

Professor Nagler

Notes on Ordinary Least Squares Estimates

Copyright - Jonathan Nagler; Updated: Oct 18, 2006

1 The True Model

We start out with the following “*true*” linear model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (1)$$

Each i denotes an observation. Hence if you have the following data:

Y	X
10	1
20	3
30	4
50	6

We would have: $Y_1 = 10$, $X_1 = 1$, $Y_2 = 20$, $X_2 = 3$, and so on.

This model is assumed to be an accurate representation of the real world. β_0 and β_1 are the *population parameters* of interest.

ϵ_i is the disturbance, or error, term. We speak of the model as having a systemic component ($\beta_0 + \beta_1 X_i$) and a stochastic, or random, component (ϵ_i).

$$E(Y|X_i) = \beta_0 + \beta_1 X_i$$

This suggests that for any value of X , the model predicts the expected value of Y . It also suggests that if X changes by some amount ΔX , we can compute the corresponding change ΔY .

We want to find their “*true*” values, i.e., the values of β_0 and β_1 that we would get if we had all the data there has been and all the data there will ever be on X and Y .

2 The Concept of Estimates

Our estimates of β_0 and β_1 will be called $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively. (They are pronounced ‘Beta-nought-hat’ and ‘Beta-one-hat’.) Once we have estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$, we can compute a predicted, or estimated, value for Y .

That predicted value is called \hat{Y} , and is generated by the following equation:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad (2)$$

Now this predicted value of Y will not necessarily be equal to the true (or observed) value of Y . In general we will be off by some amount e , called the residual (or error term). So we can alternatively write:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i \quad (3)$$

or:

$$Y_i - \hat{Y}_i = e_i \quad (4)$$

Major Point: $e \neq \epsilon$

The residuals (e) are things we can observe. The disturbances (u) are not observed.

3 Criteria for Estimates

To compute our estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ we pick the $\hat{\beta}_0$ and $\hat{\beta}_1$ that will minimize the sum of the squares of our errors; i.e., the $\hat{\beta}_0$ and $\hat{\beta}_1$ that will minimize:

$$\sum e_i^2 \quad (5)$$

For our purposes we do not need to know how to find the $\hat{\beta}_0$ and $\hat{\beta}_1$ that satisfy that criteria, a statistics package (STATA, Limdep, whatever) will do it for us. Just for information though,

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (6)$$

where

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \quad (7)$$

These are the “*OLS estimators*” $\hat{\beta}_0$ and $\hat{\beta}_1$.

Notice that $\hat{\beta}_0$ and $\hat{\beta}_1$ are linear functions of the observed data: the random variables X and Y . Since $\hat{\beta}_0$ and $\hat{\beta}_1$ are linear combinations of existing random variables, they themselves are random variables with some straightforward properties

4 Properties of the OLS Estimators

These 5 properties do not depend upon any assumptions - they are simply algebraic facts that $\hat{\beta}_0$ and $\hat{\beta}_1$ will satisfy if we compute them as given in equation 6 and equation 7 above. Just as $2 + 2 = 4$, these properties will always be true if you calculate $\hat{\beta}_0$ and $\hat{\beta}_1$ as above.

The properties are:

1. The regression line defined by $\hat{\beta}_0$ and $\hat{\beta}_1$ passes through the means of the observed values (\bar{X} and \bar{Y}).
[$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1\bar{X}$]
2. The mean of the predicted Y's for the sample will equal the mean of the observed Ys for the sample.
[$\hat{\bar{Y}} = \bar{Y}$]
3. The sample mean of the residuals will be 0. [$\bar{e}_i = 0$]
4. The correlation between the residuals and the predicted values of Y will be 0. [$\rho(e_i, \hat{Y}_i) = 0$]
5. The correlation between the residuals and the observed values of X will be 0. [$\rho(e_i, X_i) = 0$]

So far we know nothing about $\hat{\beta}_0$ and $\hat{\beta}_1$ except that they satisfy all of the properties discussed above. We need to make some assumptions about the true model in order to make inferences regarding β_0 and β_1 , the true parameters, from $\hat{\beta}_0$ and $\hat{\beta}_1$. $\hat{\beta}_0$ and $\hat{\beta}_1$ just come from our sample; the goal is to learn about the *true* parameters.

