NBER TECHNICAL PAPER SERIES

ERRORS IN VARIABLES IN PANEL DATA

Zvi Griliches

Jerry A. Hausman

Technical Working Paper No. 37

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138
May 1984

We would like to thank the NSF and NBER for research support. Adam Jaffe and Bruce Meyer provided reserarch assistance. The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Paper #37
May 1984

Errors in Variables in Panel Data

ABSTRACT

Panel data based on various longitudinal surveys have become ubiquitous in economics in recent years. Estimation using the analysis of covariance approach allows for control of various "individual effects" by estimation of the relevant relationships from the "within" dimension of the data. Quite often, however, the "within" results are unsatisfactory, "too low" and insignificant. Errors of measurement in the independent variables whose relative importance gets magnified in the within dimension are often blamed for this outcome.

However, the standard errors-in-variables model has not been applied widely, partly because in the usual micro data context it requires extraneous information to identify the parameters of interest. In the panel data context a variety of errors-in-variables models may be identifiable and estimable without the use of external instruments. We develop this idea and illustrate its application in a relatively simple but not uninteresting case: the estimation of "labor demand" relationships, also known as the "short run increasing returns to scale" puzzle.

Zvi Griliches
National Bureau of
Economic Research
1050 Massachusetts Ave.
Cambridge, MA 02138

Jerry A. Hausman
Department of Economics
M.I.T.
E52-271A
Cambridge, MA 02139

April 1984
First Draft
Comments Welcome

Errors in Variables in Panel Data

Zvi Griliches and Jerry A. Hausman*

Panel data based on various longitudinal surveys have become ubiquitous in economics in recent years. Their popularity stems in part from their ability to allow and control for various "individual effects" and other relatively slowly changing left-out-variables. Using the analysis of covariance approach, one can estimate the relevant relationships from the "within" dimension of the data. Quite often, however, the "within" results are unsatisfactory, "too low" and insignificant. The tendency is then to blame this unhappy outcome, among other things, on errors of measurement in the independent variables whose relative importance gets magnified in the within dimension.

That errors of measurement are important in micro data is well known but has had little influence on econometric practice.[1] The standard errors-in-variables model has not been applied widely, partly because in the usual context it requires extraneous information to identify the parameters of interest. It is rather obvious but does not appear to be widely known that in the panel data context a variety of errors-in-variables models may be identifiable and estimable without the use of external instruments. We exposit and develop this idea and illustrate its application in a relatively simple but not uninteresting case: the estimation of "labor demand" relationships, also known as the "short run increasing returns to scale" puzzle; see Solow (1967), and Medoff and Fay (1983).

---

* Harvard and MIT and the NBER. We would like to thank the NSF and NBER for research support. Adam Jaffe and Bruce Meyer provided research assistance.

1. Matters are somewhat better in sociology: see Griliches (1984) for general discussion, and Hauser, Bielby, and Featherman, (1977) for an applied example.

In the next four sections we first outline our approach in a very simple context; next we present and discuss the algebra for the more general case and discuss different estimation strategies; then we turn to a description and discussion of our empirical example, and conclude with recommendations for a particular empirical strategy which should be followed when analyzing such data.

I. The Problem of Errors in Variables

It is clear that once one has a time series and one is willing to assume that errors of measurement are serially uncorrelated then one can use lagged values of the relevant variables as instruments. The problem in panel data is that because one is likely to assume the presence of correlated individual effects, lagged values are not valid instruments without further analysis. But, because the errors of measurement are assumed to have a particular time series structure (usually uncorrelated over time), different transformations of the data will induce different and deduceable changes in the biases induced by such errors of measurement in the estimated parameters, which can be used to identify the importance of such errors and recover the "true" parameters.

The following simple model will serve to illustrate our main ideas. Let the true equation be of the form

$$y_{it} = \alpha_i + \beta z_{it} + \eta_{it} \tag{1.1}$$

where the α_i 's are unobserved individual effects which may be correlated with the true independent variable of interest, the z_{it}. The η_{it} are the standard "best case" disturbances: i.i.d., with mean zero and variance σ_η^2.[1] The z_{it} are not observed directly, however. Only their erroneous reflection, the x_{it}

$$x_{it} = z_{it} + v_{it}, \tag{1.2}$$

are observed, where v_{it} is an i.i.d. measurement error with variance σ_v^2.

If OLS is applied to the observed variables, the equation to be estimated is

$$y_{it} = \overline{\alpha} + \beta x_{it} - \beta v_{it} + \eta_{it} + (\alpha_i - \overline{\alpha}) \tag{1.3}$$

and the resulting parameters will be biased for two distinct reasons: (1) because of the correlation of the x_{it} with the left-out individual effects,

---

1. These assumptions correspond to the usual random effects specification. We relax this assumption in our estimation and test procedures.

(usually upward) and (2) downward because of the negative correlation between the observed x_{it} and the new composite disturbance term.

It is clear that in panel data one can eliminate the first source of bias by going "within" by analyzing deviations around individual means. It is also reasonably well known that going within might exacerbate the second source of bias and make things worse rather than better.[1] What is less obvious is that there are different ways of eliminating the first source of bias, that they imply different consequences for the magnitudes for the second type of bias, and hence provide an opportunity for identifying its magnitude and recovering the "true" coefficients.

An alternative to the "within" estimator is a first difference estimator, which also sweeps out the individual effects. Let us contrast then the consequence of errors of measurement for these two different ways of dealing with the "unobserved individual effects" problem.

First difference model

$$y_{it}-y_{t-1}=\beta(x_{it}-x_{it-1})-\beta(v_{it}-v_{it-1})+(\eta_{it}-\eta_{it-1})$$

$$dy_{it} = \beta dx_{it}- \beta dv_{it}+ d\eta_{it}$$

where the

$$dy_{it}= y_{it}-y_{it-1}$$

and similarily for the other variables.

