as an instrument for X_1, and $X_{1i.}$ can be an instrument for Z_{2i}.

4.2 Estimation

If the parameters of equation (2.1) are identified by means of a specified set of exogenous 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 $\mathrm{var}(\varepsilon_{it}) = \Omega = \sigma_\eta^2 I_{TN} + T\sigma_\alpha^2 P_V$ 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, two stage least squares (2SLS) estimates of (β, γ) in

$$(4.3) \quad \Omega^{-1/2} Y_{it} = \Omega^{-1/2} X_{it}\beta + \Omega^{-1/2} Z_i \gamma + \Omega^{-1/2} \varepsilon_{it},$$

taking as exogenous, X_{1it} and Z_{1i} would be asymptotically efficient, in the sense of converging in distribution to the limited information maximum likelihood estimator.

Alternatively, the information embodied in equations (2.1) and (2.2) can be written as a single structural equation and two multivariate reduced form equations:

$$Y_{it} = X_{it}\beta + Z_i\gamma + \varepsilon_{it}$$

$$X_{2it} = X_{1it}\pi_{11} + Z_{1i}\pi_{12} + Q_V\pi_{13} + v_{1it}$$

$$Z_{2i} = X_{1it}\pi_{21} + Z_{1i}\pi_{22} + Q_V\pi_{23} + v_{2it}$$

where X_1, Z_1, and Q_V are exogenous, v_1 and v_2 are correlated with α_i and thus with ε_{it}, and $\pi_{23} = 0$. Transforming the structural equation by θ-differencing the data, we can rewrite the system as

$$\begin{aligned} &\Omega^{-1/2}Y_{it} = \Omega^{-1/2}X_{it}\beta + \Omega^{-1/2}Z_i\gamma + \Omega^{-1/2}\varepsilon_{it} \\ (4.4) \qquad &X_{2it} = X_{1it}\pi_{11} + Z_{1i}\pi_{12} + Q_V\pi_{13} + v_{1it} \\ &Z_{2i} = X_{1it}\pi_{21} + Z_{1i}\pi_{22} + Q_V\pi_{23} + v_{2it} \end{aligned}$$

again assuming the variance components - and thus Ω - to be known.

This system represents the information in equations (2.1-2.2) in a form convenient for discussing efficiency of estimators for β and γ. In particular, equations (4.4) 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 all - by definition - just identified. Since the disturbance covariance matrix in equations (4.4) is unknown, the results of Lahiri and Schmidt (1978) imply that OLS is inconsistent but 3SLS is fully efficient. Finally, since the reduced forms are just identified, 3SLS estimates of (β,γ) in the entire system are identical to 3SLS estimator of (β,γ) in the first equation alone (Narayanan, 1969), and these are, of course, just the 2SLS estimators. Thus 2SLS estimates of (β,γ) in equation (4.3) are fully efficient, given the prior information in equation (2.2), in the sense that they coincide asymptotically with FIML estimators from the system (4.4).

Continue the assumption that Ω is known. 2SLS estimates of β and γ in equation (4.3) are equivalent to OLS estimates of β and γ in

$$P_W\Omega^{-1/2}Y_{it} = P_W\Omega^{-1/2}X_{it}\beta + P_W\Omega^{-1/2}Z_i\gamma + P_W\Omega^{-1/2}\varepsilon_{it}, \tag{4.5}$$

where P_W is the orthogonal projection operator onto the column space of instruments $W = [X_{1it} \vdots Z_{1i} \vdots Q_V]$. Least squares applied to this equation is computationally convenient:

(i) the transformation $\Omega^{-1/2}$ can be done by differencing the data, since $\Omega^{-1/2}X_{it} = X_{it} - (1-\theta)X_{i.}$, where $\theta = \left[\sigma_\eta^2/(\sigma_\eta^2+T\sigma_\alpha^2)\right]^{1/2}$, as shown in equation (2.5),

(ii) the projection of the exogenous variables onto the column space of W yields the variables themselves, and

(iii) the projection of the endogenous variables onto the column space of W can be calculated using only time averages, rather than the entire TN vectors of observations, as shown in section six.

