Econometrica, Vol. 65, No. 5 (September, 1997), 1201–1213

NOTES AND COMMENTS

CONSTRUCTING INSTRUMENTS FOR REGRESSIONS WITH MEASUREMENT ERROR WHEN NO ADDITIONAL DATA ARE AVAILABLE, WITH AN APPLICATION TO PATENTS AND R&D

BY ARTHUR LEWBEL[1]

1. INTRODUCTION

GIVEN A LINEAR REGRESSION MODEL with measurement errors in variables, this paper shows how simple functions of the model data can be used as instruments for two staged least squares (TSLS) estimation, exploiting third moments of the data. These instruments can be used when no other data are available, or they can supplement outside instruments to improve efficiency. The distribution of the errors is not required to be normal (or known), and the method readily extends to regressions containing more than one mismeasured regressor.

These results build on earlier work proposing functions of mismeasured variables as instruments, e.g., Wald (1940), Madansky (1959), and Durbin (1954) (problems with these earlier methods are documented by Pakes (1982) and Aigner, Hsiao, Kapteyn, and Wansbeek (1984)); see also Fuller (1987). There is a separate long literature on the use of higher order moments of observed data to estimate the coefficients in linear regression models with measurement error. See, e.g., Geary (1943), Scott (1950), Reiersøl (1950), Drion (1951), Durbin (1954), Kendall and Stuart (1979), Pal (1980), Kapteyn and Wansbeek (1983), and Ahn and Schmidt (1995). Recent work that combines these two approaches includes Dagenais and Dagenais (1994, 1995), and Cragg (1995). The present paper extends these results and provides an empirical application.

The empirical model concerns estimation of the elasticity of patent applications with respect to Research and Development (R&D) expenditures. Simple OLS estimates indicate substantial decreasing returns to scale, but are subject to the usual attenuation bias toward zero in the presence of measurement error. Using a variety of more structural models, most empirical research points to constant returns, i.e., an elasticity close to one (see Griliches (1990) for a survey). The simple moment based TSLS estimator proposed here also yields estimates very close to one, and so seems to work as intended to mitigate the effects of measurement error.

2. THE MODEL AND PROPOSED INSTRUMENTS

Consider the standard linear regression model with measurement error:

$$Y_i = a + b'W_i + cX_i + e_i, \tag{1}$$

$$Z_i = d + X_i + v_i, \tag{2}$$

[1] This research was supported in part by the National Science Foundation, through Grant SBR-9514977. I would like to thank Adam Jaffe for providing the data and Gary Chamberlain, Adam Jaffee, and two anonymous referees for many helpful suggestions. Any errors are my own.

where i from 1 to n indexes observations and $'$ denotes transpose. The parameters a, b, c, and d are constants. W_i and b are J vectors of elements W_{ji} and b_j, while all of the other variables and constants are scalars. The observed data consist of Y_i, W_i, and Z_i for $i = 1, \ldots, n$, while X_i, e_i, and v_i are unobserved. The mean zero model error is e_i, and $d + v_i$ is the measurement error (with d being the mean measurement error so the mean of v_i is zero). The observed Z_i equals the unobserved underlying variable X_i plus the measurement error $d + v_i$. The goal is estimation of b and c, though a can also be estimated if $d = 0$.

Equations (1) and (2) assume only one variable is measured with error. The simple extension to more than one mismeasured variable is described later.

Equations (1) and (2) imply that

$$Y_i = \alpha + b'W_i + cZ_i + \varepsilon_i \tag{3}$$

where $\alpha = a - cd$ and $\varepsilon_i = e_i - cv_i$. Estimation of b and c by applying OLS to equation (3) is inconsistent because the error ε_i is correlated with Z_i, since both depend on the measurement error v_i.

A standard cure for this inconsistency is to estimate (3) using TSLS with instruments 1, W_i, and q_i, where q_i is some vector of instruments that are correlated with X_i and uncorrelated with v_i and e_i. The difficulty is that no outside data may be available for use as instruments.

Let $\overline{S}$ denote the sample mean of a variable S_i, and let $G_i = G(W_i)$ for any given function G. This paper shows that TSLS is consistent using

$$q_{1i} = (G_i - \overline{G}), \tag{4a}$$

$$q_{2i} = (G_i - \overline{G})(Z_i - \overline{Z}), \tag{4b}$$

$$q_{3i} = (G_i - \overline{G})(Y_i - \overline{Y}), \tag{4c}$$

$$q_{4i} = (Y_i - \overline{Y})(Z_i - \overline{Z}) \tag{4d}$$

as instruments. These instruments will have correlation with the unobserved X_i that depends on third moments of the joint distribution of X_i, W_i, and G_i. The number of instruments given above is potentially large, especially since each $G(W_i)$ can be any function having finite third own and cross moments. Of course the regressors $G(w) = w$ would be included as instruments, so $G(w)$ should not be linear in w in equation (4a).

The use of these instruments does not depend on the errors having any specific distribution; for example, normality is not required or assumed.

For the case of models having only one regressor ($b = 0$), some of these TSLS estimators using only one instrument, e.g., equation (4d), will be equivalent to the method of moments based estimator proposed by Geary (1943) and Pal (1980).

