Introductory Econometrics

S2 2018

Monash Econometrics and Business Statistics

Semester 2, 2018

Welcome

- This is the first unit in econometrics and a pre-requisite for most third year units in econometrics, business analytics and actuarial studies
- You are certainly in the right place to learn econometrics

What is Econometrics?

- “Metrics” as a suffix means measuring things and analysing those measurements. “Econo” tells us that we are talking about measuring and analysing economic things.
- Other examples of quantitative analytics in other fields are biometrics and psychometrics
- Remember that this was coined in 1930, when many of the business and commerce disciplines of today were at their infancy or did not exist, so the context was ”Econo” not “Busino” (“Businometrics” does not sounds very good to me)
- Econometric methods are used in economics, finance, marketing and management
- Econometrics uses mathematics and statistics

Outline of unit

- Week 1: Introduction
- Week 2: Review of Statistical Concepts
- Week 3: Linear Regression (OLS)
- Week 4: More on Linear Regression Analysis
- Week 5: Inference
- Week 6: Model Selection / Prediction
- Week 7: Binary Variables
- Week 8: Heteroskedasticity
- Week 9: Serial Correlation
- Week 10: Persistence in Time Series Data
- Week 11: Large Sample Properties of OLS
- Week 12: Revision

Logistics

- Refer to the Unit Guide and the Moodle site

Overall goal

Evidence based (aka empirical) analysis of business and economic problems quotes. This comprises the following stages:

1. Understanding the problem
2. Formulating an appropriate conceptual model to tackle the problem
3. Collecting appropriate data
4. Looking at data (Descriptive Analytics)
5. Estimating the model, making inference, predictions and policy prescriptions as appropriate (Predictive and Prescriptive Analytics)
6. Evaluating, learning and improving each of the previous steps, and iterating until the problem is solved

Two purposes of econometric modelling

- We use data and econometric methods for two purposes:
  1. Prediction: To predict a target variable based on available information, e.g.,
     - future return to a stock based on historical returns
     - school quality based on house prices
     - rain and its intensity based on the number of people carrying umbrellas
  2. Policy prescription: To understand the causes of variation in a target variable, so that we can control that variable with suitable policy prescriptions, e.g.
     - what determines the volatility of stock prices?
     - what determines the demand for umbrellas?
     - how does university education affect earnings?
     - what is the effect of greenhouse gases on global temperature?

Two purposes of econometric modelling: 1. Prediction

- When do you think a variable x will be useful for predicting a target variable y?

Two purposes of econometric modelling: 1. Prediction

- When do you think a variable x will be useful for predicting a target variable y?
- Can we predict how many people will be carrying umbrellas using the weather bureau's forecast of incidence and intensity of rain?

Two purposes of econometric modelling: 1. Prediction

- When do you think a variable x will be useful for predicting a target variable y?
- Can we predict how many people will be carrying umbrellas using the weather bureau's forecast of incidence and intensity of rain?
- Can we predict if it is raining outside by counting how many people enter the lecture theatre with umbrellas?

Two purposes of econometric modelling: 1. Prediction

- When do you think a variable x will be useful for predicting a target variable y?
- Can we predict how many people will be carrying umbrellas using the weather bureau's forecast of incidence and intensity of rain?
- Can we predict if it is raining outside by counting how many people enter the lecture theatre with umbrellas?
- The answer to both questions is 'yes'. We can exploit correlation for prediction, regardless of the direction of causation.

Two purposes of econometric modelling: 1. Prediction

Predictive models

- The tools we learn allow us to predict y using an x that is correlated with y.
- If there are several variables that are correlated with y, these tools will allow us to use a combination of these predictors to predict y.
- When we have time series data, and today's value is correlated with yesterday's value, these tools allow us to use yesterday's value to predict today.
- Prediction of future values is called forecasting. The tools we learn allow us to build forecasting models.

Two purposes of econometric modelling: 2. Policy Prescription

Causal models

- When do you think we can manipulate a target variable y by changing a variable x?

