The Bootstrap

February 24, 2015

Good standard errors can be hard to find.

Recall heteroskedasticity in OLS:

$$Var[\mathbf{b}^{robust}] = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\hat{\mathbf{V}}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1} \quad (1)$$

We impose a great deal of structure on the model errors to get to this equation. We do similar things for deriving standard errors in Random Effects. Even models with simple variance covariance matrices (OLS) are simple because of assumptions made about errors.

An easy way to get the 'correct' standard errors

- Applicable to every model we have covered in class
- For a wide range of advanced models, *the only way* to get standard errors on coefficients.

Introduction to Bootstrapping

The type of Bootstrapping we consider here is termed the Non-Parametric Bootstrap. In this context, Non-Parametric means we do not need to assume anything about the distribution of the errors for inference (ie. to determine if a parameter is statistically different from zero)

Key Assumptions:

Assumption 1: We need to have a random sample of observations from the population.

Assumption 2: Independence of observations.

Same ideas in a regression context

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_N \end{bmatrix}_{N\times 1} \qquad \mathbf{x} = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1i} & x_{2i} & x_{3i} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1N} & x_{2N} & x_{3N} \end{bmatrix}_{N\times 4} \tag{2}$$

Steps:

- Sample (with replacement) from the rows of $\mathbf{y}, \mathbf{x}$- making sure to keep y_i and x_i together if sampled).
- Collect S such subsamples
- For each subsample s, calculate $b_s = (x_s' x_s)^{-1} x_s' y_s$ and store the estimated parameters
- Sort each column of estimated b_k and find the upper and lower 2.5% from the S regressions. For example, if S=1000, look at rows 25 and 975. This is the upper and lower 95% confidence interval of your parameter.

Can be extended to Time-Series and Panel Data

Use the block bootstrap.

- Panel Data: Sample *with replacement* from individuals.
- Time Series Data: Sample *with replacement* from Z periods of data beginning in period t_z.

These techniques are valid so long as the resampling unit is independent from other units and we have a random sample of the population.

Good Introductory Text

Bootstrapping: A Nonparametric Approach to Statistical Inference (Quantitative Applications in the Social Sciences) Mooney and Duval Sage University Press (1993).