In addition to the above instruments, the variables

$$q_{5i} = (Z_i - \overline{Z})^2, \tag{4e}$$

$$q_{6i} = (Y_i - \overline{Y})^2 \tag{4f}$$

can also be used as instruments if the measurement error and model error, respectively, are symmetrically distributed (see Dagenais and Dagenais (1995)).

If other instruments are available, then the constructed instruments above can be added to the list of outside instruments in a TSLS regression, thereby using the third moment information to improve estimation efficiency.

3. TWO STAGED LEAST SQUARES

For any variable, say S_i, let $\bar{S}$ denote the sample mean $\bar{S} = \Sigma_{i=1}^n S_i/n$; let the variable in small letters denote deviation from the sample mean, so $s_i = S_i - \bar{S}$. Formally we should write $\bar{S}_n$ and s_{ni}, but the dependence on sample size n is obvious and so the n is dropped for clarity. For any double array f_{ni}, say, let $E(f)$ denote $\text{plim}_{n\to\infty} n^{-1}\Sigma_{i=1}^n f_{ni}$ so, e.g., $E(S) = \text{plim}_{n\to\infty} n^{-1}\Sigma_{i=1}^n S_i$ and $E(s^2) = \text{plim}_{n\to\infty} n^{-1}\Sigma_{i=1}^n[S_i - n^{-1}\Sigma_{j=1}^n S_j]^2$.

If each S_i is a draw from the same distribution, then $E(S)$ and $E(s^2)$ will equal the mean and variance of that distribution, though only limits in the above weaker sense are required. In particular, observations are not constrained to be independently or identically distributed.

ASSUMPTION 1: *$E[(1, W', X)'(1, W', X)]$ exists and is nonsingular.*

Consider TSLS estimation of equation (3), using 1, W_i, and q_i as instruments, where q_i is some vector of variables q_{ki}. Given Assumption 1 and $E(v) = E(vX) = E(vW) = 0$ (the latter follows from Assumption 2 below), this TSLS estimation will be consistent if

$$(5) \qquad E(qe) = 0, \qquad E(qv) = 0, \qquad \text{and} \quad E(q\tilde{z}) \neq 0,$$

where $\tilde{z}_i$ is the residual of the projection of Z on 1 and W. Note that $\tilde{z}_i = \tilde{x}_i + v_i$ where $\tilde{x}_i$ is the residual of the projection of X on 1 and W. The condition $E(q\tilde{z}) \neq 0$ means that at least one element of $E(q\tilde{z})$ is nonzero.

The instruments proposed in equation (4) are each of the form

$$(6) \qquad q_{ki} = g_i^M z_i^L y_i^K = g_i^M (x_i + v_i)^L (b'w_i + cx_i + e_i)^K$$
$$= \sum_{\lambda=0}^{L}\sum_{\kappa=0}^{K}\binom{L}{\lambda}\binom{K}{\kappa} g_i^M (b'w_i + cx)^\kappa x_i^\lambda v_i^{I-\lambda} e_i^{K-\kappa}$$
$$= \sum_{\lambda=0}^{L}\sum_{\kappa=0}^{K}\sum_{\iota=0}^{\kappa}\binom{L}{\lambda}\binom{K}{\kappa}\binom{\kappa}{\iota} c^{\kappa-\iota}(b'w_i)^\iota g_i^M x_i^{\lambda+\kappa-\iota} v_i^{L-\lambda} e_i^{K-\kappa}$$

for M equal to zero or one and L and K each equal to zero, one, or two. For a given q_{ki}, it follows from (6) that the conditions (5) will hold if

$$(7a) \qquad E(w_{1i}^{\iota 1} w_{2i}^{\iota 2} \cdots w_{Ji}^{\iota J} g_i^M x_i^{\lambda+\kappa-\iota} v_i^{L-\lambda} e_i^{K-\kappa+1}) = 0,$$

$$(7b) \qquad E(w_{1i}^{\iota 1} w_{2i}^{\iota 2} \cdots w_{Ji}^{\iota J} g_i^M x_i^{\lambda+\kappa-\iota} v_i^{L-\lambda+1} e_i^{K-\kappa}) = 0,$$

$$(7c) \qquad \sum_{\lambda=0}^{L}\sum_{\kappa=0}^{K}\binom{L}{\lambda}\binom{K}{\kappa} E[g_i^M (b'w_i + cx)^\kappa x_i^\lambda \tilde{x} v_i^{L-\lambda} e_i^{K-\kappa}] \neq 0,$$

for the range of indices defined by the sums in (6), where each ιj is a nonnegative integer and $\Sigma_{j=1}^J \iota j = \iota$. For each instrument q_{ki} in (4), the following assumptions are sufficient to make (7) hold.