Two purposes of econometric modelling: 2. Policy Prescription

Causal models

- When do you think we can manipulate a target variable y by changing a variable x?
  - Example 1: CEO salary is positively correlated with a company's share price. Can we increase a company's share price by offering a large salary to its CEO?

Two purposes of econometric modelling: 2. Policy Prescription

Causal models

- When do you think we can manipulate a target variable y by changing a variable x?
  - Example 1: CEO salary is positively correlated with a company's share price. Can we increase a company's share price by offering a large salary to its CEO?
  - Example 2: Poor countries have high infant mortality. Does this mean that the only way to reduce infant mortality is to make the country rich?
- We need to have a causal model to enable us to give policy advice. For example, after taking account for all confounding factors, we establish that better sanitation reduces child mortality. Then we can advise the country or NGOs to implement programs to improve sanitation in order to reduce child mortality.

The role of economics and finance theories

- Theories in economics and finance suggest how variables are related to each other. Hence they are potentially useful for developing causal and also predictive models. Examples:
  - Economic theory suggests that quantity demanded is negatively related to own price and positively related to income of consumers.

The role of economics and finance theories

- Theories in economics and finance suggest how variables are related to each other. Hence they are potentially useful for developing causal and also predictive models. Examples:
  - Economic theory suggests that quantity demanded is negatively related to own price and positively related to income of consumers.
  - Finance theory suggests that bond prices depend negatively on inflation. This implies that predictors that are useful for predicting inflation should also help in predicting bond prices.

The role of economics and finance theories

- Theories in economics and finance suggest how variables are related to each other. Hence they are potentially useful for developing causal and also predictive models. Examples:
  - Economic theory suggests that quantity demanded is negatively related to own price and positively related to income of consumers.
  - Finance theory suggests that bond prices depend negatively on inflation. This implies that predictors that are useful for predicting inflation should also help in predicting bond prices.
- Also, we can use econometrics to test economics and finance theories. For example:
  - The efficient market hypothesis implies that all important information for next period's equity prices is in this period's price. Hence equity returns should be unpredictable. If we can find a significant predictor for returns, we can reject this hypothesis.

Some examples

- What predicts stock returns?
- What is the effect of reducing class size on student achievement?
- How does another year of education change earnings?
- What determines if a person chooses to buy private health insurance?
- What is the price elasticity of cigarettes?
- What is the effect on output growth of a 0.25 percentage point increase in interest rates?

Let's think of bigger issues: What do YOU think?

The most important problem facing us in the next 50 years is:

1. Global warming
2. Inequality and poverty
3. Ageing of the population
4. Threat of terrorism

Go to mars.mu, click on + and then enter Feed Code: Z3DKLD

Econometric Analysis: Data

- The gold standard for measuring causal effects is using data from randomised control trials (RCTs), similar to the way the effectiveness of new drugs or vaccines are measured
- Such data is called *experimental* data
- Unfortunately, in business and economics we cannot run experiments
- Most often econometric analysis has to be carried out using *observational* (i.e. non-experimental) data.
  - Returns to education: data on wage and educational attainment of a sample of individuals.
- Observational data pose major challenges:
  - Confounding effects (omitted factors): smarter individuals go to university. University graduates get higher wages. Is this wage differential a return to university education or a return to smartness?

Econometric Analysis: Structure of data

- Three data structures that are commonly encountered in econometrics:

1. Cross-sectional data: observations on one or more variables taken at the same point in time.
    - e.g. Observations on infant mortality and GDP per capita in a group of countries in 2015.
2. Time series data: observations on one or more variables taken at different points in time.
    - e.g. Monthly observations on returns and volume of trade of Qantas shares from January 2000 to June 2018.
3. Panel data (or longitudinal): observations on the *same* cross-section units at different points in time.
    - e.g. Daily returns and volumes for ASX200 companies from 2/1/2017 to 30/6/2018.

