Multiple Regression Analysis: Asymptotics

ECONOMETRICS (ECON 360)

BEN VAN KAMMEN, PHD

Introduction

There is not a lot of new material in this chapter, unless one wants to get into proofs of the Central Limit Theorem, probability limits, and *convergence in distribution*—which I prefer not to.

Instead my emphasis is on explaining why some of the Assumptions in the CLM are not so restrictive and that inference according to Chapter 4 methods is still possible under weaker assumptions about the distribution of the error term.

Outline

Consistency.

Asymptotic Normality and Large Sample Inference.

Asymptotic Efficiency of OLS.

Consistency

(Consistent Estimator) Defined: "An estimator that *converges in probability* to the population parameter as the sample size grows without bound."

This is stated formally by expressing the probability that the estimator falls outside an interval, ε, and that the probability approaches zero as the sample size increases, for any ε.

Convergence in probability states that:

$$\Pr[|estimator_n - parameter| > \varepsilon] = 0 \text{ as } n \to \infty.$$

Consistency (continued)

If one can collect an arbitrarily large amount of observations, he ought to be able to obtain an estimate that gets closer and closer to the true parameter value.

If this is not the case, the estimator is inconsistent and not of much use.

Fortunately, under Assumptions MLR.1 through MLR. 4, the OLS estimators ($\hat{\beta}_0$ through $\hat{\beta}_k$) are consistent estimators of their corresponding parameters.

Consistency (continued)

One can show this fairly easily for the simple regression model, using the estimator and the definition of the model:

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i, \text{ and}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_{i1} - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x})y_i}{\sum_{i=1}^{n}(x_{i1} - \bar{x})^2}.$$

So,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x})(\beta_0 + \beta_1 x_{i1} + u_i)}{\sum_{i=1}^{n}(x_{i1} - \bar{x})^2} = \beta_1 + \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x})u_i}{\sum_{i=1}^{n}(x_{i1} - \bar{x})^2}.$$

Consistency (continued)

This expression should be familiar from deriving the unbiasedness of OLS.

To show the consistency $\hat{\beta}_1$, make a small modification, dividing the numerator and denominator of the second term by the sample size.

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

Taking the probability limit ("plim") of this, as $n \rightarrow \infty$, you find that the numerator converges to the covariance of x_1 and u, and the denominator converges to the variance of x_1.

Consistency (concluded)

And the properties of probability limits state that the *plim* of a ratio of two estimators equals the ratio of their *plims*:

$$plim\left(\hat{\beta}_1\right) = \beta_1 + \frac{plim[n^{-1}\sum_{i=1}^{n}(x_{i1} - \bar{x})u_i]}{plim\left[n^{-1}\sum_{i=1}^{n}(x_{i1} - \bar{x})^2\right]} = \beta_1 + \frac{Cov(x_1, u)}{Var(x_1)}.$$

MLR.4 (SLR.4) states that x_1 and u are mean independent, which implies that their covariance is zero. So,

$$plim\left(\hat{\beta}_1\right) = \beta_1, \text{and}$$

OLS is consistent as long as the error term is not correlated with the “x” variable(s).

OLS is consistent under weaker assumptions

This is the weaker version of the fourth Assumption, MLR.4’, which states:

$$E(u) = 0 \text{ and } Cov(x_j, u) = 0 \; \forall \; j.$$

It is weaker because assuming merely that they are uncorrelated linearly does not rule out higher order relationships between x_j and u.

- The latter can make OLS biased (but still consistent), so if unbiasedness and consistency are *both* desired, you still need (the stronger) Assumption MLR.4.

Mis-specified models are still inconsistent

Inconsistency can be shown in a manner very similar to biasedness in the model with 2 explanatory variables.

If one estimates ($\tilde{\beta}_1$) a regression that excludes x_2, such that:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + v_i \text{ and } \tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x})y_i}{\sum_{i=1}^{n}(x_{i1} - \bar{x})^2},$$

$$\Leftrightarrow \tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x})(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + v_i)}{\sum_{i=1}^{n}(x_{i1} - \bar{x})^2} = \beta_1 + \beta_2 \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x})x_{i2}}{\sum_{i=1}^{n}(x_{i1} - \bar{x})^2},$$

Mis-specified models are still inconsistent (continued)

the $plim$ of the estimator is $plim\left(\tilde{\beta}_1\right) = \beta_1 + \beta_2\delta_1;\ \delta_1 \equiv \frac{Cov(x_1, x_2)}{Var(x_1)}$.

