BLS WORKING PAPERS

U.S. DEPARTMENT OF LABOR
Bureau of Labor Statistics

OFFICE OF PRICES AND LIVING CONDITIONS

Constructing Instruments for Regressions with Measurement Error When No Additional Variables Are Available: Comment

Timothy Erickson, U.S. Bureau of Labor Statistics

Working Paper 321
June 1999

The author thanks Steve Holden, Rob McClelland, Keith Waehrer, and Toni Whited for many valuable comments. The views expressed are those of the author and do not reflect the policies of the U.S. Bureau of Labor Statistics or the views of other staff members.

Abstract

Lewbel (1997) has ingeniously shown that linear instrumental variables estimators for the errors-in-variables model can be constructed using functions of the dependent variable, proxy, and perfectly measured regressors as instruments. He proves consistency for the estimator and then asserts that "standard limiting distribution theory for TSLS can now be applied." In this note I assume that "standard theory" is given by White (1982), the source of the standard errors used by Professor Lewbel in his empirical application. I show that when White's formulas are applied to Lewbel's instruments, they give an inefficient estimator, an incorrect asymptotic covariance matrix, and an inconsistent covariance matrix estimator. These results stem from a subtle violation of the familiar instrumental variable orthogonality condition. Specifically, only one of Lewbel's instruments can be measured from an arbitrary origin and satisfy the orthogonality condition; the remaining instruments satisfy orthogonality only if measured as deviations from their population means. The substitution of sample means therefore generates a nonstandard asymptotic covariance matrix of the type described by Newey and McFadden (1994) in their discussion of "plug-in" estimators. I apply the theory for such estimators to Lewbel's instruments to obtain an efficient estimator, the correct asymptotic covariance matrix, and a consistent covariance matrix estimator.

I. Introduction

Lewbel (1997) has ingeniously shown that linear instrumental variables estimators for the errors-in-variables model can be constructed using functions of the dependent variable, proxy, and perfectly measured regressors as instruments. He proves consistency for the estimator and then asserts that "standard limiting distribution theory for TSLS can now be applied." In this note I assume that "standard theory" is given by White (1982), the source of the standard errors used by Lewbel in his empirical application. I show that when White's formulas are applied to Lewbel's instruments, they give an inefficient estimator, an incorrect asymptotic covariance matrix, and an inconsistent covariance matrix estimator. These results stem from a subtle violation of the familiar instrumental variable orthogonality condition. Specifically, under *i.i.d.* sampling only one of Lewbel's instruments can be measured from an arbitrary origin and satisfy the orthogonality condition; the remaining instruments satisfy orthogonality only if measured as deviations from their population means. The substitution of sample means therefore generates a nonstandard asymptotic covariance matrix of the type described by Newey and McFadden (1994), who show that the covariance matrix of a feasible estimator differs from that of its infeasible counterpart whenever the consistency of the feasible estimator requires consistency of the nuisance parameter estimates. I apply the theory of these authors to Lewbel's instruments to obtain an an efficient estimator, the correct asymptotic covariance matrix, and a consistent covariance matrix estimator.

II. The Lewbel estimator and its asymptotic distribution

Lewbel considers the model

$$\begin{aligned} Y_i &= a + b' W_i + cX_i + e_i \qquad (1) \\ Z_i &= d + X_i + v_i, \qquad (2) \end{aligned}$$

where only (W_i, Y_i, Z_i) are observable, $c \neq 0$, and all variables are scalar except for the $J \times 1$ vectors W_i and b. Equations (1)-(2) imply

$$Y_i = \alpha + b' W_i + cZ_i + \varepsilon_i, \qquad (3)$$

where $\alpha = a - cd$ and

$$\varepsilon_i = e_i - cv_i, \tag{4}$$

the last equality combining with (2) to generate the correlated error and regressor problem associated with classical measurement error. Lewbel shows that it is possible to use nonlinear transformations of the observed variables as instruments. Specifically, if $G_i = G(W_i)$ is a nonlinear function of W_i, and $\bar{G}, \bar{Z}$, and $\bar{Y}$ denote sample means, then under certain conditions

$$\begin{aligned} q_{1i} &= G_i - \bar{G} \\ q_{2i} &= (G_i - \bar{G})(Z_i - \bar{Z}) \\ q_{3i} &= (G_i - \bar{G})(Y_i - \bar{Y}) \end{aligned}$$

are instrumental variables that yield consistent Two Stage Least Squares (TSLS) estimates of α, b, and c. Lewbel notes that these new instruments can be combined with the long-recognized instruments

$$\begin{aligned} q_{4i} &= (Y_i - \bar{Y})(Z_i - \bar{Z}) \\ q_{5i} &= (Z_i - \bar{Z})^2 \\ q_{6i} &= (Y_i - \bar{Y})^2, \end{aligned}$$

the last two being applicable if $E\left(e_i^3\right) = E\left(v_i^3\right) = 0$.

