The sampling distribution of the OLS estimator, and its statistical properties in finite and large samples

Econometrics for minor Finance, Lecture 4

Tunga Kantarcı, Fall 2025

The OLS estimator

$$\hat{\beta}$$

is a random variable and has a sampling distribution.

We now turn to the statistical properties of this distribution, which describe its mean, variance, and overall behavior under the assumptions of the classical linear regression model.

In introductory econometrics courses, the statistical properties of the OLS estimator are typically studied through theoretical derivations. This approach can feel abstract, and it is not always clear what is being taught.

However, these statistical properties are simply features of the sampling distribution of the OLS estimator. This distribution is unobserved in practice, but we approximated it using simulation. We will use the simulated distribution to study the statistical properties of the OLS estimator in a more intuitive way.
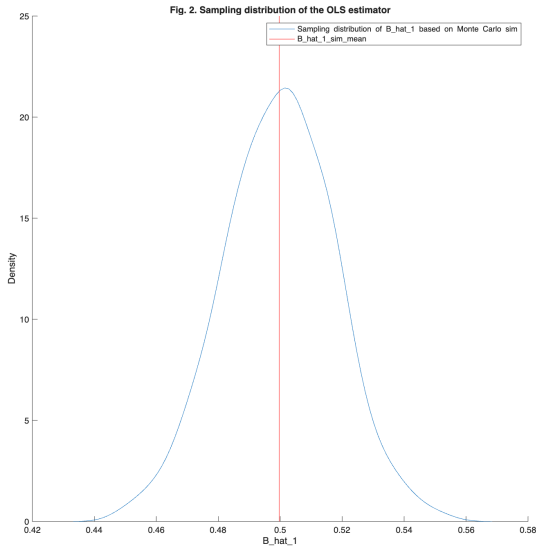
Recall the population regression function

$$y = \beta_0 + \beta_1 x_1 + u$$

and the model assumptions we made. Recall the OLS estimator of the slope parameter

$$\hat{\beta}_1$$

Recall that it has a sampling distribution, and how we demonstrated it.

Sampling distribution of the OLS estimator

Fig. 2. Sampling distribution of the OLS estimator

This is the sampling distribution of our estimator as a tool. How do want our tool to perfom?

Specifically, what should we expect about its mean and its variance?

The statistical properties of any estimator depend on the sample size. Therefore, we distinguish between:

- Finite sample: How the sampling distribution behaves in a given sample of limited size.
- Large sample or asymptotic: How the sampling distribution behaves as the sample size grows without bound.

In finite samples, we want the sampling distribution of the OLS estimator

$$\hat{\beta}$$

to be centered around the true parameter

$$\beta$$

on average.

The mean of the sampling distribution of $\hat{\beta}$ conditional on x is

$$\mathsf{E}\left[\hat{\beta} \mid x\right]$$

What does this say? Repeatedly take all possible samples of a same size from the population. Obtain the OLS estimates in each sample. Take the average of all estimates.

Sampling distribution of the OLS estimator in finite samples: The mean of the distribution: Desirable property: Unbiasedness

Recall that the mean in the population is the expected value.
Expected value is a population term, not a sample term.

Sampling distribution of the OLS estimator in finite samples: The mean of the distribution: Desirable property: Unbiasedness

In the classical linear regression framework, we treat the regressors x as fixed, or equivalently we condition on them. This means that the only source of randomness in the OLS estimator $\hat{\beta}$ comes from the error term u, which drives the variation in y. By writing expectations as

$$\mathsf{E}\left[\hat{\beta} \ \middle| \ x\right]$$

we focus on the randomness from the error, and ask whether, given the observed regressors, the estimator is centered on the true parameter β. Conditioning on x therefore clarifies that unbiasedness is about the estimator not systematically missing the true value for any given set of regressors.

The criterion we want the sampling distribution of $\hat{\beta}$ to satisfy is

$$\mathsf{E}\left[\hat{\beta} \mid x\right] = \beta$$

This in fact holds, and we say that

$$\hat{\beta}$$

is an unbiased estimator of

$$\beta$$

If the exogeneity assumption holds.

Technically,
$$\mathsf{E}\left[\hat{\beta} \mid x\right] = \beta$$
says that, given the regressors x, the mean of the sampling distribution of $\hat{\beta}$ equals the true parameter β.

