

# Econometrics book

Ben Lambert

January 11, 2015



# Contents

<b>I</b>	<b>The basics</b>	<b>7</b>
<b>1</b>	<b>How to best use this book</b>	<b>9</b>
<b>2</b>	<b>What is econometrics?</b>	<b>11</b>
<b>3</b>	<b>Estimators and their purpose</b>	<b>13</b>
3.1	Chapter mission statement . . . . .	13
3.2	The goal of this chapter . . . . .	13
3.3	What is an estimator, and why should we care? . . . . .	13
3.4	Models . . . . .	14
3.5	Sampling distributions . . . . .	18
3.6	Good properties of an estimator . . . . .	19
3.7	The central limit theorem . . . . .	19
3.8	Econometrics: GM conditions . . . . .	19
<b>II</b>	<b>Cross sectional data: useful and important</b>	<b>21</b>
<b>4</b>	<b>Ordinary Least Squares: what is it, and when to use it?</b>	<b>23</b>
<b>5</b>	<b>How to make conclusions - an introduction to hypothesis testing</b>	<b>25</b>

6	How to interpret regression results	27
7	Testing a model - does it work?	29
7.1	Hypothesis tests . . . . .	29
7.2	Replicate data generation . . . . .	29
8	Testing the Gauss-Markov assumptions, and what to do if they are violated	31
9	Instrumental variables: allowing inference in difficult circumstances	33
10	Monte Carlo: How to test the quality of an estimator	35
III	Time series: harder to master, but necessary	37
11	Why and how do we need to think about time series differently to cross sectional?	39
12	The basic building blocks of time series models: autoregressive and moving averages	41
13	Testing for stationarity and what to do with non-stationary data	43
14	Cointegration: allowing for realism in time series models	45
15	An introduction to models for real processes: partial adjustment and error-correction models	47
IV	Panel data: the best of both worlds	49
16	The benefits of panel data	51

<i>CONTENTS</i>	5
<b>17 Why do we need more estimators? An introduction to First Differences and Fixed Effects</b>	<b>53</b>
<b>18 The poor relation: Random Effects</b>	<b>55</b>
 <b>V A simple new paradigm in estimation: Maximum Likelihood</b>	 <b>57</b>
<b>19 The flaws in the Linear Probability Model</b>	<b>59</b>
<b>20 Beautifully simple: An introduction to Maximum Likelihood</b>	<b>61</b>
<b>21 Draw conclusions by likelihood: the Wald, the Score and the LM tests</b>	<b>63</b>



## **Part I**

# **The basics**





## **Chapter 1**

### **How to best use this book**



## **Chapter 2**

### **What is econometrics?**



## Chapter 3

# Estimators and their purpose

### 3.1 Chapter mission statement

After reading this chapter the student will understand the concept of an estimator, and its use in statistics.

### 3.2 The goal of this chapter

Statistical inference is the process of drawing conclusions about a population from a sample of observations. Estimators are the tools which statisticians use on samples of data; resulting in estimates of population-wide quantities which can be used to test hypotheses about the wider world. In this chapter the reader will be introduced to the concept of a *sampling distribution*, and how these can be examined to gauge the quality of an estimator.

### 3.3 What is an estimator, and why should we care?

We rarely in life have all relevant data available to us before we make a decision. When we decide where to go on holiday, we don't travel to a country, speak first-hand with locals, and taste the local food prior to deciding on a final destination for the family. We don't know perfectly

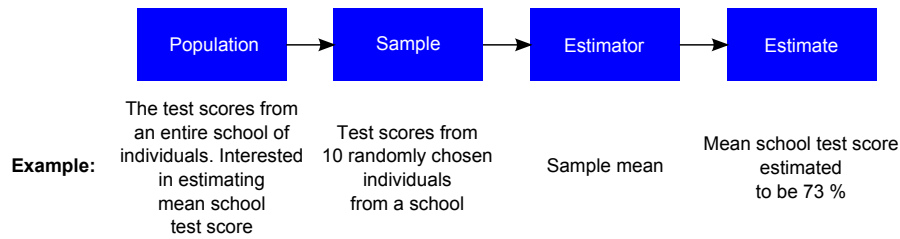


Figure 3.1: The estimation process.

what the weather will be like this afternoon when we get dressed for work. Instead, we extrapolate based on samples of data, and use these as predictive windows into the unknown. Statistical inference is the logical framework that allows us to make reasonable decisions based on our limited access to the facts.