ASSUMPTION 2: *$E(e) = E(v) = 0$. $E(w_1^{\iota 1} w_2^{\iota 2} \cdots w_J^{\iota J} g^M x^\psi v^\lambda e^\kappa) = E(w_1^{\iota 1} w_2^{\iota 2} \cdots w_J^{\iota J} g^M x^\psi)$ $E(v^\lambda)E(e^\kappa)$ and exists, for $M \in \{0,1\}$, $\iota \in \{0,1,2\}$, $\psi \in \{0,1,2)$, $\kappa \in \{0,1\}$ and $\lambda \in \{0,1\}$, where each ιj is a nonnegative integer and $\Sigma_{j=1}^J \iota j = \iota$.*

ASSUMPTION 3: *If q_i contains $g_i z_i$, $y_i z_i$, or z_i^2 (equations 4b, 4d, or 4e), then Assumption 2 also holds with $\lambda = 2$. If q_i contains $g_i y_i$, $y_i z_i$, or y_i^2 (4c, 4d, or 4f), then Assumption 2 also holds with $\kappa = 2$. If q_i contains z_i^2 (4e), then $E(v^3) = 0$. If q_i contains y_i^2 (4f), then $E(e^3) = 0$.*

ASSUMPTION 4: *Either q_i contains g_i (4a) and $E(g\tilde{x}) \neq 0$, or q_i contains $g_i z_i$ (4b) and $E(g_j x\tilde{x}) \neq 0$, or q_i contains $g_i y_i$ (4c) and $b'E(wg\tilde{x}) + cE(xg\tilde{x}) \neq 0$, or q_i contains $y_i z_i$ (4d) and $b'E(wx\tilde{x}) + cE(x^2\tilde{x}) \neq 0$, or q_i contains z_i^2 (4e) and $E(x^2\tilde{x}) \neq 0$, or q_i contains y_i^2 (4f) and $E[(b'w + cx)^2\tilde{x}] \neq 0$.*

To see how these assumptions work, for the instrument $q_{ki} = z_i^2$ (equation 4e), $M = K = 0$ and $L = 2$, so (7) reduces to $E(x_i^\lambda v_i^{2-\lambda} e_i) = 0$ and $E(x_i^\lambda v_i^{2-\lambda+1}) = 0$ for $\lambda = 0$, 1, and 2, and $\sum_{\lambda=0}^{L} \binom{L}{\lambda} E(x_i^\lambda \tilde{x} v_i^{L-\lambda}) \neq 0$. The assumptions are then applied to check that they make these conditions hold. The other instruments are analyzed in exactly the same way, resulting in the following theorem:

THEOREM 1: *Let q_i equal a vector of one or more elements of the form q_{ki} in equation (4) for $k \in \{1, 2, \ldots, 6\}$. Let Assumptions 1, 2, 3, and 4 and equations (1) and (2) hold. Then TSLS estimation of equation (3) using 1, W_i, and q_i as instruments yields consistent estimates of α, b, and c.*

Standard limiting distribution theory for TSLS can now be applied here.

4. PRECISION OF ESTIMATES

The proposed instruments will be most useful in large cross section data sets, both because estimates based on higher moments can be erratic in small samples (see, e.g., Aigner, Hsiao, Kapteyn, and Wansbeek (1984), Dagenais and Dagenais (1995), and footnote 3 of Hausman, Newey, and Powell (1995)), and because with time series data other instruments can often be found, e.g., lagged regressors.

For the most part, identification here comes from skewness in x or $\tilde{x}$. The greater the skewness, the better the quality of the proposed instruments. $E(z^3) = E(x^3) + E(v^3)$, so unless measurement error is substantially skewed in the opposite direction from x, the proposed instruments will generally work best when the sample distribution of the observed Z is strongly skewed.

To assess the quality of the estimator, consider the simple model $y_i = bx_i + e_i$ and $z_i = x_i + v_i$ with normal errors and X_i lognormally distributed (with $E(X) = E(x^2) = 1$) to give x_i the necessary skewness. By Gibrat's law, many economic variables have approximately log normal cross section distributions, e.g., income and firm size. Exact formulas for the asymptotic bias and variance of the estimated coefficient $\hat{b}$ as functions of b, σ_e^2, and σ_v^2 are reported in Table I.

For comparison, suppose we had some "real" instrument, i.e., some variable other than a function of y_i or z_i that is correlated with x_i and uncorrelated with v_i and e_i. The best possible real instrument would be x_i itself. Table I shows that the asymptotic

TABLE I

RELATIVE ESTIMATOR PRECISION IN A SAMPLE MODEL

			$b=\sigma_e^2=\sigma_v^2=1$		
Instrument	Variance	Bias	Bias	Std. dev.	MSE
$q_i=z_iy_i$	$(41+\sigma_v^2+\sigma_e^2/b+\sigma_v^2\sigma_e^2/b)\sigma_e^2/(16n)$	0	0	$1.66/\sqrt{n}$	$2.75/n$
$q_i=z_i^2$	$(41+6\sigma_v^2+3\sigma_v^4)\sigma_e^2/(16n)$	0	0	$1.77/\sqrt{n}$	$3.13/n$
$q_i=y_i^2$	$(41+6\sigma_e^2/b+3\sigma_e^4/b^2)\sigma_e^2/(16n)$	0	0	$1.77/\sqrt{n}$	$3.13/n$
$q_i=x_i$ [a]	σ_e^2/n	0	0	$1/\sqrt{n}$	$1/n$
$q_i=z_i$ [b]	$\sigma_e^2/[(1+\sigma_v^2)n]$	$b\sigma_v^2/(1+\sigma_v^2)$	.5	$.25/\sqrt{n}$	$.25+.5/n$

Notes: Reported are the asymptotic variance, bias, mean squared error (MSE), and standard deviation of $\hat{b}=(\Sigma_{i=1}^n q_iy_i)/(\Sigma_{i=1}^n q_iz_i)$, using different instruments q_i. The model is $y_i=bx_i+e_i$, $z_i=x_i+v_i$, $x_i=X_i-1$ where X_i has mean and variance equal to one and is lognormally distributed, and e_i and v_i are independent mean zero normals with variances σ_e^2 and σ_v^2.

