1 Introduction

Pleasure in the job puts perfection in the work.

—Aristotle

AT A GLANCE

This book deals with applications of econometric methods and techniques to a variety of questions, often and on purpose outside the realm of economics. Usually, econometric methods are considered for questions in macroeconomics, labor economics, or finance, among others. However, admittedly, when you are a novice to the field, the examples and illustrations in econometric textbooks may not be that entertaining, and hence the very fact that illustrations are chosen as such may reduce your wish to master the methods yourselves. Well, in plain English, several examples in textbooks do not look very challenging.

At the same time, econometric textbooks also seem to have some knowledge content that is difficult to grasp for a novice. These books usually introduce a method and then they apply it, always with great success, to a case at hand. Now, how did the authors know in advance which method would work best for a certain question? And, even more so, is it not the case that, *first*, there is a question and *then* there is a method that could come in handy? And could it be that someone first has tried a variety of methods and that then after some time it is found that one method is more useful than another?

That brings me to the main idea of this book, and that is that questions come first and econometric methods come next. As a consequence and as you will see, the chapters in this book therefore have a format opposite to the one that is commonly used. The main

text of each chapter discusses empirical questions, and each appendix collects the relevant methods and techniques for these questions. The questions collected in the chapters all have a topic in common. Chapters therefore have titles like "Money" and "Fashion, Art, and Music" instead of what you can commonly find like "Generalized Least Squares" and "Autocorrelation." Of course, this book will not claim to be a general introduction to econometric methods and techniques. The main claim is that this book hopes to arouse enthusiasm for the application of those methods and techniques and in fact to illuminate a little bit how econometric science proceeds. New methods have always been designed and developed just because someone believed that a question or a type of data required a new method. Textbooks usually turn it around, and there may be good reasons for that. But, in this very introductory book, the ordering is first the question and then the method.

MORE DETAIL

This first chapter provides the motivation and a brief outline. Econometric methods and techniques can often be applied in practice, but they usually come in third place. First, there is a question. Then there need to be data to be analyzed. Textbooks usually present an overwhelming array of methods and techniques but rarely address how to collect relevant data, let alone that the books provide creative inspiration for the formulation of questions. This book seeks to provide some such inspiration by examining a range of questions, many of which one perhaps would not have thought about in the first place, and then to illustrate how relevant data can be acquired such that basic econometric methods and techniques can be applied. It is not difficult to make econometric methods more difficult. It is more difficult to try to make matters easier.

Motivation

The title of this book may sound like an oxymoron. Wiktionary (consulted March 11, 2016) gives for "enjoyable":

> Enjoyable (comparative more enjoyable, superlative most enjoyable)
>
> 1. Pleasant, capable of giving pleasure.
>
> That was an enjoyable day; I had a lot of fun.

At the same time, the Wikipedia lemma of "Econometrics" (also consulted March 11, 2016) reads as

> Econometrics is the application of mathematics, statistical methods, and computer science, to economic data and is described as the branch of economics that aims to give empirical content to economic relations. More precisely, it is "the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference."

Well, I know that this summary describes matters well, but I can readily appreciate that it does not arouse much warmth and enthusiasm. For many students, academics, and practitioners, it therefore may seem that the two words in the book title are disconnected. The "econo" part of the second word seems OK, but the "metrics" part makes the combined term somewhat daunting. One could wonder whether "economics," which to many people already seems like an applied mathematics discipline, can be made even more quantitative by adding even more metrics.

One cause for the seeming disconnection may be a cursory look into an introductory textbook on econometrics, where one may bump into "text" like the following:

The final expression for b is

$$b = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

With $a = \overline{y} - b\overline{x}$, you can now compute the residuals

$$e_i = y_i - a - bx_i$$

with their variance

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2$$

and

$$E(s^2) = \sigma^2.$$

Now, what could this possibly mean? And, notably, this kind of text usually comes in one of the first chapters, while then there still are hundreds of pages to come in a textbook. So, why do we need all this?