These limited facts are what is known as a *sample*; by definition a subset of data from a *population* of interest. A population of interest might be the UK 18-30 females. However, it needn't just be applied to people, and could just as easily be defined to be the countries within the EU. All that matters is that the population, as a concept, is a wider entity on which we would like to make well-informed statements.

*Estimators* are tools that we use on samples that enable us to make *estimates* of quantities within the wider population being investigated. We then use these estimates to test hypotheses about the population, which allow us to better understand the functioning of the world around us.

Before we choose an estimator to use, we need to specify a *model* of the population which is proposed as a simplified and tractable representation of real life. Estimators can then be used to estimate components of these models, which when complete, can be deployed to help us explain some key features of a process we observe. Fortunately, statistical inference allows us to test the assumptions on which models are built, as well as interrogate their implications.

## 3.4 Models

Essentially, all models are wrong, but some are useful.

---

George Box

Real life is complex. It contains so many, seemingly-independently-moving things, that it can seem difficult/impossible to explain or predict anything effectively. To make life simpler for ourselves, we frequently build *models*. When we are planning how much to budget for electricity this month, we might think that electricity spend in the corresponding month last year might be a reasonable guide. When we are determining how long to spend cramming for an exam, we implicitly have an internal model of the trade off between test score and effort. When we set our alarm for the morning after, we believe that it will roughly take 10 minutes for us to snooze, and another 20 minutes to get showered and ready before leaving the house. These are all examples of models.

All models are abstractions from reality, which allow us to isolate, and concentrate on what we believe are the important parts of a system of interest. They are necessarily simplifications, and hence are not exactly representative of reality. However, as George Box's quote suggests, they can *sometimes* be useful.

The aforementioned are models which we most likely generate internally, but nonetheless are *implicitly* used to help us make decisions. Often for self-betterment, and to allow interrogation of thought however, we want to *explicitly* describe our models. The language of mathematics provides us with a logical framework with which to adequately describe these abstractions.

Before we talk through examples, it is worth reflecting on the various purposes for building models in the first place. Joshua Epstein in his article 'Why model?' lists amongst others, the following motivations for writing down *explicit* models:

- Prediction
- Explanation

- Guide data collection
- Discover new questions
- Bound outcomes to plausible ranges
- Illuminate uncertainties
- Challenge the robustness of prevailing theory through perturbations
- Reveal the apparently simple (complex) to be complex (simple)

There are no doubt other reasons, but I believe that the above covers the majority of rationales.

Imagine we are interested in determining the average test score for a particular standardised exam within the US. We don't have access to test scores for all individuals who take the exam in a specific year however. Instead we might suppose that test scores for individuals in the population are normally distributed about this theoretic mean,  $\mu$ , with some theoretic variance  $\sigma^2$ :

$$IQ_i \sim \mathcal{N}(\mu, \sigma^2) \quad (3.1)$$

In (3.1), the  $i$  subscript on  $IQ$  represents the individuals in our population. So if we have a population size of 10,000,  $i$  runs from 1 to 10,000. It is important to stress that, whilst in this example, we can suppose that  $\mu$  exists, this is not generally the case for quantities of interest in statistics/econometrics. The assumption inherent in the above model is that test scores are normally distributed around this tangible quantity with some theoretic variance. We do not actually believe that test scores are exactly normally distributed<sup>1</sup>, however to make life easier and more palatable, and thus easier to deal with, we make this approximation.

Another model which we might choose to state is that there is a linear relationship the number of years of experience and the wage which an individual commands, on average:

$$wage_i = \alpha + \beta experience_i + \epsilon_i \quad (3.2)$$

---

<sup>1</sup>However, the central limit theorem provides some justification for making this approximation. See section 3.7.



