Journal of Econometrics 31 (1986) 93–118. North-Holland

ERRORS IN VARIABLES IN PANEL DATA

Zvi GRILICHES

Harvard University and NBER, Cambridge, MA 02138, USA

Jerry A. HAUSMAN

MIT and NBER, Cambridge, MA 02139, USA

Received May 1984, final version received November 1985

Panel data based studies in econometrics use the analysis of covariance approach to control for various 'individual effects' by estimating coefficients from the 'within' dimension of the data. Often, however, the results are unsatisfactory, with 'too low' and insignificant coefficients. Errors of measurement in the independent variables whose relative importance gets magnified in the within dimension are then blamed for this outcome.

Errors-in-variables models have not been used widely, in part because they seem to require extraneous information to be identified. We show how a variety of errors-in-variables models may be identifiable and estimable in panel data without the use of external instruments and apply it to a relatively simple but not uninteresting case: the estimation of 'labor demand' relationships, also known as the 'short-run increasing returns to scale' puzzle.

1. Introduction

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-

*We would like to thank the NSF and NBER for research support; Gary Chamberlain, Bronwyn H. Hall, Adrian Pagan, Ariel Pakes, Whitney Newey, William Nordhaus and the anonymous referees for a number of very helpful comments; and John Bound, Adam Jaffe and Bruce Meyer for excellent research assistance.

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

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.

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 deducible changes in the biases due to such errors which can be used then to identify the importance of these errors and recover the 'true' parameters. 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 (1964) and Medoff and Fay (1985).

In the next four sections we first outline our approach in a very simple context; next we present the algebra for the more general case and discuss the different estimation strategies; we turn then 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.

2. The problem of errors in variables in panel data: An introduction

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

$$y_{it} = \alpha_i + \beta z_{it} + \eta_{it}, \tag{1}$$

where the α_i 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. The z_{it} are not observed directly, however. Only their erroneous reflection, the x_{it},

$$x_{it} = z_{it} + v_{it}, \tag{2}$$

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

$$y_{it} = \bar{\alpha} + \beta x_{it} - \beta v_{it} + \eta_{it} + (\alpha_i - \bar{\alpha}), \tag{3}$$

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 (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.[2] What is less obvious is that there are different ways of eliminating the first source of bias, that they imply different consequences for the size of the second 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. We shall show in the next section that assuming stationary and uncorrelated measurement errors the plims of the difference and within estimators are

$$\operatorname{plim} b_d = \beta\left(1 - \frac{2\sigma_v^2}{\operatorname{var}(\mathrm{d}x)}\right), \qquad \operatorname{plim} b_w = \beta\left(1 - \frac{T-1}{T}\frac{\sigma_v^2}{\operatorname{var}\tilde{x}}\right), \tag{4}$$

where

$$\mathrm{d}y_{it} = y_{it} - y_{it-1}, \qquad \tilde{y}_{it} = y_{it} - \bar{y}_i,$$

and similarly for the other variables.

For the most likely case in economics: positively serially correlated 'true' x's (z's) with a declining correlogram and, for $T > 2$,

$$\operatorname{var}(\mathrm{d}x) < \frac{2T}{T-1}\operatorname{var}\tilde{x}, \tag{5}$$

and hence $|\text{bias}\, b_d| > |\text{bias}\, b_w|$. That is, errors of measurement will usually bias the first difference estimators downward (toward 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 $\operatorname{var}\mathrm{d}x$ and $\operatorname{var}\tilde{x}$ 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

$$\begin{aligned}\beta = {} & \left[2b_w/\operatorname{var}(\mathrm{d}x) - (T-1)b_d/T\operatorname{var}\tilde{x}\right] \\ & /\left[2/\operatorname{var}\mathrm{d}x - (T-1)/T\operatorname{var}\tilde{x}\right],\end{aligned} \tag{6}$$

$$\sigma_v^2 = (\beta - b_d)\operatorname{var}(\mathrm{d}x)/2\beta. \tag{7}$$

[2] See Griliches (1979) for a related discussion in the context of the analysis of sibling data and Freeman (1984) and Chowdhury and Nickell (1985) in the zero–one variable (impact of unionization) context.

The first difference and within estimators are not the only ones that can yield such implicit estimates of the bias. If we define $d^j = 1 - L^j$ as the difference operator 'j periods apart', where L is the lag operator, then $\mathrm{d}^j y_t = y_t - y_{t-j}$ (and similarly for $\mathrm{d}^j x_t$) and, for example, $\mathrm{d}^{T-1}y = y_T - y_1$ is the 'longest' difference possible in a particular panel. For $T > 2$ several more slope estimators can be computed based on 'different lengths' differences with

$$\operatorname{plim} b_j = \beta - 2\beta\sigma_v^2/\operatorname{var}(\mathrm{d}^j x), \tag{8}$$

implying additional estimates of β and σ_v^2. These take the form of

$$\tilde{\beta}_{jh} = (\omega_j - \omega_h)/(s_j^2 - s_h^2), \tag{9}$$

where ω_j is the covariance between $\mathrm{d}^j y$ and $\mathrm{d}^j x$ while s_j^2 is the comparable variance of $\mathrm{d}^j x$. Such estimates can be combined optimally to improve upon their individual efficiency.

It is worth noting that these 'contrast' or 'moments' estimators can be given a straightforward instrumental variables interpretation. Let us take as an example $\tilde{\beta}_{21}$, i.e., an estimate derived by solving the implied bias relationship from OLS estimates computed using differences two periods apart $(y_T - y_{T-2})$ and one period apart $(y_T - y_{T-1})$. The numerator of (9) can be rewritten as

$$x'F_2'F_2 y - x'F_1'F_1 y = x'[F_2'F_2 - F_1'F_1]y = x'Qy = w'y,$$

and similarly also for the denominator of (9), where F_2 and F_1 are appropriate 'differencing' matrices. In the simplest $T = 3$ case, where we are using three cross-sectional equations to estimate one β, these matrices take the $(-1\ 0\ 1)$ and $(0\ -1\ 1)$ form, respectively, and the resulting instrumental variable can be seen to equal $w = (x_1 - x_3\ \ x_3 - x_2\ \ x_2 - x_1)$.[3] If the measurement errors are stationary and uncorrelated ($\sigma_{vt}^2 = \sigma_v^2$, $\sigma_{v_t v_{t+h}} = 0$ for $h \neq 0$), then $\mathrm{E}w'x = w'z$ and w is a valid instrument for x.

The above 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 are willing to make one of the following three types of assumptions:

(1) The correlation is of the moving average form of order k and the available sample is long enough in the time dimension, $T > k + 1$, to allow the use of more distant instruments.