Within model

$$(y_{it}-\overline{y}_i)=\beta(x_{it}-\overline{x}_{i.})-\beta(v_{it}-\overline{v}_{i.})+(\eta_{it}-\overline{\eta}_{i.})$$

$$\tilde{y}_{it} = \beta\tilde{x}_{it} - \beta\tilde{v}_{it} + \tilde{\eta}_{it}$$

$$\tilde{y}_{it}= y_{it} - \overline{y}_{i.}$$

Given our assumptions

$$(1.4) \quad \text{plim } b_{dy\bullet dx} = \beta\left(1- \frac{2\sigma_v^2}{Var(dx)}\right), \quad \text{plim } b_{\tilde{y}\tilde{x}} = \beta\left(1- \frac{T-1}{T}\,\frac{\sigma_v^2}{Var\ \tilde{x}}\right)$$

---

1. See Hausman (1978) and Griliches (1979) for a related discussion in the context of the analysis of sibling data.

It will be shown in the next section, that in the most likely cases in economics: positively serially correlated "true" x's (z's) with a declining correlogram and for $T > 2$,

$$(1.5) \qquad \mathrm{Var}(dx) < \frac{2(T)}{T-1} \mathrm{var}\ \tilde{x}$$

and hence $|\text{bias } b_d| > |\text{bias } b_w|$. That is, errors of measurement will usually bias the first difference estimators downward (towards zero) by more than they will bias the within estimators.

Note, however, that if we have estimated both b_d and b_w we have already computed Var dx and Var $\tilde{x}$ and hence have all the ingredients to solve out for the unknown σ_v^2 and β. In fact, consistent estimates can be had from

$$(1.6) \qquad \hat{\beta} = [2\ b_w / \mathrm{Var}(dx) - (T-1)b_d / T\mathrm{Var}\ \tilde{x}]/[2/\mathrm{Var}\ dx - (T-1)/T\mathrm{Var}\ \tilde{x}]$$

$$(1.7) \qquad \sigma_v^2 = \frac{\hat{\beta}-b_d}{\hat{\beta}} \bullet \frac{\mathrm{Var}\ dx}{2}$$

Several points are worth noting here:

1. These results were derived assuming that the measurement errors were not serially correlated, while the true z's are. It is possible to allow for serial correlation in the measurement errors, the v's, provided that we know its magnitude. For example, if x is a gross capital stock measure based on a 20 year life assumption and the i.i.d. measurement errors occur in the measurement of investment, then the first order serial correlation of the errors in the capital stock is 19/20 = .95. The first difference bias formula now becomes

$$(1.8) \qquad \mathrm{plim}(b_d-\beta) = \frac{-\beta 2\sigma_v^2(1-\rho_\varepsilon)}{\mathrm{Var}\ dx} = \frac{-0.1\beta\cdot\sigma_v^2}{\mathrm{Var}\ dx}$$

and a similar expression can be worked out for the bias of the within estimator. Other forms of serial dependence in the errors of measurement can be both tested for and consistent estimators can be derived in their presence.

2. We have derived these results assuming only one independent variable. If there are more independent variables in the equation, but they are not subject to error, they can be swept out from all the other variables and the formulae reinterpreted in terms of the variances of residuals from regressions on these other variables. If some of them are also subject to measurement error, the formulae become more complex but can be similarly derived provided that these measurement errors are mutually uncorrelated (or correlated with a known correlation).

3. The first difference and the within estimators are not the only ones that will give us an implicit estimate of the bias. We could define, for example, the "longest difference" $\ell dy = y_T - y_1$, and similarly $\ell dx = x_T - x_1$, which lead to

(1.9) $$\text{plim } b_{\ell d} = \beta - \beta 2\sigma_v^2 / \text{Var}(\ell dx)$$

and allows for another estimate of β and σ_v^2. In fact, there are T/2 such independent estimates which can be combined optimally to improve upon the efficiency of the estimators outlined above. For a 6 period cross-section, we can compute estimates of from $y_6 - y_1$, $y_5 - y_2$, and $y_4 - y_3$ and derive β and σ_v^2 from another round of estimation on using the relationships

(1.10) (a) $$b_{61} = \beta - 2\beta\sigma_v^2 / \text{Var } dx_{61}$$

(b) $$b_{52} = \beta - 2\beta\sigma_v^2 / \text{Var } dx_{52}$$

(c) $$b_{43} = \beta - 2\beta\sigma_v^2 / \text{Var } dx_{43}$$

with β given by the "constant" of such a "regression" and σ_v^2 derivable from its slope. More general optimal restricted estimators are developed below.

4. While we have expressed these formulae in terms of the observable Var dx, Var $\tilde{x}$, and so forth, in many contexts it may be more interesting to reinterpret them in terms of the parameters for the "true" unobserved z variable, its variance and serial correlation structure. This will be done in the next section, where we shall present a more rigorous derivation of some of these results.

II. Derivation of Results

We now reconsider our basic model of equation (1.1) and derive the relationship among the various estimators. The specification of the equation is

$$(2.1)\qquad y_{it} = \alpha_i + \beta x_{it} + \eta_{it} - \beta v_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it} \quad i=1,\ldots,N, t=1,T$$

where v_{it} arises from the errors in variables. For ease in derivation in the results, we assume that all the random variables are jointly covariance stationary. For the present we also assume that both η_{it} and v_{it} are not serially correlated, $E\eta_{it}\eta_{i\tau} = Ev_{it}v_{i\tau} = 0$ for $t \neq \tau$. We first calculate the probability limit of the first difference estimator.

$$(2.2)\qquad \operatorname{plim} b_d - \beta = \operatorname{plim}\left(\frac{1}{NT} dx'dx\right)^{-1} \operatorname{plim}\left(\frac{1}{NT} dx'd\varepsilon\right)$$

$$= \left(2\sigma_z^2(1-\rho_1) + 2\sigma_v^2\right)^{-1} \left(-2\sigma_v^2\beta\right) = -\left(\sigma_z^2(1-\rho_1)+\sigma_v^2\right)^{-1}\sigma_v^2\beta$$

where ρ_1 is the correlation between the true regression variables z since $x = z+v$. Note that as expected the inconsistency increases as the correlation increases so long as it is positive. First differencing "removes more of the signal" for given σ_z^2 and σ_v^2 the higher is ρ_1, which exacerbates the errors in variables problem.