To state the asymptotic distribution for the TSLS estimators, let

$$\theta = \begin{pmatrix} \alpha \\ b \\ c \end{pmatrix} \quad R_i = \begin{pmatrix} 1 \\ W_i \\ Z_i \end{pmatrix} \quad Q_i(m) = \begin{pmatrix} 1 \\ W_i \\ q_i(m) \end{pmatrix}, \tag{5}$$

where $q_i(m)$ is a column vector containing one or more of the following functions of $m = (m_G, m_Z, m_Y)'$:

$$G_i - m_G \tag{6}$$

$$(G_i - m_G)(Z_i - m_Z) \tag{7}$$

$$(G_i - m_G)(Y_i - m_Y) \tag{8}$$

$$(Y_i - m_Y)(Z_i - m_Z) \tag{9}$$

$$(Z_i - m_Z)^2 \tag{10}$$

$$(Y_i - m_Y)^2. \tag{11}$$

In this notation (3) is

$$Y_i = R_i'\theta + \varepsilon_i, \tag{12}$$

and TSLS estimators can be written as

$$\hat{\theta} = \left(\bar{M}' A_n \bar{M}\right)^{-1} \bar{M}' A_n \left(\frac{1}{n}\sum_{i=1}^{n} \hat{Q}_i Y_i\right), \tag{13}$$

where A_n is any appropriately-dimensioned positive definite matrix, $\bar{M} \equiv (1/n)\sum_{i=1}^{n} \hat{Q}_i R_i'$, $\hat{Q}_i \equiv Q_i(\hat{\mu})$, and $\hat{\mu} \equiv \left(\bar{G}, \bar{Z}, \bar{Y}\right)'$.

Assumption 1: *(i) (W_i, X_i, e_i, v_i) is i.i.d. for $i = 1, \ldots, n$, the errors e_i and v_i are independent of each other and of (W_i, X_i) and satisfy $E(e_i) = E(v_i) = E(e_i^3) = E(v_i^3) = 0$, and $(G_i, W_i, X_i, e_i, v_i)$ has finite moments of every order; (ii) A_n converges in probability to a positive definite matrix A.*

Note that part (i) satisfies the weaker assumptions given by Lewbel. Let $\psi_i(m) \equiv ((G_i - m_G), (Z_i - m_Z), (Y_i - m_Y))'$, $\mu \equiv (E(G_i), E(Z_i), E(Y_i))'$, $\psi_i \equiv \psi_i(\mu)$, $Q_i \equiv Q_i(\mu)$, $M \equiv E\left[Q_i R_i'\right]$, and

$$D = E\left(\left.\frac{\partial Q_i(m)\varepsilon_i}{\partial m}\right|_{m=\mu}\right). \tag{14}$$

Proposition 1 *If Assumption 1 holds, then $\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} N\left(0, (M'AM)^{-1}M'A\Omega AM(M'AM)^{-1}\right)$, where $\Omega = \operatorname{var}\left[Q_i\varepsilon_i + D\psi_i\right]$.*[1]

The proof is in the appendix, where one can see that Ω is the covariance matrix of the limiting distribution for $(1/\sqrt{n})\sum_{i=1}^{n} \hat{Q}_i\varepsilon_i$. The above result is therefore "nonstandard" in

[1]The object ψ_i is called the influence function for $\hat{\mu}$, because it gives the influence of the i-th observation on $\hat{\mu}$ up to an $o_p(1)$ remainder term. Newey and McFadden (1994) discuss influence functions and their use in deriving the asymptotic distributions of "plug-in" estimators.