[3] This is one of two such possible instruments. The other can be derived using the alternative $F_1 = (-1\ 1\ 0)$. A. Pagan suggested this simplification.

(2) The measurement errors are stationary while the true underlying variables (the z's) are not.[4] For example, $x_4 - x_1$ can be used as an instrument for $x_3 - x_2$ as long as $\mathrm{E}v_4v_3 = \mathrm{E}v_2v_1$ but $\operatorname{cov}(z_4, z_3) \neq \operatorname{cov}(z_2, z_1)$. We shall come back to this point in the next section.

(3) More generally, one can always allow for correlation in the measurement errors if its magnitude is know a priori. For example, if x is a gross capital stock measure based on a twenty-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 approximately $19/20 = 0.95$. The first difference bias formula now becomes

$$\operatorname{plim}(b_d - \beta) = \frac{-2\beta\sigma_v^2(1-\rho_\varepsilon)}{\operatorname{var} dx} = \frac{-0.1\beta\sigma_v^2}{\operatorname{var} dx}, \tag{10}$$

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 tested for and consistent estimators can be derived in their presence.

These results have been derived 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 structure).

We now turn to the consideration of the general case and the formulation of optimal estimators for it.

3. Derivation of results

We now reconsider our basic model, eq. (1), and derive the relationship among the various estimators. The general specification of (1) for an observation on an individual unit (there are N such units or individuals) is

$$y = z\beta + l\alpha + \eta, \tag{11a}$$

$$x = z + v, \tag{11b}$$

$$y = x\beta + l\alpha + \eta - v'\beta = x\beta + l\alpha + \varepsilon, \tag{11c}$$

[4] This was suggested to us by A. Pakes.

with

$$V(\eta)=\Sigma, \qquad V(v)=\Omega,$$

where y, z, x, y and v are $T\times 1$ vectors, l is a vector of ones, β is a constant, but α varies across the N units or individuals, while Σ and Ω are $T\times T$ symmetric matrices.

The within estimator and difference estimator can be seen to arise from the transformation of eq. (11c),

$$Ry = Rx\beta + R\alpha + R\varepsilon, \tag{12}$$

where each row of R must sum to zero to eliminate α. For instance, the R for the first difference estimator will be a bi-diagonal matrix with -1 on the diagonal and $+1$ on the superdiagonal. The within estimator has $R=I_T-J$ where J is $1/T$ times a matrix of all ones. We calculate for this class of estimators

$$\begin{aligned}\operatorname{plim}(\beta_T) &= [\mathrm{E}(x'R'Rx)]^{-1}\mathrm{E}(x'R'Ry)\\ &= [\mathrm{E}(z'Qz)+\mathrm{E}(v'Qv)]^{-1}\mathrm{E}[v'Qv]\beta,\end{aligned} \tag{13}$$

where

$$Q=R'R, \qquad \mathrm{E}(v'Qv)=\operatorname{tr}[Q\Omega],$$

and all plims in the paper are taken as $N\to\infty$. For ease in derivation of 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 and that Σ and Ω are both diagonal matrices. (We shall relax all of these assumptions below.) With these assumptions we calculate the probability limit of the first difference estimator as

$$\begin{aligned}\operatorname{plim} b_d-\beta &= \left[2\sigma_z^2(1-\rho_1)+2\sigma_v^2\right]^{-1}\left(-2\sigma_v^2\beta\right)\\ &= -\left[\sigma_z^2(1-\rho_1)+\sigma_v^2\right]^{-1}\sigma_v^2\beta,\end{aligned} \tag{14}$$

where ρ_j is the jth serial correlation coefficient between the true regression variables z. 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.

Table 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	$-(\sigma_x^2 - \sigma_z^2\rho_1)^{-1}\sigma_v^2\beta$	$-(\sigma_x^2 - (\sigma_z^2/3)(2\rho_1 + \rho_2))^{-1}\sigma_v^2\beta$	$\frac{2}{3}\rho_1 + \frac{1}{3}\rho_2 < \rho_1$
4	$-(\sigma_x^2 - \sigma_z^2\rho_1)^{-1}\sigma_v^2\beta$	$-(\sigma_x^2 - (\sigma_z^2/6)(3\rho_1 + 2\rho_2 + \rho_3))^{-1}\sigma_v^2\beta$	$\frac{1}{2}\rho_1 + \frac{1}{3}\rho_2 + \frac{1}{6}\rho_3 < \rho_1$
$\vdots$			
$T \to \infty$	$-(\sigma_x^2 - \sigma_z^2\rho_1)^{-1}\sigma_v^2\beta$	$-(\sigma_x^2 - (2\sigma_z^2/(T-1))\Sigma((T-j)/T)\rho_j)^{-1}\sigma_v^2\beta$	$(2/T)(\rho_1 + \rho_2 + \cdots) < \rho_1$

For the within estimator, the probability limit is

$$\operatorname{plim} b_w - \beta = -\left[\sigma_x^2 - \frac{2\sigma_z^2}{T(T-1)}\sum(T-j)\rho_j\right]^{-1}\sigma_v^2\beta. \tag{15}$$

The formula is the same as in the usual OLS with errors of measurement 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. The conditions which cause the within estimator to be less inconsistent, which we expect to be the usual case, are given in table 1.

For $T = 2$, the estimators give numerically identical results since the within transformation and first differences are related by the formula $\frac{1}{2}\,\mathrm{d}x = \tilde{x}$. For $T = 3$, the condition for b_w to be less biased than b_d 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. Then the probability limit of the least squares estimator based on differences j

periods apart is

$$\operatorname{plim} b_j - \beta = -\left(\sigma_z^2(1-\rho_j)+\sigma_v^2\right)^{-1}\sigma_v^2\beta. \tag{16}$$

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_1. 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 that 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.

In the more general case of serial correlation in the v's and η's, Σ and Ω are not diagonal anymore. For our two leading special cases,

$$\operatorname{tr}[Q\Omega \otimes I] = 2N(T-1)\sigma_v^2(1-r), \tag{17}$$

for first differences, and

$$\operatorname{tr}[Q\Omega \otimes I] = N\left[\frac{T-1}{T}\sigma_v^2 - \frac{2\sigma_v^2}{T^2}\sum_{j=1}^{T}(T-j)r_j\right],$$

for the within estimator, where r_j is the jth serial correlation in the measurement error v.

The inconsistency of the first difference estimator is therefore

$$\operatorname{plim} b_d - \beta = -\left[\sigma_z^2(1-\rho_1)+\sigma_v^2(1-r_1)\right]^{-1}\sigma_v^2(1-r_1)\beta. \tag{18}$$

The bias of the first difference estimation here as compared to eq. (12) 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 $\rho_1 > 0$.