[a] $q_i=x_i$ is the hypothetically best possible instrument, which would not be available in practice, and hence bounds any feasible estimator.

[b] $q_i=z_i$ corresponds to estimating the model using OLS instead of TSLS.

standard error for $\hat{b}$ using any of the available higher moment instruments in equation (4) is only about 1.7 times the standard error that would be obtained if x_i itself were observed and used. This shows that the asymptotic standard error of $\hat{b}$ using the higher moment instruments (4) is at worst 70% larger than would be obtained with any set of "real" instruments. This 70% is slightly reduced or increased when σ_e^2/b or σ_v^2 are less than or greater than $\sigma_x^2=1$, e.g., for $q_i=z_iy_i$ the ratio is 1.63 when $\sigma_e^2/b=\sigma_v^2=.5$, and 1.75 when $\sigma_e^2/b=\sigma_v^2=2$.

Table I also reports the asymptotic bias and variance of using $q_i=z_i$, which corresponds to estimating the model using OLS instead of TSLS. For the case where $b=\sigma_e^2=\sigma_v^2=\sigma_x^2=1$, the higher moment TSLS estimator has a lower tabulated mean squared error than OLS for $n>9$ observations. This means that in this model TSLS will on average (across independent samples) be more accurate than OLS as long as each sample's size n is greater than nine.

5. EXTENSIONS

In addition to equation (4), by suitable generalization of Assumptions 2 and 3 instruments of the form $g_iz_i^Ly_i^M$ for additional integers L and M can be rationalized, though the resulting higher than third moments may need even larger sample sizes to behave well.

The results of the previous section extend with little change to the case in which more than one regressor is measured with error by replacing cx_i with $c'x_i=\Sigma_kc_kx_{ki}$ in equation (1), and equation (2) becomes a vector of equalities. The instruments (4) work as before, replacing Z_i with each Z_{ki}. In addition, cross products $z_{ki}z_{\kappa i}$ for all k and κ can also be used as instruments if $E(v_{ki}^2v_{\kappa i})=E(v_{ki}^2)\,E(v_{\kappa i})$ for all k and κ.

Another extension is the partly linear structural model

$$Y_i=a+\gamma(b,W_i)+cX_i+e_i \tag{8}$$

where γ is some known function of a vector of parameters b and regressors W_i. Assumptions 2 and 3 make $E(qe)=E(qv)=0$ for q_i as defined by equation (4). Since

$a + cX_i = \alpha + cZ_\iota - cv_i$, it follows that

(9) $$E\{[Y - \alpha - \gamma(b, W) - cZ]q\} = 0.$$

The Generalized Method of Moments (GMM) estimator can then be applied to the vector of moment conditions (9) to estimate α, b, and c, provided that the moments are sufficient to identify these parameters. The TSLS in Theorem 1 is a special case of this GMM, where Assumptions 1 to 4 provide identification.

GMM can also be used to exploit fourth and higher moment information. For example, letting w_i be a scalar (the extension to more regressors is notationally cumbersome but conceptually the same) gives, by equation (6),

(10) $$E(g^M z^L y^K) = \sum_{\lambda=0}^{L} \sum_{\kappa=0}^{K} \sum_{\iota=0}^{\kappa} \binom{L}{\lambda} \binom{K}{\kappa} \binom{\kappa}{\iota} c^{\kappa-\iota} b^{\iota} E(w^{\iota} g^M x^{\lambda+\kappa-\iota}) E(v^{L-\lambda}) E(e^{K-\kappa})$$

where M, L, and K are now any nonnegative integers and Assumption 2 is extended to hold for these higher exponents. Let each expectation on the right side of equation (10) be a nuisance parameter, some of which are known, e.g. $E(e^0) = 1$ and $E(e^1) = 0$. Define θ to be the vector of these nuisance parameters, and define the function $h_{MLK}(b, c, \theta)$ as the right side of equation (10). Equation (10) can then be written as the moment condition

(11) $$E[g^M z^L y^K - h_{MLK}(b, c, \theta)] = 0.$$

Each choice of M, L, and K provides another moment condition (11). GMM can then be used to estimate b and c from a sufficient collection of such moment conditions. This method can be used to extend the previous sections' TSLS estimator to incorporate fourth and higher moment information. See also Aigner, Hsiao, Kapteyn, and Wansbeek (1984), Pal (1980), Dagenais and Dagenais (1994, 1995), and Cragg (1995).

As a further generalization, g in equations (10) and (11) can include outside instruments, and y could be a quadratic or higher polynomial in x or w. The estimator in Hausman, Newey, and Powell (1995) is an example.

