Large Sample Properties of Multiple Regression Model

Christopher Taber

Department of Economics
University of Wisconsin-Madison

March 23, 2011

Background

Lets begin with a little background from Appendix C.3 of Wooldridge

We are worried about what happens to OLS estimators as our sample gets large

The first concept to think about is *Consistency* which Wooldridge defines as

Consistency

Let W_N be an estimator of θ based on a sample $y_1, y_2, ..., y_N$ of size N. Then W_N is a *consistent estimator* of θ if for every $\varepsilon > 0$

$$Pr\left(|W_N - \theta| > \varepsilon\right) \rightarrow 0$$

as $N \rightarrow 0$.

Not particularly intuitive.

What it means is that as N gets large W_N gets closer and closer to θ.

Wooldridge also uses the notation that consistency means that

$$\text{plim}\,(W_n) = \theta.$$

Law of Large Numbers

The most important property for consistency is the **Law of Large Numbers**

This states that

Law of Large Numbers.

Let $y_1, y_2, ..., y_N$ be independent, identically distributed random variables with mean μ. Then

$$\text{plim}\,(\bar{y}) = \mu.$$

This is really useful and also quite intuitive.

It says if you rolled a die forever and kept taking the mean value you would get closer and closer to 3.5.

Lets look at an example in stata.

The key result is that for OLS:

Theorem 5.1 (Consistency of OLS)

Under Assumptions MLR.1 through MLR.4 the OLS estimator $\hat{\beta}_j$ is consistent for β_j, for all $j = 0, 1, 2, ..., K$.

This basically is just a result of the law of large numbers and Assumption MLR.4.

To see why, focus on the slope coefficient from the simple regression model

We know that

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{N}\sum_{i=1}^{N}\left(x_{i1} - \bar{x}_1\right)u_i}{\frac{1}{N}\sum_{i=1}^{N}\left(x_{i1} - \bar{x}_1\right)^2}$$

Now using the law of large numbers

$$\begin{aligned}
plim\left[\frac{1}{N}\sum_{i=1}^{N}\left(x_{i1} - \bar{x}_1\right)u_i\right] &= E\left[\left(x_{i1} - \bar{x}_1\right)u_i\right] \\
&= cov(x_1, u) \\
plim\left[\frac{1}{N}\sum_{i=1}^{N}\left(x_{i1} - \bar{x}_1\right)^2\right] &= E\left[\left(x_{i1} - \bar{x}_1\right)^2\right] \\
&= Var\left(x_1\right)
\end{aligned}$$

Putting it together

$$plim\left(\hat{\beta}_1\right) = \beta_1 + \frac{cov(x_1, u)}{var(x_1)}$$

Under Assumption SLR.4, $cov(x_1, u) = 0$ so $\hat{\beta}_1$ is consistent.

However, note as well that if $cov(x_1, u_1) \neq 0$ then $\hat{\beta}_1$ will not be consistent.

Again, we can look at this in stata.

Central Limit Theorem

While I think the law of large numbers is intuitive, the central limit theorem is not.

I will use the notation

$$\hat{\theta} \approx N(0, \sigma^2)$$

to mean that as the sample size gets large, $\hat{\theta}$ is approximately normally distributed with expected value 0 and variance σ^2.

It turns out that under the assumptions above

$$\sqrt{N}\,(\overline{y} - \mu) \approx N(0, \sigma^2)$$

This is the central limit theorem

Lets just think of this as a gift and not try to understand it.

We can look at this in stata as well.

To do this I used the following procedure

1. Start with a Uniform distribution
2. Draw a sample of size N
3. Take the sample mean of that
4. Repeat steps 1-3 many times to get a sample where each observation is a mean
5. Lets plot the distribution of these

The OLS estimator is a bit more complicated, but is like a mean because

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{N}\sum_{i=1}^{N}\left(x_{i1} - \bar{x}_1\right)u_i}{\frac{1}{N}\sum_{i=1}^{N}\left(x_{i1} - \bar{x}_1\right)^2}$$

It turns out that we can use the central limit theorem

We state it as in Wooldridge

Theorem 5.2 (Asymptotic Normality of OLS)

Under the Gauss-Markov Assumptions MLR.1 through MLR.5,

1. $\sqrt{N}\left(\hat{\beta}_j - \beta_j\right) \approx N(0, \sigma^2/a_j^2)$, where $\sigma^2/a_j^2 > 0$ is the asymptotic variance of $\sqrt{N}\left(\hat{\beta}_j - \beta_j\right)$; defined in the text. We say that $\hat{\beta}_j$ is asymptotically normally distributed.
2. $\hat{\sigma}^2$ is a consistent estimator of $\sigma^2 = var(u)$.
3. For each j,

$$\frac{\hat{\beta}_j - \beta_j}{se\left(\hat{\beta}_j\right)} \approx N(0,1)$$

   where $se\left(\hat{\beta}_j\right)$ is the usual OLS standard error (which stata also produces)

While this may be difficult to understand, it is really really useful.

This means that we can relax the normality assumption which seemed really strong, but we can still use the normal distribution to construct confidence intervals and test null hypotheses.

I did something to give you an idea how this works.

1. Assume that

$$y = \beta_0 + \beta_1 x + u$$

   where x and u both have a uniform distribution.
2. Draw a sample of size N on x and y
3. Regress y on x and keep track of $\hat{\beta}_1$
4. Repeat steps 1-3 many times to get a sample where each observation is a $\hat{\beta}_1$
5. Lets plot the distribution of these