The inconsistency of the within estimator is

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

$$\times\left[\sigma_v^2 - \frac{2\sigma_v^2}{T(T-1)}\sum_j(T-j)r_j\right]\beta. \tag{19}$$

As with first differences, the inconsistency in the within 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. To compare the bias of the first difference and within estimators, first note that, for $T = 2$, they are identical as before. For $T \geq 3$, it may be reasonable to assume that $\rho_j > r_j > 0$ 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 than the first difference estimator if $(\rho_1 - \rho_2)/(r_1 - r_2) > (1 - \rho_1)/(1 - r_1)$ which holds if the serial correlation in the true variable decreases less slowly than the serial correlation in the measurement error.[5] 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 z's be less than for the v'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 in eq. (18) 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 follow AR1 processes with $\rho_1 > r_1$, then $j = T - 1$ will minimize the inconsistency. On the other hand, if z follows 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 correlation coefficients.

It would be interesting to know which combinations of estimators of the form $(X'Qx)^{-1}X'Qy$, used to eliminate α from eq. (11), have good properties with respect to errors of measurement. Even if we choose the minimization of the inconsistency as our criterion, the optimal combination will depend on the properties of Σ and Ω. 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 non-diagonal Ω, the optimal estimator depends on both Σ and Ω. A potential topic for future research would be to characterize this dependence for interesting classes of stochastic processes determining Σ and Ω.[6]

We now turn to the question of consistent estimation. In the general correlated case, with Ω unrestricted, the problem remains unidentified. That is, external instruments uncorrelated with the measurement error are required for consistent estimation. While the assumption of stationarity of the measurement

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

[6] 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.

errors in combination with non-stationarity of the underlying true variables (the z's) does allow identification also in the correlated errors case, we do not find it especially compelling and focus first on the non-stationary z's and uncorrelated v's case. The procedures we develop can be used also in the 'partial' correlation case, e.g., if v follows the MA(m) process with $m < T$, and can be also extended to the stationary case. The strategy we propose here is to take advantage of the existence of alternative consistent estimators, i.e. over-identification, to test the assumption of no correlation in the v's and their stationarity. If the alternative estimates of β are mutually coherent, then the researcher can have some confidence that his assumptions hold true.

Our estimation strategy starts with the original eq. (11c),

$$y = x\beta + l\alpha + \varepsilon,$$

and looks for instrumental variables of the form

$$w = Px,$$

where the matrix P must satisfy three sets of conditions:

$$\begin{aligned} &l'P = 0, \\ &\mathrm{E}w'x = \mathrm{E}w'z, \quad \text{i.e.,} \quad \mathrm{E}x'P'v = 0 \rightarrow \operatorname{tr} \Omega P = 0, \qquad (20) \\ &\mathrm{E}x'P'x \neq 0. \end{aligned}$$

The first requirement assures the elimination of α, the individual effects, from (11). The second requirement allows the use of those x's and their combinations which are uncorrelated with the particular measurement error v_t. Its content will change depending on the assumptions made about the serial correlation structure of the v's. The final requirement is one of non-zero correlation between the instrument w and x.

The simplest and least restrictive set of assumptions to start with are (1) that the z's are non-stationary and (2) the v's are non-stationary and white (uncorrelated) to some order. Under these assumptions we need to *difference* the y's to get rid of the α's, and we can use non-corresponding adjacent x levels to instrument the x difference. Starting with the case of correlated errors we can either relax the restrictions further by allowing for serial correlation in the errors up to some order or restrict the model further by imposing the assumption of stationarity on the measurement errors. In the latter case additional combinations of differences in the x's become valid instruments.[7]

[7] If the z's were also stationary, there would be no gain in instruments here since the additional instruments would have zero variance. On the other hand, it is remarkable that under stationarity of (x_{it}, α_i) the original specification $y = X\beta + l\alpha + \eta$ can be estimated using differenced x's as instruments since, e.g., $\mathrm{E}[(x_{i2} - x_{i1})\alpha_i] = 0$. This approach would allow an application of the Hausman–Taylor (1981) procedure where other right-hand-side variables are included, say D, and $(D - \bar{D}.)$ are used as additional instruments.

Table 2

Potential list of instruments for $T=4$ under different assumptions about Ω; $y = z\beta + l\alpha + \eta$, $x = z + v$.

Equations to be estimated	Valid instruments[a]: Non-stationary v's: No correlation	Non-stationary v's: MA(1)	Stationary v's: No correlation	Stationary v's: MA(1)	Stationary v's: MA(2) & (3)
$(y_2 - y_1) = \beta(x_2 - x_1)$	x_4	x_4	$x_1 + x_2$, x_4	$x_1 + x_2, x_4$	$x_1 + x_2$
$(y_3 - y_2) = \beta(x_3 - x_2)$	x_1, x_4		$x_1, x_2 + x_3, x_4$	$x_2 + x_3$	$x_2 + x_3$
$(y_4 - y_3) = \beta(x_4 - x_3)$	x_1, x_2	x_1	x_1, $x_3 + x_4$	$x_1, x_3 + x_4$	$x_3 + x_4$
$(y_3 - y_1) = \beta(x_3 - x_1)$	x_2		x_2	x_2	x_2
$(y_4 - y_2) = \beta(x_4 - x_2)$	x_3		x_3	x_3	x_3
$(y_4 - y_1) = \beta(x_4 - x_1)$	x_2, x_3		x_2, x_3		
$\begin{pmatrix} y_3 - y_2 \\ y_4 - y_1 \end{pmatrix} = \beta\begin{pmatrix} x_3 - x_2 \\ x_4 - x_1 \end{pmatrix}$		$\begin{pmatrix} -x_1 \\ x_2 \end{pmatrix}\begin{pmatrix} -x_4 \\ x_3 \end{pmatrix}$		$\begin{pmatrix} -x_1 \\ x_2 \end{pmatrix}\begin{pmatrix} -x_4 \\ x_3 \end{pmatrix}$	$\begin{pmatrix} -x_1 \\ x_2 \end{pmatrix}\begin{pmatrix} -x_4 \\ x_3 \end{pmatrix}$
$\begin{pmatrix} y_2 - y_1 \\ y_4 - y_3 \end{pmatrix} = \beta\begin{pmatrix} x_2 - x_1 \\ x_4 - x_3 \end{pmatrix}$		$\begin{pmatrix} x_3 \\ x_2 \end{pmatrix}$		$\begin{pmatrix} x_3 \\ x_2 \end{pmatrix}$	$\begin{pmatrix} x_3 \\ x_2 \end{pmatrix}\begin{pmatrix} x_4 \\ x_1 \end{pmatrix}$
Total number of instruments	8	5	11	10	9

[a]Only one possible list of instruments which exhausts all available information is presented. For some equations, valid instruments are not listed. For example, x_3 is valid for the $y_2 - y_1$ difference in the no-correlation case. However, the information from this instrument is redundant. The information is already included in the instruments for the other equations.