For Ω known, then, the calculation of asymptotically efficient estimators of (β,γ) is straightforward. But the only case of practical interest is where Ω (i.e., the variance components σ_α^2 and σ_η^2) is unknown and must be estimated. The question that immediately arises is how Ω should be estimated when the only concern is the asymptotic efficiency of the derived estimators of (β,γ): Consider the equation

$$P_W\hat{\Omega}^{-1/2}Y_{it} = P_W\hat{\Omega}^{-1/2}X_{it}\beta + P_W\hat{\Omega}^{-1/2}Z_i\gamma + P_W\hat{\Omega}^{-1/2}\varepsilon_{it} \tag{4.6}$$

where $\hat{\Omega}$ is any consistent estimator for Ω. Then

<u>Proposition 4.3</u>: For any consistent estimator $\hat{\Omega}$ of Ω, least squares estimates of (β,γ) in equation (4.6) have the same limiting distribution as the least squares estimates of (β,γ) in equation (4.5), based upon a known Ω.

<u>Proof</u>: For notational convenience, absorb the $Z_i\gamma$ into the $X_{it}\beta$. We shall show that for fixed T, $\sqrt{N}[\hat{\beta}(\hat{\Omega})-\hat{\beta}(\Omega)] \xrightarrow{P} 0$. Adding and substracting β, we can write

$$\sqrt{N}[\hat{\beta}(\hat{\Omega})-\hat{\beta}(\Omega)] = \left\{\left(\frac{1}{N}\, X'\hat{\Omega}^{-1/2}P_W\hat{\Omega}^{-1/2}X\right)^{-1} \frac{1}{N}\, X'\hat{\Omega}^{-1/2}W^*\left(\frac{1}{N}\, W^{*\prime}W^*\right)^{-1}\right\}$$

$$\times \frac{1}{\sqrt{N}}\, W^{*\prime}\hat{\Omega}^{-1/2}(\alpha_i+\eta_{it})$$

$$- \left\{\left(\frac{1}{N}\, X'\Omega^{-1/2}P_W\Omega^{-1/2}X\right)^{-1} \frac{1}{N}\, X'\Omega^{-1/2}W^*\left(\frac{1}{N}\, W^{*\prime}W^*\right)^{-1}\right\}$$

$$\times \frac{1}{\sqrt{N}}\, W^{*\prime}\Omega^{-1/2}(\alpha_i+\eta_{it})$$

where the columns of the $TN\times(TN-N+k_1+g_1)$ matrix W^* span the column space of W. Since $\hat{\Omega}$ is consistent for Ω, the terms in brackets converge in probability to the same matrix. Expanding $\hat{\Omega}^{-1/2}(\alpha_i+\eta_{it}) = \hat{\theta}\alpha_i + \eta_{it} - \hat{\theta}\eta_{i.}$, the last terms reduce to

$$\sqrt{N}\;\hat{\theta}\left[\frac{1}{N}\, W^{*\prime}\alpha_i - \frac{1}{N}\, W^{*\prime}\eta_{i.}\right] + \frac{1}{\sqrt{N}}\, W^{*\prime}\eta_{it}, \quad \text{and}$$

$$\sqrt{N}\;\theta\left[\frac{1}{N}\, W^{*\prime}\alpha_i - \frac{1}{N}\, W^{*\prime}\eta_{i.}\right] + \frac{1}{\sqrt{N}}\, W^{*\prime}\eta_{it}$$

respectively. Since $\text{plim}\ \frac{1}{N} W^{*\prime}\alpha_i = \underset{\sim}{0}$, and $\text{plim}\ \frac{1}{N} W^{*\prime}\eta_{i.} = \underset{\sim}{0}$, and assuming that both $\sqrt{N}\ (\hat{\theta}-\theta)$ and $\frac{1}{\sqrt{N}} W^{*\prime}\eta_{it}$ converge in distribution to some random variable, it follows that

$$\underset{N\to\infty}{\text{plim}}\ \sqrt{N}(\hat{\beta}(\hat{\Omega})-\hat{\beta}(\Omega)) = 0$$

which completes the proof.

We have thus shown that the 2SLS estimators of the parameters in equation (4.3) - using any consistent estimator for the variance components - are asymptotically efficient. These estimators coincide with the LS estimators of β and γ in equation (4.6); for future reference, let us denote them by β^* and γ^*.

4.3 Special Cases

Depending upon the degree of identification of (β,γ) in equation (2.1), the consistent and asymptotically efficient estimators $(\hat{\beta}^*,\hat{\gamma}^*)$ exhibit some interesting peculiarities, which we examine below. First, to establish some terminology, recall the order condition for identification $k_1 \geq g_2$ and its associated rank condition in Proposition 4.1. We shall refer to the case in which these conditions hold with equality as just-identified; when the inequality is strict, the parameters will be said to be overidentified.