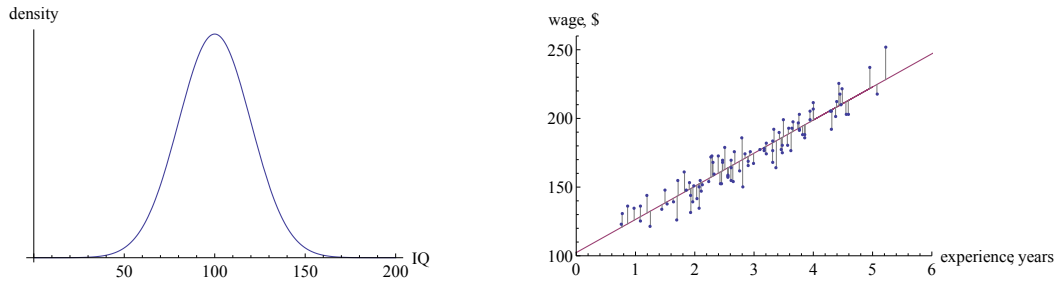


Figure 3.2: Left: the normal model for IQ. Right: the linear model between experience and wages, with the error terms  $\epsilon_i$  indicated as vertical deviations from the straight line.

In (3.2), we have chosen a straight-line<sup>2</sup> relationship between the experience and wages, apart from a catch-all variable which encapsulates the various other idiosyncratic factors<sup>3</sup> which might impact the wage an individual obtains. For example, the type of job undertaken, the number of hours worked, or their level of education. We assume here that conditional on the level of experience for individuals, the average effect of these other factors is zero.

A model is only as good as the assumptions on which it rests. It might be that IQ test scores are more variable in the US population, than a normal distribution allows. It could be that there are diminishing returns to experience, meaning that the increment to wage for an extra year of work diminishes, dependent on the stage in the particular individual's career; invalidating the assumption of *linearity*.

Much like models, assumptions are necessarily simplifying, and therefore *wrong*. However, there is a spectrum of *wrong*. An assumption is good so long as it captures the essence of reality sufficiently to allow the model to be used as is required. If the assumption is too *wrong*, then the model will cease to be useful, and we are forced to go back and examine its foundations.

The good thing about statistical models is that we can frequently test their foundations (see chapter 7). One way to do this is by asking what is the

<sup>2</sup>The model stated in (3.2) is, (with the exception of  $\epsilon_i$  term), of the form  $y = mx + c$ , which is taught in high school.

<sup>3</sup>Individual attributes.

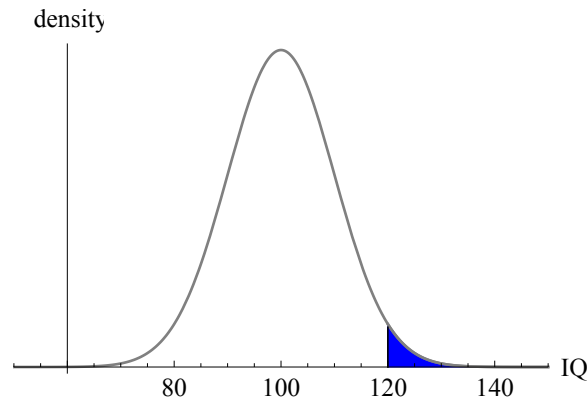


Figure 3.3: A normal distribution for IQ, with mean 100 and variance 100. The blue area represents the probability of obtaining an observation more extreme than 120.

probability that we would have obtained a sample of data as extreme as the one we happened upon. For example, if we estimated a model for individuals' test scores; resulting in a normal distribution with a mean of 100, and a variance of 100. If we then obtained a further observation for an individual who obtained a score of 120, we could ask, how likely is it that we would have obtained a score more extreme than this from our model? To do this, we would need to work out the corresponding area under such a normal distribution<sup>4</sup> (see figure 3.3), which yields a probability of about 2%. If we had a sample of 100, we would hence expect, if our model was correct, to get results as extreme as ours for only for 1 person in 50. If this pattern was upheld if we collected more sample data, this would make us doubt that our model, and hence the assumptions on which it was formed, are in fact reasonable; prompting an re-examination of the latter.

### 3.5 Sampling distributions

When we carry out statistical inference we frequently wish to draw conclusions on populations of interest. The only snag is that we usually only have a sample from the wider world. In statistics, we use estimators on our sample to generate estimates of parameters within a model defined for

<sup>4</sup>Do not worry if you don't understand this, as we will cover hypothesis testing in detail in chapter 7.

the population as a whole. We then use these estimates to test hypotheses about the wider population, in a process known as *statistical inference* (see figure ).