For the within estimator we calculate the probability limit to be

$$(2.3)\qquad \operatorname{plim} b_w - \beta = \operatorname{plim}\left(\frac{1}{NT}\tilde{x}'\tilde{x}\right)^{-1} \operatorname{plim}\left(\frac{1}{NT}\tilde{x}'\tilde{\varepsilon}\right)$$

$$\left[\left(\frac{T-1}{T}\right)\sigma_x^2 - \frac{2\sigma_z^2}{T^2}\sum_{j=1}^{T}(T-j)\rho_j\right]^{-1}\left[-\left(\frac{T-1}{T}\right)\sigma_v^2\beta\right]$$

$$= -\left[\sigma_x^2 - \frac{2\sigma_z^2}{T(T-1)}\sum(T-j)\rho_j\right]^{-1}\sigma_v^2\beta$$

where all plims in the paper are taken as $N\to\infty$.

The formula is the same as the usual OLS case except for the term involving the serial correlation coefficients ρ_j for the z's which arises from the within transformation $\tilde{x}_{it} = x_{it} - \frac{1}{T} \sum x_{ij}$. To compare the inconsistencies between the first difference and within estimators note that plim b_d does not depend on T (as $N \to \infty$), but that plim b_w does depend on T because of the within transformation. We will develop the conditions which will cause the within estimator to be less inconsistent which we expect to be the usual case. The comparison is given in Table 2.1.

Table 2.1: Comparison of First Difference and Within Estimator

T	plim $b_d - \beta$	plim $b_w - \beta$	Conditions for $\lvert b_w \rvert - \lvert b_d \rvert > 0, \rho_i > 0$
2	$-(\sigma_x^2 - \sigma_z^2\rho_1)^{-1}\sigma_v^2\beta$	$-(\sigma_x^2 - \sigma_z^2\rho_1)^{-1}\sigma_v^2\beta$	same
3	"	$-(\sigma_x^2 - \frac{\sigma_z^2}{3}(2\rho_1+\rho_2))^{-1}\sigma_v^2\beta$	$\frac{2}{3}\rho_1 + \frac{1}{3}\rho_2 < \rho_1$
4	"	$-(\sigma_x^2 - \frac{\sigma_z^2}{6}(3\rho_1+2\rho_2+\rho_3)^{-1}\sigma_v^2\beta$	$\frac{1}{2}\rho_1 + \frac{1}{3}\rho_2 + \frac{1}{6}\rho_3 < \rho_1$
$T \to \infty$	"	$-(\sigma_x^2 - \frac{2\sigma_z^2}{T-1}\Sigma \frac{T-j}{T}\rho_j)^{-1}\sigma_v^2\beta$	$\frac{2}{T}(\rho_1+\rho_2+\dots) < \rho_1$

For T=2 the estimators give numerically identical results since the within transformation and first differences are related by the formula $\frac{1}{2}dx = \tilde{x}$. For T=3 the condition is $\rho_1 > \rho_2$ which is assured with a declining correlogram. For T=4 the required condition is $\rho_1 > \frac{2}{3}\rho_2 + \frac{1}{3}\rho_3$ which again follows from a declining correlogram. The general result follows by induction. This condition then is a sufficient condition for the within estimator to be less inconsistent. The steepness in the decline of the correlogram will determine the differences in magnitude, but in many cases we would expect a substantial difference.

The situation reverses if we difference the data more than one period apart. Define $d^j x = x_t - x_{t-j}$. Then the probability limit of the least squares estimator on these data is

$$\text{(2.4)} \quad \operatorname{plim} b_j - \beta = -\left(\sigma_z^2 (1-\rho_j) + \sigma_v^2\right)^{-1} \sigma_v^2 \beta.$$

For example, take T=3 and j=2. The inconsistency of b_2 is smaller than b_w so long as $\rho_1 > \rho_2$ for positive ρ_i. For T=4 and j=3 the condition is $5\rho_3 < 3\rho_1 + 2\rho_2$, which holds under the assumption of a declining correlogram. For T=4 and j=2 so that the 'longest' difference is not used the condition is $4\rho_2 < 3\rho_1 + 1\rho_3$ so that a declining correlogram is not sufficient to assure the inconsistency in b_2 is less than b_w. The general result is that for a given sample T, the estimator with j=T-1 will be less inconsistent than b_w but for intermediate $1<j<T-1$ no definite ordering can be made. Note that our comparison only involves the inconsistency in the estimators for the case $N \to \infty$. For moderate size N the mean square error may be a better comparison criterion, and the estimator with j=T-1 eliminates a non-negligible proportion (T-2) / (T-1) of the observations.

We now turn to the general case with serial correlation in the v_{it}'s and the η_{it}'s so that $V(\eta) = \Sigma \otimes I_N$ and $V(v) = \Omega \otimes I_N$ where both Σ and Ω are T x T matrices with all diagonal elements assumed equal. For the first difference estimator it is convenient to define a bidiagonal matrix $\tilde{A}$ with -1 on the diagonal and +1 on the superdiagonal. Then we define $A = \tilde{A} \otimes I_N$. We transform the model of equation (2.1) with A to find

$$(2.4) \qquad Ay = AX\beta + A\alpha + A\varepsilon$$

and estimate by OLS, $b_d = (X'A'AX)^{-1}X'A'Ay$. Now $Ev'A'Av = \text{tr}\,[(A'A)\Omega \otimes I]$ $= 2N(T-1)\sigma_v^2(1-r_1)$ by stationarity where r_1 is the correlation coefficient from Ω. Therefore we calculate the inconsistency of the first difference estimator

$$(2.5) \qquad \text{plim } b_d-\beta = \text{plim } (\tfrac{1}{NT} X'A'AX)^{-1} \text{ plim } (\tfrac{1}{NT} -v'A'Av\beta)$$

$$= [2(T-1)(\sigma_z^2(1-\rho_1) + \sigma_v^2(1-r_1))]^{-1}[-2(T-1)\sigma_v^2(1-r_1)\beta]$$

$$=-[\sigma_z^2 (1-\rho_1) + \sigma_v^2 (1-r_1)]^{-1} \sigma_v^2 (1-r_1) \beta.$$

The bias of the first difference estimation here as compared to equation (2.2) is less if $r_1>0$. The more highly positively correlated the measurement error is, the more you eliminate using first differences. However, the presumption that $r_1>0$ seems less strong in economic data than the assumption we used before that $\rho_1 > 0$.