5 The Gaus Markov Assumptions:

$$\mathbf{A1)} \ E[(\epsilon_i|X_i)] = 0$$

For any value of X , the disturbances average out to 0. Hence $E[Y_i|X_i] = \beta_0 + \beta_1 X_i$.

$$\mathbf{A2)} \ \text{Cov}(\epsilon_i, \epsilon_j) = 0 \ \forall i \neq j.$$

We haven't defined the covariance before. By definition,

$$\text{cov}(\epsilon_i, \epsilon_j) = E[(\epsilon_i - E(\epsilon_i))(\epsilon_j - E(\epsilon_j))] \quad (8)$$

Now by assumption 1, this reduces to:

$$= E(\epsilon_i \epsilon_j) \quad (9)$$

Knowing something about the disturbance term for 1 observation, tells you nothing about the disturbance term in any other observation.

$$\mathbf{A3)} \text{Var}(\epsilon_i | X_i) = \sigma^2 \quad \forall i.$$

By definition,

$$\text{Var}(\epsilon_i | X_i) = E[(\epsilon_i - E(\epsilon_i))^2] \quad (10)$$

Which, by assumption 1

$$= E(\epsilon_i^2) \quad (11)$$

This is known as the homoscedasticity assumption, the variance of ϵ_i is the same for each X_i . Naturally enough, this suggests that we refer to conditions where $\text{Var}(\epsilon_i | X_i) = \sigma_i^2$ where $\sigma_i^2 \neq \sigma_j^2$ as conditions of heteroscedasticity. Heteroscedasticity is a violation of the Gauss-Markov Assumptions, and hence is something that has to be corrected for when it is encountered.

$$\mathbf{A4)} \text{Cov}(\epsilon_i, X_i) = 0$$

Again, by definition,

$$\text{cov}(\epsilon_i, x_i) = E[(\epsilon_i - E(\epsilon_i))(X_i - E(X_i))] \quad (12)$$

Since $E(\epsilon_i) = 0$ (by assumption 1), this simplifies to

$$\text{cov}(\epsilon_i, x_i) = E(\epsilon_i X_i). \quad (13)$$

This means that the disturbances are *exogenous*.

A5) The regression model is properly specified.

This just means that the “true” model that you specified really is the true model. We will deal with this later in the course.

6 BLUE (Defn)

An estimator $\hat{\beta}_1$ is BLUE (best linear unbiased estimator) if:

- i) It is a linear function of the random variables.
- ii) $E(\hat{\beta}_1) = \beta_1$. (it is unbiased)
- iii) $\text{Var}(\hat{\beta}_1) < \text{Var}(\beta_1^*) \forall \beta_1^*$ that are linear and unbiased.

7 Gauss-Markov Theorem:

The Gauss-Markov Theorem: If the Gauss-Markov assumptions are true, then the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are BLUE.

8 Properties of the OLS Estimators Under the G-M Assumptions

If the Gaus-Markov assumptions are satisfied, then the variance of the OLS estimators is given by:

$$\sigma_{\hat{\beta}_1}^2 = \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \quad (14)$$

and:

$$\sigma_{\hat{\beta}_0}^2 = \frac{\sum X_i^2}{N \sum (X_i - \bar{X})^2} \sigma^2 \quad (15)$$

$$(16)$$

Where σ^2 is the variance of the disturbance term u .

We can not observe σ^2 , but we can estimate σ^2 by:

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{N - 2} \quad (17)$$

9 Review

STOP: At this point we know the variance of the OLS estimators*, and that they are BLUE. *We do NOT know the distribution of the OLS estimators.* Without knowing the distribution we can not use our z-table or t-table to calculate confidence intervals.

* While we ‘know’ the variance of the OLS estimators, we can only express it as a function of an *unobserved* value: σ^2 . However, we can *estimate* the variance of the OLS estimates by using $\hat{\sigma}^2$ as an estimate of σ^2 .

10 Normality Assumption:

We generally assume: $\epsilon_i \sim IN(0, \sigma^2)$, in other words: the disturbance terms are independent of each other, and normally distributed with mean 0 and variance σ^2 .

Given the Gauss-Markov assumptions, and the Normality Assumption, then the following holds:

$$\hat{\beta}_0 \sim N(\beta_0, \sigma_{\hat{\beta}_0}^2) \quad (18)$$

and

$$\hat{\beta}_1 \sim N(\beta_1, \sigma_{\hat{\beta}_1}^2) \quad (19)$$

Now we can draw inferences about the true parameters β_0 and β_1 given the estimated values $\hat{\beta}_0$ and $\hat{\beta}_1$, as well as our estimates of $\sigma_{\hat{\beta}_0}^2$ and $\sigma_{\hat{\beta}_1}^2$.