To give another example, consider the quadratic model $Y_i = a + bX_i + cX_i^2 + e_i$ where $Z_i = X_i + v_i$ and X_i is not observed. If v_i is normal with mean zero and variance σ_v^2 (or at least has the first four moments of a normal distribution), then even this quadratic model can be estimated without outside instruments. Exactly analogous to equation (11) we can write $E[Z^L Y^K - h_{LK}(a, b, c, \theta)] = 0$. Doing so for $L = 1$ to 5 with $K = 0$ and for $L = 0$ to 3 with $K = 1$ provides nine moment conditions for estimating the nine unknowns a, b, c and the nuisance parameters σ_v^2 and $E(X^K)$ for $K = 1$ to 5. If other not mismeasured regressors W_i are included in the model, then equation (11) moments provide overidentifying information, and again GMM could be applied.

6. AN APPLICATION TO PATENT DATA

The literature on sources of technological growth analyzes patents (a measure of research output) as a function of expenditures on R&D (research and development). See, e.g., Scherer (1965), Pakes and Griliches (1984), Bound et al. (1984), Hausman, Hall, and Griliches (1984), and Hall, Griliches, and Hausman (1986).

Bound et al. (1984) describe a long term NBER project of constructing a data set that matches numbers of patent applications by firm (obtained from the US patent office) with R&D expenditures by firm (from Standard and Poor's Compustat Annual Industrial Files). The current version of this data set which is analyzed here contains the number of patent applications and the R&D expenditures annually for up to 10 years (1970 to 1979) for each of the over 1000 publicly traded US firms in the Compustat data base.

Much of the research based on this and related data sets concerns the question of returns to scale in R&D. In addition to the above listed empirical analyses, see Fisher and Temin (1973), Baldwin and Scott (1987), Cohen and Levin (1989), Caballero and Jaffe (1993), and Griliches (1994).

Section 4 of Griliches' (1990) survey on patent statistics concludes that while most empirical estimates show decreasing returns, this result is likely due to data problems (including sample selection and measurement errors), with the truth being close to constant returns for most firms. For example, with a small subset of the same data used here, Bound et. al. (1984) regressed the log of patents on the log of R&D, obtaining a coefficient of .38. They then employed a variety of corrections including sample splitting, nonlinear models, and Poisson models, yielding elasticities close to one for a majority of firms in the sample.

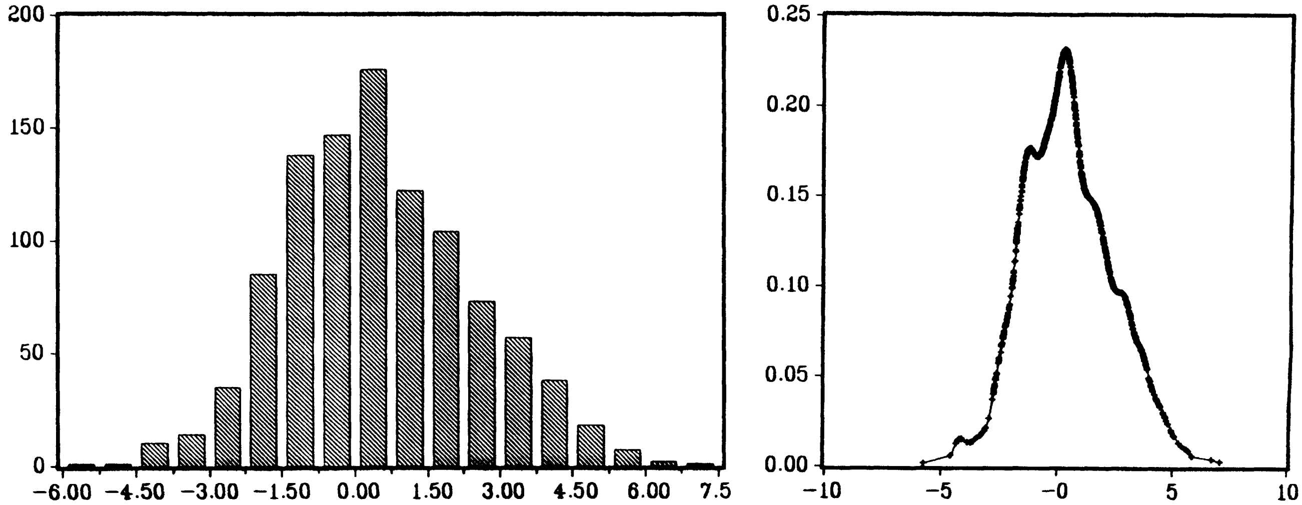
Since measurement error is likely to be a substantial source of bias in the basic OLS regression model, this paper examines the extent to which TSLS using the simple moment based instruments proposed here can reduce the bias in the OLS estimates. This application is likely to be a good one because few if any outside instruments are available, the distribution of R&D expenditures across firms is very skewed, the sample size is moderately large, and the magnitude of measurement error is likely to be substantial (which makes correcting it a priority).

Let Y be the log of the average annual number of patents for each firm, and let Z be the log of each firm's measured average annual real R&D expenditures over a ten year period. Figure 1 shows the required skewness in Z.

Model 1 in Tables II and III is a linear OLS and TSLS regression using the data shown in Figures 1 and 2. A substantial number of the firms have no R&D expenditures, no patents, or both in the observed time periods. Model 1 drops these observations from the sample. Model 2 leaves them in, assigning a value of zero for y when patents equal zero, and similarly for z when R&D is zero. The model then adds dummy variable regressors for these observations. This treatment for zeros (and the choice of other regressors considered in models 3, 4, and 5) are the same as in Bound et al. (1984).

