OLS Asymptotics

Ani Katchova

Outline

- Consistency
- Comparison of consistency versus unbiasedness
- Asymptotic normality
- Variances and standard errors in large samples

Consistency

- An estimator $\hat{\beta}_j$ is consistent if it converges in probability to the population parameter β_j as the sample size n increases.

$plim\ \hat{\beta}_j = \beta_j$, where $j = 0, 1, \dots, k$ and $n \to \infty$

Another way to express this is:

- $P\left(\left|\hat{\beta}_{jn} - \beta_j\right| < \epsilon\right) \to 1$, where ϵ is any arbitrarily small value.
- The probability *P* that the estimator $\hat{\beta}_j$ obtained from a sample size n will be arbitrarily close to the population parameter β_j goes to 1 as the sample size $n \to \infty$.

Assumption 4: zero conditional mean (review)

- Assumption 4: zero conditional mean
- $E\left(u_i | x_{ji}\right) = 0$
- Expected value of error term given independent variables is zero.
- Under assumptions 1-4 (with the zero conditional mean assumption), the OLS estimator is unbiased.
- $E\left(\hat{\beta}_j\right) = \beta_j$
- Expected value of the coefficient is the population parameter. With many samples, the average value of the coefficient will be the population parameter.

Assumption 4’: regressors are uncorrelated with error term

- Assumption 4’: regressors are uncorrelated with the error term.
- $E(u_i) = 0$ and $cov(x_{ji}, u_i) = 0$
- Assumption 4’ is weaker than assumption 4.
- Under assumptions 1-4’ (with the assumption that the regressors are uncorrelated with the error term), the OLS estimator is consistent.
- $plim\ \hat{\beta}_j = \beta_j$, where $j = 0, 1, \dots, k$
- As the sample size n increases, the sample coefficient $\hat{\beta}_j$ converges in probability to the population parameter β_j.

Consistency of the OLS estimator (proof)

- Simple regression model: $y = \beta_0 + \beta_1 x_1 + u$

- The coefficient $\hat{\beta}_1 = \frac{cov(x_1, y)}{var(x_1)} = \frac{cov(x_1, \beta_0 + \beta_1 x_1 + u\ )}{var(x_1)} =$

$$= \frac{cov(x_1, \beta_0\ )}{var(x_1)} + \frac{cov(x_1, \beta_1 x_1)}{var(x_1)} + \frac{cov(x_1, u\ )}{var(x_1)} = \beta_1 \frac{cov(x_1, x_1)}{var(x_1)} + \frac{cov(x_1, u)}{var(x_1)}$$

$$= \beta_1 + \frac{cov(x_1, u)}{var(x_1)}$$

- Using assumption 4’ that $cov(x_1, u) = 0$, then
- $plim\ \hat{\beta}_1 = \beta_1$
- The OLS estimator is consistent under assumptions 1-4’.

Consistency versus unbiasedness

Property	Undesirable property	Formula	Assump-tions	Sample size	Distribution
unbiasedness	biasedness	$E(\hat{\beta}_j) = \beta_j$	1-4	Small/any sample	t, normal
consistency	inconsistency	$plim\ \hat{\beta}_j = \beta_j$	1-4’	Large sample	asymptotically normal

- The OLS estimator may be biased in small samples with assumptions 1-4’. The stronger assumption 4 zero conditional mean is needed for the OLS estimator to be unbiased.
- Unbiasedness is ideal and holds with any sample size. But if unbiasedness cannot be achieved in a small sample, then at least consistency can be achieved with a large sample.

Omitted variable bias (review)

- The “true” population regression model is: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$
- We need to estimate: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$
- But instead we estimate a misspecified model: $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$, where x_2 is the omitted variable from this model.
- If x_1 and x_2 are correlated, there will be a relationship between them

$$x_2 = \delta_0 + \delta_1 x_1 + v$$

Substitute in above equation to get:

$$y = \beta_0 + \beta_1 x_1 + \beta_2(\delta_0 + \delta_1 x_1 + v) + u$$
$$= (\beta_0 + \beta_2\delta_0) + (\beta_1 + \beta_2\delta_1)x_1 + (\beta_2 v + u)$$

The coefficient that will be estimated for x_1 when x_2 is omitted will be biased.

Omitted variable bias (review)

- An unbiased coefficient is when $E\left(\hat{\beta}_1\right) = \beta_1$, but this coefficient is biased because $E\left(\tilde{\beta}_1\right) = \beta_1 + \beta_2\ \delta_1$, where $\beta_2\delta_1$ is the bias.
- With an omitted variable, the coefficient will not be biased if
  - $\beta_2 = 0$. Looking at $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$, this means that x_2 does not belong in the model (x_2 is irrelevant).
  - $\delta_1 = 0$. Looking at $x_2 = \delta_0 + \delta_1 x_1 + v$, this means that x_2 and x_1 are not correlated.
  - In other words, if the omitted variable x_2 is irrelevant $\beta_2 = 0$ or uncorrelated $\delta_1 = 0$, there will be no omitted variable bias.

