Handout 3 – Single-Equation LM and OLS

Chunyu Qu

Oct 18, 2021

1. Properties of OLS on Any Sample of Data

1.1. Algebraic Properties of OLS Statistics

- **Properties related with residuals**

$$\Sigma_{n=1}^{n}\,\widehat{u_\iota} = 0$$

$$\Sigma_{n=1}^{n}\,x_i\widehat{u_\iota} = 0$$

$$\Sigma_{n=1}^{n}\,\widehat{u_\iota}(\widehat{y_\iota} - \bar{y}) = 0$$

Where

$$y_i = \widehat{y_\iota} + \widehat{u_\iota}$$

1.2. Goodness of Fit

- **R-square and Adjusted R-square**

  Total Sum of Squares: $SST = \Sigma_{n=1}^{n}\,(y_i - \bar{y})^2$

  Regression Explained Sum of Squares: $SSR = \Sigma_{n=1}^{n}\,(\widehat{y_i} - \bar{y})^2$

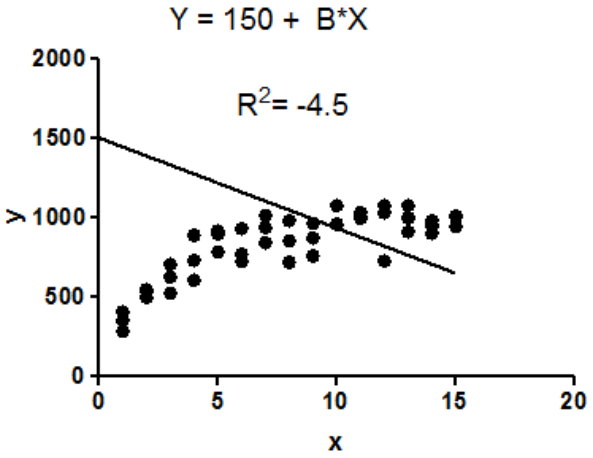
  Residual Errors Sum of Squares: $SSE = \Sigma_{n=1}^{n}\,\widehat{u_\iota}^2$

  $SST = SSE + SSR$

  $$\boldsymbol{R^2} = \frac{SSR}{SST}, R_{adj}^2 = 1 - \frac{\frac{SSR}{n}}{\frac{SST}{n}}$$

- **Key Notes**
  - Remark 1: $R^2 = \frac{SSR}{SST}$, is the ratio of the explained variation compared to the total variation; thus, it is interpreted as the fraction of the sample variation in y that is explained by x.
  - Remark 2: It is worth emphasizing now that a seemingly low R-squared does not necessarily mean that an OLS regression equation is useless. whether or not an OLS is useful does not depend directly on the size of R-squared.

- Remark 3: you may see negative R^2 sometimes, which due to the use of another definition $R^2 = 1 - \frac{SSE}{SST}$. It is negative only when the chosen model does not follow the trend of the data, so fits worse than a horizontal line. For example below, the model makes no sense at all given these data. It is clearly the wrong model, perhaps chosen by accident. A negative R^2 is not a mathematical impossibility or the sign of a computer bug. It simply means that the chosen model (with its constraints) fits the data poorly.



- Remark 4: R^2 never decreases, and it usually increases, when another independent variable is added to a regression and the same set of observations is used for both regressions. This algebraic fact follows because, by definition, the sum of squared residuals never increases when additional regressors are added to the model.
- Remark 5: An important caveat to the previous assertion about R^2 is that it assumes we do not have missing data on the explanatory variables. If two regressions use different sets of observations, then, in general, we cannot tell how the $R^{2\prime}$s will compare
- Remark 6: The adjusted R-squared ($\overline{R^2}$) is sometimes called the corrected R-squared, but this is not a good name because it implies that it is somehow better than R^2 as an estimator of the population R-squared. However, ($\overline{R^2}$ is not generally known to be a better estimator. It is tempting to think that $\overline{R^2}$ corrects the bias in $R^2$for estimating the population R-squared ρ^2, but it does not: the ratio of two unbiased estimators is not an unbiased estimator.
- Remark 7: The primary attractiveness $\overline{R^2}$ s that it imposes a penalty for adding additional independent variables to a model. We can use adjusted r-squared to choose between nonnested models.

1.3. Linear models are not simply linear

- Remark 8: Linear regression model, yet, also allows for certain nonlinear relationships. It depicts a main relationship in linear framework, not accurate sometimes but very efficiently interpretable. In reality, you may never know the true relationship. Digging into the complex methodologies help little with the bias and consistency in a interpreting framework but increase the computing burden dramatically.
- Remark 9: Whereas the mechanics of simple regression do not depend on how y and x are defined, the interpretation of the coefficients does depend on their definitions. For successful empirical

work, it is much more important to become proficient at interpreting coefficients than to become efficient at computing formulas]