The basic result, as shown in Tables II and III, is that OLS coefficient (elasticity) estimates are all significantly smaller than one, while the TSLS estimates are all very close to one. This result holds up regardless of the inclusion or exclusion of other regressors, and for different cuts of the data including five year instead of ten year averages (models 7 and 8) or using annual panel data (models 9, 10, and 11). As is common in models with measurement error, the OLS fixed effects estimator (model 11) shows greater attenuation bias than other estimators. Even in this case, TSLS appears to fix the problem, yielding estimates close to one.

The model assumes that Y is linear in X. This linearity cannot be tested directly because X is unobserved, but if the model is right then Y should be at least close to linear in the observed Z. To check this, Figure 2 shows a nonparametric kernel regression of Y on Z. The observed departure of the kernel regression from linearity for low Z's may be genuine and not just tail noise; e.g., Griliches (1990) includes evidence



Notes: The data are Z, the log of each firm's average annual real expenditures on R&D, where firms having zero R&D expenditures thoughout the entire sample period are dropped. The density estimate uses a normal kernel with bandwidth equal to .25.

FIGURE 1.—Histogram and estimated kernel density of Z.

TABLE II

ALTERNATIVE PATENT—R&D EXPENDITURE MODEL VARIANTS

Model	Data	N	Other Regressors	OLS	TSLS
1.	1970–79 avg	1029[a]		.770 (.017)	1.011 (.056)
2.	1970–79 avg	1239	patdum R&Ddum	.744 (.017)	.988 (.040)
3.	1970–79 avg	1239	patdum R&Ddum SICdums	.671 (.022)	1.044 (.068)
4.	1970–79 avg	1239	patdum R&Ddum plant	.652 (.022)	1.086 (.093)
5.	1970–79 avg	1239	patdum R&Ddum SICdums plant	.585 (.025)	1.179 (.138)
6.	1970–79 avg	1177[b]	patdum R&Ddum	.703 (.021)	1.250 (.096)
7.	1970–74 avg	1120	patdum R&Ddum	.726 (.018)	1.022 (.040)
8.	1975–79 avg	1120	patdum R&Ddum	.709 (.018)	1.022 (.046)
9.	annual 70–79	9849	patdum R&Ddum	.557 (.006)	1.071 (.017)
10.	annual 70–79	6299[a]		.670 (.007)	1.130 (.034)
11.	annual 70–79	6299[a]	firm fixed effects model	.126 (.019)	.965 (.205)

Notes. The models are linear regressions of Y on a constant, on Z, and on other regressors as defined below. Reported are the OLS and TSLS estimated coefficients of Z, with White corrected standard errors in parentheses. The instruments for TSLS are the constant, all included regressors except Z, demeaned Y times demeaned Z, and demeaned plant times demeaned Z in models 4 and 5 Due to the nature of the constructed data, instruments that require symmetry of errors were not used. N is the sample size.

The data are from 1970 to 1979, though data do not exist for all firms in all years. Each observation is a firm (for models 9, 10, and 11, each observation is a firm in a year) The dependent variable Y is the log of the firm's average annual number of patents (over all years for which the firm's data are available) or zero if the average number of patents is zero. Z is the log of the firm's average annual real (millions of 1974 dollars) reported expenditures on R&D, or zero if the expenditures are zero. Patdum equals one when the number of patents is zero. R&Ddum equals one when the R&D expenditures are zero. SICdums are 21 industry dummies (based on firm SIC codes) as defined in the Appendix of Bound et al. (1984) Plant is the log of the firm's expenditures on gross plant and equipment in millions of dollars in 1972. The fixed effects model 11 is equivalent to having a separate dummy for every firm.

[a] In models 1, 10, and 11, any observation of zero R&D expenditures or zero patents is excluded from the sample.

[b] In model 6, firms in the top 5% of R&D expenditures are excluded from the sample.

that the patent R&D relationship may be different for the smallest firms. Still, the relationship in Figure 2 looks close to linear for the bulk of the data.

Table III reports more detailed results for the basic models 1 and 2. As reported there, a Hausman test comparing the OLS and TSLS models strongly rejects the OLS, assuming that the TSLS is consistent.

Let $\beta=(a,b',c)'$ be the vector of coefficients to be estimated, let $\hat{\beta}$ and $\hat{\beta}_{ols}$ denote the TSLS and OLS estimates, respectively, and let $\beta_{ols}=\text{plim } \hat{\beta}_{ols}$. It can be easily shown (based, e.g., on Greene (1993, p. 283)) that $E(v^2)=E(zW)'(\beta-\beta_{ols})/c$. Given Theorem 1, we may therefore estimate the variance of the measurement error $E(v^2)$ as $n^{-1}\sum_{i=1}^{n} z_i W_i'(\hat{\beta}-\hat{\beta}_{ols})/\hat{c}$. In finite samples there is no guarantee that this estimate will be reasonable, e.g., it can easily be negative or be larger than the sample variance of Z. Table II reports the sample variance of Z and this implied estimate of the measurement error variance based on the OLS and TSLS coefficient estimates. In both models the implied measurement error variance is a plausible 23% of the total estimated variance of Z. However, unlike the Z coefficient estimates in Table I, this variance estimate was not robust across the other specifications in Table II, in some cases coming in negative and in others being larger than the sample variance of Z itself. Also, this calculation assumes $d=0$, which may not hold and is not otherwise required for the estimation.