Secondly, we shall be interested in estimating β and γ separately from equation (4.6), and two generic formulae will prove convenient. Let $Y = X_1\beta_1 + X_2\beta_2 + e$. Then

<u>Lemma</u>: The following two expressions for the LS estimator of β_1 are identical:

(i) "parse out" X_2 by premultiplying by $Q_2 = I - X_2(X_2'X_2)^{-1}X_2'$ and run LS on $Q_2Y = Q_2X_1\beta_1 + Q_2e$. This yields $\hat{\beta}_1 = (X_1'Q_2X_1)^{-1}X_1'Q_2Y$.

(ii) remove the LS estimates of $X_2\beta_2$ from Y and regress that on X_1: i.e., run LS on $Y - X_2\hat{\beta}_2 = X_1\beta_1 + e_1$. This yields $\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'[Y-X_2(X_2'Q_1X_2)^{-1}X_2'Q_1Y]$.

<u>Proof</u>: The first statement follows immediately from the formula for a partitioned inverse. The second expression can be derived from the first by tedious algebra, and both formulas are probably well-known.

Now, suppose the parameters in equation (2.1) are underidentified. Suppose

(1) <u>$k_1 = g_1 = 0$</u>. Here there are no exogenous variables among the X_{it}'s or the Z_i's and the set of instruments is only $W = [Q_V]$. Since $\Omega^{-1/2} = I_{NT} - (1-\theta)P_V$, $P_W\Omega^{-1/2} = Q_V[I_{NT}-(1-\theta)P_V] = Q_V$ in this case, and $\hat{\beta}^*$ is exactly the within-groups estimator $\hat{\beta}_W$ for β. For the general underidentified case,

(2) $k_1 < g_2$; $k_1 > 0$, $g_1 > 0$. Here, the instruments are $[Q_V \vdots X_1 \vdots Z_1]$ which we write as $[Q_V \vdots H]$ for convenience. The model is

$$(4.7) \qquad Y^*_{it} = P_W X^*_{it} \beta + P_W Z^*_i \gamma + \varepsilon^*_{it}$$

where $\Omega^{-1/2} X_{it}$ is denoted X^*_{it} etc. Note that $P_W Z^*_i = P_H Z^*_i$ since $Q_V Z^*_i = \underset{\sim}{0}$. When $k_1 < g_2$, $P_H Z^*_i$ 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 $P_W Z^*_i \xi = 0$ and the g vector γ cannot be identified since γ and $(\gamma+\xi)$ are observationally equivalent in equation (4.7). To calculate $\hat{\beta}^*$ we "parse out" $P_H Z^*_i$ in equation (4.7). The column space of $P_H Z^*_i$ is precisely the column space of H_i; projecting $P_W X^*_{it}$ onto the orthocomplement of $P_H Z^*_i$ yields $Q_V X_{it}$. Thus $\hat{\beta}^*$ in the generic underidentified case is the within-groups estimator $\hat{\beta}_W$ and there is no consistent estimator for γ.

Suppose the parameters of (2.1) are identified. Let

(3) $\underline{k_1 = g_2;\ k_1 > 0,\ g_2 > 0}$, which is just-identified. Here again, the rank of $P_H Z^*_i$ equals the rank of H, so that $\hat{\beta}^* = \hat{\beta}_W$. To see this algebraically, note that the parse operator in (4.7) is $I_{NT} - P_H Z^*(Z^{*\prime} P_H Z^*)^{-1} Z^{*\prime} P_H$, which

simplifies to $I_{NT} - P_H$ since $H'Z*$ and $Z*'H$ are square, non-singular matrices in the just-identified case. Thus since $[I_{NT}-P_H]P_W X^*_{it} = Q_V X^*_{it}$, the LS estimate of β in (4.7), is $\hat{\beta}_W$. Now, however, $P_H Z^*_i$ has full column rank, and the LS estimate of γ in equation (4.7) is identical to the LS estimate of γ in

$$Y_{it} - P_W X^*_{it}\hat{\beta}_W = P_W Z^*_i \gamma + \varepsilon^*_{it}$$

by the previous Lemma, since $\hat{\beta}_W$ is the LS estimate of β in equation (4.7). Thus $\hat{\gamma}*$ can be written as

$$\hat{\gamma}* = (Z^{*\prime}_i P_W Z^*_i)^{-1} Z^{*\prime}_i P_W (Y_{it} - P_W X^*_{it}\hat{\beta}_W)$$

$$= (Z^{*\prime}_i P_W Z^*_i)^{-1} Z^{*\prime}_i P_W (Y_{i\cdot} - X_{i\cdot}\hat{\beta}_W) = (Z'_i P_W Z_i)^{-1} Z'_i P_W \hat{d}_{it}$$

which is the within-groups estimator of γ defined in equation (2.7). For the just-identified case, then, our 2SLS estimators coincide with the within-groups estimators of both β and γ.