To calculate the inconsistency of the within estimator define the projection matrix $\tilde{J}$ as $\frac{1}{T}$ times a matrix of all ones. The complement is $\tilde{Q} = I_T-\tilde{J}$ and transform using $Q = \tilde{Q} \otimes I_N$, so that

$$(2.6) \qquad Qy = QX\beta + Q\alpha + Q\varepsilon$$

The within estimator follows from applying OLS to equation (2.6). Note that $E\,\frac{1}{T}\, v'QQv = \frac{1}{T}\,\text{tr}(Q\Omega \otimes I) = N[\frac{T-1}{T}\,\sigma_v^2 - \frac{2\sigma_v^2}{T^2} \sum_{j=1}^{T} (T-j)\, r_j]$. To calculate the inconsistency we take plims

$$(2.7) \qquad \text{plim } b_w-\beta = \text{plim } [\tfrac{1}{NT} X'Q'QX]^{-1} \text{plim } \tfrac{1}{NT} [-X'Q'Qv\beta]$$

$$= -[\sigma_x^2 - \frac{2}{T(T-1)} \sum [\sigma_z^2 (T-j)\rho_j + \sigma_v^2 (T-j)\, r_j]^{-1} \, [\sigma_v^2 - \frac{2\sigma_v^2}{T(T-1)} \sum_j (T-j) r_j] \beta$$

As with first differences, the inconsistency in the first difference estimator decreases with respect to the uncorrelated case so long as all $r_j > 0$. Again this assumption is not as compelling as the analogous assumption about the ρ_j's. To compare the bias of the first difference and within estimators, first note that for T=2 they are identical as before. For T>3 it may be reasonable to assume that $\rho_j > r_j > o$ for all j. That is, serial correlation is higher in the true variable than in the measurement error. Then for the case T=3 the within estimator is less biased then the first difference estimator if $\frac{\rho_1 - \rho_2}{r_1 - r_2} > \frac{1-\rho_1}{1-r_1}$ [1] which holds if the serial correlation in the true variable decreases less slowly than the serial correlation in the measurement error. This type of condition generalizes to values of T larger than 3. While the condition seems plausible, that $\rho_j > r_j$ and that the decrease in the serial correlation of the ρ_j's be less than for the r_j's, it is not overwhelming. Counterexamples are easy to construct. The particular case under consideration would need to be examined.

The 'long' difference estimator is the same as equation (2.5) with ρ_1 and r_1 replaced by ρ_j and r_j, respectively. Note that the most favorable case need no longer be j=T-1 because r_j decreases along with ρ_j. The j which minimizes the inconsistency maximizes the ratio $(1-\rho_j) / (1-r_j)$. For a positive and declining correlogram for both ρ_j and r_j the tradeoff is between removing too much signal and removing some of the noise. If both z and v follows AR1 processes with $\rho_1 > r_1$ then j=T-1 will minimize the inconsistency. On the other hand, if z follows

---

1. The necessary and sufficient condition is $(1-\rho_1)/(1-r_1) < (1 - \frac{2}{3}\rho_1 - \frac{1}{3}\rho_2)/(1 - \frac{2}{3} r_1 - \frac{1}{3} r_2)$.

an AR1 process and v follows an MA1 process, then j=1 can be optimal. The optimal choice depends on both the type of process as well as the particular coefficient values.

A question which now arises is within a family of transformations $Ry = RX\beta + R\alpha + R\varepsilon$ to eliminate α, which transformation will have good properties with respect to errors in measurement? If we write $R = \tilde{R} \otimes I$ note that each row of $\tilde{R}$ must sum to zero to eliminate α from equation (2.1). Even if we choose as the criterion function the minimization of the inconsistency, the optimal $\tilde{R}$ depends on the properties of Σ and Ω.[1] In the uncorrelated case, diagonal Ω, with a declining correlogram for z, the long difference estimator, j=T-1, minimizes minus the inconsistency. For the correlated case of nondiagonal Ω, the optimal estimator depends on both Σ and Ω. A potentially interesting topic for future research would be to determine the optimal $\tilde{R}$ for interesting stochastic processes which determine Σ and Ω.

We now turn to the question of consistent estimation. In the general correlated case, the problem remains unidentified. That is, external instruments uncorrelated with the measurement error would be necessary for consistent estimation. We therefore concentrate on the uncorrelated case. The procedures we develop can also be applied in the 'partial' correlation case, e.g. if v follows an MAm process with m<T. The strategy we propose here is to take advantage of the existence of alternative consistent estimators, i.e. overidentification, to test the assumption of no correlation in the v's. If the

---

1. The inconsistency of the estimator may well constitute the major part of a criterion such as asymptotic mean square error given the quite large samples, in N, which are often present with panel data.

alternative estimates of β are mutually coherent, than the researcher can have some confidence that his assumptions hold true. The basic formula (2.2) leads to our estimation strategy which is a consideration of the differenced estimator. The appropriate weighted average of two differenced estimators leads to a consistent estimate of β since

$$(2.8) \quad \text{plim } \left[\tfrac{1}{NT}(d^jXd^jX' - dXdX')\right]^{-1} \quad \left[\tfrac{1}{NT}\,(d^jX\, d^jy - dXdy\right] =$$

$$\left(\sigma_j^2 - \sigma_1^2\right)^{-1}\left(\sigma_j^2\,\beta - 2\sigma_v^2\,\beta - \sigma_1^2\,\beta + 2\sigma_v^2\beta\right) = \beta$$

Therefore, a consistent estimator follows from dividing the differences of the variances into the differences of the covariances. Of course, for $T>3$ more than one such estimator exists.

A convenient estimation and testing scheme arises since $m = T-2$ such estimators exist given our assumption of stationarity. Consider the stacked set of equations

$$(2.9) \qquad \begin{aligned} \beta_1 &= \left(\sigma_2^2 - \sigma_1^2\right)^{-1}\left(\omega_2 - \omega_1\right) \\ \beta_2 &= \left(\sigma_3^2 - \sigma_1^2\right)^{-1}\left(\omega_3 - \omega_1\right) \\ &\vdots \\ \beta_m &= \left(\sigma_j^2 - \sigma_1^2\right)^{-1}\left(\omega_j - \omega_1\right) \end{aligned}$$

where σ_k^2 are the variances of the k differenced X's, ω_k is the covariance of the k differenced X's and y's , and the subscript on the β_l denotes the separation of the differencing when j is the longest possible differences. The variances of the β_j's can be calculated straightforwardly, if somewhat tediously. However, a more clever approach is to recognize that estimation of say β_1 by equation (2.9) is equivalent asymptotically to estimation of the first difference equation using adjacent X's as instruments. For example, if T=3 we would use $y_2-y_1 = (X_2-X_1)\,\beta + (v_2-v_1)\beta + \eta_2-\eta_1$ with X_3 as an instrument and similarly $y_3-y_2 = (X_3- X_2)\beta+(v_3-v_2)\beta + \eta_2-\eta_1$ with X_1 as an instrument. Under the assumption that the errors of measurement are uncorrelated the adjacent X's provide valid instruments.[1] Perhaps more importantly, this approach does not depend on the stationarity assumption so that it has general applicability to the panel data case.