Cross-sectional versus Time series

- Below are two important differences in the properties between cross-sectional and time series data:
  - P1 There is no natural ordering of observations in cross-sectional data, i.e. there is no reason to assume that the characteristics of Helen should be studied before those of Paul and vice versa.
  - P1' Conversely, observations in time series data are *ordered:* There is a natural ordering in time in the sense that GDP in quarter 1 of the year precedes quarter 2 and quarter 2 precedes quarter 3
  - P2 Cross-sectional data are generally assumed to be independent, i.e. information collected on individual 1 does not provide any information on individual 2
  - P2' Conversely, time series data are generally characterized by some form of *temporal dependence*, i.e. an observation on cigarette consumption in July 2018 can be informative about cigarette consumption in August 2018.

Cross-sectional versus Time series

- Time series data can be used to accomplish two important tasks for which cross-sectional data are inadequate. These are to:
  - Forecast future values of a variable: eg. stock prices, consumer price index, gross domestic product, annual homicide rates one or several days /months /quarters / years ahead.
  - Estimate the dynamic causal effect of one variable x on another variable y: eg. estimate the effect on alcohol consumption of an increase in the tax on alcohol, both initially and subsequently as consumers adjust to the new tax.

Cross-sectional versus Time series: Notation

- Conventionally, a regression model for cross section analysis is written as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i, \text{i=1,2,}\ldots\text{,n}$$

  where i denotes units such as individuals, households, firms, schools, cities, states, countries etc

- For time series analysis a similar regression model is written as:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + u_t, \text{t=1,2,}\ldots\text{,n}$$

  where t denotes units in time such as days, months, quarters, years etc

- All that changes in these two expressions is the notation for units (from i to t)
- The remaining parts of these equations will be studied in detail in subsequent lectures

Econometric Analysis: Modelling

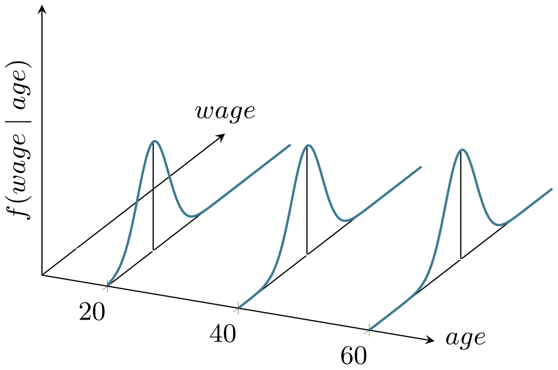
- The fundamental step is to think of all variables as random variables (variables with several possible outcomes that have a probability distribution)
- Equally as important is to remember that with no information, the best prediction of a variable is the centre of its distribution (typically its mean)
- Example: If we have a random sample of weekly wages in Australia and based on that want to predict the wage of a random person in Australia (not in our sample), what do we do?
- Answer: ...

Econometric Analysis: Modelling

- The fundamental step is to think of all variables as random variables (variables with several possible outcomes that have a probability distribution)
- Equally as important is to remember that with no information, the best prediction of a variable is the centre of its distribution (typically its mean)
- Example: If we have a random sample of weekly wages in Australia and based on that want to predict the wage of a random person in Australia (not in our sample), what do we do?
- Answer: ...
- If we group Australian population by age, the centre of the wage distribution for different ages will not be the same
- To learn about mean of wage for each age, what should we do?
- Calculating the average wage for each age in our sample is not feasible (our sample may not have all ages in it) go to data

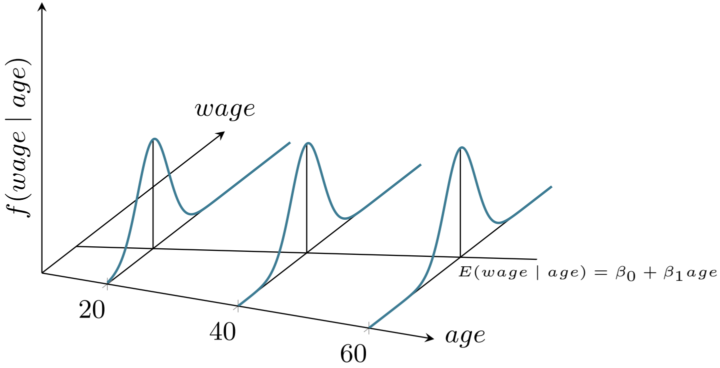
Econometric Analysis: Modelling - interaction of theory and practice

