An example to illustrate the asymptotics of IVLS — 15 December 2006

DA Freedman — Statistics Department, UC Berkeley

Let (Z_i, X_i, δ_i) be IID triplets for $i = 1, \ldots, n$; each random variable has a fourth moment, $E(\delta_i) = 0$, and $E(\delta_i^2) = \sigma^2 > 0$. We assume $Z_i \perp\!\!\!\perp \delta_i$. Let $a = E(X_i Z_i) > 0$ and $b = E(X_i \delta_i)$. To simplify the notation, take $E(Z_i^2) = 1$. Let

$$Y_i = \beta X_i + \delta_i$$

Here, a, b, β, σ^2 are parameters. We wish to estimate β. In this model, X_i is endogeneous if $b > 0$. On the other hand, we can instrument X_i by Z_i, because $Z_i \perp\!\!\!\perp \delta_i$ and $a > 0$.

The object here is to show that the IVLS estimator differs from β by a random error of order $1/\sqrt{n}$, with asymptotic bias of order $1/n$. Based on a sample of size n, the IVLS estimator is

$$\hat{\beta}_n = \Big(\sum_{i=1}^{n} Z_i Y_i\Big)\Big/\Big(\sum_{i=1}^{n} Z_i X_i\Big) = \beta + \eta_n \tag{1}$$

where

$$\eta_n = N_n/D_n, \quad N_n = \sum_{i=1}^{n} Z_i \delta_i, \quad D_n = \sum_{i=1}^{n} Z_i X_i \tag{2}$$

Let

$$\zeta_n = N_n/\sqrt{n} \tag{3}$$

By the central limit theorem, $\zeta_n \to N(0, \sigma^2)$: the $Z_i\delta_i$ are IID with mean 0, and the variance is σ^2 because $E(Z_i^2\delta_i^2) = E(Z_i^2)E(\delta_i^2) = 1$. Next, $E(Z_i X_i) = a$. So $D_n = na(1 + \xi_n)$, where

$$\xi_n = \frac{1}{n}\sum_{i=1}^{n}(a^{-1} Z_i X_i - 1) \tag{4}$$

is of order $1/\sqrt{n}$ by the central limit theorem. Thus

$$\hat{\beta}_n - \beta = \eta_n = \frac{\sqrt{n}}{na}\,\frac{\zeta_n}{1 + \xi_n} \tag{5}$$

and—the next being a little informal—

$$\eta_n \doteq \frac{\zeta_n - \zeta_n \xi_n}{a\sqrt{n}} \tag{6}$$

The step from (5) to (6) is "the delta-method," i.e., a one-term Taylor expansion of $1/(1 + \xi_n)$. A more rigorous argument will be given, below. We conclude that $\hat{\beta}_n - \beta$ is asymptotically normal, with mean 0 and an SE of $1/(a\sqrt{n})$. However, there is asymptotic bias of order $1/n$, because

$$
\begin{aligned}
a^{-1}n^{-1/2}E\big\{\zeta_n\xi_n\big\} &= a^{-1}n^{-1/2}E\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(Z_i\delta_i\Big)\,\frac{1}{n}\sum_{i=1}^{n}\Big(a^{-1}Z_iX_i-1\Big)\Big\}\\
&= a^{-1}n^{-2}E\Big\{\sum_{i=1}^{n}\Big(Z_i\delta_i\Big)\,\sum_{i=1}^{n}\Big(a^{-1}Z_iX_i-1\Big)\Big\}\\
&= a^{-2}n^{-2}E\big\{\sum_{i=1}^{n}Z_i^2X_i\delta_i\big\}\\
&= a^{-2}n^{-1}E\{Z_i^2X_i\delta_i\}
\end{aligned}
$$

For the third equality, expand the product of the two sums as a double sum

$$\sum_{ij}(Z_i\delta_i)(a^{-1}Z_jX_j-1)$$

When $i \neq j$, factors are independent, and products have expectation zero because $E(Z_i\delta_i) = E(Z_i)E(\delta_i) = 0$. Similarly, $E(Z_i\delta_i) = 0$ when $i = j$. The only terms with (possibly) non-zero expectation are $a^{-1}Z_i^2X_i\delta_i$.

We continue with previous assumptions and notation, but give a formal theorem and proof.

Theorem. $\hat{\beta}_n - \beta = \zeta_n/(a\sqrt{n}) - \Delta_n/(an)$, where ζ_n converges in distribution to $N(0,\sigma^2)$, and Δ_n converges in distribution to a random variable with expectation k/a, where $k = E(Z_i^2X_i\delta_i)$ may be positive, negative, or zero.

Proof. Keep in mind that ζ_n and $\sqrt{n}\xi_n$ are asymptotically normal, with expectation 0: the notation is therefore a little misleading. Start the argument from (5) above: $1/(1+x) = 1-1/(1+x)$ unless $x = -1$, so

$$a\sqrt{n}(\hat{\beta}_n - \beta) = \frac{\zeta_n}{1+\xi_n} = \zeta_n - \frac{\zeta_n\xi_n}{1+\xi_n} = \zeta_n - \frac{\Delta_n}{\sqrt{n}}$$

where

$$\Delta_n = \frac{\zeta_n\sqrt{n}\xi_n}{1+\xi_n} \tag{7}$$

The pairs

$$Z_i\delta_i,\ a^{-1}Z_iX_i - 1$$

are IID, with expectation 0 and covariance matrix

$$K = \begin{pmatrix} \sigma^2 & k/a \\ k/a & a^{-2}E(Z_i^2X_i^2) - 1 \end{pmatrix} \tag{8}$$

The central limit theorem shows that $(\zeta_n,\ \sqrt{n}\xi_n)$ converges in distribution to bivariate normal, with expectation 0 and covariance matrix K. In the denominator of (7), $\xi_n \to 0$, so Δ_n has the same limiting behavior as $\zeta_n\sqrt{n}\xi_n$. QED

REMARKS

(i) If $E(Z_iX_i) < 0$, replace Z_i by $-Z_i$ or a by $|a|$; if $E(Z_iX_i) \neq 0$, then $E(Z_i^2) > 0$ and $E(X_i^2) > 0$.