In summary, simple OLS regressions of Y on Z show substantial decreasing returns to scale in patent application output from R&D expenditure input. A variety of more sophisticated modeling methods summarized in, e.g., Griliches (1990), indicate that the true relationship is close to constant returns. Here simple TSLS regressions of Y on Z

TABLE III

BASIC PATENT—R&D EXPENDITURE DATA AND MODEL SUMMARY

Zero Patents and Zero R&D Excluded (Model 1 in Table I):	Y	Z	Constant
Mean	1.182	.506	
Variance	3.410	3.828	
Skewness	.327	.271	
OLS coefs		.770	.792
OLS std errs		(.017)	(.034)
OLS White std errs		(.017)	(.036)
TSLS coefs		1.011	.671
TSLS std errs		(.053)	(.045)
TSLS White std errs		(.056)	(.051)

OLS $R^2 = .666$, TSLS $R^2 = .601$, Hausman test $\chi_4^2 = 23.27$
Implied measurement error variance = .882

Zero Patents and Zero R&D Included (Model 2 in Table I):	Log (Patents)	Log (R&D)	R&Ddum	Patdum	Constant
Mean	.949	.341	.119	.075	
Variance	3.276	3.511	.105	.069	
Skewness	.488	.344	2.344	3.222	
OLS coefs		.744	−1.154	.359	.806
OLS std errs		(.017)	(.097)	(.122)	(.035)
OLS White std errs		(.017)	(.111)	(.111)	(.035)
TSLS coefs		.988	−1.114	.718	.691
TSLS std errs		(.040)	(.105)	(.142)	(.041)
TSLS White std errs		(.040)	(.114)	(.155)	(.044)

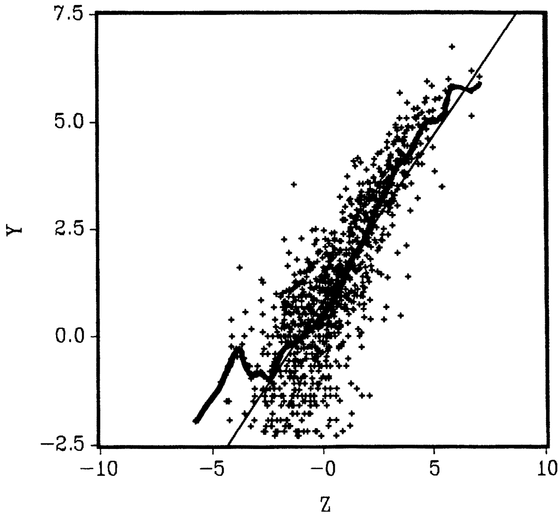
OLS $R^2 = .638$, TSLS $R^2 = .577$, Hausman test $\chi_2^2 = 44.86$
Implied measurement error variance = .806

using moment based instruments also indicate close to constant returns, that are stable across different choices of W and different cuts of the data. The moment based instruments are especially useful in contexts like this one where other more structural instruments are difficult or impossible to obtain.

7. CONCLUDING REMARKS

This paper extends Dagenais and Dagenais (1995) to show that each of the simple functions of regression data given by equation (4) can be used as instruments in TSLS estimation. These instruments can be used for identification and estimation when no other instruments are available, or can be used to augment the list of available instruments for a given model.

The method involves estimates of third moments of the data, which can be very sensitive to outliers. To mitigate this problem TSLS components like $(\sum_{i=1}^{n} q_i z_i/n)$ could be replaced with robust estimates of these moments if desired. Similarly, Fuller's modified IV estimator (Fuller (1987, Chapter 2)) can be used to guarantee existence of the required moments in finite samples.



Notes: Y and Z are the logs of each firm's average annual patents and real R&D expenditures, respectively, where firms having zero patents or zero R&D expenditures thoughout the entire sample period are dropped. The kernel regression (thick line) uses a normal kernel with bandwidth equal to .25.

FIGURE 2.—Scatter plot, OLS, and kernel regression of Y on Z.

The moment based estimator depends on skewness in the data, and so can be sensitive to data transformations. Similarly, the distribution of some variable Z in subsets of the sample (e.g., in the subsamples comprising a Chow test that splits the sample into low and high Z values) can differ substantially from the distribution of Z for the whole sample, and so greatly affect the subsample TSLS estimates.

In some cases, such as when the mismeasured regressors are not strongly skewed or the model has other weakly correlated regressors, the method may yield many relatively weak instruments. When this happens, results concerning the use of TSLS with many weak instruments, such as Staiger and Stock (1993), Angrist (1994), and Kitamura (1994), could be applicable.

Dept. of Economics, Brandeis University, Waltham, MA 02254, U.S.A.

Manuscript received November, 1995; final revision received January, 1997.

REFERENCES

AIGNER, D. J., C. HSIAO, A. KAPTEYN, AND T. WANSBEEK (1984): "Latent Variable Models in Econometrics," in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. D. Intriligator. New York: Elsevier.