Therefore, the recommended approach is to stack the differenced equations which correspond to the β_j's in equation (2.9)

$$\text{(2.10)}\qquad \begin{aligned} dy &= dX\beta - dv\beta + d\eta = dX\beta + d\varepsilon \\ d^2y &= d^2X\beta - d^2v\beta + d^2\eta = d^2X\beta + d^2\varepsilon \\ &\vdots \\ d^my &= d^mX\beta - d^mv\beta + d^m\eta = d^mX\beta + d^m\varepsilon \end{aligned}$$

Note that each difference equation has N observations. The X's provide the instrumental variables for each equation where all X's not involved in the difference are used as instruments. It is important to note that future X's as well as past X's provide instrumental variables. Indeed when the underlying Z variable process is close to a random walk, the future realizations of X provide much of the identifying information especially as T becomes large and the effect of the initial condition diminishes. However, if the future Z's are not exogenous with respect to current and past η's than a one-sided estimator can be used with only past X's serving as instrumental variables.

We estimate β from equation (2.10) using instrumental variable methods. The asymptotically efficient estimator is a 'system' estimator where the estimated β's are constrained to be equal and the covariance of the stochastic disturbances $\sum$ is taken account of when the variance of the estimated coefficient is

---

1. It is remarkable that under stationarity the original specification $y = X\beta + \alpha + \eta$ can be estimated using differenced X's as instruments since e.g. $E[(X_{ij} - X_{i1})\,\alpha_i] = 0$. This approach would allow an application of the Hausman-Taylor (1981) procedure where other right hand side variables are included, say $W\gamma$ and $(W_{it} - \overline{W}_{i\cdot})$ are used as additional instruments.

calculated. It is important to note that a 3SLS or GLS type estimator is inconsistent because instruments from a given equation are not orthogonal to the disturbances in another equation unless they are contained in the instrument set of that equation.

Therefore, when system 2SLS is done, different instrumental variables are used in each equation. Of course, any single equation of the set of equations from (2.10) can be estimated alone. Conditional heteroscedasticity can also be allowed for at some increase in computational complexity with the estimator that Hansen (1982) and White (1982) proposed. The estimator is

$$(2.12) \qquad \beta^* = [\tilde{X}'\tilde{Q}W^{-1}\tilde{Q}'\tilde{X}]^{-1}\tilde{X}'\tilde{Q}W^{-1}\tilde{Q}'\tilde{y} \text{ for } W = \frac{1}{NT}\sum_{j=1}^{NT} \tilde{Q}_j{}'\tilde{\varepsilon}_j\tilde{\varepsilon}_j\tilde{Q}_j$$

where $\tilde{\varepsilon}$ are the stacked $d^j\varepsilon$'s and $\tilde{\varepsilon}$ is calculated from an initial consistent estimate of β and $\tilde{X}$ are the stacked d^jX's, $\tilde{y}$ is the stacked d^jy's, and $\tilde{Q}$ is the stacked matrix of instruments.[1] The asymptotic covariance matrices of the estimator is $V(\beta^*) = [QW^{-1}\tilde{Q}'\tilde{X}]^{-1}$.

We now turn to the question of whether the no correlation assumption in the errors in measurement is valid. This assumption or an assumption about the form of the process generating the measurement is needed because the general correlation case is unidentified without the use of other variables as instruments. Note that if we applied least squares (OLS) to equation (2.10), equation by equation, we expect the estimates of β to differ according to our previous formulae. Similarly, it can be demonstrated that in the IV case with correlated errors in measurement that different the estimates of β will have different probability limits. Therefore, a testing procedure is to estimate the system of equations (2.10) in unrestricted form so that each equation is allowed to have its own β.

---

1. Note that $\tilde{Q}$ is a block diagonal matrix.

A large sample A large sample χ^2 test with m-1 degress of freedom for equality of the β_j's is equivalent to the implicit test in equation (2.9) that the β_j's are equal. An alternative specification test is to take an equation, say the first, and restrict the set of instruments. For $X_{it}-X_{it-1}$ instead of using all other X's as instrumental variables we could restrict the instrumental variables to be those X's which are at least 2 time periods away. The test statistic proposed by Hausman (1978) or Hausman and Taylor (1981) provides a large sample χ^2 test with one degree of freedom. Lastly, tests of the overidentification type of Sargan (1958) and Hansen (1982) can be used. Under stationarity assumptions these various tests are closely related. In the general case of nonstationarity of the X's they will differ although Newey (1983) provides a partial guide to their comparability.

The general approach that we suggest then goes as follows:

(i) Estimate equation (2.1) by GLS (variance components) and by the within estimator. Do a test for equality of the estimates using a Hausman (1978) or Hausman-Taylor (1981) type test.

(ii) If you reject the hypothesis in (i) then calculate some differenced estimates by OLS. If they differ significantly, errors in measurement may well be present. A joint test of all the differenced estimates can be made by using GLS on the system of equations in (2.10).[1] (iii) Estimate the equations in

(2.10) by IV. Then do a specification test(s) of the no correlation assumption in the errors in measurement. If the different estimates of β do not differ significantly you are done. If they do differ significantly, the specification of a correlated errors in measurement process, use of outside instruments, or respecification of the original model (2.1) seems to be called for.

---

1. The within estimator can be neglected since it is a linear combination of the differenced equation estimators of (2.10).