2. Functional Form Misspecification

2.1. Omitted Variable Bias

- *Ex 1: Simple Case. suppose that wage is determined by*

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + u$$

Because ability is not observed, we instead estimate the model

$$wage = \widetilde{\beta_0} + \widetilde{\beta_1} educ + v$$

$where\ v = \beta_2 abli + u$

When $educ$ and $abil$ are correlated, there is omitted bias. For example, the bias in $\widetilde{\beta_1}$ is positive if $\beta_2 > 0$, where the bias is defined as $Bias(\widetilde{\beta_1}) = E[\widetilde{\beta_1}] - \beta_1$

Here is a summary

	$\text{Corr}(x_1, x_2) > 0$	$\text{Corr}(x_1, x_2) < 0$
$\beta_2 > 0$	Positive bias	Negative bias
$\beta_2 < 0$	Negative bias	Positive bias

- If $E[\widetilde{\beta_1}] > \beta_1$, then we say that $\widetilde{\beta_1}$ has an upward bias. When $E[\widetilde{\beta_1}] < \beta_1$, then we say that $\widetilde{\beta_1}$ has an downward bias.

- *Ex 2: Multiple regressors. Suppose the population model is*

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

But we omit x_3 and estimate the model as

$$\tilde{y} = \widetilde{\beta_0} + \widetilde{\beta_1} x_1 + \widetilde{\beta_2} x_2$$

Now, suppose that x_2 and x_3 are uncorrelated, but x_1 and x_3 are correlated. Our first intuition that $\widetilde{\beta_1}$ is biased but $\widetilde{\beta_2}$ is unbiased is invalid. Both $\widetilde{\beta_1}$ and $\widetilde{\beta_2}$ ill normally be biased. The only exception to this is when x_1 and x_2 are uncorrelated as well. It can be difficult to obtain the direction of bias in $\widetilde{\beta_1}$ and $\widetilde{\beta_2}$. This is because x1, x2, and x3 can all be pairwise correlated.

- *Ex 3: Omitted Terms. Suppose the population model is*

$$\log wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 female + \beta_5 female \times edu + u$$

If we omit the interaction term, female # educ, then we are misspecifying the functional form. In general, we will not get unbiased estimators of any of the other parameters, and because the return to education depends on gender, it is not clear what return we would be estimating by omitting the interaction term.

2.2. Using Proxy Variables for Unobserved Explanatory Variables

- *Ex 4: Omitted Terms. Suppose the population model is*

$$\log wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 abil + u$$

  - This model shows explicitly that we want to hold ability fixed when measuring the return to educ and exper. Our primary interest is in the slope parameters β_1 and β_2. We do not really care whether we get an unbiased or consistent estimator of the intercept. Also, we can never hope to estimateβ_3 because abil is not observed; in fact, we would not know how to interpret β_3.
    One possibility is to obtain a proxy variable for the omitted variable. Loosely speaking, a proxy variable is something that is related to the unobserved variable that we would like to control for in our analysis.
    In the wage equation, one possibility is to use the intelligence quotient, or IQ, as a proxy for ability. This does not require IQ to be the same thing as ability; what we need is for IQ to be correlated with ability, something we clarify in the following discussion. All of the key ideas can be illustrated in a model with three independent variables, two of which are observed:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3^* + u$$

    Where we know that IQ is related to ability as

$$x_3^* = \delta_0 + \delta_3 x_3 + v_3$$

  - If $\delta_3 = 0$, then x_3 is not a proper proxy for x_3.
  - How can we use x3 to get unbiased (or at least consistent) estimators of b1 and b2?
    The proposal is to pretend that treat x_3 and x_3^* the same. We call this the plug-in solution to the omitted variables problem. Since they not the same, we should determine when this procedure does in fact give consistent estimators of β_1 and β_2.
  - We need two conditions

$$E[abil|edu, exper, IQ] = E[abil|IQ] = \delta_0 + \delta_3 IQ$$

    And

$$\delta_3 \neq 0$$

  - The effect of IQ on socioeconomic outcomes has been documented in the controversial book The Bell Curve, by Herrnstein and Murray (1994). Column (2) shows that IQ does have a statistically significant, positive effect on earnings, after controlling for several other factors. Everything else

being equal, an increase of 10 IQ points is predicted to raise monthly earnings by 3.6%. The standard deviation of IQ in the U.S. population is 15, so a one standard deviation increase in IQ is associated with higher earnings of 5.4%.

- *Ex 5: Using Lagged Dependent Variables as Proxy Variables*

  Using a lagged dependent variable in a cross-sectional equation increases the data requirements, but it also provides a simple way to account for historical factors that cause current differences in the dependent variable that are difficult to account for in other ways. For example, some cities have had high crime rates in the past. Many of the same unobserved factors contribute to both high current and past crime rates. Likewise, some universities are traditionally better in academics than other universities. Inertial effects are also captured by putting in lags of y.
  Consider a simple equation to explain city crime rates:

$$crime = \beta_0 + \beta_1 unem + \beta_2 expend + \beta_3 crime_{-1} + u$$

  where crime is a measure of per capita crime, unem is the city unemployment rate, expend is per capita and $crime_{-1}$ indicates the crime rate measured in some earlier year (this could be the past year or several years ago). We are interested in the effects of unem on crime, as well as of law enforcement expenditures on crime. We expect that $\beta_3 > 0$ since crime has inertia.