If the parameters are overidentified, the within-groups procedure is no longer appropriate. In the extreme, suppose there are no endogenous variables.

(4) <u>$k_2 = g_2 = 0$</u>, which is overidentified. Here, the set of instruments coincides with the right hand variables in equation (4.3), so that the 2SLS estimator coincides with LS.

For estimated $\Omega^{-1/2}$, this is identical to the GLS procedure in equation (2.5); for known $\Omega^{-1/2}$, it is Gauss-Markov.

If endogenous variables are present, consider

(5) $k_1 > g_2$; $k_2 > 0$, $g_2 > 0$, which is the general overidentified model. In equation (4.7), the column rank of $P_W Z_i^*$ is now g and the column space of $P_W Z_i^*$ no longer coincides with that of H . Thus $\hat{\beta}^*$ will differ from $\hat{\beta}_W$ in the overidentified case. Since $\hat{\gamma}^*$ is derived from the regression of $Y - X\hat{\beta}^*$ on $P_W Z_i^*$, $\hat{\gamma}^*$ will differ from $\hat{\gamma}_W$, which we derived from the regression of $Y - X\hat{\beta}_W$ on $P_W Z_i^*$.

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

4.4 Testing the Identifying Restrictions

More efficient estimates of β and consistent estimates of γ require prior knowledge that certain columns of X_{it} and Z_i are uncorrelated with the latent α_i. An important feature of our model is that when the parameters are overidentified, all of these prior restrictions can be tested. This is an extremely unusual and useful 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 so that $\hat{\beta}_W$ provides a consistent benchmark against which all (or some) of the restrictions in equation (2.2) can be tested by comparing $\hat{\beta}_W$ with $\hat{\beta}^*$. The principles of such tests are outlined in Hausman (1978) and extended in Hausman and Taylor (1980).

Following the latter analysis, we compare our efficient estimator for β (which uses equation (2.2)) with the within-group estimator (which does not require this information for consistency). The null hypothesis is of the form

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

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 upon H_0.

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

$$\begin{aligned} \hat{q} &= [(X^{*\prime}P_WX^*)^{-1}X^{*\prime}P_WQ_Z - (X^{*\prime}Q_VX^*)^{-1}X^{*\prime}Q_VQ_Z]\Omega^{-1/2}Y \\ &= DY^* \end{aligned} \tag{4.8}$$

where $Y^* = \Omega^{-1/2}Y$ is multivariate normal with a scalar covariance matrix and mean 0 under H_0. From the asymptotic Rao-Blackwell argument in Hausman (1978),

$$\mathrm{Var}(\hat{q}) = \mathrm{Var}(\hat{\beta}_W) - \mathrm{Var}(\hat{\beta}^*)$$

since we have shown that $\hat{\beta}^*$ is asymptotically efficient under H_0. From equation (4.8), we can write $\mathrm{Var}(\hat{q}) = DD'$. Since $\hat{q}$ is being used to test k_1+g_1 restrictions - which may be bigger or smaller than its dimension k - the rank of the kxk matrix DD' and the degrees of freedom for the χ^2 test may be less than k. Following Hausman and Taylor (1980),

Lemma: Under the null hypothesis, $\hat{q}'(DD')^{-}\hat{q} \sim \chi^2_d$, where $d = \mathrm{rank}(D)$ and $(DD')^{-1}$ denotes a generalized inverse of the covariance matrix of $\hat{q}$.

Observing in equation (4.8) that Q_V projects onto a proper subspace of the column space of P_W,

Proposition 4.4: Rank (D) = min $[k_1-g_2, k, TN-k]$.

Proof: From Hausman and Taylor (1980),

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

since $P_W = P_H + Q_V$. Under the usual linear independence assumptions, the second term in brackets equals TN-k. For the first term

$$\mathrm{rank}\ (X^{*\prime}P_H) = \min\ [k,\ \mathrm{rank}\ (Q_ZP_H)\ ]$$

and since $P_H = P_Z P_H + Q_Z P_H$, $\text{rank}(Q_Z P_H) = (k_1+g_1) - (g_1+g_2) = (k_1-g_2)$, which we called the degree of overidentification in the previous section.