AHN, S. C., AND P. SCHMIDT (1995): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5–27.

ANGRIST, J. (1994): "Jackknife and Split Sample Instrumental Variables," Unpublished manuscript, Hebrew University.

BALDWIN, W., AND J. T. SCOTT (1987): *Market Structure and Technological Change*. NY: Harwood Academic Publishers.

BOUND, J., C. CUMMINS, Z. GRILICHES, B. H. HALL, AND A. B. JAFFE (1984): "Who Does R&D and Who Patents?" in *R&D, Patents, and Productivity*, ed. by Z. Griliches. Chicago: University of Chicago Press.

CABALLERO, R. J., AND A. B. JAFFE (1993): "How High are the 'Giants' Shoulders?" *NBER Macroeconomics Annual*, ed. by O. J. Blanchard and S. Fisher. Cambridge: MIT Press, 15–74.

COHEN, W. M., AND R. C. LEVIN (1989): "Empirical Studies of Innovation and Market Structure," in *Handbook of Industrial Organization*, Vol. II, ed. by R. Schmalensee and R. Willig. Amsterdam: North-Holland, 1059–1107.

CRAGG, J. G. (1995): "Using Higher Moments to Estimate the Simple Errors-in-the-Variables Model," Unpublished manuscript.

DAGENAIS, M. G., AND D. L. DAGENAIS (1994): "GMM Estimators for Linear Regression Models with Errors in the Variables," Discussion Paper #0594, Université de Montréal.

——— (1995): "Higher Moment Estimators for Linear Regression Models with Errors in the Variables," *Journal of Econometrics*, forthcoming.

DRION, E. F. (1951): "Estimation of the Parameters of a Straight Line and of the Variances of the Variables, if They are Both Subject to Error," *Indagationes Mathematicae*, 13, 256–260.

DURBIN, J. (1954): "Errors in Variables," *Review of the International Statistics Institute*, 1, 23–32.

FISHER, F. M., AND P. TEMIN (1973): "Returns to Scale in Research and Development: What Does the Schumpeterian Hypothesis Imply?" *Journal of Political Economy*, 87, 386–389.

FULLER, W. A. (1987): *Measurement Error Models*. New York: Wiley.

GEARY, R. C. (1943): "Inherent Relations Between Random Variables," Proceedings of the Royal Irish Academy, Series A, 47, 63–64.

GREENE, W. H. (1993): *Econometric Analysis*, 2nd Ed. Englewood Cliffs: Prentice Hall.

GRILICHES, Z. (1990): "Patent Statistics as Economic Indicators," *Journal of Economic Literature*, 28, 1661–1707.

——— (1994): "Productivity, R&D, and the Data Constraint," *American Economic Review*, 84, 1–23.

HALL, B., Z. GRILICHES, AND J. A. HAUSMAN (1986): "Patents and R&D: Is There a Lag?" *International Economic Review*, 27, 265–283.

HAUSMAN, J. A., B. HALL, AND Z. GRILICHES (1984): "Econometric Models for Count Data With an Application to the Patents—R&D Relationship," *Econometrica*, 52, 909–938.

HAUSMAN, J. A., W. K. NEWEY, AND J. L. POWELL (1995): "Nonlinear Errors in Variables: Estimation of Some Engel Curves," *Journal of Econometrics*, 65, 205–233.

KAPTEYN, A., AND T. J. WANSBEEK (1983): "Identification in the Linear Errors in Variables Model," *Econometrica*, 51, 1847–1849.

KENDALL, M. G., AND A. STUART (1979): *The Advanced Theory of Statistics*, 4th Edition. New York: Macmillan.

KITAMURA, Y. (1994): "Misspecification Tests and Poor Quality Instruments," Unpublished manuscript, University of Minnesota.

MADANSKY, A. (1959): "The Fitting of Straight Lines When Both Variables Are Subject to Error," *Journal of the American Statistical Association*, 54, 173–205.

PAKES, A. (1982): "On the Asymptotic Bias of the Wald-Type Estimators of a Straight Line When Both Variables Are Subject to Error," *International Economic Review*, 23, 491–497.

PAKES, A., AND Z. GRILICHES (1984): "Patents and R&D at the Firm Level: A First Look," in *R&D, Patents, and Productivity*, ed. by Z. Griliches. Chicago: University of Chicago Press.

PAL, M. (1980): "Consistent Moment Estimators of Regression Coefficients in the Presence of Errors in Variables," *Journal of Econometrics*, 14, 349–364.

REIERSØL, O. (1950): "Identifiability of a Linear Relation Between Variables Which Are Subject to Error," *Econometrica*, 18, 375–389.

SCHERER, F. M. (1965): "Firm Size, Market Structure, Opportunity, and the Output of Patented Inventions," *American Economic Review*, 55, 1097–1125.

SCOTT, E. L. (1950): "Note on Consistent Estimates of the Linear Structural Relation Between Two Variables," *Annals of Mathematical Statistics*, 21, 284–288.

STAIGER, D., AND J. STOCK (1993): "Instrumental Variables Estimation with Weak Instruments," Unpublished manuscript, Harvard University.

WALD, A. (1940): "The Fitting of Straight Lines if Both Variables are Subject to Errors," *Annals of Mathematical Statistics*, 11, 284–300.