The second term is the inconsistency, and this estimator converges closer to the inaccurate value $(\beta_1 + \beta_2\delta_1)$ as the sample size grows.

In the $k > 2$ case, this result is general to all of the explanatory variables; none of the estimators is consistent if the model is mis-specified like above.

Asymptotic normality and large sample inference

This is the most consequential lesson from Chapter 5.

Knowing that an estimator is consistent is satisfying, but it doesn't imply anything about the distribution of the estimator, which is necessary for inference.

- The OLS estimators are normally distributed if the errors are assumed to be (with constant variance σ^2), as well as the values of $(y|x_1 ... x_k)$.
- But what if the errors are *not* normally distributed?
  - Consequently neither are the values of *y*.
- As the text points out, there are numerous such examples, e.g., when y is bound by a range (like 0-100), or in which it's skewed (example 3.5), and the normality assumption is unrealistic.

Asymptotic normality and large sample inference (continued)

However inference is based on the estimators have a constant mean ($\hat{\beta}_j$) and variance. When they are standardized, they have mean zero and standard deviation 1 (note: we maintain the homoskedasticity assumption).

Crucially, as the sample size approaches infinity, the distribution of the standardized estimator converges to standard normal.

This property applies to all averages from random samples, and is known as the Central Limit Theorem (CLT). Its implication is that:

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \overset{d}{\rightarrow} Normal(0,1)\ \forall\ j;\ \overset{d}{\rightarrow} \text{ means } converges\ in\ distribution.$$

Asymptotic normality and large sample inference (continued)

Another way of saying it is that the distribution of the OLS estimator is asymptotically normal.

One more feature of the OLS asymptotics is that the estimator, $\hat{\sigma}^2$, consistently estimates σ^2, the population error variance, so it no longer matters that the parameter is replaced by its consistent estimator.

- Nor is it necessary to make a distinction between the standard normal and the t distribution for inference—because in large samples the t distribution converges to standard normal anyway.
- For the sake of precision, however, t_{n-k-1} is the exact distribution for the estimators.

Asymptotic normality and large sample inference (concluded)

Assumption MLR.6 has been replaced with a much weaker assumption—merely that the error term has finite and homoskedastic variance.

As long as the sample size is "large", inference can be conducted the same way as under Assumption MLR.6, however.

- How many observations constitutes "large" is an open question.
- The requisite in some cases can be as low as 30 for the CLT to provide a good approximation, but if the errors are highly skewed ("non-normal") or if there are many regressors in the model (k "eats up" a lot of degrees of freedom) reliable inference with 30 observations is overly optimistic.

Precision of the OLS estimates

Finally we investigate "how fast" the standard error shrinks as the sample size increases. The variance of $\hat{\beta}_j$ (square root is the standard error) is:

$$\widehat{Var}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{SST_j(1 - R_j^2)} = \frac{\hat{\sigma}^2}{n s_j^2 (1 - R_j^2)}, \text{where}$$

the total sum of squares of $x_j (SST_j)$ can be replaced according to the definition of x_j's sample variance (s_j^2):

$$s_j^2 = \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{n} = \frac{SST_j}{n}.$$

Precision of the OLS estimates (continued)

As n gets large, these sample statistics each approach their population values.

$$plim(\hat{\sigma}^2) = \sigma^2, plim(s_j^2) = \sigma_j^2, \text{and } plim(R_j^2) = \rho_j^2, \text{and}$$

none of these parameters depends on sample size. Variance gets smaller at the rate $(\frac{1}{n})$ because of the explicit "n" term in the denominator. I.e.,

$$\widehat{Var}(\hat{\beta}_j) = \frac{\sigma^2}{n\sigma_j^2(1-\rho_j^2)}; \ \frac{\partial \widehat{Var}(\hat{\beta}_j)}{\partial n} = -\frac{\widehat{Var}(\hat{\beta}_j)}{n}.$$

Precision of the OLS estimated (concluded)

The asymptotic standard error is just the square root and it get smaller at the rate of $(n^{-\frac{1}{2}})$.

$$se(\hat{\beta}_j) = \frac{1}{\sqrt{n}} \frac{\sigma}{\sigma_j (1 - \rho_j^2)^{\frac{1}{2}}}.$$

F tests for exclusion restrictions, as well as *t* tests, can be conducted—for large samples—as you learned in Chapter 4 under the assumption of normally distributed errors.

