Econ 452 Section 8 - Unbiasedness and Consistency

Connor Cole

October 30, 2015

An Aside on Unbiasedness and Consistency

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

We have learned in class that assuming that MLR1-4 hold (especially the assumption that the error term is conditionally mean independent of X, or, $\mathbb{E}[\epsilon|X] = 0$) ensures that the OLS is both unbiased and consistent. Note that we do not assume normal errors or even homoskedasticity for this conclusion to hold. It can be hard to appreciate what is meant by consistency and unbiasedness.

Unbiasedness, technically, means that:

$$\mathbb{E}[\hat{\beta}_j] = \beta_j$$

For all j. We interpret this claim as meaning that, if we were able to repeatedly sample data with different observations but the same sample size as the data we are working with, then on average the estimated $\hat{\beta}_j$ values would be equal to the true β_j value.

To appreciate what this means, let's consider what we call a Monte Carlo simulation in econometrics, or a setting where we simulate data and then look at the performance of an estimator. Since we know how the data is constructed, we know what the true values of different parameters is. Below, I have repeatedly created data from the following data generating process:

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

Where $\epsilon \sim N(0, 1)$ and $x \sim N(0, 1)$. Note that, by construction, ϵ is independent of x, so $\mathbb{E}[\epsilon|X] = 0$, and OLS is unbiased and consistent. So, I draw data from this data generating process repeatedly, and then estimate a model of the following form:

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

First, let's consider a graph of outcomes. Figure 1 is a graph that on the horizontal axis measures the number of observations in the dataset (or, n) and on the vertical axis measures the mean $\hat{\beta}_1$ over 1000 different datasets drawn with the same number of observations. Now, note in particular that the average $\hat{\beta}_1$ is almost precisely the true value of β_1, or 2, and the pattern doesn't seem to vary with n.

Figure 1: Unbiasedness of $\hat{\beta}_1$ (Average Over 1000 Repetitions)

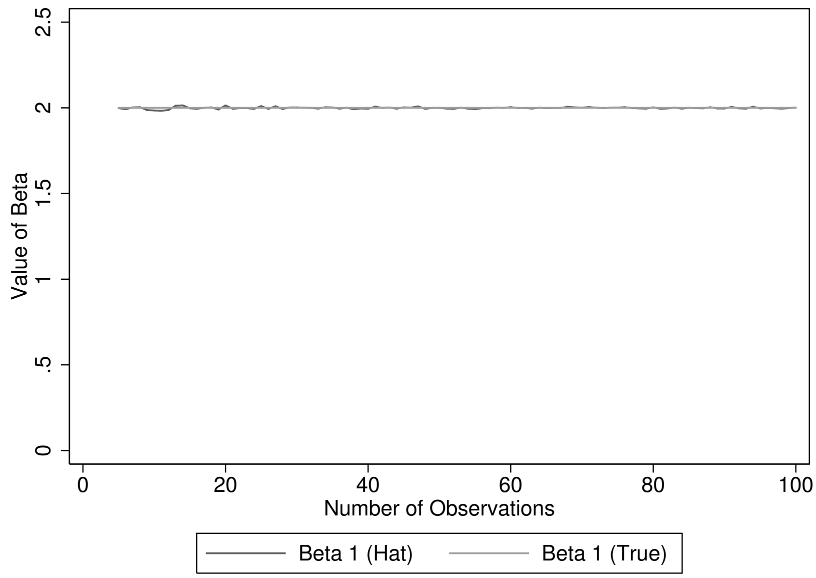
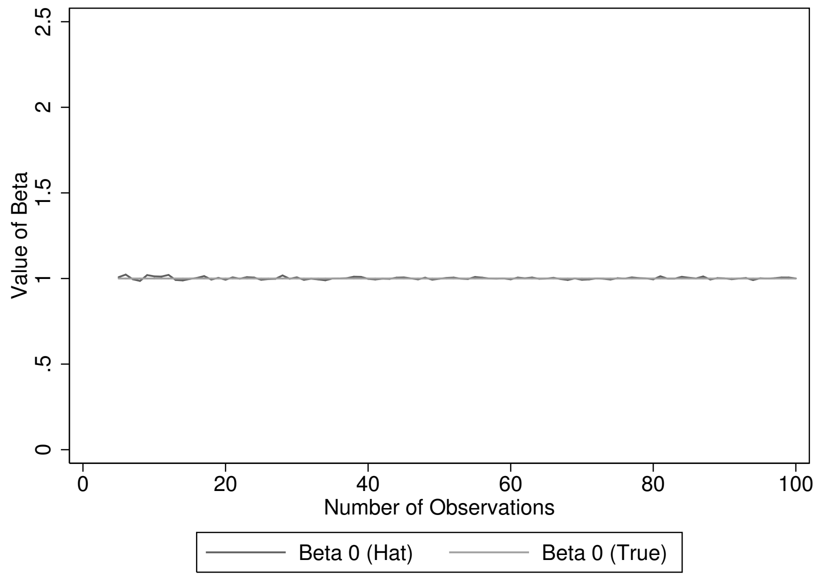


