How to use this book …

A minority of students relish the prospect of studying econometrics, as many are deterred perhaps by the technical jargon e.g. 'autocorrelated disturbances', 'heteroscedasticity' and 'identification problems'. The aim of *Running Regressions* is to apply the practical aspects of elementary regression analysis to some interesting issues and without adopting an excessively esoteric and/or mathematical approach. *Running Regressions* shows that statistics and econometrics can be straightforward, useful and interesting.

A view to the real world

Running Regressions focuses on some topical, real-world issues. Econometric analyses may not seem to be the focus of articles in *The Economist*, the front page of the *Financial Times*, *New York Times* or in OXFAM's latest newsletter. But the professional application of econometric techniques underlies the insights expressed in the business press. Applied econometrics can earn you a living and win you arguments; it gives empirical 'flesh' to what might otherwise be dismissed as subjective opinion or abstract theory. Econometric analysis is used by policy-makers in deciding how much of the budget to devote to health or education, and/or what the tax rate should be. It is used by central bankers deciding whether to raise or lower interest rates. It is used by non-governmental organisations (NGOs) in illustrating the benefits of debt relief and/or how world trade rules can be designed better to advance the interests of developing countries. And it is used by hedge fund analysts to forecast stock market movements and currency fluctuations. Applied econometrics is relevant to every part of economics, finance and development studies and the topics selected for analysis in *Running Regressions* are chosen

accordingly; for example, we have including chapters on international relations (world poverty and armed conflict), local and personal issues (housing, divorce) and cutting-edge fields such as behavioural economics.

Use it simply

The book is not about excessively complex econometric techniques. Our present goal is to explain economic phenomenon using elementary econometric techniques of the 'plain vanilla' variety – one at a time to highlight the key aspects. Good econometricians will focus on parsimony: they will attempt to abstract from the complexity to strip theoretical models to their bare essentials. In selecting techniques, the best techniques will be the most straightforward and intuitive as long as these techniques are also relatively reliable and accurate. Adopting a relatively simple approach does not reduce the power of the analysis. Simplicity is often very effective. So the aim of *Running Regressions* is to present analyses in which there is clear connection between the phenomenon observed, the underlying theory, the raw data and the econometric techniques.

Use it with versatility: econometrics for the social sciences

The econometric tools outlined in *Running Regressions* can be used to analyse a wide range of issues from economics, finance, business studies and the other social sciences. Social scientists examine complex aspects of human decision-making and so the methods of pure science cannot be adopted without adaptation. Good applied econometrics requires a judicious mix of sound economic theory, reliable data and correctly applied statistical techniques. These ingredients do not always readily combine. Finding the data with which to examine theory is usually the hardest part, in practical terms. Social scientists do not deal with a purely physical world and we rarely have laboratories in which we can run controlled, replicable experiments to generate reliable data sets. We may not have data that is directly related to the phenomenon we want to study. It is so often as if we had started off with the plan to make *duck à l'orange* for dinner, only to discover that there is no duck and we must substitute chicken; that we have only lemon rind when orange is called for; parsley in place of *fines herbes* and so on. This does not mean that the analysis is going to be a flop – indeed, if we select our techniques carefully we will

come up with useful insights as long as we realise that we may be able only roughly to approximate the reality we want to understand.

It follows from this that applied econometrics involves some subjective judgements. Which elements of the real world should we include and highlight? Which should we subdue or leave out entirely? This decision will be influenced by preference and perception. Just as Monet's representation of exactly the same tree would be different from, say, Vincent van Gogh's, so two econometricians can differ in their representation of social science. Different features of reality can be emphasised or downplayed according to the perspective that seems most relevant, topical or exciting – or which best fits a view of the world. It is this aspect of discretion and personal choice that contributes to making econometrics as much an art as it is a science. It is useful to remember, when carrying out one's own analysis or interpreting the work of others, that there are many different ways of approaching the same question.

Use it frequently

We toyed with the idea of calling this book *An Econometrician's Toolbox* because the methods presented in this book are used regularly by social scientists using applied econometric techniques. Whether one works in business (estimating the demand curve for airline ticket sales), in development (estimating the effectiveness of various policies on poverty reduction), in law (fighting a monopoly pricing suit, or trying to prove unfair business practices) … the applications are as infinite as the questions that social scientists set out to answer. Of course it is rare that one would run regressions every single day of the week, but the point remains that researchers are required most days to give empirical evidence to back up their arguments and to persuade colleagues as well as critics. Not least, applied econometrics is about providing empirical examples that support (or refute) hypotheses, illustrate trends and allow us to draw inferences about relationships between phenomena. Even if working life consists more of reading the results of studies conducted by others, we must nonetheless understand the implications of what went into the models (and what was left out) and be able to judge whether or not the conclusions are robust. Why was one theory chosen over another? Why those variables selected and not others? Does the data used adequately represent the problem at hand? Has the appropriate yardstick been chosen to assess statistical significance? We need to understand the mechanics and implications of all these elements.