To the extent that we have enough data (time periods) some versions of this model will be heavily overidentified and hence some of these restrictions are testable. We shall illustrate our approach and list the different possible sets of instruments for the $T=4$ case in table 2. Consider, for example, the second line of table 2. It corresponds to estimating β from the cross-sectional difference between y_3 and y_2. (Note that each such cross-section has N degrees of freedom.) For uncorrelated measurement errors, x_1 and x_4 are both valid instruments. If the errors are correlated according to a MA(1) or higher order scheme there are no valid instruments for this difference without imposing further restrictions. If we are willing to assume stationarity for the v's, we gain x_2+x_3 as an instrument. The quality or contribution of such an instrument is dubious, however, since it depends on $\operatorname{var}(z_2)-\operatorname{var}(z_3)\neq 0$ and non-negligible. As we relax the no-correlation assumption in the stationary v's case, we lose some instruments, but can form other 'combinations' of difference estimators which use all of the available information in the data. They are shown in the southeast corner of this table where the corner $\tilde{\beta}$ is given by $[x_3'(y_2-y_1)+x_2'(y_4-y_3)]/[x_3'(x_2-x_1)+x_2'(x_4-x_3)]$. A more general version of table 2 for arbitrary T is given in the appendix (co-authored with Bruce Meyer) with a discussion of the rules for constructing and finding the relevant number of such instruments.

The asymptotically efficient way of combining all the different $\tilde{\beta}$ estimators in table 2 is using a 'system' estimator where the estimated β's are constrained to be equal and the covariance of the stochastic disturbances Σ is taken into account in weighting them and in computing the variance of the resulting coefficient. 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 also contained in the instrument set of that equation.

It is useful, as a start, to estimate each equation in table 2 separately by 2SLS using different instrumental variables in each equation. But to combine such estimators optimally we use the Generalized Method of Moments estimator, developed by Hansen (1982) and White (1982), and allow also for conditional heteroscedasticity. The estimator is

$$\beta^* = [\tilde{x}'\tilde{w}U^{-1}\tilde{w}'\tilde{x}]^{-1}\tilde{x}'\tilde{w}U^{-1}\tilde{w}'y \quad \text{with} \quad U=\frac{1}{N}\sum_i^N \tilde{w}_i'\tilde{\varepsilon}_i\tilde{\varepsilon}_i'\tilde{w}_i, \tag{21}$$

where $\tilde{\varepsilon}$ are the stacked $\mathrm{d}^j\varepsilon$'s and $\tilde{\varepsilon}$ is calculated from an initial consistent estimate of β, $\tilde{x}$ are the stacked d^jx's, $\tilde{y}$ are the stacked d^jy's, and $\tilde{w}$ is the matrix of instruments.[8] The asymptotic covariance matrices of the estimator is $V(\beta^*)=[\tilde{x}'\tilde{w}U^{-1}\tilde{w}'\tilde{x}]^{-1}$. This asymptotic covariance matrix is different from

[8] Note that w is a block-diagonal matrix.

a simple weighting of β_{IV} from 2SLS. Effectively we estimate each instrument equation combination (or orthogonality condition) separately, compute their associated estimated variance–covariance matrix, and then pool the individual β's using this matrix for weighting them.

We now turn to the question of whether the no-correlation assumption in the errors in measurement is valid. Some such assumption about the form of the process generating the measurement error is needed because the general correlated case is unidentified without the use of special stationarity assumptions or other extraneous variables as instruments. Note that if we applied least squares (OLS) to the equations in table 2, equation by equation, we would 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 different estimates of β will have different probability limits. Therefore, a testing procedure is to estimate a system of equations based on different d^j's in unrestricted form so that each equation is allowed to have its own β. A large sample χ^2 test with $m-1$ degrees of freedom for equality of the β_j's is equivalent to the implicit test in table 2 that all 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 the other x's as instrumental variables we could restrict the list to those x's which are at least two time periods away. I.e., use only the instruments given in column 2 or 4 of table 2. 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, overidentification tests of the Sargan (1958) and Hansen (1982) type can also be used. Under stationarity assumptions these various tests are closely related but have different operating characteristics because they are based on different degrees of freedom. In the general case of non-stationarity of the x's they will differ, although Newey (1983) provides a partial guide to their comparability.

There is an alternative, asymptotically equivalent approach to estimation based on treating the y and x as jointly normally distributed and using maximum likelihood estimators to estimate the 'unobservables': Ω and the variance–covariance matrices of the z's. We can rewrite our original (11) model as

$$\begin{aligned} y_1 &= z_1\beta + l\alpha + \eta_1 = x_1\beta + l\alpha + \eta_1 - \beta v_1, \\ \mathrm{d}y &= (\mathrm{d}z)\beta + \mathrm{d}\eta, \\ x &= z + v, \end{aligned} \tag{22}$$

where $\mathrm{d}y$ and $\mathrm{d}z$ are now $(T-1)\times 1$ vectors. Since the relationship between α and z_1 is free, the y_1 equation is unconstrained and can be ignored in what follows. We consider then the $(2T-1)\times(2T-1)$ matrix of observable moments of all the data:

$$S = \begin{pmatrix} \mathrm{d}y \\ x \end{pmatrix} (\mathrm{d}y \quad x) = \begin{pmatrix} \mathrm{d}y \cdot \mathrm{d}y' & \mathrm{d}y \cdot x' \\ \mathrm{d}y' \cdot x & xx' \end{pmatrix},$$

whose expectation, given our model, is

$$M(\Theta) = \begin{pmatrix} \beta^2 \mathrm{d}z \cdot \mathrm{d}z' + \tilde{\Sigma} & \beta \mathrm{d}z \cdot z' \\ \beta \mathrm{d}z' \cdot z & zz' + \Omega \end{pmatrix}, \tag{23}$$

where

$$\tilde{\Sigma} = \mathrm{E}\,\mathrm{d}\eta \cdot \mathrm{d}\eta'.$$

We leave $\tilde{\Sigma}$ unconstrained and hence the identification of β hinges on our assumptions about Ω and the order of zz'. For $T=4$ and Ω diagonal, the interesting right half of $M(\Theta)$ can be written as

$$\beta \begin{bmatrix} z_{21} - z_{11} & z_{22} - z_{12} & z_{23} - z_{13} & z_{24} - z_{14} \\ z_{31} - z_{21} & z_{32} - z_{22} & z_{33} - z_{23} & z_{34} - z_{24} \\ z_{41} - z_{31} & z_{42} - z_{32} & z_{43} - z_{33} & z_{44} - z_{34} \end{bmatrix}$$