β has lots of *consistent* estimators

The OLS estimator, $\hat{\beta}$, also has the lowest asymptotic variance among estimators that are linear in parameters and rely on functions of x, e.g., $g(x)$.

An estimator that uses an alternative to $g(x) = x$ can be called $\tilde{\beta}_1$, and has the form:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(z_i - z)y_i}{\sum_{i=1}^{n}(z_i - \bar{z})x_i}; z_i \equiv g(x_i); g \not\equiv f(x) = x.$$

As long as z and x are correlated, this estimator converges in probability to the true value of β_1, i.e., it is consistent.

β has lots of consistent estimators (continued)

Depending on what kind of non-linear function "g" is, this can fail because correlation only measures linear relationships.

And since x and u are mean independent,

$$E(u|x_1) = E(u|g(x)) = E(u|z) = 0; \text{so are } u \text{ and } z.$$

$$\tilde{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(z_i - z)u_i}{\sum_{i=1}^{n}(z_i - \bar{z})x_i}; plim(\tilde{\beta}_1) = \beta_1 + \frac{Cov(z,u)}{Cov(z,x)} = \beta_1.$$

Asymptotic efficiency of OLS

But the variance of $\tilde{\beta}_1$ is no less than the variance of $\hat{\beta}_1$.

$$Var(\tilde{\beta}_1) = E(\tilde{\beta}_1 - \beta_1)^2 = E\left(\frac{\sum_{i=1}^{n}(z_i - z)u_i}{\sum_{i=1}^{n}(z_i - \bar{z})x_i}\right)^2 = \frac{\sigma^2 Var(z)}{[Cov(z,x)]^2}, \text{since}$$

only the "own" products show up in the numerator. And,

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{Var(x)}, \text{as before.}$$

Asymptotic efficiency of OLS (continued)

So in order for $Var(\hat{\beta}_1) \leq Var(\tilde{\beta}_1)$,

$$\frac{\sigma^2}{Var(x)} \leq \frac{\sigma^2 Var(z)}{[Cov(z,x)]^2} \Leftrightarrow [Cov(z,x)]^2 \leq Var(x)Var(z).$$

This property is satisfied by the Cauchy-Schwartz Inequality, which states that there cannot be more covariance between two variables than there is overall variance in them.

So the OLS estimator, $\hat{\beta}_1$, has a smaller variance than any other estimator with the same form:

$$Avar(\hat{\beta}_1) \leq Avar(\tilde{\beta}_1); Avar \text{ denotes asymptotic variance.}$$

Conclusion

According to the asymptotic properties of the OLS estimator:

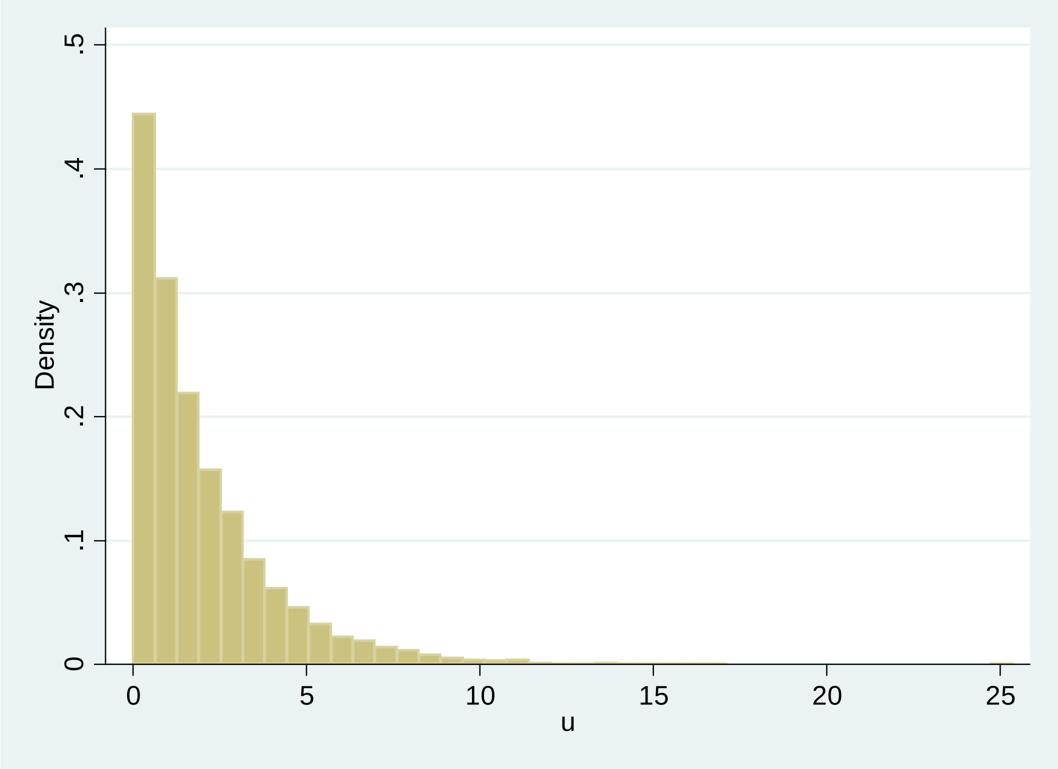
- OLS is consistent,
- The estimator converges in distribution to standard normal,
- Inference can be performed based on the asymptotic convergence to the standard normal, and
- OLS is the most efficient among many consistent estimators of β.

A non-normal error term

```
. clear

. drawnorm x, n(10000) means(12) sds(2) clear
(obs 10000)

. generate u = rgamma(1,2)

. gen y=2+x+u

. reg y x
```

The error term is definitely not normally distributed.

As the histogram (right) shows.



Bootstrapping

To reveal the distribution of $\hat{\beta}_1$ in the regression, $y = \beta_0 + \beta_1 x + u$, I resample my 10,000 observations many (2000) times.

- This would take a long time, were it not for the software.

Stata code, for $n = 10$:

```
bootstrap, reps(2000) size(10) saving(U:\ECON 360 - Spring 2015\BS 10.dta, every(1) replace) : reg y x
```

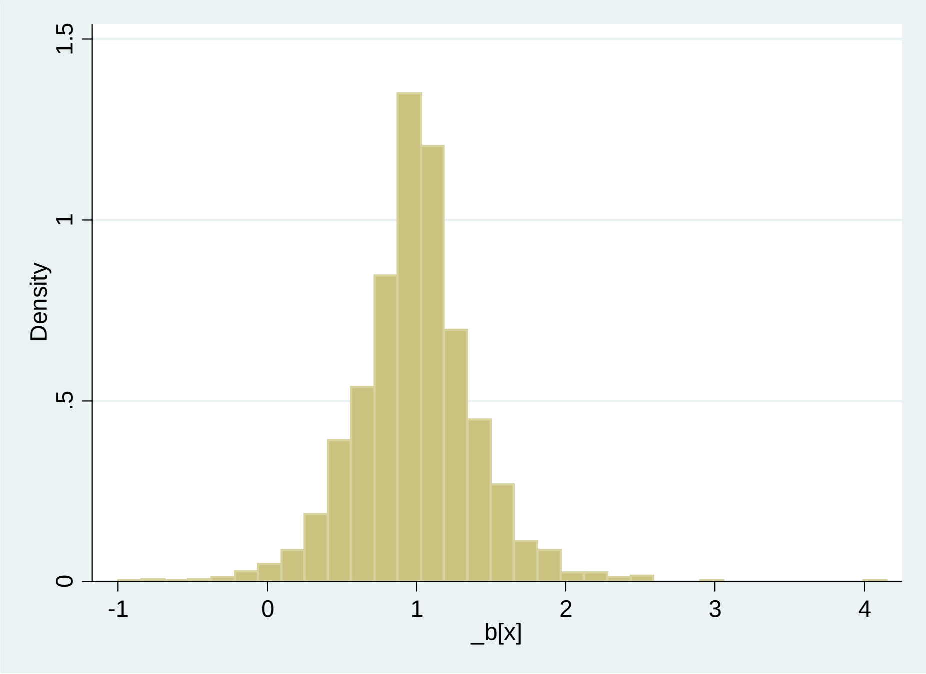
Normality?