Intuitively,

$$\mathsf{E}\left[\hat{\beta} \mid x\right] = \beta$$

says that, across repeated samples, $\hat{\beta}$ will on average hit the true value. In other words, the estimates we obtain from different samples do not systematically miss the true value.

Recall the simulation exercise. We have set the population parameter

$$\beta_1$$

to

$$0.5$$

Sampling distribution of the OLS estimator in finite samples: The mean of the distribution: Desirable property: Unbiasedness Illustration

```
N_sim = 1000
N_obs = 9000
B_0 = 0.2
B_1 = 0.5
x = random('Uniform', -1, 1, [N_obs 1])
B_hat_0_sim = NaN(1, N_sim)
B_hat_1_sim = NaN(1, N_sim)
for i = 1:N_sim
    u = random('Normal', 0, 1, [N_obs 1])
    y = B_0 + B_1 * x + u
    B_hat_1 = sum((x-mean(x)).*(y-mean(y))) /
              sum((x-mean(x)).^2);
    B_hat_0 = mean(y) - B_hat_1 * mean(x);
    B_hat_1_sim(1,i) = B_hat_1
    B_hat_0_sim(1,i) = B_hat_0
end
```
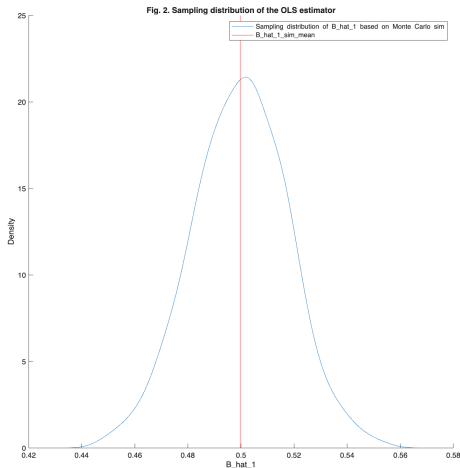
Let us plot the sampling distribution of

$$\hat{\beta}_1$$

and mark the mean of the estimates form repeated samples. it is

$$0.499$$

Sampling distribution of the OLS estimator in finite samples: The mean of the distribution: Desirable property: Unbiasedness: Illustration

Fig. 2. Sampling distribution of the OLS estimator

This demonstrates that the OLS estimator is unbiased. On average, it gives the population slope, almost.

Almost because there is simulation noise. If we increase the number of simulations, we will converge to the true parameter.

Sampling distribution of the OLS estimator in finite samples: The mean of the distribution: Desirable property: Unbiasedness

What does unbiasedness imply in practice?

Suppose that you draw an unlucky sample from the population, and obtain a bad $\hat{\beta}$. Or think of our simulation experiment. Suppose that in a generated sample, draws of u happen to be extreme.

Then,

$$\hat{\beta}$$

based on that sample will be far from its population mean

$$\mathsf{E}\left[\hat{\beta} \mid x\right] = \beta$$

Hence, in practice, to satisfy unbiasedness as much as possible, we hope that the sample at hand is typical.

Sampling distribution of the OLS estimator in finite samples: The mean of the distribution: Desirable property: Unbiasedness

So in practice, we could be unlucky with the sample data at hand.

The limitation by the unbiasedness criterion is summarized nicely by the story of three econometricians who go duck hunting. The first shoots about a foot in front of the duck, the second about a foot behind. The third yells: "We got him!"

In reality, the random sample we take is usually representative of the population. Hence, the likelihood of being mocked by this joke is not very high.

From the last lecture slides we have

$$\hat{\beta}_1 = \beta_1 + \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})u_i}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}$$

Take expectations conditional on x to obtain

$$\mathsf{E}\left[\hat{\beta}_1 \,\middle|\, x\right] = \beta_1 + \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})\,\mathsf{E}\left[u_i \mid x\right]}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}$$

If the exogeneity assumption

$$E[u_i \mid x] = 0$$

holds, the numerator vanishes. We obtain

$$E\left[\hat{\beta}_1 \mid x\right] = \beta_1$$

The exogeneity assumption is a requirement for unbiasedness.

