

ECON 3740: INTRODUCTION TO ECONOMETRICS

INSTRUCTOR: CHAOYI CHEN
Department of Economics and Finance, University of Guelph

Lecture 7

Last lecture, we derived the variance of $\hat{\beta}_1$, and the error variance σ^2 . Today, we will

- summarize topic four - the simple Regression Model
- topic five - carrying out an empirical project
 - finding interesting research questions
 - literature review
 - data collection
 - econometric analysis

Simple linear regression model: summary

- In this topic, we have introduced the simple linear regression model.
- We have learned using OLS method to derive the estimators of the slope and the intercept parameters.
- We have demonstrated the algebra of the OLS regression line, including computation of fitted values and residuals, and the obtaining of predicted changes in the dependent variable for a given change in the independent variable.
- We have discussed two issues of practical importance: (1) change the units of measurement (2) the use of the natural log to model nonlinearity.

Simple linear regression model: summary

- We showed that under four assumptions, the OLS estimators are unbiased. (The expected value of the OLS estimator equals to the true parameter)
- By imposing one more assumption, "homoskedastic error", we showed the simpler expressions of the variance of the OLS estimators. As we saw, the variance of the slope estimator $\hat{\beta}_1$ increases as the error variance increases, and it decreases when there is more sample variation in the independent variable.
- We showed an unbiased estimator of the σ^2 , $\hat{\sigma}^2 = \frac{SSE}{(n-2)}$. Then, we use $\hat{\sigma}^2$ to get the unbiased estimator of $Var(\hat{\beta}_1)$.
- At last, we briefly discussed regression through the origin and regression on a constant

Carrying out an empirical project: finding interesting research questions

- Posing a question
 - Knowing precisely what question you want to answer is essential
 - You can only collect your data if you exactly know your question
 - You can only know whether you can complete your project in the allotted time if you know whether the necessary data is available
 - You can only know if your research question is of interest to someone if you can precisely state it and discuss it with your class mates/instructor

Carrying out an empirical project: finding interesting research questions

- Finding interesting research questions
 - Choose the area of economics/social sciences you are interested in
 - Examples for typical research questions
 - Labor Economics: Explaining wage differentials
 - Public Economics: Effect of taxes on economic activity
 - Education Economics: Effect of spending on school performance
 - Macroeconomics: Effect of investment on GNP growth
 - Look for published papers on the chosen topic using tools such as EconLit, Google Scholar, the Journal of Economic Literature (JEL) etc.

Carrying out an empirical project: finding interesting research questions

- Your research project should add something new
 - Add a new variable whose influence has not been studied before
 - Expand economic questions to include factors from other sciences
 - Study an existing question for more recent data (may be boring)
 - Use a new data set or study a question for a different country
 - Try out new/alternative methods to study an old question
 - Find a completely new question (hard but possible)
 - It helps if your research question is policy relevant or of local interest

- Literature review
 - A literature review is important to place your paper into context
 - Use online search services to systematically search for literature
 - When searching, think of related topics that may also be relevant
 - A literature review can be part of the introduction or a separate section
- Data collection
 - Most questions can be addressed using alternative types of data (pure cross-sections, repeated cross-sections, time series, panels)

Carrying out an empirical project: data

- Deciding on the appropriate data set
 - Many questions can in principle be studied using a single cross-section
 - But for a reasonable ceteris paribus analysis one needs enough controls
 - Panel data provides more possibilities for convincing ceteris paribus analyses as one can control for time-invariant unobserved effects
 - Examples for panel data sets: PSID (individuals), Compustat (firms)
 - Panel data for cities, counties, states etc. are often publicly available
 - Data sets are often available online, in journal archives, or from authors

Carrying out an empirical project: data

- Entering and storing your data
 - Data formats: 1) printed, 2) ASCII, 3) spreadsheet, 4) software specific
 - Important identifiers: 1) observational unit, 2) time period
 - Time series must be ordered according to time period
 - Panel data are conveniently ordered as blocks of individual data
 - It is always important to correctly identify and handle missing values
 - Nonnumerical data also have to be handled with great care
 - Software specific formats often provide good ways of documentation

Carrying out an empirical project: data

- Inspecting, cleaning, and summarizing your data
 - It is extremely important to become familiar with your data set
 - Even data sets that were used before may contain problems/errors
 - Look at individual entries/try to understand the structure of your data
 - Understand how missing values are coded; if they are coded as 999 or -1, this can be extremely dangerous for your analysis
 - It is better to use nonnumerical values for missing values
 - Understand the units of measurement of your variables

Carrying out an empirical project: data

- Inspecting, cleaning, and summarizing your data
 - Know whether your data is real/nominal, seasonally adjusted/unadjusted
 - Check if means, std.dev., mins, and maxs of your data are plausible
 - Clean your data of implausible values and obvious coding errors
 - When making data transformations (differencing, growth rates)
 - make sure your data is correctly ordered and no wrong operations result
 - For example, in a panel data set, be aware that the first observation of each cross-sectional unit has no predecessor

- Econometric Analysis

- Given your research question and the data available, you have to decide on the appropriate econometric methods to use

- Some general guidelines

- OLS is still the most widely used method and often appropriate
- Make sure the key assumptions are satisfied in your model
- Always check for possible problems of omitted variables, self-selection, measurement error, and simultaneity

- Some general guidelines
 - Carefully choose functional form specifications (logs, squares etc.)
 - Beginners mistake: do not include variables that are listed as numerical values but have no quantitative meaning (e.g., 3-digit occupations)
 - Transform such variables to dummy variables representing categories
 - Handle ordinal regressors in a similar way (e.g., job satisfaction)
 - One can also reduce ordered variables to binary variables

Carrying out an empirical project: econometric analysis

- Some general guidelines
 - Think of secondary complications such as heteroskedasticity
 - Specific problems in time series regressions: 1) levels vs. differences, 2) trends and seasonality, 3) unit roots and cointegration¹
 - Carry out misspecification tests and think about possible biases
 - Sensitivity analysis: look at variations of your specification/method
 - Hopefully, results do not change in a substantial way
 - Are there problems with outliers/influential observations?

¹These are problems relating to time series data, which may not be investigated in this course. However, if you have interest or are necessarily to use the time series data, please let me know.