Econ 452 Section 8 - Asymptotics Examples

Connor Cole

October 30, 2015

An Aside on Asymptotics

NOTE: This material is meant to elaborate on what was discussed in class and is completely optional.

Previously, we began talking about regression inference by claiming that we assume $\epsilon \sim N(0, \sigma^2)$. As we have learned this week, our basic tools of statistical inference still work without normally distributed errors as long as MLR1-5 hold and we let the sample size go to infinity. Technically speaking, we refer to these kinds of results as regression asymtotics. While the prospect of sending a sample size to infinity may seem unrealistic, in practice it takes a surprisingly small sample size for asymptotics to work in a relatively believable fashion.

Before seeing some evidence of how asymptotics work, let's review what theory says about the asymptotic behavior of regression. We already know that $\hat{\beta}_j$ is consistent in OLS under MLR 1-4, or that it converges in probability to the true value of β_j for all j. Hence, informally speaking, we can say that $\hat{\beta}_j$ by itself 'collapses' into a single point at the true value of β_j as we let the sample size go to infinity. However, to do regression inference, we want to preserve some information about the distribution of $\hat{\beta}_j$ without letting that distribution collapse into a single point. So, we generally look at test statistics of the form $\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)}$ or $\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}$. As we will see, these test statistics do not collapse on a single point and instead converge to smooth distributions that make regression inference possible.

Now, as we know from class, if we assume errors are *iid* from some distribution, homoskedastic, and $\mathbb{E}[\epsilon|X] = 0$, then the test statistic we've looked at before where we take the observed coefficient $\hat{\beta}_j$ from OLS, subtract off the true value of β_j and divide by the true standard deviation of $\hat{\beta}_j$ (without an estimated $\hat{\sigma}^2$) converges in distribution to a standard normal distribution:[1]

[1] We will later in class relax assumptions even further and not require homoskedasticity to make these asymptotic results work. For now, with the tools we have available, we need to assume homoskedasticity.

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \xrightarrow{d} N(0,1)$$

Professor Smith in class described this result as follows:

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim AN(0,1)$$

Remember that:

$$sd(\hat{\beta}_j) = \sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2(1 - R_j^2)}}$$

Where R_j^2 is the R^2 measure from a regression of x_j on all other covariates (including a constant).

It is relatively straightforward using the Slutsky Theorem to show that even when $\hat{\sigma}^2$ is estimated, then:

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim AN(0,1)$$

Where:

$$se(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2(1 - R_j^2)}}$$

Let's take a minute to talk about what we actually mean when saying something has an asymptotic distribution, or that the distribution of some random variable converges in distribution to some other distribution. Formally speaking, we say that the distribution of some random variable T_n, denoted by $F_n(x)$, converges in distribution to $F(x)$ if $F_n(x)$ converges in limit at all points x to $F(x)$.

$$\lim_{n \to \infty} F_n(x) = F(x)$$

In the case of the test statistics we considering, convergence in distribution of the test statistics to a standard normal implies that if we were to look at a single possible value of the test statistic and consider the probability of seeing that test statistic or some value less as we add more observations n to our calculation, then that probability would converge to the probability of seeing x or some value less in a standard normal.

Now, note that since the distribution of the test statistics converges to a known distribution as n goes to infinity, then assuming that we have 'enough' n so that we're 'sufficiently close' to the asymptotic distribution of the test statistic, we can use this fact to do hypothesis testing using the normal distribution. For example, let's say that I estimate $\hat{\beta}_j = 2.05$ and my null hypothesis I would like to test is that $\beta_j = 2$. Then, let's say that $se(\hat{\beta}_j) = .01$ and $n = 300$. Since n seems 'large enough' I can assume that, roughly:

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

Hence, under my null hypothesis, the probability of seeing an outcome as extreme as 2.05 or greater would be:

$$2 \cdot (1 - \Phi(\frac{2.05 - 2}{.01})) \approx 0$$

Hence, my p-value for this null hypothesis is very low, and I would reject the null hypothesis at the 1% confidence level. It's worth taking a moment to appreciate how this result differs from our previous finite sample claims about normal errors. Under the assumption of *iid* and homoskedastic normal errors, then:

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim N(0,1)$$

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

Note that the distribution of these values is exact in finite samples. In the previous example, the regression inference logic worked because I assumed that a certain number of observations meant that the distribution of test statistics would be 'close enough' to the normal distribution to make regression inference work. Here, I need make no such assumption, as I know the exact finite sampling distribution of the test statistics. While it is theoretically possible that we could make claims about the exact finite sample distribution of test statistics when we look at non-normal errors, we do not in general do so because we don't want to assume we know the distribution of the error term, and, perhaps more importantly, working out these exact finite sample distributions is hard.

It can be difficult to appreciate these facts when hearing about them in the abstract as it is hard to picture the difference between a finite sample result and a result that depends on sample sizes going to infinity. As previously in the note on unbiasedness and consistency, a good way to see how an estimator behaves is to run a Monte Carlo simulation where we generate data under some known generating process and then apply our estimation process to the data. Since we know how the data was constructed, we know what the true values of various parameters must be. Let's look first at a model generated with standard normal errors. Consider data generated from the following data generating process:

$$y = 2x + 1 + \epsilon$$

Where $x \sim N(0,1)$ and $\epsilon \sim N(0,1)$. As in the previous note on unbiasedness, by construction, it must be that $\mathbb{E}[\epsilon|X] = 0$, and the error terms are normally distributed. Let's estimate this model by fitting the following model with regression:

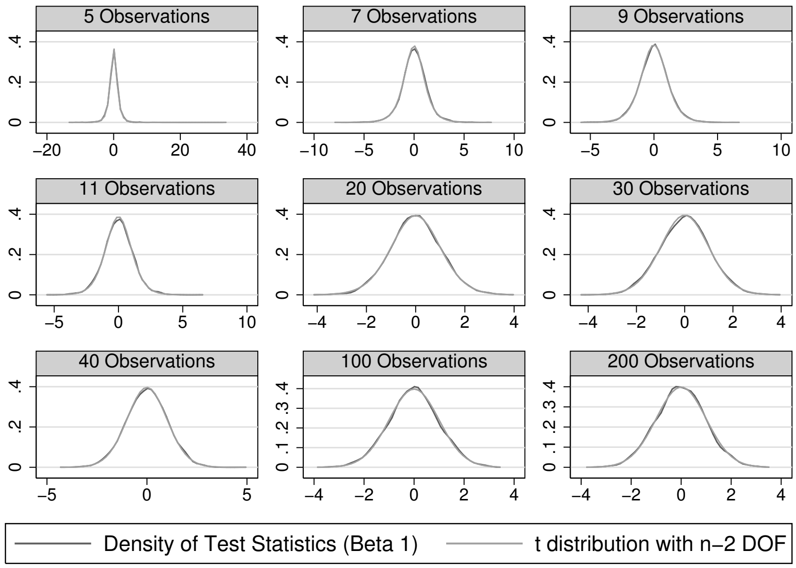
$$y = \beta_1 x + \beta_0 + \epsilon$$

Since MLR1-6 are met, we know that the regression estimation procedure results in an exact finite sample distribution of our test statistics in regression of the form:

$$T = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-2}$$

Thus, our test statistics T for $j = 1, 0$ have an exact t distribution with $n-2$ degrees of freedom since one parameter is estimated and a constant is included in the regression. Figure 1 shows the distribution of test statistics T generated from datasets with the listed number of observations. For each category, 5,000 datasets were generated and 5,000 coefficients were estimated to create a distribution of test statistics. In addition to this distribution of test statistics, the graph below also shows the density of a t distribution with $n-2$ degrees of freedom. As is clear, the distributions nearly line up precisely at all sample sizes, with only minor variations due to sampling variation in the datasets we've created. This figure is rather striking evidence for the believability of the finite sample hypothesis testing results we have talked about in class.

Figure 1: Distribution of Test Statistics under Normal Errors

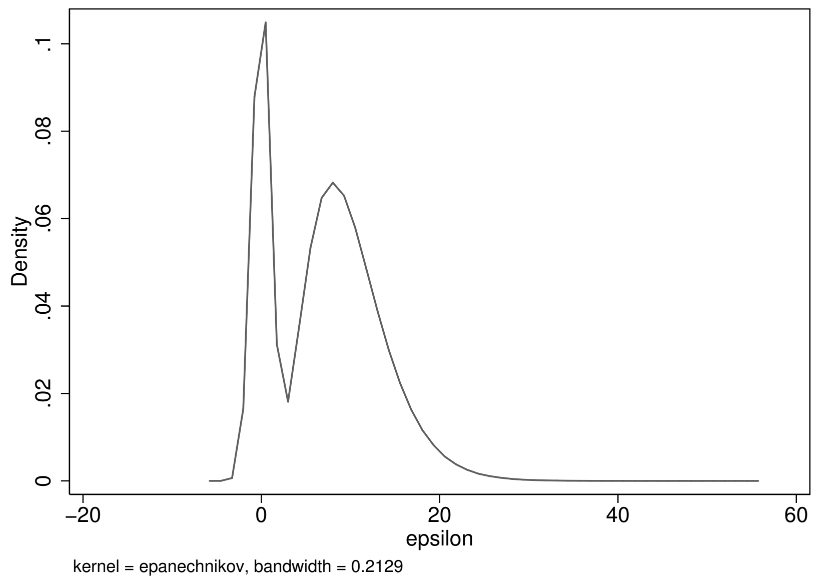


Now, let's consider what happens when we don't have normal errors. Consider data generated from the following data generating process:

$$y = 2x + 1 + \epsilon$$

Where ϵ is an error term that takes on values from $N(0,1)$ distribution with probability .3 and takes on values from a $\Gamma(5,2)$ distribution with probability .7. The error term will have the density depicted in Figure 2. As is apparent, the distribution of the error term is bimodal and not shaped at all like a normal distribution.

Figure 2: Distribution of Error Term



Now, consider the model we are estimating here. The error terms are not normal, but by construction $\mathbb{E}[\epsilon|X] = 0$, and furthermore the variance of the error term does not depend on X. Hence, MLR 1-5 are met, and the distribution of test statics of the form $T = \frac{\hat{\beta}_j - \beta_j}{\hat{se}(\beta_j)}$ will be asymptotically standard normal, although we do not know what the finite sample distributional properties of these test statistics are.

So, let's estimate the following model using regression as we did before:

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

Figure 3 shows the distribution of test statistics T generated from datasets with the listed number of observations. For each category, 5,000 datasets were generated and 5,000 coefficients were estimated to create a distribution of test statistics. As is clear, the distribution of test statistics seems to change as the number of observations goes up. For a relatively small number of observations, the distribution of test statistics is not especially close to the normal distribution. Even at a relatively large number of observations, like 100, the distribution of test statistics is close to the standard normal but just a little off. However, by 200 observations, the fit seems reasonably close.

This last observation raises an important question - how many observations are 'enough' in order to invoke asymptotic results? In practice, most economists don't question the use of asymptotics to justify hypothesis testing when looking at more than a few hundred observations in a regression. As is apparent here, even with a seemingly large number of results, the fit is not quite precise, as strictly speaking asymptotic approximations require thinking of sending the sample size to infinity.

Figure 3: Distribution of Test Statistics under Non-Normal Errors