Distribution of wage conditional on different values of age



Econometric Analysis: Modelling - interaction of theory and practice

Distribution of wage conditional on different values of age



Econometric Analysis: Modelling - interaction of theory and practice

- The problem is reduced to estimating the intercept and slope of the conditional expectation line from the data
- Find the best fitting (regression) line to the scatter plot of the data
- We study the regression line and establish why it is the best we can do

Econometric Analysis: Modelling - interaction of theory and practice

- There are other predictors of wage, like education and IQ
- In the theory universe, we generalise the model to $E(wage \mid age, educ, IQ) = \beta_0 + \beta_1 age + \beta_2 educ + \beta_3 IQ$
- In the data universe, we find the combination of age, education and IQ that fits the wage data best
- We study why this is the best we can do
- We also learn how to incorporate information that is qualitative, e.g. gender and occupation.

Econometric Analysis: Modelling - The Main Focus

We study **regression modelling** in this unit, which is a powerful **tool** for empirical analysis. We learn how to **estimate** regression models, how to **interpret** them, and how to make **inference**, **predict** and prescribe policy advice on the basis of these models.

Summary

- The goals of econometric modelling is either to predict or to prescribe policy
- The data available in business and economics are often observational data
- Observational data sets are either cross sectional, or time series or panel (longitudinal) data sets
- Regression modelling is a powerful tool for econometric analysis

Quotes by Jim Barksdale

Jim Barksdale was the CEO of Netscape, and is now the co-chairman of Spread Networks. He is known for his managerial skills.

> *"You cannot manage that which you cannot measure."*
>
> *"If we have data, lets look at data. If all we have are opinions, lets go with mine."*

Back to overview

Global Warming

- Global temperatures have been on a warming trend
- Question: Can we control this warming by reducing the human activities that produce green house gases?
- This translates to econometric questions such as:

Global Warming

- Global temperatures have been on a warming trend
- Question: Can we control this warming by reducing the human activities that produce green house gases?
- This translates to econometric questions such as:
  - What is the nature of the warming trend in temperatures?

Global Warming

- Global temperatures have been on a warming trend
- Question: Can we control this warming by reducing the human activities that produce green house gases?
- This translates to econometric questions such as:
  - What is the nature of the warming trend in temperatures?
  - What causes a warming trend in temperatures?

Global Warming

- Global temperatures have been on a warming trend
- Question: Can we control this warming by reducing the human activities that produce green house gases?
- This translates to econometric questions such as:
  - What is the nature of the warming trend in temperatures?
  - What causes a warming trend in temperatures?
  - What is the sensitivity of global temperatures to green house gas emissions?

Global Warming: Data

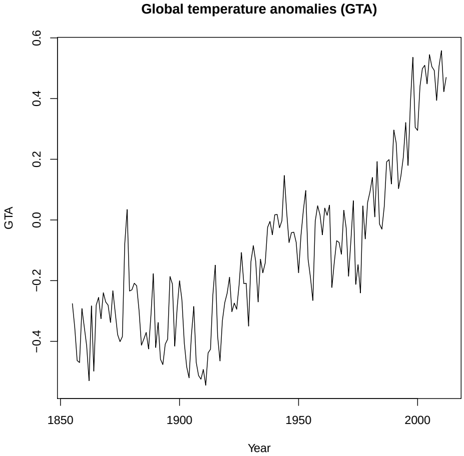


Figure: GTA: Global temperature "anomalies" (actual average annual temperatures minus the 1961-1990 average)