Is the OLS estimator the only unbiased estimator? Consider a competitor estimator

$$\hat{\beta}_0 = Cy$$

C is some matrix that depends on x. Taking the expectation conditional on x,

$$\begin{aligned}
\mathsf{E}\left[\hat{\beta}_0 \mid x\right] &= \mathsf{E}\left[Cy \mid x\right] \\
&= \mathsf{E}\left[C(\beta x + u) \mid x\right] \\
&= C\beta x + C\mathsf{E}\left[u \mid x\right] \\
&= C\beta x \\
&= \beta
\end{aligned}$$

if $\mathsf{E}\left[u \mid X\right] = 0$, and if $Cx = I$. The OLS estimator is not the only unbiased estimator!

The OLS estimator is not the only unbiased estimator. We chose unbiasedness as a criterion to trust an estimator. But once we see that more than one estimator can be unbiased, a question arises:

Why should we still believe in the OLS estimator?

We need to evaluate $\hat{\beta}$ using a criterion beyond unbiasedness in order to justify our preference for it. The additional criterion is

$$\text{Var}\left[\hat{\beta} \mid X\right] \leq \text{Var}\left[\hat{\beta}_0 \mid X\right]$$

This states that the variance of the sampling distribution of the unbiased OLS estimator

$$\hat{\beta}$$

is smaller than or equal to the variance of the sampling distribution of any other competing unbiased estimator

$$\hat{\beta}_0$$

Hence, OLS estimator is the most efficient estimator in its class.

We skip the proof. The OLS estimator is efficient if the exogeneity assumption holds. The proof requires this.

The OLS estimator has the smallest variance within the class of estimators that are both unbiased and linear.

Efficiency means attaining the minimum possible variance among all unbiased estimators in a given class.

Therefore, the OLS estimator is the most efficient estimator in this class.

An estimator that is most efficient is called the 'best' in its class.

As with unbiasedness, one could illustrate efficiency using simulation.

The idea is to compare the sampling distribution of the OLS estimator with that of an alternative estimator. For a fixed sample size, the OLS distribution is more peaked, reflecting its smaller variance. This demonstrates the desirable property of efficiency.

An illustration will be added later.

We claimed that the OLS estimator a linear estimator. Let's prove this.

If an estimator is a linear function of the dependent variable, it is a linear estimator. Is $\hat{\beta}$ a linear estimator? Yes. The values of y are linearly combined using weights that are a non-linear function of the values of x. Hence, $\hat{\beta}$ is a linear estimator with respect to how it uses the values of the dependent variable only, irrespective of how it uses the values of the regressors.

Let us consider our earlier bivariate model

$$y = \beta_0 + \beta_1 x_1 + u$$

In this model

$$\hat{\beta}_1 = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} = \left(\frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} \right) y_i - \left(\frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} \right) \bar{y}$$

$\hat{\beta}_1$ is a linear function of the values of y.

Recall our assumption that u follows a normal distribution:

$$u \mid x \sim N\left[0, \sigma^2\right]$$

with mean

$$E\left[u \mid x\right] = 0$$

and variance

$$\text{Var}\left[u \mid x\right] = \sigma^2$$

Does this imply that

$$\hat{\beta}$$

follows a normal distribution?

Recall from last lecture:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})u_i}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}$$

Conditioning on x_i, the denominator is fixed. The numerator is a linear combination of the error terms u_i. Since each $u_i \sim N(0, \sigma^2)$ and linear combinations of normal random variables are normal, it follows that
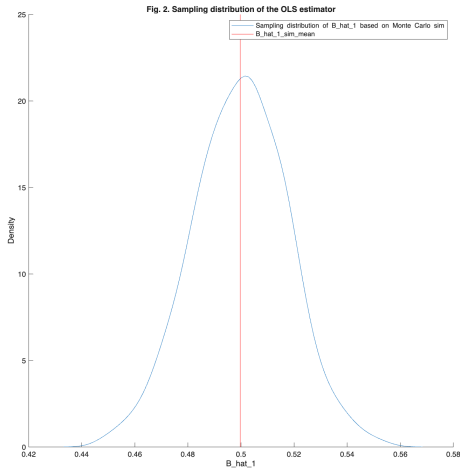
$$\hat{\beta}_1 \sim N\left[\beta_1, \ \frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\right]$$

$$\hat{\beta}_1 \sim N\left[\beta_1, \; \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$$

Notice that the mean comes from unbiasedness as we have just demonstrated. The variance comes from the previous lecture. They are the mean and variance in the population.