Figure 2 reports a similar graph (with similar axes) for $\hat{\beta}_0$. Note that the average $\hat{\beta}_0$ is also almost precisely the true value of β_0 or 1, with the pattern again seemingly unrelated to n.

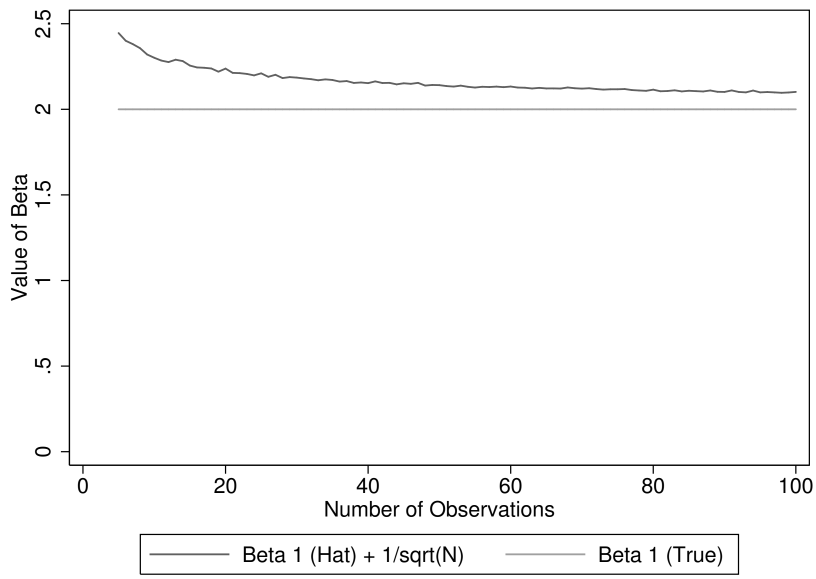
Figure 2: Unbiasedness of $\hat{\beta}_0$ (Average Over 1000 Repetitions)



In both of these figures, given that we have only run 1000 simulated datasets and not an infinitely large one, it's not surprising that there is some slight variation in the average $\hat{\beta}$ values, but the averages are strikingly close for all values of n. This pattern is a direct result of the unbiasedness of $\hat{\beta}$ in OLS given conditional mean independence of the error term.

Now, let's consider a clearly biased estimator. For example we could consider an estimator $\hat{\beta}_1 + \frac{1}{\sqrt{n}}$. Given that $\hat{\beta}_1$ is unbiased in OLS, this estimator will have bias $\frac{1}{\sqrt{n}}$. This estimator is graphed in Figure 3.

Figure 3: Biased Estimator of $\hat{\beta}_1$ (Average Over 1000 Repetitions)



Now we see a clear bias on average, but it seems to attenuate as n grows larger.

Let's consider another biased estimator. It turns out that $se(\hat{\beta}_j)$ is actually a biased estimator of $sd(\hat{\beta}_j)$ in OLS. Remember that:

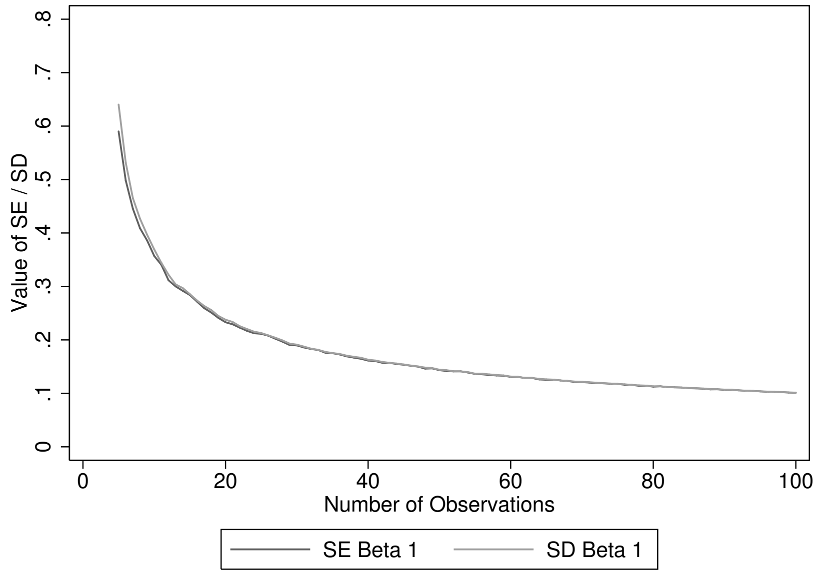
$$se(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})}}$$

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

Note that the only difference between the two is that the standard error has an estimated variance of the error term, while the standard deviation does not.[1] Figure 4 is a graph of the estimated standard error with the true standard deviation, which we know because we know the true variance of the error term.

[1] It is worthwhile to note that $\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})}$ is an unbiased estimator of $\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})}$, or the estimated variance of $\hat{\beta}$ is an unbiased estimator of the true variance of $\hat{\beta}$. Remember that nonlinear functions of unbiased estimators are not usually themselves unbiased.

Figure 4: Standard Error and Standard Deviation of $\hat{\beta}_1$ (Average Over 1000 Repetitions)



As with Figure 3 above, however, the bias seems to shrink as sample size increases. The bias seems to vanish in both of these estimates because they are both consistent.

Consistency

Consistency is a different property than unbiasedness. You can have estimators that are biased but consistent, for example. As we learned in class, $\hat{\beta}_j$ is a consistent estimator for β_j in OLS for all j assuming MLR 1-4 hold. Formally speaking, this means that:

$$\lim_{n \to \infty} P(|\hat{\beta}_j - \beta_j| > \epsilon)$$

For every $\epsilon > 0$. Sometimes you see this result stated as:

$$\hat{\beta}_j \xrightarrow{p} \beta$$

Or that $\hat{\beta}_j$ converges to β in probability limit.

Similar to unbiasedness, we have a general way of interpreting this result which is less formal. If $\hat{\beta}_j$ converges to β_j in probability limit, then as we increase n for a given realization of data, the observed value of $\hat{\beta}_j$ becomes closer and closer to the true value such that the probability of seeing a realization of $\hat{\beta}_j$ that is not arbitrarily close to β_j becomes vanishingly small in OLS. Keep in mind - we are not saying that the exact value of $\hat{\beta}_j$ becomes close to β_j as there is always sampling variation that might pull us away, but the probability of seeing data that pulls $\hat{\beta}_j$ away from β_j becomes smaller and smaller as the sample size increases.

Let's return to our previous data generating process:

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

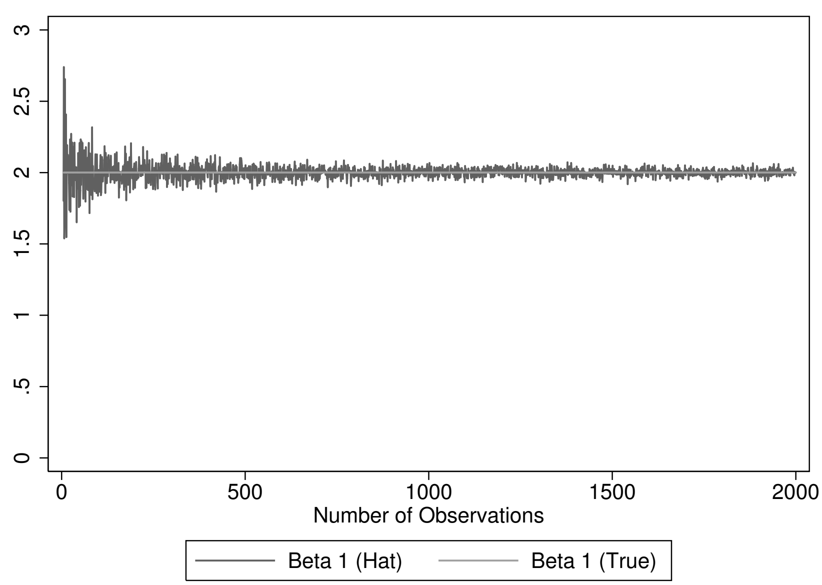
Where $\epsilon \sim N(0, 1)$ and $x \sim N(0, 1)$.