Global Warming: Data

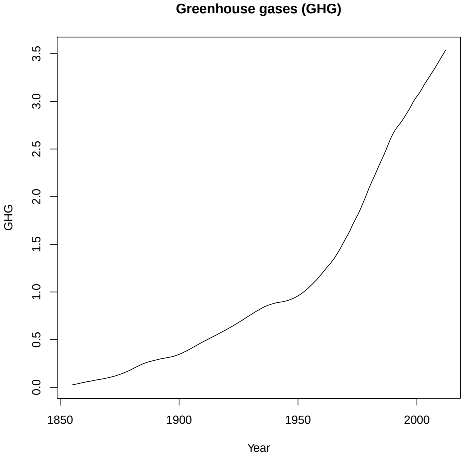


Figure: GHG: Green House Gasses (An index that includes CO_2, CH_4, N_2O and CFC constructed by NASA)

Global Warming: Data Analysis

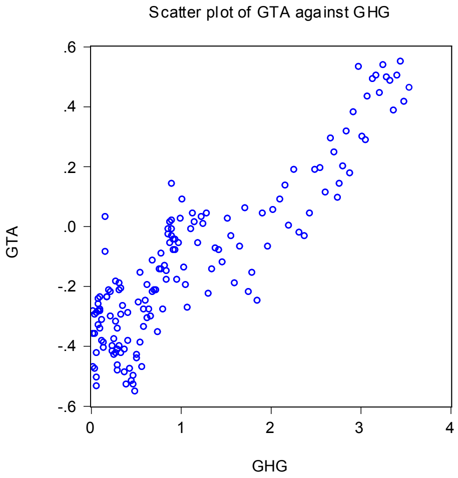


Figure: The scatter plot of GTA against GHG

Global Warming: Data Analysis

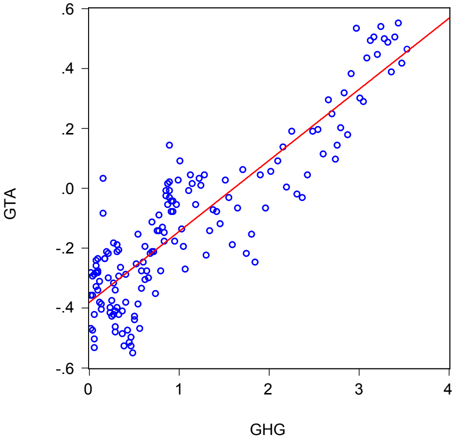


Figure: The scatter plot of GTA against GHG

Global Warming: Data Analysis

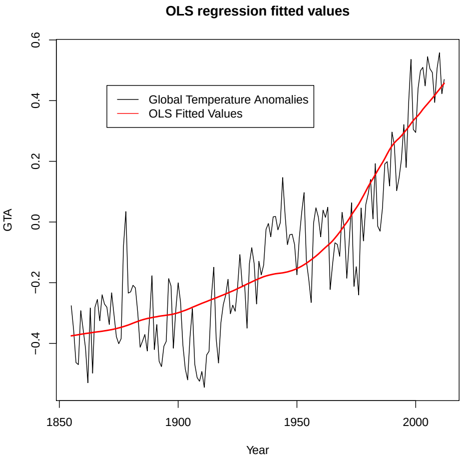


Figure: Predicted GTA using $\widehat{GTA}_t = 0.381 + 0.237\, GHG_t$

Global Warming: Data Analysis

- We learn how to estimate the parameters using data.

Global Warming: Data Analysis

- We learn how to estimate the parameters using data.
- Things we need to pay attention to:
  1. Possibility of spurious regression due to trend.
  2. Possibility of better estimation of this relationship exploiting the persistence in errors.
  3. Feedback of temperatures on greenhouse gases.
- We learn how to pay attention to some but not all of the above in this unit.

Back to problems

Poverty and Inequality

- Reduction of poverty and inequality has been a millennium development goal.
- A blatant evidence of poverty and inequality among nations is variation in child mortality.
- Obviously poorer nations have higher child mortality. But is that the best policy advice we can provide: increase GDP per capita and that will reduce child mortality?