This specification test of the restrictions embodied 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 (2.2) is false, $\hat{q}$ should differ from zero. Yet the degrees of freedom for the test depend upon the number of overidentifying restrictions (k_1-g_2). Moreover, the degrees of freedom cannot exceed the dimension of β (k) or the degrees of freedom in the original regression (TN-k), whichever is smaller. When the model is just-identified, $\hat{\beta}_W = \hat{\beta}^*$ (see section 4.3); in this case, the degrees of freedom are zero and $\hat{q} = \underset{\sim}{0}$. 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 excgeneity information about X and Z is subject to test by this procedure.[6]

---

[6]This test compares instrumental variables estimators under two nested subsets of instruments: $\hat{\beta}^*$ uses $[Q_V \vdots X_1 \vdots Z_1]$ and $\hat{\beta}_W$ uses $[Q_V]$. If one wished to test particular columns of X_1 and Z_1 for correlation with α_i while maintaining a particular set of identifying assumptions, a test - similar to the above - can be constructed by comparing $(\hat{\beta}^*, \hat{\gamma}^*)$ with $(\hat{\beta}_W, \hat{\gamma}_W)$ where $\hat{\gamma}_W$ is given in equation (2.7). For details, see Hausman and Taylor (1980).

5. Mundlak's Model

A final special case is the model discussed at length by Mundlak (1978), in which no time-invariant observables are present and all explanatory variables are correlated with α_i:

$$Y_{it} = X_{it}\beta + \alpha_i + \eta_{it}. \tag{5.1}$$

The relationship between α and X is expressed by Mundlak through the "auxillary" regression $\alpha_i = X_{i.}\pi + \omega_i$ where no prior information is assumed about π. Mundlak shows that

(i) if α_i is correlated with every column of $X_{i.}$ (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 $X_{i.}$ ($\pi = \underset{\sim}{0}$), 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 (3) and (4) 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 $X_{i.}$ and correlated with others, (π obeys some linear restrictions), the model is overidentified ($k_1 > g_2 = 0$) and case (5) 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 pre-multiply Mundlak's model (5.1) by $\Omega^{-1/2}$ and re-parameterize for convenience:

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

(5.2)

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

where $M_{it} = X_{it}S$, $\xi = S^{-1}\beta$, $\varepsilon^*_{it} = \theta\alpha_i + \eta_{it} - (1-\theta)\eta_{i.} = \alpha^*_i + \eta^*_{it}$ 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

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

which yield r homogeneous restrictions on h_M. Note that

(i) the exogeneity information in equations (2.2) can be expressed as $Rh_x = \underset{\sim}{0}$ 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 exogenous where R_i $(i = 1,\ldots,r)$ is a row of R;

(iii) homogenous restrictions on h_x correspond uniquely to homogenous restrictions on π in Mundlak's specification; i.e., $Rh_x = \underset{\sim}{0} \Rightarrow \operatorname*{plim}_{N\to\infty} \frac{1}{N} R(X_{i\cdot}'X_{i\cdot}) \times (X_{i\cdot}'X_{i\cdot})^{-1}X_{i\cdot}'\alpha_i \Rightarrow \tilde{R}\,\pi = 0$ where $\tilde{R} = R(X_{i\cdot}'X_{i\cdot})$.

In the model (5.2), then, certain linear combinations of the columns of M_{it} are assumed uncorrelated with α_i^* and all of the columns of M_{it} are orthogonal.

<u>Proposition 5.1</u>: The 2SLS estimator $\hat{\xi}^*$ in equation (5.4) is Gauss-Markov for ξ.

Proof: Let F denote the k×k non-singular matrix

$$F = [R' \vdots B']$$

where the columns of B' (k×k-r) are k-r basis vectors for the column space of $I_k - R'(RR')^{-1}R$. Now, reparameterize equation (5.4) as

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

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

which we write as

$$(5.3)\quad \Omega^{-1/2}Y_{it} = \Omega^{-1/2}L_{1it}\delta_1 + \Omega^{-1/2}L_{2it}\delta_2 + \varepsilon_{it}^{*}$$

where $\delta = [\delta_1' \vdots \delta_2'] = F^{-1}\xi$. Consider 2SLS estimates of δ in equation (5.3), using as instruments $W = [Q_V \vdots L_{1it}]$ since $\operatorname{plim}_{N\to\infty} \frac{1}{N} L_{1it}'\alpha_i^{*} = \operatorname{plim} \frac{1}{N} R\, M_{it}'\Omega^{-1}\alpha_i = \underset{\sim}{0}$ by assumption. By construction, $\Omega^{-1/2}L_1$ and $\Omega^{-1/2}L_2$ are orthogonal, and $P_W L_1 = L_1$, so the 2SLS estimator

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

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

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

since $P_W L_2 = Q_V L_2$. 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 estimators were biased and thus not Gauss-Markov?

Proposition 5.2: The 2SLS estimator $\hat{\beta}^*$ coincides with $S\hat{\xi}^*$ but $\hat{\beta}^*$ is biased for β and not a Gauss-Markov estimator.