Another possible cause for the felt disconnection can be that most textbooks on econometrics typically focus on economic relations in the macroeconomic world or in finance. So, the focus often is on gross domestic product, unemployment, and inflation or on the New York Stock Exchange and the Nikkei index. In the present book, where the intention is to give meaning to the title, this second potential cause for believing enjoyable econometrics is an oxymoron is addressed in a particular way. As you may have seen in the table of contents, some of the questions that I will address are not the very standard ones.

Clearly, anyone who has a glance at the content of an introductory textbook of econometrics will surely agree that mathematics is important for econometrics because most of these textbooks are full of mathematical notation. In fact, also in the forthcoming Chapter 2, some of such notation will be used too, although the intention is to keep it to a minimum. So, at first sight, econometrics is considered a highly technical subject, which demands substantial skills from those who practice it in real-life situations. At the same time, it is *not* true that econometric methods are only designed for the sake of creating ever more complicated tools. No, in many situations the techniques are really very helpful, for example, when it comes to forecasting economic variables like, indeed, gross domestic product (GDP) growth, inflation, and unemployment or when it comes to evaluate, for example, the merits of educational programs or tax reform programs.

Yet a closer look at the empirical applications of econometric methods, not only in textbooks but also in scientific articles in international academic journals, also reveals that many applications apparently address issues in macroeconomics and finance, with some exceptions concerning labor economics or microeconomics. These areas are usually characterized by the abundant availability of data, as national statistical institutes and financial institutions collect a wealth of information on a daily to an annual basis.

This suggested abundance of accessible data is not common in all economic areas, and in many cases students and practitioners have to collect their own data. And then rarely it is told in the textbooks how one should collect such data. How do you start? Which data are the proper ones? How many do you need?

Now, here is the key idea. The collection of data is preceded by a question. Students may believe that the model comes first because textbooks seem to only discuss the ins and outs of models, but in the real world it does not work that way. In fact, in the real world it starts with a question. It is important that people learn to formulate such a question. Textbooks usually assume that there was a question from the onset and then start with methods to find some apparently relevant estimates. Examples of questions for macroeconomics and finance could be "Is turmoil on American stock markets contagious for European stock markets?" or "Do lower inflation levels spur economic growth?," but seldom are they formulated in this way.

When a question is well articulated, the next step is to collect the relevant data, and then only at the end does an econometric method or model come into play. Sometimes these data are available in a clean and nice format (again in macro and finance), or they are available in a messy and noisy format (discourses on social media, the mentioning of the word "recession" in Google Trends), but quite often they are not available at all. Or worse, there are data collected but for another purpose or inappropriately, and then the subsequent analysis could be impossible or difficult.

A typical response of students is that everything seems to work well in those nice and shiny textbooks, but, and I quote one of my own undergraduate students, when "I analyze my data, nothing seems to work." Another response is that the nice and simple regression model (more on that in Chapter 2) never adequately fits the data; there always seems to be a problem with the data. So, here is the challenge!

Combining technical mathematics with data that do not work well and perhaps also questions that are less interesting to the novice can make people to feel that econometrics can hardly be enjoyable.

This book intends to change this perspective by doing a couple of things. First, Chapter 2 outlines the two main tools of econometrics, that is, correlation and the regression model. We do need these two tools, as almost everything proceeds from there. With various illustrations, I will show that basic tools can already be quite handy for interesting research questions, even such a basic tool as an average value of various numbers. One such question is at what age creative individuals make their best work. It seems that, on average, novelists, composers, and painters make their best work around the age of 40, but that is not so interesting, in particular because various artists have already died by that age – think of Jimi Hendrix and Amy Winehouse. So, perhaps a more interesting figure is not the absolute age but the relative age, that is, at which fraction of their lives did these creative individuals make their best work? There is a surprising outcome to this question, and join me later on to have a look in Chapter 2 at what it is. An illustration of correlation in that chapter concerns the link between fouls and yellow cards in football. When all fouls would be punished with a yellow card, then this correlation would be equal to 1, right? Well, is that the case? We will see. A third illustration in Chapter 2 concerns the regression model and seeks to answer the question of whether men and women could possibly run the marathon in less than two hours, sometime or ever in the future. We will see that this might happen for men, but it most likely will take a while, and we will have to wait for a few more Olympics or championships.

