

ECON 5340 Class Notes

Chapter 4. Finite-Sample Properties of the LS Estimator

1 Introduction

We now wish to compare the properties of the LS estimator to other potential estimators. This analysis will be exact and will hold for any sample size.

2 Gauss-Markov Theorem

The **Gauss-Markov Theorem** states that, provided the Classical assumptions hold, the ordinary least squares (OLS) estimator b is the minimum variance estimator among all linear unbiased estimators. Sometimes it is said that the OLS estimator is BLUE (Best Linear Unbiased Estimator).

Proof.

First, we need to show that b is unbiased. We know

$$b = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + \epsilon) = \beta + (X'X)^{-1}X'\epsilon.$$

Taking expectations gives

$$E[b] = \beta + (X'X)^{-1}X'E[\epsilon] = \beta$$

because β and X are not random variables (recall X is assumed to be fixed in repeated sampling). Therefore, b is an unbiased estimator of β .

Second, we need to show that b has the smallest variance (among all linear unbiased estimators). Begin by noting that b is a linear estimator because it is linear in Y (or alternatively ϵ). Now consider all other possible linear unbiased estimators $b_0 = CY$, where C is a fixed $(k \times n)$ matrix. For b_0 to be unbiased, it must be that $CX = I$ because

$$E[b_0] = E[CX\beta + C\epsilon] = CX\beta.$$

The variance of b is

$$\begin{aligned} \text{var}(b) &= E[(b - \beta)(b - \beta)'] \\ &= E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] \\ &= (X'X)^{-1}X'\sigma^2IX(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}. \end{aligned}$$

The variance of b_0 is

$$\begin{aligned} \text{var}(b_0) &= E[(b_0 - \beta)(b_0 - \beta)'] \\ &= E[C\epsilon\epsilon'C'] \\ &= \sigma^2CC'. \end{aligned}$$

The question is now whether $(X'X)^{-1}$ or CC' is bigger (in a matrix sense). Toward that end, define $D \equiv C - (X'X)^{-1}X'$ so that $DX = 0$. Using this, we can write

$$\begin{aligned} \text{var}(b_0) &= \sigma^2(D + (X'X)^{-1}X')(D + X'X)^{-1}X' \\ &= \sigma^2(X'X)^{-1} + \sigma^2DD' \\ &= \text{var}(b) + \sigma^2DD'. \end{aligned}$$

Finally, we note that DD' is a non-negative definite matrix (Greene A-114) so that the variance of b is no larger than the variance b_0 . ▼

A few notes about the Gauss-Markov Theorem:

- Notice that no distributional assumptions were necessary.
- b is a random variable with a sampling distribution and associated sampling variance, $\text{var}(b)$.
- If $\epsilon|X \sim N(0, \sigma^2I)$ then $b \sim N(\beta, \sigma^2(X'X)^{-1})$ and is the best among all linear and nonlinear estimators.
- Keep in mind that some biased estimators may have smaller MSEs.
- If b is BLUE for β , then any linear combination of b_i s is BLUE for the same linear combination of β_i s.

- The Gauss-Markov Theorem holds for stochastic regressors as well, provided the standard assumptions for X and ϵ are made.

3 Estimating the Variance of the LS Estimator

We know $\text{var}(b) = \sigma^2(X'X)^{-1}$, but σ^2 is an unknown parameter. Therefore in order to find $\widehat{\text{var}}(b)$, we need to find a good estimator for σ^2 .

Start by defining $M = I - X(X'X)^{-1}X'$, which is symmetric and idempotent. This matrix can be used to relate the residuals e to the errors ϵ

$$e = MY = M(X\beta + \epsilon) = M\epsilon.$$

Using this relation, we can then find an unbiased estimator for σ^2 . Begin by finding the expectation of the inner product of e

$$E[e'e] = E[\epsilon'M\epsilon] = E[\text{tr}(\epsilon'M\epsilon)].$$

The last equality uses the fact that the trace (tr) of a scalar is simply the scalar. This can be further manipulated to give

$$E[\text{tr}(\epsilon'M\epsilon)] = E[\text{tr}(\epsilon\epsilon'M)]$$

using (Greene A-94). Taking expectations through the trace operator then gives

$$E[\text{tr}(\epsilon\epsilon'M)] = \sigma^2 \text{tr}(M) = \sigma^2 [\text{tr}(I_n) - \text{tr}(X(X'X)^{-1}X')]$$

which after using (Greene A-94) again produces

$$\sigma^2 [\text{tr}(I_n) - \text{tr}(X(X'X)^{-1}X')] = \sigma^2 [n - \text{tr}((X'X)^{-1}(X'X))] = \sigma^2 [n - k].$$

Therefore, if we define $s^2 = e'e/(n - k)$, we know it will be an unbiased estimator for σ^2 and $\widehat{\text{var}}(b) = s^2(X'X)^{-1}$. The square root of the estimated variance of b is often called the **standard error of b** .

