Practical Econometrics

Non-Parametric Econometrics

Ibrahim Ahamada
Emmanuel Flachaire

Non-Parametric Econometrics

Ibrahim Ahamada
Emmanuel Flachaire

Translated by
Andrew Clark

OXFORD
UNIVERSITY PRESS

Preface

The role of econometrics is to bring together economic theory, statistical methods, and observed data. One way of doing so is via regression. In this context, one of the overriding preoccupations of econometricians is finding the right specification of their regression model. The simple reason for doing so is that bad model specifications increase the risk of drawing the wrong conclusions from the analysis of data, with all of the negative side-effects that this would imply.

Standard econometric textbooks set out the properties of various estimators for different specifications of regression models. It is very commonly supposed in these presentations that the form of the relationship linking the dependent to the explanatory variables is known, and that the regression model is correctly specified: we are here in the world of parametric regression models, both linear and non-linear. However, the retained specification may be very dissimilar to the real underlying relationship in certain cases. The appeal of non-parametric and semi-parametric methods is that they relax the specification constraint: no *a priori* relationship is imposed between the explanatory and the dependent variables. In the context of the search for the best specification, these methods have brought new information to the debate, and have caused established regression specifications to be questioned in a number of empirical domains.

Recent developments in non- and semi-parametric estimation have opened up a wide vista of empirical applications across all the domains in which econometrics is applied: Macroeconomics, Microeconomics, Finance, and so on. These new methods can equally be applied to all of the different types of data used by researchers: cross-section, time-series, panel, and qualitative. The goal of this book is to present these new empirical tools so as to encourage their wider use in empirical applications.

Chapter 1 starts by setting out the principles of the estimation of a density. Chapters 2, 3, and 4 cover the simplest of the regression cases, with one dependent variable and one explanatory variable. These chapters describe the different methods of non-parametric estimation: kernel, splines, and wavelets. However, most empirical applications include a number of explanatory variables. Chapter 5 acknowledges this reality, and describes the estimation of semi-parametric regressions. Finally, mixture models are presented in Chapter 6. These introduce unobserved heterogeneity into the estimation of both densities and regression models. Sections marked $*$ are more technical and can be skipped if required.

This book sets out to be a pedagogical piece of work, illustrated by a number of original applications of the techniques at hand. We particularly emphasize the understanding of the methods, the underlying intuition, and how these techniques can be applied in practice. Each chapter includes a certain number of concrete applications of the methods described, as a way of illustrating them and of underlining both the advantages of these new techniques and their limitations. The applications described cover a wide variety of economic topics: economic convergence, income inequality, earnings, the Phillips curve, interest rates, financial markets, inflation, and house prices. An appendix describes how to apply these methods using statistical software. The reader can then reproduce the results in the applications described in this book using the original data.

This book is aimed at advanced undergraduate and graduate students in Economics, Business and Statistics, academic and non-academic readers, and researchers and analysts who use econometrics in firms and public organizations.

We have benefitted enormously from the advice and comments of a number of colleagues while writing this book. In particular, we thank Jean-Marc Bardet, Luc Bauwens, Philippe Bertrand, Russell Davidson, Véronique Delouille, Alain Desdoigts, Jean-Pierre Florens, Philippe Jolivaldt, Hubert Kempf, Mohamad Khaled, Michel Lubrano, Olivier Nũnez, Anne Péguin-Feissolle, Véronique Simonnet, Antoine Terracol, and the editors Jurgen Doornik and Bronwyn Hall. The students in the Nonparametric Econometrics course at the University of Paris 1 Panthéon-Sorbonne also played a crucial part in the development of this work. We are very grateful to them. We are particularly grateful to Andrew Clark, who translated this book into English and, with relevant remarks, largely contributed to improving its quality and readability. We also thank the

Centre d'Economie de la Sorbonne and the Région Provence-Alpes-Côte d'Azur for financial support.

Those who, unbeknownst to them, contributed the most to this book are our families and our children. They provided us with both the desire and the opportunity to carry out our work. This book is dedicated to them: to Muriel, Haris, and Adam, and to Valérie, Rémi, and Tom.

Contents

List of Figures

List of Tables

1

Kernel Density Estimation

1.1 Introduction

Has the gap in wealth between rich and poor countries narrowed or widened over time? Is investment in shares risky? How does the tax system affect the distribution of income between households? A preliminary analysis of the distribution of the variable of interest is often extremely useful to address this kind of question. The probability density function is a fundamental concept in statistics, providing a natural description of a variable's distribution. The analysis of this density clearly reveals certain properties of the variable in question, and yields information that is very likely useful in guiding and deepening the empirical analysis. This chapter is mainly devoted to the estimation of this density function from a data sample and its analysis.