$$\begin{bmatrix} z_{11} + \tau_1 & z_{12} & z_{13} & z_{14} \\ & z_{22} + \tau_2 & z_{23} & z_{24} \\ & & z_{33} + \tau_3 & z_{34} \\ & & & z_{44} + \tau_4 \end{bmatrix}$$

where the z_{ii}'s are the appropriate second-order moments of the unobservable z's and the τ_t's are the diagonal elements of Ω. We have then the twenty-four distinct observable moments of the $\mathrm{d}y \cdot x'$ and $x \cdot x'$ variety and thirteen unknowns [ten components of zz', β, and four τ's, where $\tau_t = \sigma_v^2(t)$] and the model is heavily over-identified. It can be constrained further by assuming $\tau_t = \tau$, that the v's are stationary, and/or it can be relaxed by adding another set of τ_{ij} terms to allow for MA(1) or MA(2) correlation between these measurement errors.

More generally, under our assumptions, S is a sufficient statistic for $M(\Theta)$ and we can estimate the unknown parameter vector Θ by 'fitting' the model $M(\Theta)$ to the observed matrix of variances and covariances S using either the LISREL or MOMENTS statistical packages and different versions of the model can then be compared using likelihood ratio tests.[9]

[9]See Jöreskog and Sörbom (1981) and Hall (1979), respectively, and Aigner et al. (1984) and Griliches (1984) for additional discussions of such models. Note that fourth moments are not being used in either estimation or inference here.

4. 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:

$$l = \frac{1}{\alpha}q - \frac{1-\alpha}{\alpha}k, \tag{24a}$$

$$l = \log\alpha + q - w', \tag{24b}$$

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 (1974)]. 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 1,242 U.S. manufacturing firms for the six years, 1972–1977, from the NBER R&D panel [Cummins, Hall and Laderman (1982)], and adopt the second interpretation of the equation to be estimated. In this model,

$$l_{it} = d_t + q_{it}^* + \{-w_{it}' + (\log\alpha_i - \log\bar{\alpha}) + \eta_{it}\}, \tag{25}$$

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; specifically, we assume that

$$w_{it}' = \mu_i + \gamma_t + \tau_{it} \tag{26}$$

has a variance component structure with γ_t subsumed in the d_t and τ_{it}

assumed to be uncorrelated with q_{it}; (2) a term associated with the fact that the labor elasticities α_i 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

$$q_{it} = q_{it}^* + v_{it}, \tag{27}$$

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.[10] We can rewrite the model in terms of observables as

$$l_{it} = a_i + \beta q_{it} + d_t + (-\beta v_{it} - \tau_{it} + \eta_{it}), \tag{28}$$

where we expect $\beta = 1$ and the a_i 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}.

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.[11] 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 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, or 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 estimate is greater than both the first difference estimate and the within estimate as the derivations in section 3 predicted.

There are two ways of interpreting these results. The first would maintain the assumption that $\beta = 1$, ignore the potential presence of correlated individ-

[10] 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.

[11] 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. We experimented also with the use of distributed lags in certain of our models. While lagged q is present when errors in variables are not accounted for, only current q has any statistical or 'economic' significance in the errors in variables models. Thus, 'costs of adjustment' do not seem important on an annual basis in our model once errors in variables are accounted for.

Table 3

Estimates of the employment–output relationship for 1,242 U.S. manufacturing firms, 1972–77; $l_{it} = \alpha_i + \beta q_{it} + d_t$.[a]

Estimation method and degrees of freedom	Net variance of q	β	MSE	Instrumental variables estimates β	MSE
(1) Total d.f. = 7,425	2.265	0.966 (0.003)	0.158	0.994 (0.003)	0.159
(2) Within d.f. = 6,203	0.0313	0.643 (0.008)	0.011	n.c.	
(3) First difference d.f. = 6,204	0.0246	0.480 (0.010)	0.015	0.868 (0.153)	0.019
(4) 'Long' difference d.f. = 1,240	0.1359	0.731 (0.016)	0.047	1.063 (0.039)	0.062

[a]The terms in parentheses are the estimated standard errors. Total regressions contain also five year and twenty-two industry dummy variables. The within and first difference regressions include also year dummies. 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.

ual effects, and accept the instrumental variable results as vindicating this position. There are difficulties with this view, however. The implied 'error' in variables variance of σ_v^2 is 0.038, 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 the estimation of β and σ_v^2 using only 'internal' instruments, i.e., adjacent 'non-corresponding' x's and their combinations. The results, based on the differenced form of our model, are summarized in table 4. The individual firm effects, the α_i, are eliminated by the differencing operation which eliminates the correlation between the individual firm effects and output. We list eleven such differences in this table: five first differences, four differences two periods apart, and two differences four periods apart. The remaining four possible differences (three three periods and one five periods apart) can be derived, in the non-stationary errors case, as linear combinations of the listed (IV) estimates.[12] In the first column we give the simple OLS results for all of these differences. They again yield quite low estimates of β, which rise significantly as the period of differencing is lengthened. Also, the data

[12]Even this system has redundancies as far as the total number of valid orthogonality conditions. It will suffice below to use a shortened list of instruments and to impose all of the twenty-four distinct orthogonality conditions. With $T = 6$, the no correlation of errors assumption yields twenty-five orthogonality conditions of which one is lost in the elimination of the α's. See the appendix for additional discussion of the 'right' number of instruments.

seems distinctly non-stationary, especially during the steep recession year of 1975. Column 2 gives the instrumental variable results under the assumption of no correlation in measurement error [labeled MA(0)]. Here, for example, the instrumental variables used for the second line, the difference between 1974 and 1973, are the levels of output in 1972, 1975, 1976 and 1977. Again, the 1975–74 difference gives a significantly lower estimate than do the other years, as does also the 1973–72 difference. But the latter has a much higher standard error due to the absence of good (early) instruments for it. In lines 12 and 13 we give two different restricted IV results where β is restricted to be the same across all the combination of years. The first is based on combining optimally the eleven different 2SLS estimates of β, using the estimated 2SLS residuals to form the appropriate weighting matrix which allows, also, for conditional heteroskedasticity. They are not fully efficient because they do not allow a free correlation structure between all the different possible instrumental variable estimators in such a set-up, and because of the redundancy in the list of instruments and equations. Line 14 gives the Generalized Method of Moments estimator which takes the individual single instruments as its 'elemental' set and uses the 2SLS residuals to form the appropriate weighting matrix (of larger dimension). The two different system estimators yield similar results. The null hypothesis of equality of β across these different combinations of years leads to a χ^2 variable with the number of distinct instruments minus one as its degree of freedom which is listed in line 14 of table 4. For the column 2, MA(0)–IV estimators, it is 80 with 10 degrees of freedom, and implies a rejection of the null hypothesis, although one should recognize that our rather large sample size makes rejection of most such hypotheses likely at the usual significance levels.