Chapter 2 concludes with a lucid discussion of what else is in econometric textbooks by mainly using very simple notation. Correlation and regression will always be useful, but the form of the tools will change depending on the question and on the properties of the data. In all further chapters, I will put some of those adapted versions in the appendix so that you can get an impression of how that works. If you decide to skip each appendix, it is perfectly fine, as the main text can be read without the appendix. On the other hand, the appendices do give you an indication how questions and data help to shape econometric models and methods. The level of mathematics is always at a high school level, as it basically involves the operators "adding up," "subtracting," "multiplying," "dividing," and "raising to the power."

The key message of this book is, with emphasis, that it all starts with asking a question. Thus, this book will be full of questions. As is evident from the just-mentioned Chapter 2 questions, the questions will not be the mainstream questions that one typically encounters in econometric articles and books. Deliberately, questions will be considered that sometimes may seem awkward, and it is hoped that this partly may help to introduce the joy into applying econometric methods. Indeed, who does not want to know if hemlines have any predictive value for the economic cycle? Well, Chapter 6 will show that they do not, by the way, but it is good to know anyway. This book also highlights the subsequent collection of relevant data. It is important for anyone applying econometric methods to make, so to speak, dirty hands. This means, for example, to collect data and then to find out that there are all kinds of issues with the data (typos and missing data). To have personal practical experience is very relevant because it helps to learn how and when basic tools need further refinement.

When you look at textbooks, it seems that the ordering is "method," "data," and "illustration." That is, here we have a model, and let us illustrate it on the following data. This sequence obviously suggests that the model or method comes first. Of course, in reality,

the ordering is "question," "data," "model," and "answer to the question," and this is what I follow here too.

A potential consequence of the first ordering of "method, data, and illustration" is that students and practitioners may believe that more complicated methods also require more sophisticated questions. If the econometric method is highbrow, so should be the question, right? Well, I can tell you that this is not necessarily the case. In fact, usually when models get more complicated, it is because the data have certain features. For example, sales data typically obey an *S*-shaped pattern; that is, first the innovators take a try with the new product, then an early majority, which makes the sales levels take off, and way at the end there is a group of laggards who adopt the product when it has almost disappeared from the stores. Such an *S*-shaped pattern does suggest an upward trend in the data, and a trend that does not move on forever and eventually dies out. We will see that econometric models exist that allow for this feature and that these models are a little more involved than the simple regression model. So, features of the data can suggest new models and methods.

CHAPTERS 3 TO 8

The six chapters after Chapter 2 all address topical questions, all arranged in chapters with a specific theme. Chapter 3 deals with money, quite literally in fact. In early 2002, something happened in Germany that triggered some of my questions in this chapter. What happened was the exchange of a banknote of 300 euros for payment at a gas station, where the customer received 250 euros in notes of 100 and 50 in return. The euro had just been introduced, and people apparently needed time to adjust to the new notes, and certainly that one of 300 euros. Well, as you might have guessed, there is no such thing as a 300-euro banknote! This smart but illegal construction, however, triggered me. Why do we have 10, 20, and 50 on our banknotes and not 30 or 40? And how do we make payments anyway? What is it that we do when we want to make a cash payment and look into our wallets to make a choice among the available notes and coins?

In Chapter 3, we will also ask ourselves whether we suffer from money illusion, that is, does it matter if we have banknotes of 10 or of 10,000, for example? Next we move to the issue of how we pay with cash. I will show that, even before data are collected, it helps to think deeply about the choice process when you look into your wallet and make a decision. You may of course always choose the largest banknote, if it is available, but after some transactions, you will always end up with a wallet containing smaller notes and coins. And what do you do when some notes are missing? That is, what happens if only notes of, say, 10 and 50 are available and no notes of 20? And, rounding, does that matter? Anyway, Chapter 3 has lots of very practical questions.