Sampling distribution of the OLS estimator in finite samples: The shape of the distribution: Normality: Illustration

Fig. 2. Sampling distribution of the OLS estimator

The simulated sampling distribution of $\hat{\beta}$ appears normal. This is because, in our data generating process, we assumed normally distributed errors!

With a finite number of simulations we considered, the shape is not exactly normal. However, if we increase the amount of repeated sampling, the sampling distribution of $\hat{\beta}$ will converge to normality. This demonstrates that our simulated sampling distribution is an approximation of the true sampling distribution.

Sampling distribution of the OLS estimator in finite samples in finite samples: The shape of the distribution: Normality

In finite sample analysis, the normality of $\hat{\beta}$ follows from the assumption that u is normal and x is fixed.

In large sample analysis, later in these slides, we will obtain an approximate normal distribution for $\hat{\beta}$ without assuming normal errors or fixed x.

So far, we judged the OLS estimator by unbiasedness and efficiency. These properties do not depend on the sample size N.

Now we introduce new criteria that depend on N.

But why do we care about N when evaluating the OLS estimator?

Sampling distribution of the OLS estimator in large samples: Intuition

First reason. As we increase N, $\hat{\beta}$ should come closer to the true population value β.

Who would want an estimator that fails to improve with larger samples?

Sampling distribution of the OLS estimator in large samples: Intuition

Second reason. In small samples, many estimators are biased. A key question is whether such bias disappears when we increase N.

For example, the OLS estimator is biased when a lagged dependent variable appears as a regressor. Yet the bias vanishes asymptotically, when N is very large.

Sampling distribution of the OLS estimator in large samples: Intuition

Third reason. In small samples, it is often very hard to work out the exact properties of an estimator. The problem is that non-linear transformations behave unpredictably with limited data. However, when we increase N, these transformations behave much more regularly: What happens to the statistic also happens to the function of that statistic. This makes large-sample analysis far easier and more useful.

Recall the sampling distribution of $\hat{\beta}$ obtained in the simulation experiment.

Imagine creating a sequence of sampling distributions of $\hat{\beta}$ with successively larger N. If the distributions in this sequence become more and more similar in form to some specific distribution as N becomes very large, this specific distribution is called the asymptotic distribution of $\hat{\beta}$.

If the asymptotic distribution of $\hat{\beta}$ becomes concentrated on the particular value β as N approaches infinity, β is said to be the probability limit of $\hat{\beta}$. We write this in shorthand as

$$\text{plim } \hat{\beta} = \beta$$

This holds. We then say that the OLS estimator

$$\hat{\beta}$$

is a consistent estimator. If the exogeneity assumption holds.

Sampling distribution of the OLS estimator in large samples: Convergence of the distribution: Desirable property: Consistency: Proof

Recall our result

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

Sampling distribution of the OLS estimator in large samples: Convergence of the distribution: Desirable property: Consistency: Proof

Take the probability limit

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{plim } \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}) u_i}{\text{plim } \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Applying the weak law of large numbers to the numerator gives

$$\text{plim}\,\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})u_i = \text{E}\left[(x_i - \mu_x)u_i\right]$$

where μ_x is the population mean of x_i.

Expand the product

$$E\left[(x_i - \mu_x)u_i\right] = E[x_i u_i] - \mu_x E[u_i]$$

Recall from slides of an earlier lecture that

$$E\left[x_i u_i\right] = 0$$

and

$$E\left[u_i\right] = 0$$

if

$$E\left[u_i \mid x_i\right] = 0$$

which is our famous exogeneity assumption.

The probability limit of the denominator is the non-zero constant:

$$E[(x_i - \mu_x)^2]$$

Hence,

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{0}{E[(x_i - \mu_x)^2]} = \beta_1$$

if

$$E[u_i \mid x_i] = 0$$

That is, exogeneity is required for consistency.