For example, imagine that we are wanting to investigate whether parental education positively influences the wage of an individual in the population, on average. We propose that there ‘exists’ some population process<sup>5</sup> which generates the weekly \$-wages of an individual  $i$ , of the linear form:

$$wage_i = \alpha + \beta educ_i + \epsilon_i \quad (3.3)$$

In (3.3),  $educ_i$  refers to the parental education of person  $i$ , perhaps measured in the number of full-time years of schooling completed. The term  $\epsilon_i$  refers to the myriad of other individual factors which likely impact an individual’s wage. Furthermore, we hypothesise that  $\beta > 0$ , since we expect the impact of parental education to likely raise the prospects of their children, either directly through knowledge passed on, or indirectly through better schooling, and other opportunities.

Not wanting to exceed our research budget, we are forced to use a sample<sup>6</sup> of 100 individuals’ data as a window to test whether parental education is an important determinant of an individual’s wage they go on to earn. We can then use our sample, along with a particular estimator, for example least squares<sup>7</sup>, to estimate the parameter  $\beta$  in the population model given in (3.3). Notwithstanding the many reasons which likely make this estimation methodology flawed, which we will discuss in detail in later chapters, we obtain an estimate of  $\beta = 50$ ; meaning that for each extra year of full-time education completed by parents, an individual on average earns \$50 more on average.

Not being content with the results obtained from the first sample, we then go on to obtain a second sample of size 100, on which to estimate  $\beta$ . Confusingly, we now obtain an estimate of  $\beta = 5$ , which represents an effect only  $\frac{1}{10}$  the size of our previous result!

Undeterred, we go back to the population of interest, each time generating a new sample, and again and again we obtain different estimates of the

---

<sup>5</sup>Sometimes called *estimands*.

<sup>6</sup>We suppose that each sample can be thought of as being a *random sample*. We will discuss this in chapter 4.

<sup>7</sup>We will cover this in chapter 4.

effect of parental education on individual wages. One time we even obtain a value that is negative!

You have now probably guessed what has been happening. The relation (3.3) is only meant to represent the *average* population model; meaning some of the populous gained significantly from their parental education level, whereas for others, the effect was less marked. When we choose a sample of individuals from our population, we are implicitly picking individuals who exhibit heterogeneity in their personal impacts of parental education. Furthermore, this process of sample selection means that the each vat of people is slightly different in terms of their *average* impact of parental education; reflected in the array of estimates we obtained for  $\beta$ , in each of the different samples.

The process which leads an estimator to produce a variety of parameter estimates is called *sampling error*.

### 3.6 Good properties of an estimator

### 3.7 The central limit theorem

### 3.8 Econometrics: GM conditions

## **Part II**

# **Cross sectional data: useful and important**



## **Chapter 4**

# **Ordinary Least Squares: what is it, and when to use it?**





## **Chapter 5**

# **How to make conclusions - an introduction to hypothesis testing**



## **Chapter 6**

# **How to interpret regression results**



## **Chapter 7**

# **Testing a model - does it work?**

### **7.1 Hypothesis tests**

### **7.2 Replicate data generation**



## **Chapter 8**

# **Testing the Gauss-Markov assumptions, and what to do if they are violated**





## **Chapter 9**

**Instrumental variables:  
allowing inference in difficult  
circumstances**



## **Chapter 10**

# **Monte Carlo: How to test the quality of an estimator**



## **Part III**

**Time series: harder to master,  
but necessary**



## **Chapter 11**

**Why and how do we need to  
think about time series  
differently to cross sectional?**





## **Chapter 12**

**The basic building blocks of  
time series models:  
autoregressive and moving  
averages**



## **Chapter 13**

# **Testing for stationarity and what to do with non-stationary data**



## **Chapter 14**

# **Cointegration: allowing for realism in time series models**



## **Chapter 15**

# **An introduction to models for real processes: partial adjustment and error-correction models**





## **Part IV**

# **Panel data: the best of both worlds**



## **Chapter 16**

# **The benefits of panel data**



## **Chapter 17**

# **Why do we need more estimators? An introduction to First Differences and Fixed Effects**



## **Chapter 18**

# **The poor relation: Random Effects**





## **Part V**

# **A simple new paradigm in estimation: Maximum Likelihood**



## **Chapter 19**

# **The flaws in the Linear Probability Model**



## **Chapter 20**

# **Beautifully simple: An introduction to Maximum Likelihood**



## **Chapter 21**

**Draw conclusions by  
likelihood: the Wald, the Score  
and the LM tests**