You can judge whether it *looks* like the normal distribution.

But a normal distribution is supposed to have 0 skewness (symmetry) and a kurtosis of 3.

- This one has 0.25 (right) skewness and 6.746 kurtosis.

```
. summ _b_x, detail

                            _b[x]
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -.0410598      -.9976056
 5%      .375233      -.7887098
10%     .5267513      -.7529542       Obs                2000
25%     .7741854      -.5366573       Sum of Wgt.        2000

50%     .9984144                      Mean           .9962621
                        Largest       Std. Dev.      .3981664
75%     1.201218       2.497071
90%     1.459825       2.522474       Variance       .1585365
95%     1.634268       2.966994       Skewness       .2467498
99%     2.093876       4.149807       Kurtosis       6.746221
```



Histogram of 2000 estimates (10 obs. each) of $\hat{\beta}_1$.

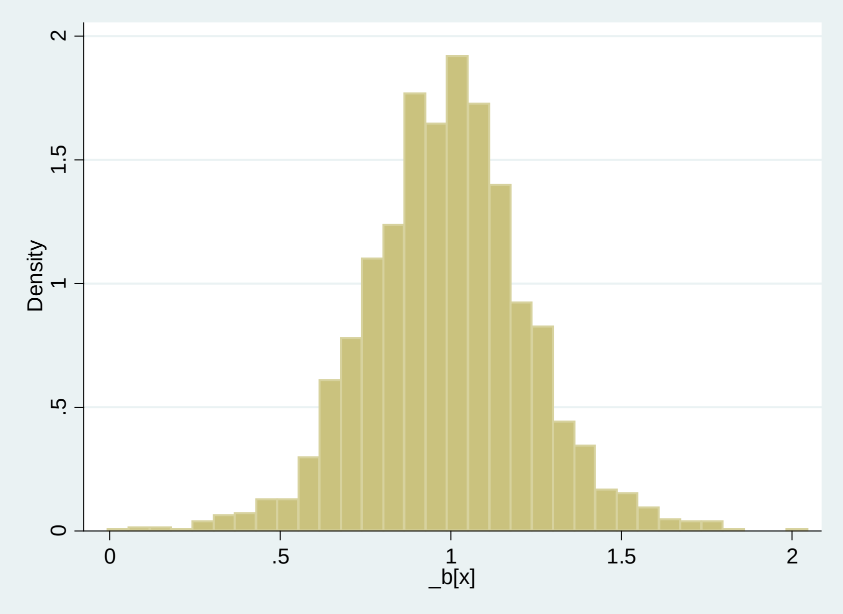
Non-Normality

The statistical test for whether the distribution of beta hats is Normal, called theJarque-Bera statistic, rejects the null that the distribution is Normal.

- Code is: *sktest _b_x*
- Similar to a joint hypothesis with 2 restrictions. $H_0: skewness = 0$ and $kurtosis = 3$.
- In this case, the *p* value is <0.0001.