4 Inference in Least Squares Regressions

Now reintroduce Classical assumption #6: $\epsilon \sim N(0, \sigma^2 I)$. Because b is a linear function of ϵ , we then know that $b \sim N(\beta, \sigma^2(X'X)^{-1})$. In this section, we will take a first pass at testing some simple hypotheses regarding the population regression equation.

4.1 Single Coefficient Tests

Consider two different cases – σ^2 is known to the econometrician, and more realistically, σ^2 is not known to the econometrician. We wish to test whether a single element of the β vector is equal to zero or not.

- σ^2 known. Assume we are testing the null hypothesis, $H_0: \beta_k = 0$, against the alternative hypothesis, $H_A: \beta_k \neq 0$. We can form the statistic

$$z_k = \frac{b_k - \beta_k}{\sigma \sqrt{s^{kk}}},$$

where s^{kk} is the k^{th} diagonal element of $(X'X)^{-1}$. Under the null hypothesis, z_k will have a standard normal $N(0, 1)$ distribution.

- σ^2 unknown. Testing the same hypothesis, we want to replace σ with s . Doing so, means z_k no longer has a standard normal distribution. Instead, we form the statistic

$$t_k = \frac{b_k - \beta_k}{s_{bk}}$$

where $s_{bk} = s\sqrt{s^{kk}}$, which has a student's t distribution with $n - k$ degrees of freedom. This is motivated by the fact that the ratio of z_k (standard normal) to the square root of $(n - k)s^2/\sigma^2$ (chi-squared distribution) divided by its degrees of freedom has a student's t distribution. This also relies on independence between the z_k and $(n - k)s^2/\sigma^2$.

4.2 Confidence Intervals

Confidence intervals provide a convenient method of presenting the same information as in the hypotheses tests above. Below are the confidence intervals for two population parameters, β_k and σ^2 .

- The $(1 - \lambda) * 100$ confidence interval for β_k can be found by noting that $\Pr(-t_{\lambda/2} < (b_k - \beta_k)/s_{bk} < t_{\lambda/2}) = 1 - \lambda$. Rearranging, the confidence interval for β_k is $[b_k - t_{\lambda/2}s_{bk}, b_k + t_{\lambda/2}s_{bk}]$.

- The $(1 - \lambda) * 100$ confidence interval for σ^2 can be found by noting that $\Pr(\chi_{1-\lambda/2}^2 < (n - k)s^2/\sigma^2 < \chi_{\lambda/2}^2) = 1 - \lambda$. Rearranging, the confidence interval for σ^2 is $[(n - k)s^2/\chi_{\lambda/2}^2, (n - k)s^2/\chi_{1-\lambda/2}^2]$.

4.3 Overall Significance

Now assume that we wish to assess the overall significance of the regression model. That is, we want to test whether or not the explanatory variables explain a significant amount of the variation in the dependent variable. The null hypothesis in this case will be $H_0: \beta_2 = \dots = \beta_k = 0$. The alternative hypothesis is that the null is false. Not surprisingly, we can use the R^2 to execute the test. The test statistic is

$$F = \frac{R^2/(k - 1)}{(1 - R^2)/(n - k)}$$

which under the null hypothesis has an F distribution with $k - 1$ and $n - k$ degrees of freedom in the numerator and denominator, respectively. A couple of notes.

- If the true model is $y_i = \beta_1 + \epsilon_i$, then no variation in y_i around its mean of β_1 is explained and $R^2 = 0$.
- It is possible for all individual coefficients to be insignificant while jointly they are significant.

4.4 Example. Prison Population - Violent Crimes cont.

In this example, we will use Gauss to perform three different hypothesis tests on the regression equation

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i$$

where the new variable x_{3i} is the unemployment rate for the i^{th} state.

1. Since more violent crimes and higher unemployment should lead to a higher prison population, we test

$$H_0 : \beta_2 \leq 0 \text{ versus } H_A : \beta_2 > 0$$

$$H_0 : \beta_3 \leq 0 \text{ versus } H_A : \beta_3 > 0.$$

The critical t value with $df = n - k = 48$ at the 5% significance level is 1.68. The estimated t values are 10.32 and -1.36 respectively. As a result, we reject the first null in favor of a positive relationship

between violent crimes and the prison population, and fail to reject the second null that there is a non-positive relationship between the unemployment rate and the prison population.

2. The 95% confidence interval for β_2 is $[0.038, 0.056]$. The interval is centered on the point estimate $b_2 = 0.047$, which implies that each violent crime leads to a 0.47 increase in the prison population, holding constant the unemployment rate.
3. The overall significance test is

$$H_0 : \beta_2 = \beta_3 = 0$$

$$H_A : H_0 \text{ false.}$$

The 5% critical F value with $k - 1 = 2$ degrees of freedom in the numerator and $n - k = 48$ degrees of freedom in the denominator is 3.21. The estimated F value is 53.46. Therefore, we reject the null that all the slope coefficients are jointly zero. The model does indeed explain a significant amount of the variation in the prison population across states.

See [Gauss example 4.1](#) for more details.

5 Data Problems

Data problems such as multicollinearity, missing observations and outliers will be covered later.