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

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

where $W = [Q_V \vdots L_{1it}]$ is the appropriate set of instruments here, as well as in equation (5.3). Then

$$\hat{\beta}^* = [X_{it}'\Omega^{-1/2}P_W\Omega^{-1/2}X_{it}]^{-1}X_{it}'\Omega^{-1/2}P_W\Omega^{-1/2}Y_{it}$$

$$= S[SX'\Omega^{-1/2}P_W\Omega^{-1/2}XS]^{-1}SX'\Omega^{-1/2}P_W\Omega^{-1/2}Y$$

$$= S\hat{\xi}^*.$$

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 $\text{cov}[\hat{\beta}^*-\beta] = \text{cov}[S(\hat{\xi}^*-\xi)] \neq S[\text{cov}(\hat{\xi}^*-\xi)]S'$ where $\text{cov}(\hat{\xi}^*-\xi)$ 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 <u>all</u> the explanatory variables are correlated with α_i are identical to the within-groups estimators, and these are unbiased in finite samples.

To see this, recall that the set of instruments in this case is just the columns of Q_V, and Q_V is orthogonal to α_i in small samples, not simply as a probability limit.

6. Estimating the Returns to Schooling

In this section, we apply our estimation and testing techniques to a returns to schooling example. This problem has received extensive attention since many analysts have felt that the unobserved individual component α_i may contain an ability component which is correlated with schooling. Since our sample 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 (1977) 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 of the schooling coefficient, 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 (1977) and Griliches, Hall, and Hausman (1978) find that treating schooling as endogenous with family background variables as instruments leads to a *rise* in the estimated schooling coefficient of about 50%.[7] Thus, we would like to investigate how our estimation

[7] Using a specification test of the type Wu (1973) and Hausman (1978) propose, we find a statistically significant difference between the IV and OLS estimates. Chamberlain (1978) also finds a significant increase in the schooling coefficient when he compares OLS estimates with estimates from his two factor model.

method affects the return to schooling coefficient, since we do not require excluded family background variables to serve as instruments, as did the previous estimates.

Our sample consists of 750 randomly chosen (non-SEO) prime age males, age 25-55, from the PSID sample. We consider two years, 1968 and 1972, to minimize problems of serial correlation apart from the permanent individual component.[8] 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.[9] The PSID data does not include IQ. The 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 (1978)) 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

---

[8] Lillard and Willis (1978) 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 test for an autoregressive process, but here we use a simpler case. Note that we are not investigating the dynamics of wages or earnings here.

[9] Experience was used as either experience with present employer or a measure of age - schooling - 5. Qualitatively, the results are similar, so we report results using the latter definition. As the results show, use of age also yields very similar results for the schooling coefficient. Unlike Griliches (1977), we are not attempting to separate out the influence of age from experience.

(1978)); 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, although our results should be interpreted with this exclusion in mind.

We now consider the estimation method proposed in Section 4.2 from the standpoint of computational convenience. Equation (4.5) and Proposition 4.3 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 (4.6), 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 TxN) may exceed the computational capacity of much econometric software. If this occurs, using equation (4.5), 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 TxN regression, an equivalent procedure is to form $\hat{X}_{2it} = X_{2it} - X_{2i.} + \hat{X}_{2i.}$. The last term, $X_{2i.}$, 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.[10] A similar technique works with the transformed variables in equation (4.6) which yields asymptotically efficient estimates of β and γ. The reason that calculating X_{2it} in this manner is equivalent to the more cumbersome approach of a TxN sample regression of X_{2it} on instruments as indicated in equation (4.4) is that Q_V is orthogonal to any time-invariant variable. Thus parsing out Q_V in the second and third equations of (4.4) is equivalent to premultiplying them by P_V, and $\hat{X}_{2i.}$ and $\hat{Z}_{2i}$ can be calculated from the sample size N regressions on $X_{1i.}$ and $Z_{1.}$. To get $\hat{X}_{2it}$, we must add $Q_V\hat{\pi}_{13}$ to $\hat{X}_{2i.}$, so that $\hat{X}_{2it}$ is given by $\tilde{X}_{2it} + \hat{X}_{2i.}$.

If computational capacity is not a difficulty, a standard instrumental variables package can be used, with $\tilde{X}_{1it}$, $X_{1i.}$, $\tilde{X}_{2it}$, and Z_{1i} as instruments. The variables which are time invariant have T identical entries for each individual i. So long as Proposition 4.1 is satisfied, the parameters are identified and the number of columns of $X_{1i.}$ is at least as great as the number of columns of Z_{2i} (i.e., $k_1 \geq g_2$). 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}.