Figure 1.1 shows the estimated density function of GDP per capita in 121 countries across the world in 1988. The data here come from the Penn World Table of Summers and Heston (1991), and have been divided by the mean of GDP per capita over all countries. On the X-axis, the value of 1/2 corresponds to one-half of the world mean of GDP per head, and the value 2 corresponds to twice the world mean. If we consider GDP per head to be a useful measure of a country's wealth, this figure allows us to analyze the world distribution of income. One of the first things that we note is that the density function is bimodal. The existence of two modes suggests that there are two distinct groups: one composed of the "richest" countries, and another consisting of the "poorest". The second mode is much less pronounced than the first, which indicates that the two groups are not of the same size: there are relatively few "rich" countries, and distinctly more "poor" countries. Further, the first mode is located just to the left of the value 0.5 on the X-axis, while the second is found at around 3. We can

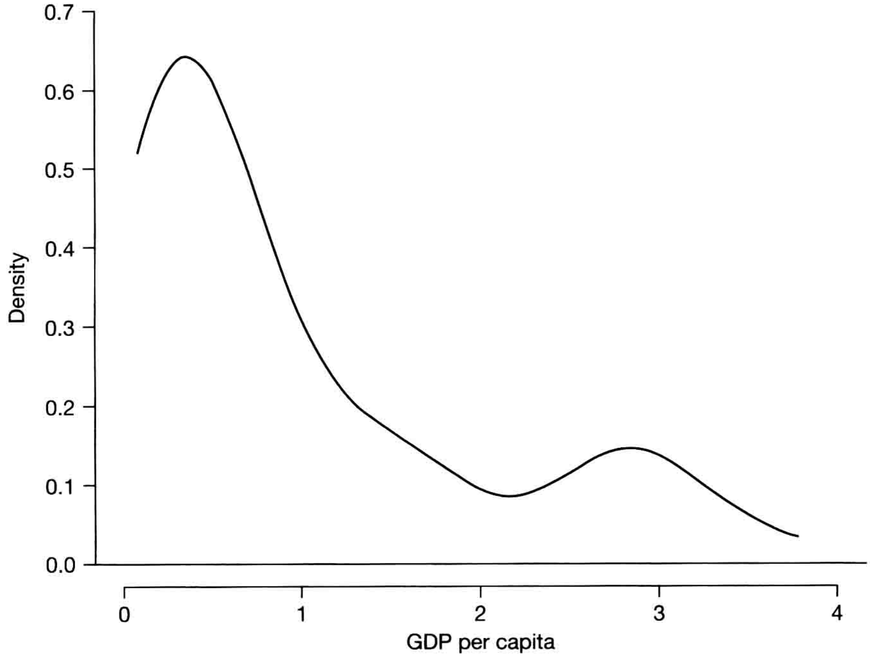

Figure 1.1: Kernel density estimation of the world distribution of income

thus conclude from this figure that, on average in 1988, "rich" countries enjoyed a level of GDP per capita that was around three times the world average, whereas that of "poor" countries was only half of the average level. We can also carry out a dynamic analysis by taking into account data from different years. For example, the same figure using data from 1960 reveals that the distance between the two modes was larger in 1988 than in 1960: the gap between the two groups of countries seems to have grown over time. The in-depth analysis in Quah (1997) considers the question of convergence between groups, but also the persistence of the gap, and the mobility of countries between different groups since the 1960s.

Figure 1.2 presents the estimated density function of daily returns of the CAC 40, the French stock market index, between March 1, 1990 and September 27, 2006. For comparison purposes, we have overlaid a Normal distribution (the dotted line). This distribution of financial returns is Leptokurtic: compared to the Normal distribution, the peak is more pointed and the extremities are thicker. As such, the probabilities at the extremes and in the middle of the distribution are greater than those in the Normal distribution. This shape is commonly observed in data from