III. An Empirical Example

The empirical example we consider is related to the old conundrum of "short run increasing returns to scale." Let ℓ = logarithm of employment and q = logarithm of output. The relationship between ℓ and q depends on what is assumed about the production function, what is held constant, and what expectational assumptions are made about the relevant prices. If the production function is assumed to be Cobb-Douglas with a labor elasticity α, then one can derive two alternative relationships: the first based on inverting the production function and the second on solving the value of marginal productivity equals the wage condition:

$$\ell = \frac{1}{\alpha} q - \frac{1-\alpha}{\alpha} k \tag{3.1a}$$

$$\ell = \log \alpha + q - w' \tag{3.1b}$$

where k is the logarithm of capital services and w' is logarithm of the real wage $\log w - \log P$, where P is the price of the product. In either form, the coefficient of q should be one or higher. In econometric practice one tends to get coefficients which are less than one, implying short-run increasing returns to labor alone (Brechling 1973, Sims 1975). Adding lags helps a little, but usually not enough. A reasonable interpretation of the data and one rationale for the introduction of lags is that labor is hired in anticipation of "normal" or expected output, while actual output is subject to unanticipated "transitory" fluctuations. Since this argument is isomorphic with the errors-in-variables model (see Friedman, 1957, Maddala 1977), we can apply our framework to it.

We shall use data on 1242 U.S. manufacturing firms for the 6 years, 1972-1977 from the NBER R&D panel (Cummins, Hall and Laderman 1983), and adopt the second interpretation of the equation to be estimated. In this model

$$(3.2)\quad \ell_{it} = d_t + q^*_{it} + \{- w'_{it} + (\log \alpha_i - \log \overline{\alpha}) + \eta_{it}\}$$

q^*_{it} is the expected or "permanent" output level, d_t is a set of individual year constants (time dummies) and the bracketed term represents a composite "disturbance" which consists of three terms: (1) a real wage term, which presumably differs in some consistent fashion across firms and moves, more or less in unison for all the firms, over time. For instance,

$$(3.3)\quad w'_{it} = \mu_i + \gamma_t + \tau_{it}$$

may have a variance component structure with γ_t subsumed in the d_t and τ_{it} assumed to be uncorrelated with q_{it}. A term associated with the fact that the labor elasticities α_i's might differ across firms; and (3) a pure i.i.d. disturbance term η_{it}. We do not observe the expected output variable q^*_{it} but only the actual output

$$(3.4)\quad q_{it} = q^*_{it} + v_{it}$$

where v_{it} is an i.i.d. "error" or transitory component in q_{it}. Note that v_{it} need not be an actual "measurement" error. Observed q_{it} may be measured correctly but relatively to the conceptual variable desired in the model, q_{it} is erroneous.[1] We can rewrite the model in terms of observables as

$$(3.5)\quad \ell_{it} = a_i + \beta q_{it} + d_t + (-\beta v_{it} - \tau_{it} + \eta_{it})$$

where we expect $\beta = 1$ and the a_i's are a set of individual firm effects incorporating both permanent real wage differences and differences in the labor elasticity across firms and hence likely to be correlated with the q_{it}.

---

1. They need not be "errors" as far as other variables are concerned. For example, both hours worked per man and materials used are likely to be related to such transitory output fluctuations.

To recapitulate the model, we assume that workers are hired in anticipation of actual demand, that actual demand is met primarily by unanticipated fluctuations in hours of work per man (which are unobservable in our data) and inventory fluctuations, and that we can subsume the real wage variable into the time dummies and the individual firm effects.[1] Our focus then is on the estimation of β and σ_v^2 , the variance of the "error" (v) in q, its unanticipated component.

Table 3.1 presents the estimated β's for different cuts of the data, total, within, first differences, and "long"differences, and the associated net variance of q (net of year and industry dummy variables). It also shows a set of parallel instrumental variable estimates of β, where data on capital are used as an external instrument for q. The validity of such external instruments depends on the lack of a firm specific short-run movement in real wages, on the non-correlation of the capital measures with τ_{it}. Note that the OLS results behave as predicted, with the first difference estimator being lower than the within one. The long difference estimates is greater than both the first difference estimate and the within estimate as the derivations in Section 2 predicted.

There are two ways of interpreting these results. The first would maintain the assumption that β = 1, ignore the potential presence of correlated individual effects, and accept the instrumental variable results as vindicating this position. There are difficulties with this view, however. The implied "error"

---

1. An alternative interpretation would divide ℓ into two components, "fixed" labor which changes only in response to permanent changes in q, and "variable" labor, which is related to v.

Table 3.1: Estimates of the Employment-Output Relationship for 1242 U. S. Manufacturing Firms, 1972-77

$$\ell_{it} = \alpha_i + \beta q_{it} + d_t$$

Estimation Method and Degrees of Freedom	Net Variance of q	β	MSE	Instrumental Variables Estimates β	Instrumental Variables Estimates MSE
1. Total d.f. = 7547	2.256	.966 (.003)	.168	.995 (.003)	.160
2. Within d.f. = 6303	.0327	.640 (.008)	.011	n.c.	
3. First differences d.f. = 6304	.0251	.481 (.010)	.015	.851 (.160)	.019
4. "Long differences 1972-77" d.f. = 1260	.1438	.730 (.016)	.048	1.050 (.037)	.063

The bracketed terms are the estimated standard errors. Total regressions contain also 5 year and 22 industry dummy variables. The within and first difference regressions include also year dummy. The instrumental variables used are the logarithm of net plant, the first difference in log net plant and the long difference in log net plant respectively.

The variables variance of σ_v^2 is .038, which is larger than the variance of the first differences of q which should contain $2\sigma_v^2$, if the model were right! Also, it is unlikely that net investment which is the first difference in net plant is independent of the unmeasured fluctuations in real wage rates. Hence, the consistency of the external instrumental variable estimates is rather suspect.

We turn, therefore, to deriving an estimate of β and σ_v^2 from the contrast between the first difference and the within estimator, both of which should be free from the correlated individual effects bias. Using the formulae given in equations (1.6) and (1.7) and the values calculated in Table 1 yields

$$\beta = .776 \text{ and } \sigma_v^2 = .0048 \tag{3.6}$$

Another estimate can be had by using the "long" differences instead of the first ones in these formulae. This calculation yields

$$\beta = .785 \quad \text{and } \sigma_v^2 = .0050 \tag{3.7}$$

which is very close and supports our final conclusion that for this equation the "true" β is about .78 and that the "unanticipated" variance output accounts for about 40 percent of the observed variance in the first differences, 18.5 percent of the within variance, and a much smaller fraction, 7 percent, of the "long" differences, where the variability in growth trends predominates.