---

[10] One note of caution, however. The estimates of the variance from the second stage are inconsistent, for the same reason as doing 2SLS in two steps yields inconsistent variance estimates in the second step. To estimate the variances consistently, one must use the estimated coefficients and the model (2.1) without the hatted variables on the right hand side.

We now turn to our log wage regressions to determine the effects on the schooling coefficient of our estimation procedure. Column 1 of Table 6.1 gives the OLS results while Column 2 gives the GLS estimates under the assumption of no correlation between the explanatory variables and α_i. The OLS and GLS estimates are reasonably close, 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 (ρ = .638) indicates the importance of the unobserved individual effect. The finding that an additional year of schooling leads to a 6.7% higher wage is very similar to other OLS results, both on PSID and other data sets.

In the third column of Table 6.1, 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 coefficient estimates change markedly from the first two columns. The effect of bad health falls by 26%, the effect of unemployment falls by 34%, while the effect of an additional year of experience rises by 59%. Comparing the within-groups and GLS estimates, using

Table 6.1 DEPENDENT VARIABLE: LOG WAGE

	OLS*	GLS	Within	IV/GLS
1. Exp	+.0132 (.0011)	.0133 (.0017)	.0241 (.0042)	.0175 (.0026)
2. Race	-.0853 (.0328)	-.0878 (.0518)	- -	-.0542 (.0588)
3. Bad Health	-.0843 (.0412)	-.0300 (.0363)	-.0388 (.0460)	-.0249 (.0399)
4. Unemp Last Yr	-.0015 (.0267)	-.0402 (.0207)	-.0560 (.0295)	-.0636 (.0245)
5. Union	+.0450 (.0191)	.0374 (.0296)	- -	.0733 (.0434)
6. Yrs School	+.0669 (.0033)	.0676 (.0052)	- -	.0927 (.0191)
Other Variables	Constant Time	Constant Time	-	Constant Time
NOBS	1500	1500	1500	1500
S.E.R.	.321	.192	.160	.193
RHO		.623		
Instruments				Dad's Educ Poor Mom's Educ

* Reported standard errors are inconsistent since they do not account for variance components.

results in Hausman (1978), 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 $m = 20.2$, we can reject the null hypothesis with any reasonable size test. This confirms Hausman's (1978) earlier finding that mis-specification was present in a similar log wage equation.

In the last column of Table 6.1, we present traditional instrumental variables estimates of the wage equation, treating schooling as endogenous. Family background variables are used as additional instruments: father's education, mother's education, and a binary variable for a poor household. The estimated schooling coefficient rises to .0915, which echoes previous results of Griliches (1977) and Griliches, Hall, and Hausman (1978) who find an increase of an almost identical amount. Under the null hypothesis that the instruments are uncorrelated with α_i, the estimated coefficients should be about the same. Note that the instrumental variables estimates are somewhat closer to the consistent 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 serving as instruments. Yet the result of the specification test is $m = 8.70$ which again indicates the presence of remaining correlation between the instruments and the latent individual effects. 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 6.2, we present the results of our estimation method. We assume that X_1 contains experience, bad health, and unemployment last year, all initially assumed to be uncorrelated with the individual effect. Z_1 is assumed to contain race and union status, while Z_2 contains schooling, which is assumed to be correlated with α_i. The estimated schooling coefficient rises to .125, which is 62% above the original OLS estimate and 32% above the traditional instrumental variables estimate. Also, note that 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 4.4, 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 test statistic is m = 2.24 which is distributed as χ^2_2 under the null hypothesis

Table 6.2 Dependent Variable: Log Wage

	HT/IV*	HT/GLS	HT/GLS	HT/GLS	HT/GLS
1. EXP	.0217 (.0027)	.0217 (.0031)	-	.0268 (.0037)	.0241 (.0045)
2. EXP^2	-	-	-	-.00012 (.00015)	-
3. AGE	-	-	.0147 (.0028)	-	-
4. RACE	-.0257 (.0531)	-.0278 (.0758)	-.0046 (.0824)	-.0014 (.0662)	-.0175 (.0764)
5. BAD HEALTH	-.0535 (.0468)	-.0294 (.0307)	-.0228 (.0318)	-.0243 (.0318)	-.0388 (.0348)
6. UNEMP LAST YEAR	-.0556 (.0311)	-.0559 (.0246)	-.0634 (.0265)	-.0634 (.0236)	-.0560 (.0279)
7. UNION	.1245 (.0560)	.1227 (.0473)	.1648 (.0721)	.1449 (.0598)	.2240 (.2863)
8. YRS SCHOOL	.1247 (.0380)	.1246 (.0434)	.1311 (.0490)	.1315 (.0319)	.2169 (.0979)
OTHER VARIABLES	Constant Time	Constant Time	Constant Time	Constant Time	Constant Time
NOBS	1500	1500	1500	1500	1500
S.E.R.	.352	.190	.196	.189	.629
RHO		.661	.678	.674	.817