Omitted variable bias – asymptotic analog

- A consistent coefficient is when $plim\ \hat{\beta}_1 = \beta_1$, but this coefficient is inconsistent because $plim\ \hat{\beta}_1 = \beta_1 + \beta_2\ \delta_1$, where $\beta_2\delta_1$ is the bias.
- With an omitted variable, the coefficient will be consistent if
  - $\beta_2 = 0$. Looking at $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$, this means that x_2 does not belong in the model (x_2 is irrelevant).
  - $\delta_1 = 0$. Looking at $x_2 = \delta_0 + \delta_1 x_1 + v$, this means that x_2 and x_1 are not correlated. Note that $\hat{\delta}_1 = \frac{cov(x_2, x_1)}{var(x_1)}$
  - In other words, if the omitted variable x_2 is irrelevant $\beta_2 = 0$ or uncorrelated $\delta_1 = 0$, the coefficient will be consistent.

Asymptotic normality

- Under assumptions 1-6 (Gauss Markov assumptions and normality), the coefficients have normal sampling distribution.
- $\hat{\beta}_j \sim normal(\beta_j, var\left(\hat{\beta}_j\right))$
- Under assumptions 1-5 (Gauss Markov assumptions), the coefficients have asymptotically normal sampling distribution.
- $\hat{\beta}_j \overset{a}{\sim} normal\left(\beta_j, var\left(\hat{\beta}_j\right)\right)$
- In other words, even without normality of errors, asymptotic normality is achieved in large samples.

Asymptotic normality

- The normality assumption 6 does not always hold in practice.
- In small samples, the normality assumption 6 is needed for the t-tests and F-tests to be valid.
- In large samples, without the normality assumption 6:
  - The OLS estimators are normal.
  - The t-tests and F-tests are valid.
  - Note that assumptions 1-5 are still needed.

OLS properties

- OLS properties that hold for any sample
  - Expected values and unbiasedness under assumptions 1-4 (linearity, random sample, no perfect collinearity, zero conditional mean)
  - Variance formulas under assumptions 1-5 (linearity, random sample, no perfect collinearity, zero conditional mean, homoscedasticity)
  - Gauss-Markov theorem (BLUE) under assumptions 1-5
  - Exact sampling distributions (t-test and F-test) under assumptions 1-6 (Gauss-Markov assumptions + normality)
- OLS properties that hold in large samples (asymptotics)
  - Consistency under assumptions 1-4’ (regressors uncorrelated with error term)
  - Asymptotic normality/tests under assumptions 1-5

Variances of OLS estimators - asymptotics

Recall, the variance of the OLS estimator:

$$var(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{SST_j(1 - R_j^2)} = \frac{\hat{\sigma}^2}{n * var(x_j)\,(1 - R_j^2)}$$

- $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2 = n * var(x_j)$ is the total sampling variance of variable x_j. As the sample size increases, it grows with n.
- R_j^2 is the R-squared from a regression of x_j on all other independent variables. Converges to a fixed number.
- $\hat{\sigma}^2$ is the variance of the residual. Converges to σ^2.
- As the sample size increases, $var(\hat{\beta}_j)$ changes by $1/n$ and $se(\hat{\beta}_j)$ changes by $\sqrt{1/n}$.

Standard errors and sample size example

- Regression model:
- $wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$
- Estimate regression model with full sample and with half sample.
- Note the standard errors for coefficient on experience and sample sizes.
- As sample size n increases, standard errors change at a rate of $\sqrt{1/n}$
- With larger sample size, standard errors are lower, leading to more significance of the coefficients.

Standard errors and sample size example

	Model with full sample	Model with half the sample
VARIABLES	wage	wage
educ	0.599***	0.732***
	(0.0513)	(0.0875)
tenure	0.169***	0.209***
	(0.0216)	(0.0359)
exper	0.0223*	0.0477**
	(0.0121)	(0.0194)
Constant	-2.873***	-4.815***
	(0.729)	(1.237)
Observations	526	262
R-squared	0.306	0.329

For the coefficient on experience, se1=0.0121, se2=0.0194, n1=526, n2=262.

The ratios se1/se2=0.62 and $\sqrt{n2/n1}$=0.71 are almost the same.

With the full sample (double the sample size), standard errors are 62% lower.

Review questions

- Define consistency.
- Compare consistency and unbiasedness.
- Define asymptotic normality.
- At what rate does the variance change when the sample size changes?