We can illustrate consistency using our simulation exercise. The figure below shows how the sampling distribution of $\hat{\beta}$ behaves as N increases from 1,000 to 10,000 and then to 100,000.

As N grows, the distribution becomes more concentrated around the true value β, demonstrating consistency.

In the simulation code, no dependence is modeled between the error u and regressor x. Therefore, the exogeneity assumption

$$\mathsf{E}[u_i \mid x_i] = 0$$

is coded to hold in the data generating process.

Sampling distribution of the OLS estimator in large samples: Convergence of the distribution: Desirable property: Consistency: Illustration
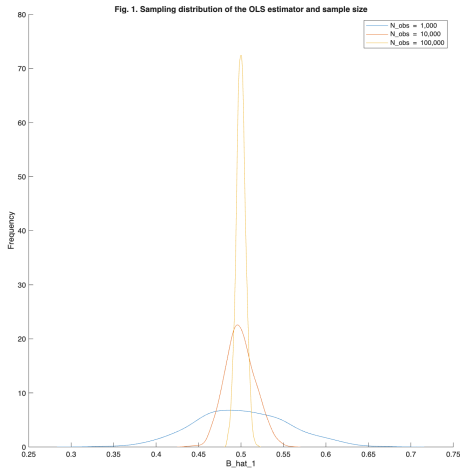
Fig. 1. Sampling distribution of the OLS estimator and sample size

The variance of the sampling distribution of $\hat{\beta}$ in a large sample is called the asymptotic variance of $\hat{\beta}$.

Among the consistent estimators, the asymptotic variance of $\hat{\beta}$ is smaller than the asymptotic variance of any other estimator. Therefore, the OLS estimator

$$\hat{\beta}$$

is asymptotically efficient. If the exogeneity assumption holds.

Sampling distribution of the OLS estimator in large samples: Convergence of the distribution: Desirable property: Most efficient in a large sample: Proof

Suppose that the asymptotic variance of a competitor estimator $\hat{\beta}_0$ is Ω. Then, under the regression model assumptions we made, it can be shown that

$$\text{Asy. Var}\left[\hat{\beta}\right] - \text{Asy. Var}\left[\hat{\beta}_0\right] = \Omega - \frac{\sigma^2}{n}\left(\text{E}\left[x_i x_i'\right]\right)^{-1}$$

$$\leq 0$$

This proof requires the exogeneity assumption holds which is not shown here for brevity. That is, exogeneity is required for asymptotic efficiency.

As with consistency, one could illustrate efficiency using simulation.

The idea is to compare the asymptotic distribution of the OLS estimator with that of an alternative consistent estimator. As N goes to infinity, the OLS sampling distribution has the smallest asymptotic variance under the classical assumptions, including exogeneity and homoskedasticity. This demonstrates the desirable property of asymptotic efficiency.

An illustration will be added later.

Assume that the random sampling assumption holds. Assume that the exogeneity assumption holds. Do not assume normality of u_i. Using the Central Limit Theorem, $\hat{\beta}_1$ is asymptotically normal:

$$\hat{\beta}_1 \overset{a}{\sim} N\left[\beta_1, \frac{1}{n} \frac{\sigma^2}{\mathrm{E}\left[(x_i - \mu_x)^2\right]}\right]$$

where $\mu_x = \mathrm{E}[x_i]$ is the population mean of x_i. This is the asymptotic distribution of $\hat{\beta}_1$. It says that as N increases, the sampling distribution of $\hat{\beta}_1$ approaches normality.

$$\sigma^2$$

and

$$\mathsf{E}\left[(x_i - \mu_x)^2\right]$$

are population quantities of the limiting distribution. They are unobserved.

In practice, they can be estimated, respectively, with

$$\hat{\sigma}^2$$

and with

$$s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

These plug-in estimates are what give us the SEE in practice.

We could conduct a simulation exercise where we increase N from 90 to 9,000. It would show that the sampling distribution of $\hat{\beta}$ becomes approximately normal as N grows. This is not because we assumed normal errors, but because of the Central Limit Theorem: Even with non-normal errors, the distribution of $\hat{\beta}$ would converge to normality in large samples. With a finite number of simulations, the shape is only approximately normal. As we increase the sample size, the distribution would tighten and approach the true asymptotic normal distribution.