We now take a closer look at the estimates to see how well the hypothesis of uncorrelated measurement error holds up. We first present the moment estimates from equation (2.9) which are derived from the variances and covariances of the differenced variables. They should be approximately equal if the no correlation assumption holds true. These moment estimates are based on sucessive differences so that the first estimate is based on variances and covariances one period apart, the second estimate uses moments two periods apart, and so on. Note that the estimates in equations (3.6) and (3.7) use the same moments but use the within estimate as the basis while here we use the first difference estimator as the basis. As can be seen in (3.8), the last 3 estimates are quite close, although they continue to rise which may

$$(3.8) \qquad \beta_1 = .674 \qquad \beta_2 = .757 \qquad \beta_3 = .783 \qquad \beta_4 = .794$$

indicate some remaining correlation in the measurement error. However, β_1 seems definitely lower. A tentative specification might therefore be an MA1 process in the measurement error. To determine whether this is the case, we now turn to instrumental variable estimation which is asymptotically equivalent to our previous moment estimators and which allows for convenient calculation of variances of estimates and of test statistics.

In Table 3.2, we present the results of estimating the model in first difference form. Thus, the individual firm effects, a_i, are eliminated so that correlation between firm effects and output no longer is present. Results from both OLS and IV (2SLS) are presented where the internal instruments are chosen subject to the no correlation in measurement error assumption. In the first column, the OLS results again yield quite low estimates of β. Furthermore, the data seems distinctly non-stationary, especially during the steep recession year of 1975. In Column 2, the instrumental variable results under the no correlation assumption for measurement error are given. Here, for example, the instrumental variables used for the first line, the difference between 1974 and 1973, are output in the years 1972, 1975, 1976, and 1977.[1] Again the 1975-1974 difference gives a significantly lower estimate than do the other years. In line 5, we give the restricted IV results where β across years is restricted to be the same. The null hypothesis of equality of β across years leads to a χ^2_3 variable. The

[1] We use 1973 as the base year in our estimates so that the initial conditions from 1972 can be used as an instrumental variable.

Table 3.2: OLS and IV Estimates of the Employment-Output Relationship

Years for Difference	OLS	No Correlation IV	MA1 IV	Test Stat. χ^2_1	MA2 IV
First Difference					
1. 1974-1973	.568 (.025)	.842 (.037)	1.25 (.292)	1.98	1.24 (.294)
2. 1975-1974	.395 (.025)	.512 (.028)	.647 (.048)	— 11.99	—
3. 1976-1975	.505 (.020)	.748 (.037)	1.45 (.322)	— 4.82	3.55 (3.56)
4. 1977-1976	.491 (.022)	.726 (.164)	1.19 (.250)	6.04	1.20 (.250)
5. Restricted Estimates	.515 (.010)	.666 (.017)	.675 (.041)	0.06	1.22 (.190)
Nested differences					
6. 1977-1972	.730 (.016)	.745 (.020)	.842 (.038)	9.01	—
7. 1976-1973	.673 (.017)	.732 (.019)	—	—	—
8. Restricted Estimates	.553 (.012)	.696 (.016)	.765 (.031)	—	—
Overlapping Diff.					
9. 1975-1973	.581 (.019)	.716 (.023)	.791 (.141)	.291	—
10. 1976-1973	.673 (.017)	.732 (.019)	—	—	—
11. 1977-1973	.703 (.017)	.753 (.020)	.224 (.247)	4.56	—
12. Restricted Estimates	.535 (.010)	.745 (.019)	.712 (.138)	0.06	

estimated statistic is 55.8 which leads to a rejection, although one should recognize that our rather large sample size makes rejection of most hypotheses likely at the usual significance levels.

In lines 6 and 7 of Table 2, we present results for the nested differences. Note that under the model specification these estimates would be independent if Σ were diagonal. The third overlapping differences estimate for 1975 minus 1974 is given in line 2. The IV estimates are quite close for 1977-1972 and for 1976-1973 with 1975-1974 again quite different. The restricted estimate is .696 which again leads to a rejection of equality across years. The test statistic is calculated to be 50.9. The estimates are distinctly non-independent with the correlation coefficients equal to .52, .21, and .29. However, our estimation procedures and estimated standard errors account for this possible non-independence in Σ. We conclude that all the serial correlation in the disturbances is not accounted for by the firm effects, industry effects, and time effects.

In lines 9-11 of Table 2 we present the overlapping difference results. The results for 1974-1973 are given in line 1. Note that the OLS estimates increase monotonically as the gap in years increases which is exactly what our errors in variables analysis for the uncorrelated measurement error model of Section 2 predicted. The IV estimate for 1974-1973 is .842. The increase in the IV estimates in lines 9-11 together with the moment estimator results in equation (3.8) raise doubts about the no correlation in the measurement errors assumption. The restricted estimate is .745 which is also quite close to the moment estimator results which we presented earlier.[1]

1. The test statistic for the restricted estimate is 8.6 which is significant at the 5% level but not at the 1% level for a χ^2_3 variable.

In column 3 of table 2, we relax the no correlation in measurement error assumption and allow for an MA1 process which was suggested by the results in equation (3.8). To do this, we use only instruments that are two or more years away from the years used to form the particular difference . First, note that the IV estimates rise in each period. On the basis of a Hausman (1978) test, the difference is significant for 3 out of 4 estimates.[1] Also, note that in 3 of the 4 equations the estimate of β exceeds one, although never by a statistically significant amount. These results are more in line with the predictions of the production function model of equation (3.1). In fact, they are what is expected with a true value of about .75 which is a commonly accepted value of α. However, the difference of 1975 minus 1974 is again much lower. Because of this one equation, the restricted estimates gives an almost identical estimate to the no correlation restricted estimate.

For nested differences, we can only estimate 1977-1972 and 1975-1974 due to the unavailability of instruments for 1976-1973. In both years, the estimates rise by a significant amount. The overlapping estimates also rise. These two additional sets of estimates again lead to severe doubts about the no correlation assumption in the measurement error.