2.3. Measurement Error

- Measurement Error in the Dependent Variable

$$y^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_k x_k + u$$

  The measurement error (in the population) is defined as the difference between the observed value and the actual value: $e_0 = y - y^*$

- When does OLS with y in place of y^* produce consistent estimators of the β?
  - It is only natural to assume that the measurement error has zero mean; if it does not, then we simply get a biased estimator of the intercept
  - The relationship between the measurement error and the explanatory variables should be independent. f this is true, then the OLS estimators from are unbiased and consistent. Further, the usual OLS inference procedures (t, F, and LM statistics) are valid.

- Measurement Error in an Explanatory Variable
  Traditionally, measurement error in an explanatory variable has been considered a much more important problem than measurement error in the dependent variable.

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

  Where $e_1 = x_1 - x_1^*$. We assume that the average measurement error in the population is zero $E[e_1] = 0$
  - We want to know the properties of OLS if we simply replace x_1^* with x_1 and run the regression of y on x_1. They depend crucially on the assumptions we make about the measurement error. Two assumptions have been the focus in econometrics literature, and they both represent polar extremes. **The first assumption** is that e1 is uncorrelated with the observed measure, x1:

$$cov(x_1, e_1) = 0$$

- Because this assumption implies that OLS has all of its nice properties, this is not usually what econometricians have in mind when they refer to measurement error in an explanatory variable. **The classical errors-in-variables (CEV) assumption** is that the measurement error is uncorrelated with the unobserved explanatory variable:

$$cov(x_1^*, e_1) = 0$$

But this will lead to $cov(x_1, e_1) \neq 0$. Thus, in the CEV case, the OLS regression of y on x_1 gives a biased and inconsistent estimator.

- $plim(\widehat{\beta_1})$ is always closer to zero than is β_1. This is called the attenuation bias in OLS due to CEV: on average (or in large samples), the estimated OLS effect will be attenuated (underestimate β).
- If the variance of x_1^* is large relative to the variance in the measurement error, then the inconsistency in OLS will be small. So measurement error need not cause large biases.

3. General Test for Functional Form Misspecification

- **Lagrange Multiplier (Score) Tests**

$$y = \beta_0 + \boldsymbol{\beta_1 x_1} + \boldsymbol{\beta_2 x_2} + \boldsymbol{u}$$
$$H_0: \boldsymbol{\beta_2} = 0$$

This test can be done simply with a standard F-test. Following the steps:
Step 1. Regress *y* on the *restricted* set of independent variables and save the residuals, $\tilde{u}$
Step 2. Regress $\tilde{u}$ on *all* of the independent variables and obtain the *R*-squared, say, R_u^2
Step 3. Compute $LM = nR_u^2$
Step 4. Compare *LM* to the appropriate critical value, *c*, in a χ_q^2 distribution. if *LM* > *c*, the null hypothesis is rejected.

- **Ramsey's (1969) regression specification error test (RESET)**

  - If the original model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_k x_k + u$$

Consider the expanded equation

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_1 \hat{y}^3 + e$$

Where $\hat{y}$ *is the OLS fitted values.*

  - The hypothesis set is then: $H_0: \delta_1 = 0, \delta_2 = 0$.
    A significant F statistic suggests some sort of functional form problem. The distribution of the F statistic is approximately $F_{2,n-k-3}$ in large samples under the null hypothesis
  - Further, the test can be made robust to heteroskedasticity

- **Tests against Nonnested Alternatives**
  It is possible to test the original model
  $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$
  against the model
  $$y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + u$$

  - **Method 1 - Mizon and Richard (1986)**
    Construct a comprehensive model that contains each model as a special case and then to test the restrictions that led to each of the models. In the current example, the comprehensive model is

    $$y = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 \log(x_1) + \gamma_4 \log(x_2) + u$$

    We can first test $H_0: \gamma_3 = 0, \gamma_4 = 0$ or test $H_0: \gamma_1 = 0, \gamma_2 = 0$

  - **Method 2 - Davidson and MacKinnon (1981)**
    They point out that if the original model $E[u|x_1, x_2] = 0$, the fitted values from the other model should be insignificant when added to equation. So we first estimate the alternative model by OLS to obtain the fitted values, $\check{y}$. The Davidson-MacKinnon test is obtained from the t statistic on $\check{y}$ in the auxiliary equation.
    $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \theta_1 \check{y} + e$$

    Because the $\check{y}$ are just nonlinear functions x_1 and x_2, they should be insignificant if the original model is correct conditional mean model. Therefore, a significant t statistic (against a two-sided alternative) is a rejection of the original model.
    Secondly, we test the alternative model by estimating the $\hat{y}$ first from the original model, then test $\theta_1 = 0$ in the following form

    $$y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \theta_1 \hat{y} + e$$

    a significant t statistic is evidence against the alternative model.