In column 3 of table 4, we relax the no correlation in measurement error assumption and allow for a MA(1) process. 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 for most of the differences (eight out of eleven). On the basis of a Hausman (1978) test, the difference is significant for eight out of the eleven estimates![13]

Also note that in five of the eleven equations the estimate of β exceeds one, although never by a statistically significant amount. However, the difference of 1975 minus 1974 is again much lower. Primarily because of this one equation, which has the lowest standard error, the restricted estimates give an almost identical estimate to the no correlation restricted estimate.

In the last column of table 4 we allow for a MA(2) process in the measurement error. Only four of the eleven equations remain individually

[13] 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 five first difference estimates in column 2 of table 4, the test statistics are: 8, 14, 15, 28 and 14. All of these statistics are higher than conventional significance levels of a χ^2_3 variable.

Table 4

OLS, IV and GMM estimates of the employment–output relationship.[a]

Years for difference	Coefficient (standard error)			
		Consistent IV estimates		
	OLS	MA(0)	MA(1)	MA(2)
(1) 1973–2	0.481 (0.024)	0.276 (1.35)	0.465 (0.145)	0.635 (1.77)
(2) 1974–3	0.569 (0.029)	0.842 (0.038)	1.257 (0.292)	1.241 (0.289)
(3) 1974–4	0.395 (0.029)	0.512 (0.021)	0.647 (0.047)	
(4) 1976–5	0.506 (0.020)	0.748 (0.037)	1.454 (0.322)	3.557 (3.562)
(5) 1977–6	0.491 (0.023)	0.726 (0.165)	1.192 (0.249)	1.203 (0.254)
(6) 1974–2	0.599 (0.022)	1.003 (0.048)	−0.040 (0.340)	1.597 (0.996)
(7) 1975–3	0.581 (0.017)	0.716 (0.027)	0.791 (0.141)	
(8) 1976–4	0.587 (0.020)	0.666 (0.026)	3.135 (2.537)	
(9) 1977–5	0.617 (0.019)	0.695 (0.042)	1.327 (0.107)	3.068 (2.096)
(10) 1976–2	0.708 (0.017)	0.751 (0.019)	−41.795 (1572.56)	
(11) 1977–3	0.703 (0.017)	0.754 (0.020)	0.223 (0.247)	
(12) Restricted β OLS & W2SLS	0.567 (0.015)	0.727 (0.021)	0.670 (0.045)	0.916 (0.161)
(13) Wald test for equality of β's (d.f.)	189 (10)	89.0 (10)	25.0 (10)	6.0 (5)
(14) GMM		0.705 (0.021)	0.678 (0.038)	0.643 (0.115)
(15) Wald test for equality of β's (d.f.)		155.0 (23)	47.0 (18)	61.0 (14)
			MA(1) vs. MA(0)	MA(2) vs. MA(1)
(16) Hausman test (using GMM)		Stationarity 5.7	0.7	0.1

[a]The MA(0) column uses all adjacent non-coinciding x levels as instruments. There are four such instruments per equation. The MA(1) and MA(2) columns use only those instruments which are one or two periods away, respectively.

In line 12 all eleven estimates are pooled using the estimated 2SLS residuals allowing for conditional heteroskedasticity. Line 14 is based on the pooling of the twenty-four individual (independent) orthogonality conditions [nineteen for MA(1), and fifteen for MA(2)]. The stationary MA(0) estimate alluded to in line 16, column 2, is computed by adding five additional instruments of the $x_t + x_{t+1}$ form. The resulting estimate is 0.66 (0.02).

identified, and in all cases the estimated β rises or remains the same, though the precision with which they are estimated declines precipitously. The restricted estimate rises to 0.92, but the GMM estimator falls to 0.64, and is not significantly different from the MA(1) one. The estimated coefficients are quite unstable at this point. A GMM estimator using a subset of five instruments yields $\beta = 1.2$ (0.2), implying that the relationship may not be stable over time (as is already indicated by the high χ^2 value for the pooling tests).

We can also test the hypothesis of stationarity of the error mechanism by using additional instruments to estimate the various models. In the MA(0) case it is rejected using the Hausman test (see line 16, column 1 of table 4). It is rather cumbersome, however, within the GMM framework to test similar hypotheses in the MA(1) and MA(2) context. Moreover, the Hausman tests may not be very powerful against such a hypothesis since they focus on the resulting changes in β rather than on the estimation of the correlation structure of the errors, which is of direct interest in this context. Maximum likelihood procedures are more convenient for this purpose since they yield explicit estimates of the various error variances and covariances.

Maximum likelihood estimates of this same range of models are given in table 5. The model given in (23) is estimated using the LISREL-V program. It differs from the GMM estimators by assuming joint normality for all of the variables and by not allowing for conditional heteroskedasticity explicitly. The latter fact explains why the computed standard errors are somewhat lower in this table. We show only the estimated β and the associated standard errors and χ^2 statistics. Because the 75–74 difference is an outlier in many respects, the residual from the fitted model being twice as large as for the other years and exceeding four times its estimated standard error, we present also results based only on using four differences (but all six x levels) in the bottom half of the table. Except for the final size of β, the results are similar to the system estimates in table 4. Here the stationarity of the v's is rejected resoundingly and so is also the no-correlation assumption. If 75–74 is eliminated from the system, the fit improves significantly and the MA(1) assumption appears to be adequate.[14] As we relax the various restrictions, the estimated β's rise towards one and beyond, though the precision with which they are estimated falls concomitantly.

Our empirical results are not easily summarized. First, we 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, in the predicted direction. Next, 2SLS and system-IV estimators reduce the magnitude of the bias. Correlation in measurement error seems present, although an MA(1) process seems adequate for our particular data set.[15]

[14] The estimated first-order correlation of the errors is positive and about 0.3.

[15] A MA(1) 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.