* Reported standard errors are inconsistent since they do not account for variance components.

of no correlation between the explanatory variables and α_i. While m is somewhat higher than its expected value, 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 present some additional results to see how robust our estimates are to specification change. Column 3 of Table 6.2 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 cause correlation. The results are quite similar to our previous findings. The effect of schooling is .120, only slightly lower than the .125 found previously. Race again has little or no effect, while the effects of bad health and unemployment are similar to those in the specification with experience. In the next column of Table 6.2, we include both experience and experienced squared as explanatory variables.[11] Again, the results are quite similar to 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

---

[11]Neither of these alternative specifications of age or both experience and experience squared pass the specification test if estimated by OLS and compared with the appropriate within-groups estimates. In both specifications, the latent individual effects continue to be correlated with the explanatory variables.

alternative specifications.

Our last specification relaxes the correlation assumptions among the explanatory variables. We now remove experience and unemployment from the X_1 category to the X_2 category, permitting 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 4.4 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 estimate. Nevertheless, it is interesting to note that the point estimate of the schooling coefficient has risen to .217. Thus all our different estimation methods have led to an increase 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.

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 - both to estimate their own coefficients and to serve as instruments for correlated time-invariant variables - allows efficient estimation of both β and γ. The method is a two-fold improvement over the within-groups estimator: it is more efficient and also produces estimates of the coefficients of time-invariant variables. It also appears to be better than traditional instrumental variables methods which rely on excluded exogenous variables for instruments. Perhaps most important is the fact that in the overidentified case ($k_1 \geq g_2$), a specification test exists which allows a test of the appropriateness of the instruments. Since the within-groups estimates of β always exist, they provide a baseline 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.

REFERENCES

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

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

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

Griliches, Z., "Estimating the Returns to Schooling: Some Econometric Problems," Econometrica 45 (1977), pp. 1-22.

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

Hausman, J., "Specification Tests in Econometrics," Econometrica 46 (1978), pp. 1251-1272.

Hausman, J. and W. Taylor, "Specification Tests for the Legitimacy of Instruments", unpublished manuscript, 1980.

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

Lahiri, K. and P. Schmidt, "On the Estimation of Triangular Structural Systems", Econometrica 46 (1978), pp. 1217-1222.

Lillard, L. and R Willis, "Dynamic Aspects of Earning Mobility," Econometrica 46 (1978), pp. 985-1012.

Maddala, G., "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," Econometrica 39 (1971), pp. 341-358.

Mundlak, Y., "Empirical Production Function Free of Management Bias," Journal of Farm Economics, 43 (1961), pp. 44-56.

Mundlak, Y., "On the Pooling of Time Series and Cross Section Data", _Econometrica_ 46 (1978), pp. 69-86.

Narayanan, R., "Computation of Zellner-Theil's Three Stage Least Squares Estimates", _Econometrica_, 37 (1969), pp. 298-306.

Nerlove, M., "A Note on Error Components Models", _Econometrica_, 39 (1971), pp. 359-382.

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

Scheffé, H., _The Analysis of Variance_, John Wiley & Sons: New York, 1959.

Taylor, W., "Small Sample Considerations in Estimation from Panel Data," _Journal of Econometrics_, forthcoming.

Wu, D., "Alternative Tests of Independence Between Stochastic Regressors and Disturbances," _Econometrica_ 41 (1973), pp. 733-750.

6654 044

Date Due

PERIODICALS ARE LENT FOR THREE DAYS WITH RENEWAL ONLY

FEB - 1 '85	FEB 20 1984	
MAY 24 '87	MAR 26 1987	
JUL 22 '87		
JE 17 '89		
AG 10 '8		
DEC 10 1990		

Lib-26-67

MIT LIBRARIES DUPL
3 9080 004 446 776
MIT LIBRARIES DUPL
3 9080 004 446 784
MIT LIBRARIES
3 9080 004 446 792
MIT LIBRARIES
3 9080 004 446 800
MIT LIBRARIES
3 9080 004 446 818
MIT LIBRARIES
3 9080 004 446 826
MIT LIBRARIES
3 9080 004 446 834
MIT LIBRARIES
3 9080 004 446 842