the sense that constructing the instruments as deviations from $\hat{\mu}$ rather than μ has generated the term $D\psi_i$ appearing in Ω. Note that D is generally nonzero if any of (7) through (11) are included in $Q_i(m)$. For example, if (7) is included, then (14) includes the element

$$\begin{aligned} E\left[\left(\frac{\partial(G_i - m_G)(Z_i - m_Z)}{\partial m_G}\bigg|_{m=\mu}\right)\varepsilon_i\right] &= E\left[-\left(Z_i - E(Z_i)\right)\varepsilon_i\right] \\ &= E\left[-\left(X_i - E(X_i) + v_i\right)\left(e_i - cv_i\right)\right] \\ &= cE\left(v_i^2\right), \end{aligned}$$

where the second equality is obtained by substituting from (2) and (4), and the last equality follows from the assumed independence of e_i, v_i, and (W_i, X_i). The reason D is nonzero is that the instrumental variable orthogonality conditions associated with (7) through (11) are satisfied if and only if $m = \mu$. To see this for the above example, note that $E\left[(G_i - m_G)(Z_i - E(Z_i))\varepsilon_i\right]$ can be rewritten as $-cE(G_i - m_G)E(v_i^2)$, which vanishes if and only if $m_G = E(G_i)$. In contrast, (6) does not generate nonzero elements in D because the orthogonality condition $E\left[(G_i - m_G)\varepsilon_i\right] = 0$ holds for all m_G. This situation exemplifies Newey and McFadden's comment that the covariance matrix of a feasible estimator requires adjustment whenever consistency of the feasible estimator depends on consistency of the "plugged-in" estimate.

Users of standard theory will estimate Ω with White's formula $\hat{\Omega}_{STD} = (1/n)\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \hat{Q}_i \hat{Q}_i'$, where $\hat{\varepsilon}_i \equiv Y_i - R_i'\hat{\theta}$. The following result shows that $\hat{\Omega}_{STD}$ is not consistent for Ω:

Proposition 2 *If Assumption 1 holds, then* $\hat{\Omega}_{STD} \xrightarrow{p} \operatorname{var}\left[Q_i\varepsilon_i\right]$.

Using $\hat{\Omega}_{STD}$ to estimate Ω therefore produces asymptotically invalid inferences. Also, since an efficient estimator of θ is obtained by choosing A_n to be a consistent estimator of Ω^{-1} (see Newey (1994), p. 1368), it follows that $A_n = \hat{\Omega}_{STD}^{-1}$ gives an inefficient estimator of θ. Valid inferences and efficient estimation are made possible by the following result, where $\hat{\psi}_i \equiv \psi_i(\hat{\mu})$ and

$$\hat{D} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial Q_i(m)\hat{\varepsilon}_i}{\partial m}\bigg|_{m=\hat{\mu}}\right).$$

Proposition 3 *If Assumption 1 holds, then* $\hat{\Omega} \equiv n^{-1} \sum_{i=1}^{n} \left(\hat{Q}_i \hat{\varepsilon}_i + \hat{D}\hat{\psi}_i \right) \left(\hat{Q}_i \hat{\varepsilon}_i + \hat{D}\hat{\psi}_i \right)' \xrightarrow{p} \Omega$.

III. A bibliographic note

Professor Lewbel also shows that high-order moments can be used to estimate regressions that are quadratic in a mismeasured regressor. It should be noted that this is a special case of the analysis of polynomial regressions given by Geary (1942, pp. 73-75) in his seminal paper on estimation using high-order cumulants.

Appendix: Proofs

Lemma 4.3 of Newey and McFadden (1994), henceforth Lemma 4.3NM, will be used repeatedly in proving Propositions 1, 2, and 3. Lemma 1 below verifies a condition assumed by Lemma 4.3NM. For notation, let $\|A\| \equiv \|\mathrm{vec}\,(A)\|$, where A is a matrix and $\|\cdot\|$ is the Euclidean norm. Two easily verified facts used below are $\|Ab\| \leq \|A\| \cdot \|b\|$ and $\|ab'\| = \|a\| \cdot \|b\|$, where a and b are column vectors and $\|\cdot\|$ can denote norms for differently dimensioned vectors.

Lemma 1 *There exist neighborhoods* $\mathcal{N}_\mu$ *of* μ *and* $\mathcal{N}$ *of* (μ', θ') *such that*
(i) $E\left[\sup_{m \in \mathcal{N}_\mu} \|Q_i(m)R_i'\|^2\right] < \infty$; *(ii)* $E\left[\sup_{m \in \mathcal{N}_\mu} \left\|\frac{\partial Q_i(m)\varepsilon_i}{\partial m'}\right\|\right] < \infty$;
(iii) $E\left[\sup_{(m,t) \in \mathcal{N}} \|Q_i(m)\left(Y_i - R_i't\right)\|^2\right] < \infty$; *(iv)* $E\left[\sup_{(m,t) \in \mathcal{N}} \|\psi_i(m)Q_i(m)'\left(Y_i - R_i't\right)\|\right] < \infty$.