Table 5

Maximum likelihood estimates of β in the employment–output relationship coefficients (standard errors) and model χ^2's.[a]

	Non-stationary v's β	Difference in χ^2's	Stationary v's β	Difference in χ^2's
(1) Full data 5 differences				
MA(0)	0.701		0.689	
	(0.015)		(0.014)	
	402		440	38 (5)
MA(1)	0.780		0.763	
	(0.028)		(0.020)	
	324	78 (5)	414	26 (1)
MA(2)	1.355		0.715	
	(0.220)		(0.022)	
	304	20 (4)	401	13 (1)
(2) Without 75–74				
MA(0)	0.789		0.739	
	(0.024)		(0.018)	
	199		261	62 (5)
MA(1)	1.274		0.906	
	(0.157)		(0.032)	
	123	76 (5)	215	46 (1)
MA(2)	1.356		0.885	
	(0.200)		(0.034)	
	117	6 (4)	213	2 (1)

[a] The numbers in the main columns are first the estimated coefficient β, its standard error (in parentheses) and the χ^2 statistic for the estimated model as a whole. The numbers in columns 2, 3 and 4 are the respective differences in the χ^2 and the associated degrees of freedom.

Lastly, we return to the puzzle: Is β really less than one? Apart from 1975–1974, the system-MA(1) results and the maximum likelihood estimates indicate that $\beta \simeq 1$, while the GMM estimates put it closer to 0.7, but the precision of these estimates is not very high. The following might be a possible interpretation. Approximately 0.006 of the variance of log of output is unanticipated.[16] This is less than 0.3% of the total variance in log output, but it accounts for close to 50% in the variance of its first differences.[17] Allowing for such errors raises the estimated β from about 0.5 to about 0.9, for most years,

[16] And unadjusted to as far as labor input is concerned. This same fluctuation may not be an 'error' as far as more flexible inputs such as materials or hours per man are concerned. In similar computations using data on French firms, Jacques Mairesse found that while the employment–output relationship behaves very similarly to what has been reported here, the materials–output relationship yields coefficients of 1 also in the within, first difference, and long difference versions. As far as material purchases are concerned, such fluctuations are not 'errors'!

[17] For the MA(1) version, the numbers are approximately 0.0095 for the variance and 0.0035 for the covariance of such errors. In the non-stationary case, these numbers vary significantly from year to year.

leaving us rather close to the expected unitary elasticity. A related interpretation of these results arises from the distinction between variable and overhead labor. If 'overhead' labor does not vary much over the horizon and range of our data and the size of shocks that we observe, our estimates would imply that it accounts for about 10 percent of manufacturing employment. Our results are also consistent with Sims (1974), whose final estimate of the 'total' β was about 0.8 for a specification which used the number of workers rather than manhours as the dependent variable.

5. A suggested research strategy

The general approach that we suggest can be summarized as follows:

(i) Estimate eq. (11) 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 the hypothesis in (i) is rejected, calculate some differenced estimates (of different lengths) by OLS. If they differ significantly, errors in measurement may well be present.

(iii) Estimate the equations in table 2 or their equivalent by Instrumental Variables, Maximum Likelihood, or the Generalized Method of Moments. Then do a specification test(s) of the no-correlation assumption in the errors in measurement. If they do differ significantly, the specification of a correlated errors in measurement process, use of outside instruments, or respecification of the original model (11) seems to be called for.

6. Notes on the literature

For work on panel data, see Maddala (1971), Mundlak (1978), Hausman (1978), Hausman and Taylor (1981) and Chamberlain (1982). For the importance of errors in such contexts, see Griliches (1979, 1984) and Freeman (1984). For an earlier effort at identifying the error variance from the contrast between levels and first differences in a single series, see Karni and Weissman (1974). For a related attempt to derive consistent estimators in error-ridden panel data in the context of zero–one variables, see Chowdhury and Nickell (1985). For estimation methods in such contexts, see Aigner et al. (1984) and Hansen (1982).

Appendix on optimal instruments

by Z. Griliches, J. Hausman and Bruce Meyer

Consider the model (11):

$$y_{it} = z_{it}\beta + \alpha_i + \eta_{it}, \tag{A.1}$$

where α_i is potentially correlated with z_{it},

$$x_{it} = z_{it} + v_{it}, \tag{A.2}$$

$$y_{it} = x_{it}\beta + \alpha_i + \eta_{it} - v_{it}\beta, \qquad i = 1,\dots,N, \quad t = 1,\dots,T. \tag{A.3}$$

Stacking the observations for a given i we have

$$\begin{aligned} y_i &= x_i\beta + l\alpha_i + \eta_i - v_i\beta \\ &= x_i\beta + l\alpha_i + \varepsilon_i, \qquad i = 1,\dots,N, \end{aligned} \tag{A.4}$$

where y_i, x_i, η_i, v_i and ε_i are all T-dimensional column vectors and l is the vector of 1's. Also let $\mathrm{var}(\eta_i) = \Sigma$ and $\mathrm{var}(v_i) = \Omega$, $i = 1,\dots, N$. The model can be stacked once more to obtain

$$y = x\beta + \alpha + \eta - v\beta, \tag{A.5}$$

where y, x, α, η and v are $NT \times 1$ column vectors.

This appendix examines instrumental variables estimators of the form $\beta_{\mathrm{IV}} = (w'x)^{-1}w'y$, where $w = (I_N \otimes P)x$ for some $T \times T$ matrix P. Note that the various difference estimators as well as the within estimator can be written in this form. For w to be a valid instrument, P must satisfy three requirements. One needs

$$l'P = 0, \tag{A.6}$$

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i'P\varepsilon_i = 0, \tag{A.7}$$

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i'Px_i \neq 0. \tag{A.8}$$

An efficient estimate of β will use all non-redundant instruments w, satisfying (A.6)–(A.8). By non-redundant we mean that each instrument contains some information about x not included in the other instruments. I.e., in the multiple regression of x on the w's, the coefficients on the w's should all be identified and asymptotically non-zero. In general, this is a stronger condition than the w's being linearly independent. However, here the two conditions will be the same. After obtaining a complete set of instruments, the efficient β is calculated as a weighted average of the β's from each w. The inverse of the variance–covariance matrix of the β's is the appropriate weighting matrix.

The number of non-redundant instruments equals the number of linearly independent P matrices satisfying (A.6)–(A.8). In most cases (A.8) is not binding. In such cases the number of linearly independent P matrices equals T^2 minus the number of unique linear restrictions (A.6) and (A.7) impose on P.

Requirement (A.6),

$$l'P = 0,$$

assures the elimination of α, the fixed effects, by having the columns of P sum to zero. This imposes T linear restrictions. Requirement (A.7),

$$\operatorname{plim} \frac{1}{N} \sum_{i=1}^{N} x_i' P \varepsilon_i = 0,$$

is the usual requirement that an instrument be uncorrelated with the composite error term. Rewriting (A.7), we have

$$\operatorname{plim} \frac{1}{N} \sum_{i=1}^{N} (z_i + v_i)' P (\eta_i - v_i \beta)$$

or

$$\operatorname{plim} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{\tau=1}^{T} v_{it} v_{i\tau} P_{t\tau} = 0.$$

The implied restrictions on P depend on Ω, the covariance of v_i. For example, if the v_{it} are not stationary or serially correlated, then $p_{t\tau}$ must equal zero whenever $t = \tau$. This imposes T additional restrictions on P. If the v_{it} are stationary, then only the sum of the diagonal elements of P must equal zero. This imposes one linear restriction. Note that the additional restriction MA($T-1$) imposes vis-a-vis MA($T-2$) is redundant.