Use it as a complement to other texts, or by itself

The advantage of *Running Regressions* is that it blends a practical approach with a holistic perspective. So *Running Regressions* is an ideal complement to mainstream econometrics texts because it illustrates a wide range of commonly taught techniques using topical and thought-provoking examples soundly based in elementary economic theory. Alternatively, it does cover all the basic econometric techniques and so it could be used as the main text for introductory applied econometrics and/or research methods courses.

The structure of *Running Regressions* and its chapters

You can read *Running Regressions* from start-to-finish, or take the *à la carte* approach, selecting each chapter from the menu of contents, according to particular need or interest. *Running Regressions* is unusual in that it incorporates two titles for each chapter: a main title referring to an econometric technique and a sub-title referring to an economic issue or theory. This division of the titles reflects *Running Regressions*' marriage of theory and applied econometrics. Both titles are listed in the table of contents. So if you are selecting on a chapter-by-chapter basis, you need to decide whether to select either from the list of econometric techniques or from the list of theoretical concepts. This might depend on whether your goal is to prepare for project work/class assignments, to refresh your understanding of a particular theory, to find out more about a particular data issue or to learn about an econometric technique via an interesting illustration of how it works. This unconventional combination of the disparate elements of econometrics into a cohesive whole was one of the book's primary goals, based on the numerous requests from students for a 'roadmap' of econometrics.

If your aim is to read a general overview of econometrics, we recommend starting at the beginning and working your way through. The book opens with fundamental principles of econometrics and concludes with more complex examples. (The last few chapters, for example would not usually be covered in first-year undergraduate courses.) *Running Regressions* covers a variety of methods – starting at the most basic (running a regression by hand using a hand-held calculator), moving through the statistical tools available via Excel and then focussing on a widely used tailor-made econometrics package – *EViews*. In early chapters, data sets are available to allow our readers to try the

techniques themselves; in later chapters website links are provided. If you use our data then there may be slight variations in the estimates you get – reflecting rounding differences or specific features of the various programmes but your results should be much the same as the ones we got ourselves.

If you are reading *Running Regressions* for guidance about how to do a research report or project in an area covered by this book then follow the 'Table of Contents' for guidance about the relevant theories and models. Core theories and their contenders are described, along with a reading list for further and more specialised reading.

How to get started

Each of the chapters in *Running Regressions* goes through a systematic series of steps to show how data and econometric analysis can be used to test a theory. So each chapter is roughly set out in the format that is conventionally expected in applied economic analysis: a theoretical model is developed, the data is presented and analysed and models are estimated and assessed. (Alternatively, one could say that each chapter introduces a different theoretical issue, and then uses an econometric analysis to illustrate it.) For each chapter, the argument will be built up as follows:

1. *The issue*
   In this preliminary section, the basic theoretical ideas, data and econometric approaches are briefly introduced.
2. *The theory*
   We set out the underlying theory or hypothesis, on the basis of existing literature. We then analyse the theory using econometrics but modifying the theory if the econometric evidence suggests that the real world does not conform to the theory. We express this theory as an econometric model. This means simplifying the hypothesis into a simple testable assertion and then re-working it into a form that can be analysed using the available data.
3. *Data*
   Each chapter will go into important details about using data. One can learn a great deal simply by evaluating and describing the empirical landscape that a dataset presents. Descriptive analysis is an oft under-estimated, critical step in the whole exercise. So data plots should be examined carefully: chart variables over time and against each other. Look for obvious trends and patterns.

4. *Empirical results and interpretation*
   The next step is to obtain numerical values, or estimates, of the relationships between the variables in the model, from the data available. In this book, we do this through regression analysis, usually using Ordinary Least Squares (OLS), as explained in Chapter 1. In explaining the detail of OLS techniques, we have kept things simple by concentrating on one set of econometric issues per chapter. But in real-world analysis, it is important to remember that we will usually have more than one step in our estimation procedure and so for your own investigations you may have to combine econometric techniques from two or more chapters. Usually a model will need improving. So it is important to consider what could be done to make the model better. Sometimes this means additional work to extend the theoretical model. Sometimes it will involve finding more appropriate or precise data. At other times it may be a matter of using more reliable and accurate econometric techniques.

   It is also important to interpret our results in the light of theoretical issues. Do the findings confirm our theoretical hypotheses or not? What do our econometric results tell us about the relationship between the variables of our model? Are the relationships between variables positive or negative; are the impacts large or small? In order to assess this we test the statistical significance of our estimates using hypothesis tests such as Student's t-test and/or F tests. What do the tests mean in terms of the underlying phenomena studied, and in terms of our understanding of the theory? Do they allow us to make some forecasts about the future? Of course, just as a defendant in a courtroom can only be declared 'not guilty' rather than 'innocent', the results can only provide support (or not) to a hypothesis. We can never 'prove' a hypothesis. This is why econometricians are cautious with their language, sometimes the statistical linguistic style is necessarily unwieldy: hypotheses are not 'accepted'; at best they are 'not rejected'. This is because the empirical findings are by their nature probabilistic: unlikely outcomes are always possible and evidence appearing to support a hypothesis may just be a coincidence.
5. *Policy implications and conclusions*
   It is not enough just to present results and state whether or not they confirm a theory. Good applied econometrics provides practical answers and policy guidance. So we need to set out our conclusions clearly but also to think about the policy implications suggested by our results. This may be where advocacy comes in, for example if an NGO wants to support arguments for policy change; or if Treasury or central bank officials see a

case for financial reform during a credit crunch; or funds managers see a warning to overhaul their corporate policy on portfolio allocations.