Would a bigger sample size fail to reject H_0? $n = 20$

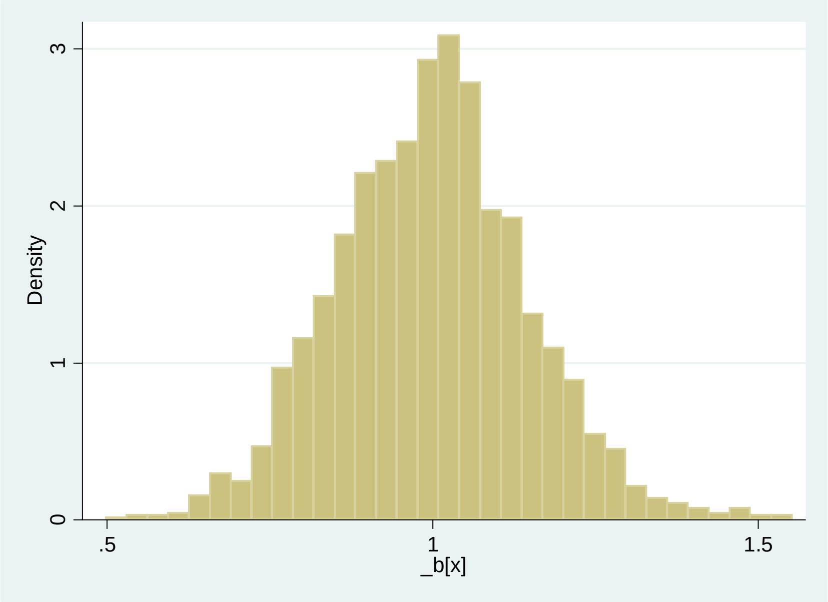
```
. summm _b_x, detail

                              _b[x]
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .3718964      -.0062484
 5%     .6183816       .1094247
10%     .6945582       .1131481       Obs                2000
25%     .8423931       .1282413       Sum of Wgt.        2000

50%     .9921745                      Mean           .9891276
                        Largest       Std. Dev.      .2382368
75%     1.133664       1.754412
90%     1.279575       1.784783       Variance       .0567568
95%     1.377411       1.837642       Skewness       .0079499
99%     1.599728       2.046466       Kurtosis       3.905129
```



The skewness is mostly gone, but the distribution is still too “peaked” to be Normal: *p* value on the J-B statistic is still <0.0001.

$n = 50?$

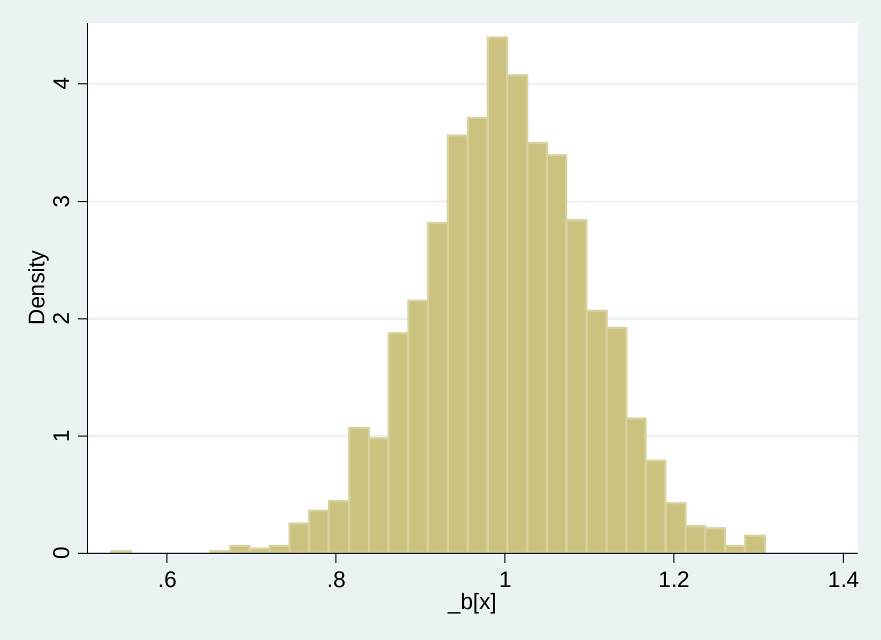
```
. summ _b_x, detail

                            _b[x]
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .6642559       .4981143
 5%     .7648403       .5408446
10%     .8089489       .5615669       Obs                2000
25%     .8994239       .5639632       Sum of Wgt.        2000

50%     .9992069                      Mean           .9980457
                        Largest       Std. Dev.      .1472866
75%     1.090446       1.502628
90%     1.181844       1.509088       Variance       .0216934
95%     1.240575       1.548912       Skewness       .1596157
99%     1.374213       1.551746       Kurtosis       3.348677
```



The skewness comes back a little, but the kurtosis is coming down now: *p* value on the J-B statistic is up to <0.0005.