And think about again estimating:

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

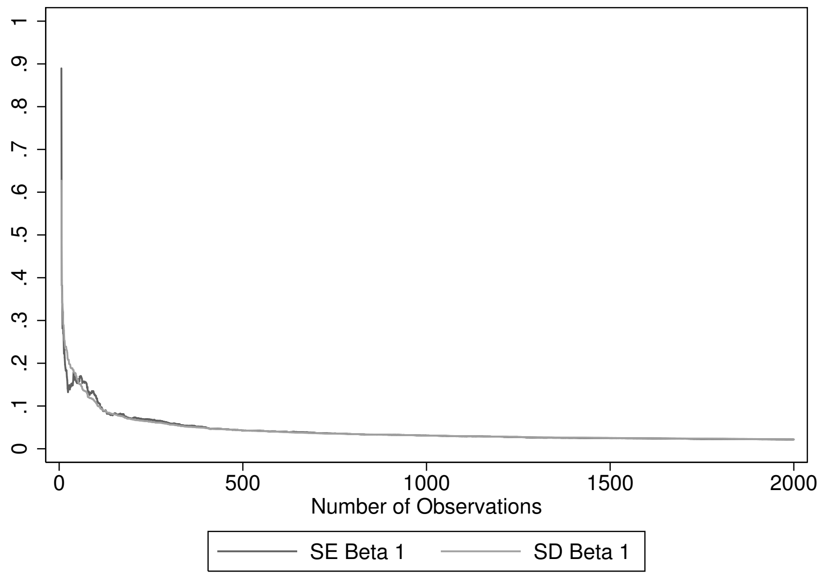
Since conditional mean independence holds by the way we have constructed this data, we know that $\hat{\beta}_j$ in OLS is consistent for β_j for all coefficients j. Previously, we approximated unbiasedness by looking at the behavior of $\hat{\beta}_1$ across repeated samples from the same distribution. Here, to look at consistency, we look at the behavior of $\hat{\beta}_1$ across a single realization of the data as we increase n. Figure 5 demonstrates what happens as we steadily increase the number of observations in a single dataset. As we can see, the observed value of $\hat{\beta}_1$ comes closer and closer to the true value of β_1. There is obvious variability in $\hat{\beta}$, but as n increases, the probability of seeing a realization of $\hat{\beta}_1$ outside of a narrow bound around 2 becomes smaller and smaller.

Figure 5: Consistency of $\hat{\beta}_1$ for β_1



Now, let's look at the estimators that we said were biased before. First, we look at the standard error of $\hat{\beta}_1$, an estimate of the standard deviation that we found was upwardly biased. As we see in Figure 6, the standard error jumps around the standard deviation some until settling at values that seem very close to the true standard deviation of $\hat{\beta}_1$. Again, the probability of seeing realizations of the data such that the standard error is far away from the true standard deviation becomes smaller and smaller as the sample size increases.

Figure 6: Consistency of Standard Error for Standard Deviation of $\hat{\beta}_1$



Now let's consider our biased estimator of β_1, $\hat{\beta}_1 + \frac{1}{\sqrt{n}}$. As n increases, we know that $\hat{\beta}_1$ is consistent for β_1 in OLS, but it should also be apparent that $\frac{1}{\sqrt{n}}$ goes to 0 as n increases, as for every ϵ we can choose an n such that $P(|\frac{1}{\sqrt{n}} - 0| > \epsilon) = 0$. Hence, $\frac{1}{\sqrt{n}}$ converges in probability to 0. Note that from the properties of consistent estimators, the sum of two consistent estimators adds up to their probability limits. Therefore, $\hat{\beta}_1 + \frac{1}{\sqrt{n}}$ is consistent for $\beta_1 + 0 = \beta_1$. We can see this by looking at a similar graph of the observed value of $\hat{\beta}_1 + \frac{1}{\sqrt{n}}$ as n increases. As is apparent in Figure 7, $\hat{\beta}_1 + \frac{1}{\sqrt{n}}$ starts off with a sizable bias that falls as n increases.

Figure 7: Consistency of $\hat{\beta}_1 + \frac{1}{\sqrt{n}}$ for β_1