Proof: To prove (i), first note that $\sup_{m \in \mathcal{N}_\mu} \|Q_i(m)R_i'\|^2 = \sup_{m \in \mathcal{N}_\mu} \|Q_i(m)\|^2 \cdot \|R_i\|^2$. In view of (5), it therefore suffices to show $\sup_{m \in \mathcal{N}_\mu} \|q_i(m)\|^2$ is integrable. Since $\|q_i(m)\|^2$ is a sum of terms of the form $\left((G_i - m_G)^\gamma (Z_i - m_Z)^\delta (Y_i - m_Y)^\tau\right)^2$, where γ, δ, and τ are nonnegative integers summing to 1 or 2, it suffices to show

$$\sup_{m \in \mathcal{N}_\mu} (G_i - m_G)^{2\gamma} (Z_i - m_Z)^{2\delta} (Y_i - m_Y)^{2\tau} \tag{15}$$

is integrable. Let $\mathcal{N}_\mu = B_G \times B_Z \times B_Y$, where B_k is an open interval of length $2\lambda_k$ centered at $E(k_i)$, $k = G, Z, Y$. Then (15) equals

$$\sup_{m_G \in B_G} (G_i - m_G)^{2\gamma} \sup_{m_Z \in B_Z} (Z_i - m_Z)^{2\delta} \sup_{m_Y \in B_Y} (Y_i - m_Y)^{2\tau}, \tag{16}$$

which is less than or equal to

$$(|G_i - E(G_i)| + \lambda_G)^{2\gamma}(|Z_i - E(Z_i)| + \lambda_Z)^{2\delta}(|Y_i - E(Y_i)| + \lambda_Y)^{2\tau}. \tag{17}$$

Assumption 1 and equations (1)-(2) imply that (17), and therefore (16) is integrable, establishing (i). The proof of (ii) follows by a similar argument, after noting that $\sup_{m \in \mathcal{N}_\mu} \left\| \frac{\partial Q_i(m)\varepsilon_i}{\partial m'} \right\| = \sup_{m \in \mathcal{N}_\mu} \left\| \frac{\partial Q_i(m)}{\partial m'} \right\| \cdot |\varepsilon_i|$ and that $\left\| \frac{\partial Q_i(m)}{\partial m'} \right\|^2$ is a sum of terms of the form $\left(\phi(G_i - m_G)^\gamma (Z_i - m_Z)^\delta (Y_i - m_Y)^\tau\right)^2$, where $(\phi, \gamma, \delta, \tau)$ are nonnegative integers. To prove (iii), let $\mathcal{N} = \mathcal{N}_\mu \times B_\theta$, where B_θ is an open ball of radius λ_θ centered at θ, and note that $\sup_{(m,t) \in \mathcal{N}} \|q_i(m)\,(Y_i - R_i' t)\|^2 = \sup_{m \in \mathcal{N}_\mu} \|q_i(m)\|^2 \sup_{t \in B_\theta} |Y_i - R_i' t|^2$. The right hand side of the last equality is less than or equal to the product of (17) and $(|Y_i - R_i'\theta| + \lambda_\theta \|R_i\|)^2$, a product that is integrable by Assumption 1. Part (iv) is proven similarly, after noting that $\|\psi_i(m) q_i(m)'\,(Y_i - R_i' t)\| = \|\psi_i(m)\| \cdot \|q_i(m)\| \cdot |Y_i - R_i' t|$, $\sup_{m \in \mathcal{N}_\mu} (\|\psi_i(m)\| \cdot \|q_i(m)\|) \leq \sup_{m \in \mathcal{N}_\mu} \|\psi_i(m)\| \sup_{m \in \mathcal{N}_\mu} \|q_i(m)\|$, and $\sup_{m \in \mathcal{N}_\mu} \|\psi_i(m)\|^2 = \sup_{m_G \in B_G} (G_i - m_G)^2 + \sup_{m_Z \in B_Z} (Z_i - m_Z)^2 + \sup_{m_Y \in B_Y} (Y_i - m_Y)^2$. ■

Proof of Proposition 1: Expressions (12) and (13) imply

$$\sqrt{n}\left(\hat{\theta} - \theta\right) = \left(\bar{M}' A_n \bar{M}\right)^{-1} \bar{M}' A_n \frac{1}{\sqrt{n}} \sum_{i=1}^n Q_i(\hat{\mu})\varepsilon_i. \tag{18}$$