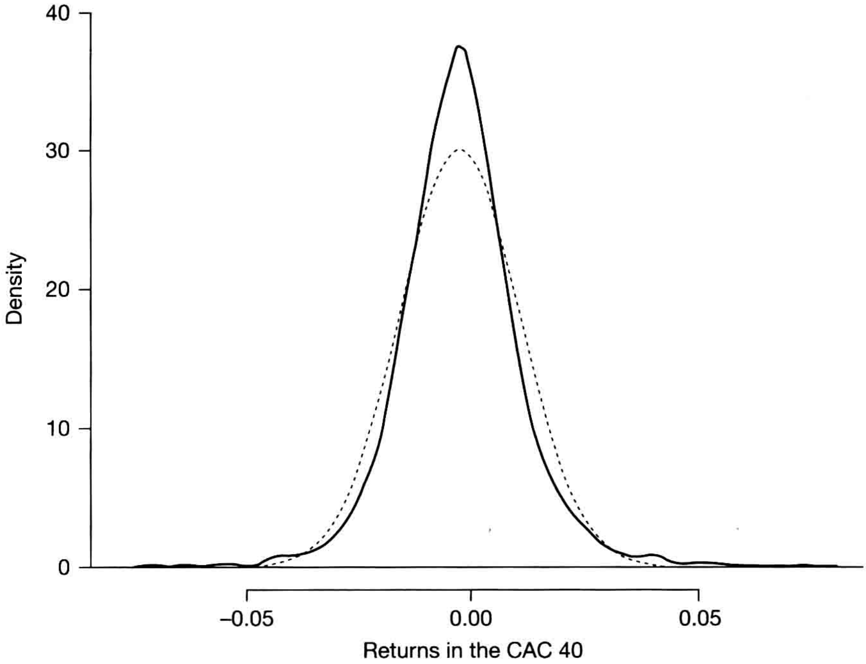

Figure 1.2: Kernel density estimation of daily returns in the CAC 40

financial markets, where extreme movements are more common than the Normal distribution would predict. This is one of the reasons for which ARCH (AutoRegressive Conditional Heteroskedasticity) regression models are often used in finance: the predicted returns from ARCH models are often Leptokurtic.

Figure 1.3 depicts the distribution of household *gross* and *net* income, that is, before and after taxes, in 1998 (equivalent income in 1000 Italian Lire). Both income measures are expressed in logarithms. In this figure, there is no very clear difference between the form of the two density functions: a simple shift would appear to suffice to move from one density function to the other. A shift in a logarithmic scale corresponds to a percentage change in the variable of interest. As such, the horizontal displacement that is evident in Figure 1.3 reveals a multiplicative effect on the distribution of income (of x%). In other words, the tax system in Italy affects the concentration of income distribution (the location of log-income distribution) but has no impact on the shape of the income distribution across the population. The tax system thus seems to deduct the same proportion of income from all households (Fiorio, 2008).