11 Precision of The OLS Estimators

What determines the precision of the Estimates?

Go back to equations (14), (15), and (16).

We would like $\sigma_{\hat{\beta}_1}^2$ to be small; i.e., we want there to be very little variance on our estimate $\hat{\beta}_1$ because then we would be relatively certain that the true β_1 is very close to $\hat{\beta}_1$.

There are *two* things that influence $\sigma_{\hat{\beta}_1}^2$: σ^2 and $\Sigma(x_i - \bar{x})^2$. This last term is the variance of the independent variable.

Notice that the *more* variance on X , the *lower* the variance about $\hat{\beta}_1$; or, the more variance about the X the more precise will be our estimate of $\hat{\beta}_1$.

The other term influencing $\sigma_{\hat{\beta}_1}^2$ is σ^2 , the variance of the regression. The lower σ^2 , the lower $\sigma_{\hat{\beta}_1}^2$.

So the better our model ‘fits’ the data, the more precisely will we be able to estimate $\hat{\beta}_1$.

This is intuitively reasonable: large σ^2 denotes a large disturbance, or lots of ‘noise.’ It should not be surprising that lots of noise makes it difficult to discern the true β_1 . Both of these traits are easy to understand graphically.

12 Common Violations of GM

12.1 Heteroscedasticity:

Symptoms: Heteroscedasticity violates **GM** assumption **A3** above. If **GM** is violated we know our OLS estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are not BLUE. But what is wrong with them? Our estimates are still unbiased, but they are not minimum variance. And employing the OLS estimation technique will give us estimates of the variance for $\hat{\beta}_0$ and $\hat{\beta}_1$ that are **wrong!** Moreover, we will not know if our estimated variances are too large or too small. Put another way, we will have no clue what the correct variances of $\hat{\beta}_0$ and $\hat{\beta}_1$ are. Obviously this makes inferences about the true parameters β_0 and β_1 impossible. **Note:** the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased in the presence of heteroscedasticity!

When to Expect Heteroscedasticity: Look for heteroscedasticity primarily when the units of analysis are of different size. Or, look for it when different time-periods or different regions are being combined in one regression.

Tests for Heteroscedasticity: There are several tests for heteroscedasticity. The most common one is to ‘eye-

ball' a plot of the residuals against the independent variables. If there is any kind of pattern then heteroscedasticity may be present. There are also analytic tests available. The Parks test (described in Gujarati) is easily implemented.

Solution: Weighted Least Squares (**WLS**) is the usual solution to heteroscedasticity.

13 Non-Violations of GM

13.1 Multicollinearity

Symptoms: Perfect multicollinearity would violate GM. However, the phrase 'multicollinearity' does not denote perfect multicollinearity (i.e., one of the independent variables is a linear combination of the other independent variables). Rather 'multicollinearity' denotes a condition where two or more of the independent variables are highly correlated. Under such a scenario the OLS estimates are still BLUE! However, they may be 'unstable'. By unstable we mean that they may be particularly sensitive to model specification, or to outliers in the data. Moreover, the standard errors of the estimates will be higher than they would be in the absence of multicollinearity. In other

words, it will be harder to produce significant coefficients. However, if in the face of this handicap you still do produce significant coefficients then perhaps you do not have to worry about multicollinearity?

Readily Observable Symptoms: If you have ‘high’ R^2 , or a low standard error of the regression, but very few of your coefficients are significant, then this indicates that you probably have multicollinearity.

Tests for Multicollinearity: You can first look at a correlation matrix comparing the bivariate correlations of all of your independent variables. Then you can proceed by estimating separate regressions of *each* of your independent variables against all of the remaining independent variables. The goal is to determine if one of your independent variables is a linear combination – – or close to it – – of your other independent variables.

Solution: You can gather more data. You can drop some variables. You can test for robustness. You can live with it.

14 The OLS Estimator in Matrix Form for Multiple Regression

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (20)$$

$$\text{Var} - \text{Cov}(\hat{\beta}) = (X'X)\sigma^2 \quad (21)$$

15 Another Look at Multicollinearity and Var-Cov($\hat{\beta}$)