In certain circumstances (A.6) and (A.7) will imply that (A.8) is not satisfied. This occurs when T^2 minus the number of restrictions implied by (A.6) and (A.7) is less than or equal to $T(T-1)/2 - (T-1)$. $T(T-1)/2 - (T-1)$ is always the dimension of the space of P matrices satisfying (A.6) and (A.7), but *not* satisfying (A.8). To see this, note that when (A.8) is not satisfied, P is of the form

$$\begin{bmatrix} 0 & -A \\ A & 0 \end{bmatrix},$$

where A is any triangular array with $T(T-1)/2$ elements. (A.6) imposes $T-1$ independent linear restrictions on A or P, since one of the restrictions is redundant. The restrictions of (A.7) always hold when P has the above form.

Table 6 gives the number of non-redundant instruments. This equals also the number of P matrices satisfying (A.6) and (A.7) unless (A.8) is violated. Then

Table 6

The optimal number of instruments.[a]

(A)	*$T = 6$ case*	
1.	Stationarity and no correlation	29
2.	No stationarity and no correlation	24
3.	Stationarity, MA(1)	28
4.	Stationarity, MA(2)	27
5.	Stationarity, MA(3)	26
6.	Stationarity, MA(4)	25
7.	Stationarity, MA(K), $K > 4$	25
8.	No stationarity, MA(1)	19
9.	No stationarity, MA(2)	15
10.	No stationarity, MA(3)	12
11.	No stationarity, MA(K), $K \geq 4$	0[b]
(B)	*Arbitrary T case*	
1.	Stationarity and no correlation	$T^2 - (T+1)$
2.	No stationarity and no correlation	$T^2 - (2T)$
3.	Stationarity and MA(K), $K < T - 2$	$T^2 - (T + K + 1)$
4.	Stationarity and MA(K), $K \geq T - 2$	$T^2 - (T + T - 1)$
5.	No stationarity and MA(K)[c]	$T^2 - [2T + (T-1) + \cdots + (T-K)]$

[a] The number of P matrices satisfying (A.6)–(A.7) and the number of valid instruments is the same as long as $[T - [2T + (T-1) + \cdots + (T-K)]] > (T-1)(T/2 - 1)$.

[b] The number of P matrices satisfying (A.6)–(A.7) is ten here, even though there are no valid instruments.

[c] 0 if $[T - [2T + (T-1) + \cdots + (T-K)]] \leq (T-1)(T/2-1)$. The number of P matrices is still $T^2 - [2T + (T-1) + \cdots + (T-K)]$.

no valid instruments are available. Both the $T = 6$ case and the arbitrary T case are presented.

In our framework, it is fairly easy to construct an optimal set of instruments under any assumptions about v. One lists linearly independent P matrices satisfying (A.6)–(A.8). Since a basis for this space can be written using P matrices with entries of 1, -1 and 0, it is fairly easy to check linear independence. As a given basis of a vector space is not unique, the set of instruments will not be either.

References

Aigner, D., C. Hsiao, A. Kapteyn and T. Wansbeek, 1984, Latent variable models in econometrics, in: Z. Griliches and M. Intriligator, eds., Handbook of econometrics, Vol. 2 (North-Holland, Amsterdam) 1321–1393.

Bhargava, A. and D. Sargan, 1983, Estimating dynamic random effects models from panel data covering short time periods, Econometrica 51, no. 6.

Bielby, W.T., R.M. Hauser and K.K. Featherman, 1977, Response errors of non-black males in models of the stratification process, in: A. Goldberger and D. Aigner, eds., Latent variables in socioeconomic models (North-Holland, Amsterdam) 227–251.

Brechling, F.P.R., 1965, The relationship between output and employment in British manufacturing industries, Review of Economic Studies 32, no. 3.

Chamberlain, G., 1982, Multivariate regression models for panel data, Journal of Econometrics 18, no. 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, no. 2.

Chowdhury, G. and S. Nickell, 1985, Hourly earnings in the U.S.: Another look at unionization, schooling, sickness and unemployment using PSID data, Journal of Labor Economics 3, 38–69.

Cummins, C., B.H. Hall and E.S. Laderman, 1982, The R&D master file: Documentation, Unpublished (Harvard University and NBER, Cambridge, MA).

Fay, J.A. and J.L. Medoff, 1985, Labor and output over the business cycle: Some direct evidence, American Economic Review 76, no. 4.

Freeman, R.B., 1984, Longitudinal analyses of the effects of trade unions, Journal of Labor Economics 2, 1–26.

Friedman, M., 1957, A theory of the consumption function, NBER general series 63 (NBER, New York).

Griliches, Z., 1979, Sibling models and data in economics: Beginnings of a survey, Journal of Political Economy 78, no. 5, part 2.

Griliches, Z., 1984, Data problems in econometrics, in: Z. Griliches and M. Intriligator, eds., Handbook of econometrics, Vol. 3 (North-Holland, Amsterdam) forthcoming.

Hall, B. H., 1979, MOMENTS: The moment matrix processor user manual (Stanford, CA).

Hansen, L., 1982, Large sample properties of generalized methods of moments, Econometrica 50, no. 4.

Hausman, J., 1978, Specification tests in econometrics, Econometrica 46, no. 6.

Hausman, J. and W. Taylor, 1981, Panel data and unobservable individual effects, Econometrica 49, no. 6.

Jöreskog, K.G. and D. Sörbom, 1981, LISRELV: Analysis of linear structural relationships by maximum likelihood and least squares method (National Education Resources, Chicago, IL).

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 Association 69.

Maddala, G. S., 1971, The use of variance components models in pooling cross-section and time series data, Econometrica 39, no. 2.

Maddala, G. S., 1977, Econometrics (McGraw-Hill, New York).

Mundlak, Y., 1978, On the pooling of time series and cross section data, Econometrica 46, no. 1.

Newey, W., 1983, Specification testing and estimation using a generalized method of moments, Ph.D. thesis (MIT, Cambridge, MA).

Sargan, J. D., 1958, The estimation of economic relationships using instrumental variables, Econometrica 26, no. 3.

Sims, C. A., 1974, Output and labor input in manufacturing, Brookings Papers on Economic Activity no. 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, no. 2.