Handout 2 – Linear Regression Framework and Asymptotic Properties

Chunyu Qu

Oct 4, 2021

1. Unbiasedness of OLS Estimators

1.1. Parameters

- **With the following four assumptions, the OLS estimators are unbiased.**

Assumption 1. Linear in Parameters

Assumption 2. Random Sampling

Assumption 3. No Perfect Collinearity

In the sample (and therefore in the population), none of the independent variables is constant, and there are no *exact linear* relationships among the independent variables.

Assumption 4. Zero conditional mean. $E[u|x] = 0$

- When Assumption 4 holds, we often say that we have exogenous explanatory variables. If some variables are correlated, we say they are endogenous explanatory variables. The terms "exogenous" and "endogenous" originated in simultaneous equations analysis.

1.2. Unbiasedness of Variance

- With the following assumption we can use $MSE = \frac{SSE}{n-2}$ as an unbiased estimator of σ^2

Assumption 5. Homoskedasticity. $var(u|x) = \sigma^2$

- It is important to remember that heteroskedasticity does not cause bias or inconsistency in the OLS estimators. The interpretation of our goodness-of-fit measures R^2 s also unaffected by the presence of heteroskedasticity. The key issue is that the usual OLS *t* statistics do not have *t* distributions in the presence of heteroskedasticity, invalid for constructing confidence intervals. At the same time, *F* statistics are no longer *F* distributed, and the *LM* statistic no longer has an asymptotic chi-square distribution. Thus, the conclusion of significant estimators is not reliable any more.
- The solution is to increase the sample, change the specification, or use the Heteroskedasticity-Robust standard errors.

1.3. Unbiasedness of different conditions

(1) Overspecify. Including one or more irrelevant variables in a multiple regression model, or overspecifying the model, does not affect the unbiasedness of the OLS estimators.
(2) Omitted Variable. One main source of biased estimators.

2. Consistency

2.1. Consistency > Unbiasedness.

- *Asymptotic unbiasedness and consistency also do not imply each other.*
- Unbiasedness of estimators, although important, cannot always be obtained. Although not all useful estimators are unbiased, virtually all economists agree that consistency is a minimal requirement for an estimator. The Nobel Prize–winning econometrician Clive W. J. Granger once remarked, "If you can't get it right as *n* goes to infinity, you shouldn't be in this business." The implication is that, if your estimator of a particular population parameter is not consistent, then you are wasting your time.
- Naturally, for any application, we have a fixed sample size, which is a major reason an asymptotic property such as consistency can be difficult to grasp. Consistency involves a thought experiment about what would happen as the sample size gets large (while, at the same time, we obtain numerous random samples for each sample size). If obtaining more and more data does not generally get us closer to the parameter value of interest, then we are using a poor estimation procedure.

2.2. Consistency of OLS

- Under Assumptions 1 through 4, the OLS estimators are consistent.

Assumption 4'. Zero conditional mean. $E[u] = 0, and\ cov(x_i, u) = 0, for\ i = 1,2, \dots k.$

- Assumption 4' is weaker than Assumption 4 in the sense that the latter implies the former.

correlation between *u* and *any* of *x* generally causes *all* of the OLS estimators to be inconsistent. This simple but important observation is often summarized as: *if the error is correlated with any of the independent variables, then OLS is biased and inconsistent*. This is very unfortunate because it means that any bias persists as the sample size grows

- An important point about inconsistency in OLS estimators is that, by definition, the problem does not go away by adding more observations to the sample.

3. Lagrange multiplier statistic - score statistic (for q Exclusion Restrictions)

To derive the *LM* statistic, consider the usual multiple regression model with *k* independent variables:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

We would like to test whether, say, the last *q* of these variables all have zero population parameters: the null hypothesis is

$$H_0: \beta_{k-q+1} = \cdots = \beta_k = 0$$

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.