(ii) The source of the bias in IVLS is randomness in ξ_n, coupled with the correlation between ξ_n and ζ_n—that is, randomness in $\sum Z_iX_i$, coupled with the correlation between $\sum Z_iX_i$ and $\sum Z_i\delta_i$. When n is large, $\xi_n \doteq 0$—the law of large numbers—and the bias is negligible. The correlation traces back to the endogeneity of X_i, i.e., the correlation between X_i and δ_i. If, e.g., $(X_i, Z_i) \perp\!\!\!\perp \delta_i$, it is straightforward to show that $E(\hat{\beta}_n|X, Z) = \beta$. Then $k = 0$ in (8).

(iii) Equations (1–2) and the strong law of large numbers show that $\hat{\beta}_n \to \beta$ almost surely.

(iv) What about estimating σ^2? In our setup, if $e = Y - X\hat{\beta}_n$ is the vector of residuals, then $e_i - \epsilon_i = X_i(\hat{\beta}_n - \beta)$ so $\|e - \epsilon\|^2 = \sum_i X_i^2(\hat{\beta}_n - \beta)^2$ and $\|e - \epsilon\|^2/n \to 0$ almost surely.

(v) The usual presentation of IVLS conditions on Z. Then $\hat{\beta}_{\text{IVLS}} - \beta = \sum_1^n Z_i\delta_i / \sum_1^n Z_iX_i$; conditionally, the numerator is essentially normal with mean 0 and variance $\sum_1^n Z_i^2 \doteq nE(Z_i^2)$. The denominator is essentially $\sum_1^n Z_iE(X_i|Z_i) \doteq nE[Z_iE(X_i|Z_i)] = nE(Z_iX_i) = na$. With some more effort, the theorem can be extended to describe the limiting conditional behavior of $(\xi_n, \sqrt{n}\zeta_n)$, given $Z_1, \ldots, Z_n$. In a little more detail, let $\phi(Z_i) = a^{-1}Z_iE(X_i|Z_i) - 1$, so $E(\sqrt{n}\zeta_n|Z_1, \ldots, Z_n) = n^{-1/2}\sum_1^n \phi(Z_i)$. The $\phi(Z_i)$ are IID, and $E\big(\phi(Z_i)\big) = a^{-1}E(Z_iX_i) - 1 = 0$ by the definition of a. Moreover, $0 \le \text{var}\big(\phi(Z_i)\big) < \infty$. If the variance is positive, the central limit theorem applies and $E(\sqrt{n}\zeta_n|Z_1, \ldots, Z_n)$ converges in distribution. In any event, we can center, considering the conditional joint distribution of

$$\xi_n,\ \sqrt{n}\big(\zeta_n - E(\zeta_n|Z_1, \ldots, Z_n)\big)$$

given $Z_1, \ldots, Z_n$. Apparently, this conditional distribution converges weak-star, along almost all sample sequences of $Z_1, Z_2, \ldots$. For example, ξ_n is $n^{-1/2}\sum_1^n Z_i\delta_i$, where the δ_i are IID with mean 0, and—conditionally—the Z_i are (almost surely) a well-behaved sequence of constants:

$$\frac{1}{n}\sum_{i=1}^{n} Z_i^2 \to 1, \quad \frac{1}{n}\sum_{i}\{Z_i^2 : 1 \le i \le n \ \&\ |Z_i| > L\} \to 0 \text{ as } n \to \infty \text{ and then } L \to \infty$$

As in many other such situations, when n is large, there would seem to be little difference between conditional and unconditional inference.

(vi) Suppose $(Z_i, \delta_i, \epsilon_i)$ are independent, with expectation 0, variance 1, and fourth moments. We can set $X_i = aZ_i + b\delta_i + c\epsilon_i$. Then $\text{cov}(Z_i, X_i) = a$, because $E(Z_i^2) = 1$; and $\text{cov}(X_i, \delta_i) = b$, because $E(\delta_i^2) = 1$. The k in the theorem is $k = E(Z_i^2X_i\delta_i) = b$, because

$$Z_i^2X_i\delta_i = aZ_i^3\delta_i + bZ_i^2\delta_i^2 + cZ_i^2\epsilon_i,$$

while $E(Z_i^3\delta_i) = E(Z_i^3)E(\delta_i) = 0$, $E(Z_i^2\delta_i^2) = E(Z_i^2)E(\delta_i^2) = 1$, $E(Z_i^2\epsilon_i) = E(Z_i^2)E(\epsilon_i) = 0$.

(vii) With, say, two instruments and one endogenous variable, the proof of consistency and asymptotic normality is about the same. However, evaluating the small-sample bias is trickier. For instance, expansions like (6) can be done in the matrix domain.