In a sense, Chapter 4 continues with money, but now about money in the future, possibly with interest. In our daily lives, we have to make decisions about current situations, like the choice among coins in our current wallet, but in many other cases, we have to make decisions that have consequences that persist for a long time in the future. For example, we can choose to pay the full amount for a new television set now, but we may also decide to postpone our payments by resorting to a scheme of monthly payments during the next few years. Therefore, we pay 480 USD now for a television set, or we pay 25 USD per month for the next two years. Both options have their benefits, but beware that the second option can mean that you still have to pay for the old television when you may have already purchased a new model. This example by the way does involve interest rate payments. Usually, when payments are postponed, the one who receives the money in the future wants to have some interest on his or her money. Now, a notoriously difficult exercise concerns computations with interest rates. This is because you need quite some numerical skills to solve interest rate equations. Such skills are usually summarized under the header of financial literacy and numeracy, which gives the title of Chapter 4. With some simple ways of data collection and with the use of two new econometric models beyond those of Chapter 2, I will show how questions on numeracy can be answered simply. Data collection is key here, and

the examples give you some idea how this works in practice. The two new econometric tools are multiple regression and the logit model, both of which are highlighted, as promised, in the appendix.

We stay close to money in the next chapter, as Chapter 5 deals with banknotes and postage stamps, where now the focus is on their potential use as collectibles. At times, people wonder what are the best ways to invest money and to have some returns at a later stage, perhaps to be used as savings for a rainy day. In economics, it is usually assumed that people invest in companies buy holding stocks or countries by having bonds, but people can also invest in art or in rare collector's items. Stocks and bonds have various specific features, and one of them is that their prices may suddenly go up, in a race to the top, and after that show a price nadir of substantial size. These patterns are usually called bubbles, and the most famous ones are the Tulip Bubble in seventeenth-century Netherlands and the South Sea Bubble in early eighteenth-century England. In Chapter 5, it is asked whether postage stamp prices also face bubbles and what might cause these bubbles. It turns out that scarcer stamps keep on having their high investment value, whether there are bubbles or not. Now, how can we estimate which stamps are more rare than others are? And how does this work for banknotes? Some banknotes were issued only for a short while, and perhaps these are scarce? As there may be a link between inflation rates and the values on banknotes, would there be any relation between these, that is, does inflation make the number of zeroes to go up, or is it the other way around? Chapter 5 introduces, yes again in the appendix, a multi-equation version of the simple regression model. This model is called a vector autoregression, and it is very often used in practice because it is simple to apply. This model can also prevent finding spurious relations across two or more variables. The appendix to Chapter 5 shows how that works.

Chapter 6 addresses a range of quite different questions. The first question deals with a well-known urban legend. It is said – or, better, it is believed – that the current state of the economy is visible from the length of a woman's dress. When the economy is doing well, skirts get

shorter (where somehow people have the roaring twenties and the hippy sixties in mind), and when the economy is going down, dress lengths increase. This phenomenon is called the hemline index, and it regularly gets attention in popular media and also in professional media. Well, one day I decided (together with a student) to collect the data and to put the urban legend to an econometric test.

In Chapter 6, we also return to the peak years of creative artists. There the focus is on whether people who peak later also get older. That would make sense from an experience point of view, but is there also some evidence for that?

Going back to the sixties and seventies of the twentieth century brings up an interesting question about music. Being over 50, I notice myself sometimes saying to students that "many things were better in the past," and then I usually refer to music. I mention Pink Floyd, Queen, the Rolling Stones (never was very fond of the Beatles, by the way), Led Zeppelin, and the like, and then they usually gaze at me with this misty look like "What is he talking about?" Well, for a long time, these rock bands dominated the charts, and to many these days, these bands are still the main favorites. In the Netherlands, there is, each year, a chart called the Top2000, and it features the 2000 most favorite songs, as indicated by and voted for by the Dutch public. Often this list is called a list for old men (where, for students, old means over 40), as for many years now, those bands mentioned have topped the lists. So, the question is, is older music better indeed? Looking at many years of data in the Top2000 lists should somehow provide an answer. In part, the analysis relies on an alternative econometric model, which is a model that allows for *S*-shaped patterns, a model that is introduced in the appendix.