Poverty and Inequality

- Reduction of poverty and inequality has been a millennium development goal.
- A blatant evidence of poverty and inequality among nations is variation in child mortality.
- Obviously poorer nations have higher child mortality. But is that the best policy advice we can provide: increase GDP per capita and that will reduce child mortality?
- We need to determine that after controlling for GDP per capita (i.e. in countries with similar GDP per capita) what other factors determine variations in child mortality.

Poverty and Inequality

- Reduction of poverty and inequality has been a millennium development goal.
- A blatant evidence of poverty and inequality among nations is variation in child mortality.
- Obviously poorer nations have higher child mortality. But is that the best policy advice we can provide: increase GDP per capita and that will reduce child mortality?
- We need to determine that after controlling for GDP per capita (i.e. in countries with similar GDP per capita) what other factors determine variations in child mortality.
- Multiple regression analysis allows us to do that.

Poverty and Inequality: Data

- Data from the World Bank Development Indicators (WDI) data set
- Countries with highest and lowest under-5 mortality rates (per 1000 live births) in 2015

COUNTRY	MORTALITY	GDP PERC...	WATER	SANITATION
Chad	130.9	2047.64	42.54	9.55
Central Afri...	128.8	626.41	54.14	25.09
Sierra Leone	118.8	1315.97	58.09	14.54
Mali	114.2	1919.23	74.27	31.27
Nigeria	108.0	5670.64	67.34	32.60
Benin	100.3	1987.17	67.02	13.93
Japan	3.0	37882.98	98.95	100.00
Sweden	2.9	45679.28	100.00	99.30
Norway	2.7	64008.29	100.00	98.06
Singapore	2.7	80892.06	100.00	100.00
Finland	2.4	38941.76	100.00	99.45
Slovenia	2.4	29037.74	99.52	99.11

Poverty and Inequality: Data Analysis

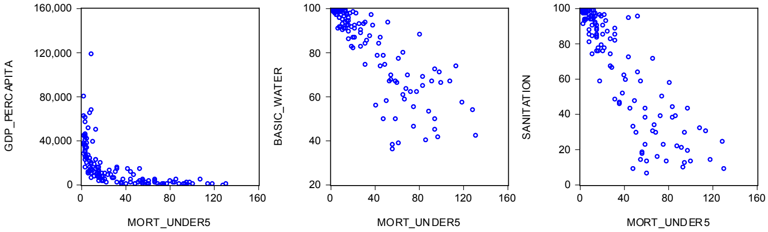


Figure: Scatter plots of mortality rate against several contributing variables

Poverty and Inequality: Data Analysis

Dependent Variable: MORT_UNDER5
Method: Least Squares
Sample: 1 136
Included observations: 136

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	155.50	13.69	11.36	0.00
LOG(GDP_PERCAPITA)	-5.51	2.00	-2.75	0.01
BASIC_WATER	-0.37	0.17	-2.22	0.03
SANITATION	-0.57	0.10	-5.64	0.00

R-squared	0.80	Mean dependent var	33.13
Adjusted R-squared	0.80	S.D. dependent var	32.92
S.E. of regression	14.75	Akaike info criterion	8.25
Sum squared resid	28722.51	Schwarz criterion	8.34
Log likelihood	-556.96	Hannan-Quinn criter.	8.28
F-statistic	180.08	Durbin-Watson stat	0.97
Prob(F-statistic)	0.00		

Figure: Multiple regression results

Poverty and Inequality: Data Analysis

- We used observational data for this analysis
- There are important issues surrounding the implementation of sanitation and water purification programs: how to persuade people to participate? How to evaluate the success of such programs?
- Modern Development Economics includes a great deal of field work on program implementation and program evaluation.
- These areas are important areas in economics where experimental data is being created and analysed (including important work by Monash staff and students in Indonesia, India and Timor Leste).

Back to problems