In the last column of Table 2, we now allow for an MA2 process in the measurement error. For 1974-1973 and 1977-1976 the results are virtually identical to the MA1 measurement error model. For 1976-1975 the estimate is so imprecise that no conclusion can be drawn. A possible conclusion here is that the MA1 measurement process is sufficient. A Hausman-type test does not reject

1. Overidentifying restriction tests on sets of instruments can also be used here. They lead to χ^2 tests with higher degrees of freedom. See Newey (1983). For instance for the 4 first difference estimates in column 2 of Table 3.2, the test statistics are: 20.8, 17.6, 32.2, and 22.7. All of these statistics are higher than conventional significance levels of a χ^2_3 variable.

this hypothesis in any of the three years.[1] Furthermore, an MA1 process in measurement error could well arise because of differences in fiscal years across firms and the change in fiscal years among many firms which took place in 1976.

Our empirical results are not easily summarized. First we have found that it is quite likely that correlation exists between firm effects and measured output. Second, traditional covariance techniques are subject to errors in variables which have a sizeable effect which goes in the predicted direction. Next moments estimators and IV estimators reduce the magnitude of the bias. Correlation in measurement error seems present, although an MA1 process seems adequate for our particular data set. Lastly, we return to the puzzle: Is $\beta < 1$?. Apart from 1975-1974, the MA1-IV results indicate that $\beta > 1$. However, the nested difference and overlapping difference results continue to indicate that β lies in the range of .75 to .85. These conflicting results are difficult to tell a completely coherent story about within the context of our model. To some extent, the $\beta > 1$ results must be discounted since they should not arise within the context of the errors in variables model which we begin with and estimates derived from the original levels specification.

The following might be a possible interpretation. Approximately .005 of the variance of log of output is unanticipated. This is less than 0.3% of the total variance in log output, but it accounts for close to 40% in the variance of its first differences. Allowing for such errors raises the estimated β from about .5 to about .75 or .8 leaving us rather far from the expected unitary elasticity.

---

1. Overidentifying restriction tests on the MA1 measurement model reject only for 1974-73, but not for the other 3 first difference estimates. The test statistics, with degrees of freedom in parentheses, are 9.9 (1), 2.0 (1), 2.6 (1), and 1.7 (2). These results lead to further confirmation about the adequacy of the model.

One interpretation of these results is the distinction between variable and overhead labor. If "overhead" labor does not vary much over the horizon and of our data and the size of shocks that we observe, our estimates imply that it accounts for about 20+ percent of manufacturing employment. Our results are also consistent with Sims (1974) whose final estimate of the 'total' β was about .8 for a specification which used the number of workers rather than manhours as the dependent variable.

IV. Conclusion

Correlation between individual and firm effects and right hand side variables has been emphasized in the recent literature on estimation in panel data. Specification tests often lead to the rejection of the hypothesis of no correlation. This paper points out that errors in variables can lead to the rejection since covariance estimators, e.g. within or fixed effects procedures, are affected more by errors in variables than are the GLS, e.g. random effects, estimators. Under reasonable assumptions about correlograms, the bias moves in a predictable direction so it can be discovered by considering the differences between first difference and long difference estimators. Consistent estimators are available from a method of moments approach and from an IV approach. The IV approach leads to specification tests which permit an assessment of whether the measurement error correlation assumptions, which provides the rationale for the validity of the instruments, hold true. Our empirical example provides a rather convincing illustration of the utility of our approach to the analysis of errors in variables in panel data. Unfortunately, we have not been able to solve the long standing $\beta \gtrless 1$ puzzle of the elasticity of labor demand with respect to output.

V. Notes on the literature: For work on panel data see Maddala (1973), Mundlak (1978), Hausman (1978), Hausman and Taylor (1981) and Chamberlain (1982). For the importance of errors in such contexts see Griliches (1974, 1979, and 1984). For an earlier effect at identifying the error variance from the contrast between levels and first differences in a single series see Karni and Weissman (1975). For a discussion of labor demand estimates see Brechling (1973), Sims (1974), Solow (1964), and Medoff and Fay (1983).

References

Bielby, W. T., R. M. Hauser and K. K. Featherman, 1977. "Response Errors of Non black Males in Models of the Stratification Process," in Goldberger and Aigner (eds.) Latent Variables in Socioeconomic Models, Amsterdam: North Holland Publishing Company 227-251.

Brechling, F. P. R., 1965. "The Relationship Between Output and Employment in British Manufacturing Industries, " Review of Economic Studies, Vol. 32.

Chamberlain, G., 1982. "Multivariate Regression Models for Panel Data," Journal of Econometrics 18 (1).

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

Cummins, C. B. H. Hall, E. S. Laderman, 1982. "The R&D Master File: Documentation," Harvard University and NBER, unpublished.

Friedman, M 1957. A Theory of the Consumption Function, New York: NBER General Series 63.

Griliches, Z., 1979. "Sibling Models and Data in Economics: Beginnings of a Survey," Journal of Political Economy 87 (5), Part 2.

————, 1984. "Data Problems in Econometrics," forthcoming in Z. Griliches and M. Intriligator (eds.) Handbook of Econometrics, Vol. III, Amsterdam: North Holland Publishing Company.

Hansen, L., 1982. "Large Sample Properties of Generalized Methods of Moments," Econometrica, 50.

Hausman, J. 1978. "Specification Tests in Econometrics", Econometrica, 46.

Hausman, J. and W. Taylor, 1981. "Panel Data and Unobservable Individual Effects," Econometrica 49.

Karni, E. and I. Weissman, 1974. "A Consistent Estimator of the Slope in a Regression Model with Errors in Variables," Journal of the American Statistical Assocation 69 (345).

Maddala, G. S. 1971. "The Use of Variance Components Models in Pooling Cross-Section and Time Series Data," Econometrica, 39 (2).

——————,1977. Econometrics, New York: McGraw Hill.

Medoff, J. L. and J. A. Fay, 1983. "Labor and Output Over the Business Cycle: Some Direct Evidence," Harvard University, unpublished.

Mundlak, Y. 1978. "On the Pooling of Time Series and Cross Section Data," Econometrica, 46.

Newey, W., 1983. Specification Testing and Estimation Using a Generalized Method of Moments , MIT Ph.D. Thesis.

Sargan, J. D., 1958. "The Estimation of Economic Relationship Using Instrumental Variables, Econometrica, 26.

Sims, C. A., 1974. "Output and Labor Input in Manufacturing," Brookings Papers on Economic Activity, Vol. 3.

Solow, R., 1964. "Draft of Presidential Address on the Short-Run Relation of Employment and Output," unpublished draft.

White, H., 1982, "Instrumental Variables Regression with Independent Observations," Econometrica, 50.