In Chapter 7, the move is toward the academic process itself. Scientists are expected to publish their work in academic journals, usually in the format of a paper. Papers can range from 5 to 40 printed pages in those academic journals, and there is a ranking of the quality of these journals. A publication in a top journal can advance one's career, and such publications usually mark the start of a successful

career. Now, getting one's work published is one thing, but to have colleagues and the media pay attention to your work is quite another thing. Among other numbers, one number that I will present in this chapter is the fraction of my own articles that effectively were ignored completely. I will not give away the fraction yet, but if I would have known it in advance, I could have saved more than 1500 pages of typing. Anyway, sometimes articles do receive attention, and it would be interesting to know which features these successful articles have. So, one question to be addressed is which types of econometrics-based articles get the most attention? Is it new data? Is it a new econometric method, or is it the research question itself? And does it help to be more productive, as an academic? Does the writing of more articles lead to more attention for those articles? And, as one part of the publication process involves judgment and suggestions by reviewers, do the implemented modifications based on their suggestions lead to more citations? In the technical appendix, I will present another model for *S*-shaped data.

To end this book, there will be chapter on trends and hypes. In economics, when we think of trends, we consider trends in economic prosperity, in populations, in prices, and in, for example, production. Often, the trends go up, and the question typically is whether these trends will go up forever and whether events can change recent trends and make the trend wander away in another direction. In the appendix to Chapter 8, I will address a few simple econometric models for trends, and I will discuss how to choose between these models, if one wishes to do so. Much academic research in econometrics in the past few decades has addressed this seemingly simple issue because it was found to not be simple. In the appendix, I will direct some attention to the underlying reason why this is the case. And, because common trends are such an important and Nobel Prize–worthy topic, the appendix contains a concise discussion on those.

Otherwise, Chapter 8 will try to answer various other questions about trends. Trends in the climate, for example, are a hotly disputed topic. But is local or global warming equally distributed throughout

the year, or could it be that only winters are getting warmer? Anecdotal evidence and personal memories about summers in the Netherlands do not suggest that summers are getting warmer.

Another trend can be found in prescribing medication, in particular for attention deficit hyperactivity disorder symptoms. Recent years have seen a tremendous growth in the use of methylphenidate and atomoxetine, in particular in young children. Are these trends going to persist?

The prices and salaries of top football players are on the rise too. In 2016, Paul Pogba came to Manchester United for 105 million euros, and one may wonder: Where will this end?

Finally, sometimes a trend turns out to be hype. By the late 1990s, there was much buzz about a phenomenon called "The New Economy." One feature of this was the declared and foreseen end of unemployment. Well, with a recent worldwide crisis just recently experienced and the rise of unemployment fresh in our memory, one may wonder: Was it hype? Moreover, if it was, when will we stop talking and writing about it?

TO CONCLUDE

One motivation for writing this book is to emphasize that econometric models and methods can also be applied to more unconventional settings, which are typically settings where the practitioner has to collect his or her own data first. Such collection can be done by carefully combining existing databases but also by holding surveys or running experiments. A nice by-product of having to collect one's own data is that this helps to make a proper and educated choice among the wealth of potential methods and techniques to be considered.

Another motivation for writing this book concerns the recommendation to engage in applying econometric methods. A casual glance at many econometric studies seems to suggest that most practitioners have a solid view of their research questions. Indeed, someone who needs an econometric model to create out-of-sample forecasts usually knows that he or she wants a tool that somehow provides accurate forecasts. And those who evaluate policy programs

want to answer the question of whether the program was successful. However, there are many more perhaps slightly unconventional research questions that also benefit from using econometric methods, and in this book I will address quite a few of those.

These two reasons for writing this book come together in the single overall goal, and that is to attract newcomers by showing that econometrics also can involve smart ways of formulating research questions and smart ways of data collection. So, while some may shy away from the mainstream applications of econometrics, it is hoped that this book, full of alternative research questions and innovative datasets, will provide a motivation to embark on a primer course of econometrics. Digging into technicalities will then prevail, so, first for now, let us have some fun!