IV, 2SLS

January 10, 2014

Asymptotics, for IV

$$\hat{\beta} = (Z'X)^{-1}Z'Y = (Z'X)^{-1}Z'X\beta + (Z'X)^{-1}Z'\varepsilon \tag{1}$$

$$\therefore \hat{\beta} - \beta = (Z'X)^{-1}Z'\varepsilon \tag{2}$$

IV is *not* unbiased.

Asymptotics, for IV

$$\hat{\beta} = (Z'X)^{-1}Z'Y = (Z'X)^{-1}Z'X\beta + (Z'X)^{-1}Z'\varepsilon \tag{1}$$

$$\therefore \hat{\beta} - \beta = (Z'X)^{-1}Z'\varepsilon \tag{2}$$

IV is *not* unbiased.

It is consistent if $\Sigma_{ZX} = E[Z_j'X_j]$ is non-singular; i.e. $\hat{\beta} \xrightarrow[n\to\infty]{P} \beta$. Because $(1/n)Z'X \xrightarrow{P} \Sigma_{ZX}$ and $(1/n)Z'\varepsilon \xrightarrow{P} E[Z_j'\varepsilon] = 0$.

Asymptotics, for IV

If $\mathrm{Var}(\varepsilon_j \mid Z_j) = \sigma^2$ for all j, the CLT tells us that

$$\frac{1}{\sqrt{n}} Z'\varepsilon \xrightarrow{D} N(0, \sigma^2 \Sigma_Z) \, ,$$

where $\Sigma_Z = E[Z_j' Z_j]$. Therefore in that case

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} N(0, \sigma^2 (\Sigma_{ZX}^{-1} \Sigma_{ZZ} (\Sigma_{XZ})^{-1}) \, .$$

In practice we estimate the covarance matrix and use

$$\hat{\beta} \sim N(\beta, s^2 (Z'X)^{-1} Z'Z (X'Z)^{-1}) \, .$$

White standard errors

If $\varepsilon_j \mid Z_j$ does not have a constant variance, We can replace $\sigma^2(Z'Z)$ in the middle of the covariance matrix expression by $E[Z'\varepsilon\varepsilon' Z]$. Just as with the SNLM, we can assume that $E[\varepsilon\varepsilon' \mid Z] = \Omega$ and model Ω, in which case we get a more efficient estimate analogous to GLS. (We omit working out this IV analogue of GLS in detail.)

We can also, in the case where we know Ω is diagonal, use $\sum Z_i' Z_i \varepsilon_i^2 / n$ as a consistent estimate of $E[Z'\varepsilon\varepsilon' Z]$, giving us "heteroskedasticity-robust" IV standard errors for the estimates. Stock and Watson suggest always using these heteroskedasticity-robust standard errors, though in fact as with GLS there is a tradeoff — if the heteroskedasticity- consistent standard errors are nearly the same as the non-robust ones, the non-robust ones are likely more accurate estimates. If there is instead a big difference, it is likely that modeling heteroskedasticy would substantially improve efficiency.

Two stage least squares

So far we have considered simple IV, where the number of instruments matches the number of X's and $Z'X$ is square and non-singular. If we have more instruments than X's, we need to consider how to use them efficiently.

In the univariate case — one instrument, one X — the asymptotic variance is $\sigma^2\sigma_Z^2/\operatorname{Cov}(Z,X)^2$. This is one over the explained sum of squares in a regression of X on Z. So we get smaller variance the more Z is correlated with X. When we have many Z's, then, it seems natural to form an instrument matrix as a linear combination of the Z's that has as much correlation with X as possible. This is the idea of 2SLS. When described as two stages it works like this:

Stage 1 Estimate θ in the regression $X = Z\theta + \nu$ by OLS. (Note that since X is $n \times k$, this is really k separate regression equations, one for each column of X.)

Stage 2 Form $\hat{X} = Z\hat{\theta}$ and use that as an instrument, i.e.

$$\hat{\beta}_{2SLS} = (\hat{X}'X)^{-1}\hat{X}'Y = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y\ .$$

The asymptotic covariance matrix is

$$\sigma^2(X'Z(Z'Z)^{-1}Z'X)^{-1}\ .$$

A practical caution

Since $\hat{X}'X = \hat{X}'\hat{X}$ (because $X - \hat{X}$ is uncorrelated in the sample with $\hat{X}$ by construction), the 2SLS estimator is exactly the OLS estimate of β in a least squares regression of Y on $\hat{X}$. But the covariance matrix for $\hat{\beta}_{2SLS}$ as standard regression output from this "second stage" is *not* a consistent estimate of the true 2SLS covariance matrix. What comes out of a standard regression program is $s^2(\hat{X}'\hat{X})^{-1}$, where s^2 is an estimate of the residual variance in a regression of y on $\hat{X}$, while what is needed is an estimate of $\mathrm{Var}(\varepsilon_i)$. This can be estimated as the sample variance of $y - X\hat{\beta}$, whereas the standard OLS output would use the sample variance of $y - \hat{X}\hat{\beta}$. These can be quite different.