6. *Further reading*
   We have included further reading for readers who would like to follow-up on the intricacies of the issues we explore. We have referred to a range of readings from textbooks, journal articles and policy reports but these references are just a taster of the range of readings available. We have selected the ones that are relatively straightforward, particularly relevant and/or readily accessible. Intentionally, we have kept these further reading lists relatively brief and we haven't been able to discuss in detail every relevant reference or reading. For further ideas on readings, don't forget to do further searches using JSTOR, Google and/or Wikipedia, or just by browsing through economics collections in libraries. The *Journal of Economic Literature* can be a good source of summary articles. Many topical economic and financial issues are the focus of articles and/or surveys in *The Economist* and so this magazine is also a useful starting point in a quest to understand topical world issues.
7. *Chapter exercises*
   For most chapters, we have provided some exercises for students who would like to explore the data and/or test and techniques in more detail.

Going on from here …

Where possible, each chapter gives references to institutions that are associated with the topic examined or to ongoing research programmes. Our goal is to stimulate readers to realise that applied econometrics can be relevant, fun, and an essential beginning in a journey of understanding the economic, financial and social world around us …

Simple regression and data analysis

In this part we outline some basic econometric techniques, including running regressions by hand using Ordinary Least Squares (OLS) estimation techniques. We focus on simple regressions, i.e. regressions on models with just one explanatory variable.

We also introduce Microsoft's Excel as a user-friendly spreadsheet package for running simple regressions.

Running regressions on models with more than one explanatory variable – Multiple Regression Analysis (MRA), will be analysed from Part II onwards.

CHAPTER

An introduction to Ordinary Least Squares

This chapter provides a brief and intuitive summary of the econometric theory that underlies OLS estimation, including:

- The data generating process and sample regression functions
- Types of data
- Ordinary Least Squares (OLS) estimation
- Measuring correlation and goodness of fit
- The Gauss–Markov theorem
- Properties of estimators
- Hypothesis testing

1.1 Introduction

Running Regressions is about the quantitative analysis of observed behaviours and phenomena using economic theory, probability and statistics. By bringing these different elements together we can improve our understanding of people, firms and countries – in the past, the present and the future. For those readers who approach the subject with trepidation, it may be worth remembering this practical goal.

The skills that are required are not intrinsically difficult if a systematic approach is followed. Not least, our intention in *Running Regressions* is to illustrate commonly used techniques in an imaginative and interesting way. This chapter provides a brief introduction to concepts and techniques that are then worked through as practical examples in the following chapters.

Finally, whilst we focus on economics, finance and development studies in our selection of topics, the approach can be used in analysing a wide range of

all real-world situations. The techniques can be, and are, used in all the social and natural sciences.

1.2 Models and data

The aim of econometric analysis is to understand the **data generating processes** (DGPs) that underlie economic systems and human behaviour. These DGPs are like the engines in a car and they propel socio-economic actions and events. For example, the interaction between income and consumption is determined by a DGP. This DGP causes consumption and income to rise together and these rises will be transmitted into the recorded data on income and consumption. Unfortunately, an econometrician rarely has direct knowledge of DGPs and so he/she has to draw inferences about it from samples of data. For a brief analysis of the types of data used in econometric analyses, see Box 1.1.

To capture how DGPs operate an econometrician will use socio-economic and/or financial theory as an initial guide, building up a testable **model** from previous knowledge about a process or system. Once a testable empirical model has been devised, the econometrician will proceed to estimate the **parameters** of their model. As explained below, estimating parameters is one of the aims in applied econometrics because the parameters of a model capture the magnitude and direction of interactions between the variables of the model. Another aim is to use econometric evidence for forecasting the future and parameter estimates are used in constructing these forecasts. Either way, estimating parameters accurately and reliably is the essence of an applied econometrician's task.

1.3 Constructing empirical models

Theoretical ideas are translated into empirical hypotheses via some assumptions about the DGP. The DGP is expressed as a mathematical function linking a **dependent variable** – the variable that the econometrician is trying to 'explain' (sometimes called the **regressand**); and the explanatory variables (sometimes called **regressors** or **independent variables**). For a lot of econometrics, the selection of a variable as the independent variable is complicated by the fact that causality may not be unidirectional, e.g. for the example used in this chapter: poverty might cause low growth; low growth might cause poverty; they may both affect each other (i.e. there may be a **simultaneous** relationship between them) and/or a third factor may be generating changes in both.