$n = 100?$

```
. summm _b_x, detail

                            _b[x]
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .7633944       .5340427
 5%     .8333735       .6678689
10%     .8760604       .6756338       Obs                2000
25%     .9362201       .6903498       Sum of Wgt.        2000

50%     .9993333                      Mean             1.0007
                        Largest       Std. Dev.      .0990559
75%     1.066902        1.29039
90%     1.127472       1.295925       Variance       .0098121
95%      1.16099       1.300377       Skewness      -.0652147
99%     1.235443       1.307498       Kurtosis       3.241735
```

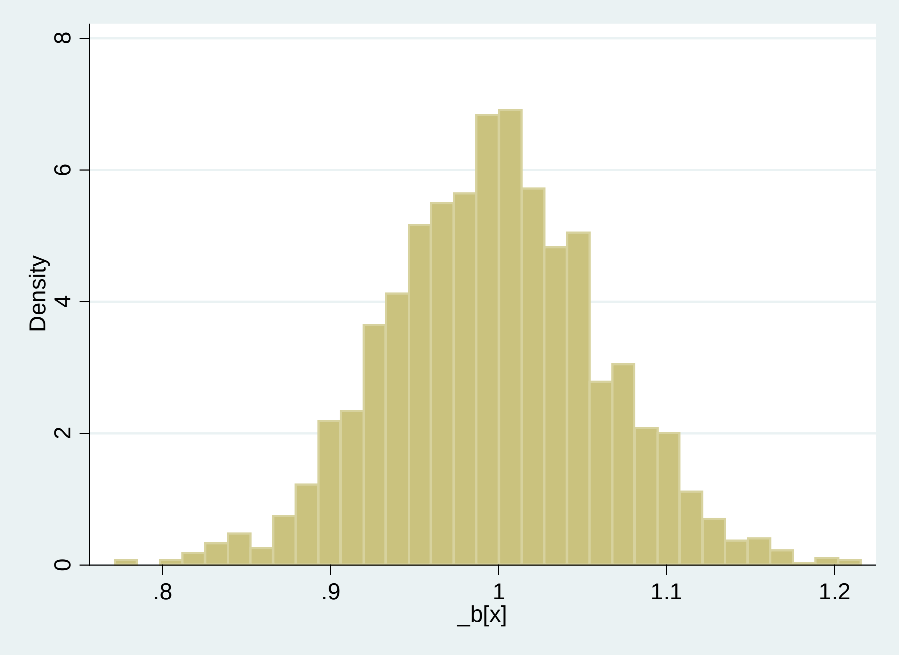



```
                  Skewness/Kurtosis tests for Normality
                                                    ------ joint ------
    Variable |    Obs   Pr(Skewness)   Pr(Kurtosis)  adj chi2(2)   Prob>chi2
-------------+---------------------------------------------------------------
        _b_x |  2.0e+03    0.2325        0.0388        5.67        0.0586
```

p>0.05; first
"fail to reject"!

$n = 250$? Normality far from rejected.

```
. summm _b_x, detail

                            _b[x]
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .8413554       .7718233
 5%     .8956159       .7786757
10%      .917204        .799898       Obs                2000
25%     .9550751       .8113453       Sum of Wgt.        2000

50%      .997361                      Mean           .9972047
                        Largest       Std. Dev.      .0636356
75%       1.0383       1.190229
90%     1.080251       1.191585       Variance       .0040495
95%     1.103124       1.210835       Skewness       .0216586
99%      1.15125       1.215675       Kurtosis        3.14645

. sktest _b_x

                    Skewness/Kurtosis tests for Normality
                                                   ------- joint ------
    Variable |       Obs   Pr(Skewness)   Pr(Kurtosis) adj chi2(2)   Prob>chi2
-------------+---------------------------------------------------------------
        _b_x |   2.0e+03     0.6915         0.1817         1.94        0.3791
```



That’s asymptotic normality

And I only had to run 10,000 regressions to show it!