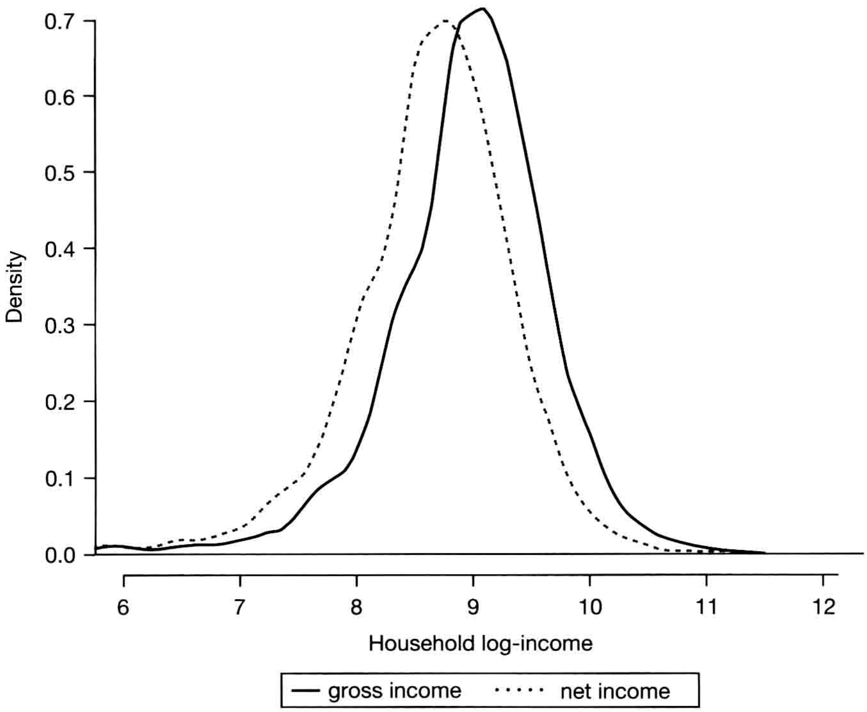

Figure 1.3: Kernel density estimation of Italian household incomes

All of the conclusions from the three examples above could have been reached by other means; however, the estimated density function is an ideal tool for the illustration of the properties of the data at hand, as it is easily understood by non-statisticians. The estimation of the density functions in the examples above was carried out by kernel methods with an automatic choice of the smoothing parameter. The following sections present this non-parametric density estimation method, and describe its statistical properties. More information is available in Silverman (1986), as well as Härdle (1989), Scott (1992), Wand and Jones (1995), Simonoff (1996), Bowman and Azzalini (1997), and Pagan and Ullah (1999).

Section 1.2 presents the standard methods for the estimation of a density, and Section 1.3 the Kernel estimation methods. The choice of the smoothing parameter is addressed in Section 1.4, and adaptive Kernel methods are described in Section 1.5. Finally, Section 1.6 considers the case of multivariate analysis.

1.2 Standard Density Estimation Methods

First of all, we restate the definition of a probability density function. If we say that a random variable Y has a probability density function of f, this means that for any real values a and b, the probability that Y falls between a and b is:

$$P(a < Y < b) = \int_a^b f(y)\,dy \quad \text{for any } a < b.$$

The function f has the following properties: it takes only non-negative values, it is integrable, and its integral from $-\infty$ to $+\infty$ is equal to 1. If we have available a sample of data, which we suppose to have come from an unknown density function, then the calculation of the estimated density function will appeal only to the observed data to hand. A number of different approaches are then possible. In this section, we consider the parametric estimation of this density function and the estimation by histogram, which are very often used in practice.

1.2.1 *Parametric Estimation*

If the density function is known up to k parameters, written as $f(y;\theta)$, the estimation of the density boils down to estimating the vector θ, which consists of k unknown parameters. This is the parametric approach, which requires the prior specification of a family of distributions from which the data sample is supposed to have been obtained.

Consider, for example, that we have a sample of n observations, $y_1, \dots, y_n$, obtained from a Normal distribution, with mean μ and variance σ^2. The density function of the Normal distribution $N(\mu, \sigma^2)$, evaluated at x, is given by:

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

The estimation of the density function is equivalent to obtaining estimates of the parameters μ and σ using the sample of data. One classic method consists in specifying the joint density of the sample and calculating the parameter values which maximize this function. In other words, we choose the parameters which maximize the likelihood of observing the actual values in the data sample. This is the method of Maximum Likelihood. If the observations are independently and identically distributed, the joint density function of $y_1, \ldots, y_n$ is equal to the product of the individual densities:

$$f(y; \mu, \sigma) = \prod_{i=1}^{n} f(y_i; \mu, \sigma).$$

The estimation of the density therefore requires the maximization of this function with respect to the parameters μ and σ. It is simpler to maximize the logarithmic transformation of this function, called the log-likelihood:

$$\ell(y; \mu, \sigma) = \log f(y; \mu, \sigma) = \sum_{i=1}^{n} \log f(y_i; \mu, \sigma).$$

This is the sum of the individual log-densities. The solution of the two maximizations is identical, as the logarithmic transformation is positive and monotonic, but the analytical resolution of the log transformation is much simpler. In our case, with a family of Normal distributions, the parameter values which maximize the log-likelihood are the mean and the variance of our data sample. Different examples can be presented with different probability distributions. The resolution of the problem may then be more complicated, but the same principle applies for other families of parametric distributions.

The main drawback of this approach is that we have to suppose the data are drawn from a given family of distributions f. If this hypothesis does not hold, then the estimated density function can differ substantially from its real value. For example, the estimated Normal distribution in Figure 1.2 is obtained from a parametric estimation of the density under the *a priori* hypothesis that the data were indeed distributed normally. This curve is markedly different from that which results from a non-parametric estimation, which latter reveals a pronounced Leptokurtic distribution. Here, the hypothesis underlying the parametric estimation seems very strong, and ignores one of the notable properties of financial markets.

The advantage of non-parametric estimation is that it makes no such *a priori* hypotheses regarding the form of the density function. As such, it is much more flexible than the parametric approach. Non-parametric estimation lets the data themselves determine the shape of the density, without constraining the function to belong to any particular family of distributions.