Applying a mean value theorem gives

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n Q_i(\hat{\mu})\varepsilon_i = \frac{1}{\sqrt{n}} \sum_{i=1}^n Q_i(\mu)\varepsilon_i + \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial Q_i(\mu^*)\varepsilon_i}{\partial m'}\right) \sqrt{n}(\hat{\mu} - \mu), \tag{19}$$

where μ^* is a weighted average of μ and $\hat{\mu}$. (Abusing notation in the usual way, μ^* differs for each individual equation in the vector equation (19).) Lemma 1(ii) and Lemma 4.3NM imply $n^{-1} \sum_{i=1}^n \left(\partial Q_i(\mu^*)\varepsilon_i / \partial m'\right) \xrightarrow{p} D$. The assumption of *i.i.d.* sampling with finite moments and the fact that $\hat{\mu}$ is a vector of sample means implies $\sqrt{n}(\hat{\mu} - \mu)$ converges in distribution to a normal distribution. These facts together with Slutsky's theorem imply

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n Q_i(\hat{\mu})\varepsilon_i = \frac{1}{\sqrt{n}} \sum_{i=1}^n Q_i(\mu)\varepsilon_i + D\sqrt{n}(\hat{\mu} - \mu) + o_p(1), \tag{20}$$

where $o_p(1)$ denotes a random vector that converges in probability to zero. Substituting the easily verified equation $\sqrt{n}(\hat{\mu}-\mu) = n^{-1/2}\sum_{i=1}^{n}\psi_i + o_p(1)$ into (20) gives

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Q_i(\hat{\mu})\varepsilon_i &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Q_i(\mu)\varepsilon_i + D\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Q_i(\mu)\varepsilon_i + D\psi_i\right) + o_p(1).\end{aligned}$$

The assumption of finite moments implies $Q_i(\mu)\varepsilon_i + D\psi_i$ has a finite second moment, so $(1/\sqrt{n})\sum_{i=1}^{n} Q_i(\hat{\mu})\varepsilon_i \xrightarrow{d} N(0,\Omega)$ by the Lindeberg-Levy central limit theorem. Proposition 1 follows from this result together with equation (18), the assumption $A_n \xrightarrow{p} A$, and the fact that Lemma 1(i) and Lemma 4.3NM imply $\bar{M} \xrightarrow{p} M$. ■

Proof of Proposition 2: Since $\hat{\Omega}_{STD} = \frac{1}{n}\sum_{i=1}^{n}\left\{Q_i(\hat{\mu})(Y_i - R_i'\hat{\theta})\left[Q_i(\hat{\mu})(Y_i - R_i'\hat{\theta})\right]'\right\}$ and $\left\|Q_i(m)(Y_i - R_i't)\left[Q_i(m)(Y_i - R_i't)\right]'\right\| = \left\|Q_i(m)\,(Y_i - R_i't)\right\|^2$, Lemmas 1(iii) and 4.3NM imply $\hat{\Omega}_{STD} \xrightarrow{p} E\left\{Q_i(\mu)(Y_i - R_i'\theta)\left[Q_i(\mu)(Y_i - R_i'\theta)\right]'\right\} \equiv \operatorname{var}\left[Q_i(\mu)\varepsilon_i\right]$. ■

Proof of Proposition 3: Note that $\Omega \equiv \operatorname{var}\left[Q_i\varepsilon_i\right] + 2DE\left[\psi_i Q_i'\varepsilon_i\right] + D\operatorname{avar}(\hat{\mu})D'$ and that $\hat{\Omega} \equiv \hat{\Omega}_{STD} + 2\hat{D}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_i\hat{Q}_i'\hat{\varepsilon}_i\right) + \hat{D}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_i\hat{\psi}_i'\right)\hat{D}'$. Lemmas 1(iv) and 4.3NM imply $\frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_i\hat{Q}_i'\hat{\varepsilon}_i \xrightarrow{p} E\left[\psi_i Q_i'\varepsilon_i\right]$. The proposition follows from this result, the limit $\hat{D} \xrightarrow{p} D$ established above, the standard result $\frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_i\hat{\psi}_i' \xrightarrow{p} \operatorname{avar}(\hat{\mu})$, and Proposition 2. ■

References

Geary, R. C. (1942): "Inherent Relations Between Random Variables," *Proceedings of the Royal Irish Academy A,* 47, 63-76.

Lewbel, A. (1997): "Constructing Instruments for Regressions With Measurement Error When No Additional Data Are Available, With an Application to Patents and R&D," *Econometrica,* 65, 1201-1213.

Newey, W. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica,* 62, 1349-1382.

Newey, W. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics,* Vol. IV, ed. by R. Engle and D. McFadden. Amsterdam: Elsevier, 2111–2245.

White, H. (1982): "Instrumental Variables Regression with Independent Observations," *Econometrica,* 50, 483-499.