Nonparametric Econometrics: A Primer
Editorial Scope

Foundations and Trends® in Econometrics will publish survey and tutorial articles in the following topics:

- Identification
- Model Choice and Specification Analysis
- Non-linear Regression Models
- Simultaneous Equation Models
- Estimation Frameworks
- Biased Estimation
- Computational Problems
- Microeconometrics
- Treatment Modeling
- Discrete Choice Modeling
- Models for Count Data
- Duration Models
- Limited Dependent Variables
- Panel Data
- Dynamic Specification
- Inference and Causality
- Continuous Time Stochastic Models
- Modeling Non-linear Time Series
- Unit Roots
- Cointegration
- Latent Variable Models
- Qualitative Response Models
- Hypothesis Testing
- Interactions-based Models
- Duration Models
- Financial Econometrics
- Measurement Error in Survey Data
- Productivity Measurement and Analysis
- Semiparametric and Nonparametric Estimation
- Bootstrap Methods
- Nonstationary Time Series
- Robust Estimation

Information for Librarians
Foundations and Trends® in Econometrics, 2008, Volume 3, 4 issues. ISSN paper version 1551-3076. ISSN online version 1551-3084. Also available as a combined paper and online subscription.
Nonparametric Econometrics: A Primer

Jeffrey S. Racine

Department of Economics, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4M4, racinej@mcmaster.ca

Abstract

This review is a primer for those who wish to familiarize themselves with nonparametric econometrics. Though the underlying theory for many of these methods can be daunting for some practitioners, this article will demonstrate how a range of nonparametric methods can in fact be deployed in a fairly straightforward manner. Rather than aiming for encyclopedic coverage of the field, we shall restrict attention to a set of touchstone topics while making liberal use of examples for illustrative purposes. We will emphasize settings in which the user may wish to model a dataset comprised of continuous, discrete, or categorical data (nominal or ordinal), or any combination thereof. We shall also consider recent developments in which some of the variables involved may in fact be irrelevant, which alters the behavior of the estimators and optimal bandwidths in a manner that deviates substantially from conventional approaches.
# Contents

1 Introduction 1

2 Density and Probability Function Estimation 5
  2.1 Parametric Density Estimators 6
  2.2 Histograms and Kernel Density Estimators 6
  2.3 Bandwidth Selection 12
  2.4 Frequency and Kernel Probability Estimators 16
  2.5 Kernel Density Estimation with Discrete and Continuous Data 18
  2.6 Constructing Error Bounds 20
  2.7 Curse-of-Dimensionality 21

3 Conditional Density Estimation 25
  3.1 Kernel Estimation of a Conditional PDF 25
  3.2 Kernel Estimation of a Conditional CDF 28
  3.3 Kernel Estimation of a Conditional Quantile 29
  3.4 Binary Choice and Count Data Models 31

4 Regression 33
  4.1 Local Constant Kernel Regression 33
  4.2 Local Polynomial Kernel Regression 38
Nonparametric methods are statistical techniques that do not require a researcher to specify functional forms for objects being estimated. Instead, the data itself informs the resulting model in a particular manner. In a regression framework this approach is known as “nonparametric regression” or “nonparametric smoothing.” The methods we survey are known as kernel methods. Such methods are becoming increasingly popular for applied data analysis; they are best suited to situations involving large data sets for which the number of variables involved is manageable. These methods are often deployed after common parametric specifications are found to be unsuitable for the problem at hand, particularly when formal rejection of a parametric model based on specification tests yields no clues as to the direction in which to search for an improved parametric model. The appeal of nonparametric methods stems from the fact that they relax the parametric assumptions imposed on the data generating process and let the data determine an appropriate model.

A “kernel” is simply a weighting function.
Introduction

Nonparametric and semiparametric methods have attracted a great deal of attention from statisticians in the past few decades, as evidenced by the vast array of texts written by statisticians including Prakasa Rao (1983), Devroye and Györfi (1985), Silverman (1986), Scott (1992), Bickel et al. (1993), Wand and Jones (1995), Fan and Gijbels (1996), Simonoff (1996), Azzalini and Bowman (1997), Hart (1997), Efromovich (1999), Eubank (1999), Ruppert et al. (2003), Härdle et al. (2004), and Fan and Yao (2005). However, the number of texts tailored to the needs of applied econometricians is relatively scarce including Härdle (1990), Horowitz (1998), Pagan and Ullah (1999), Yatchew (2003), and Li and Racine (2007a) being those of which we are currently aware.

The first published paper in kernel estimation appeared in 1956 (Rosenblatt (1956)), and the idea was proposed in an USAF technical report as a means of liberating discriminant analysis from rigid parametric specifications (Fix and Hodges (1951)). Since then, the field has undergone exponential growth and has even become a fixture in undergraduate textbooks (see, e.g., Johnston and DiNardo (1997, Chap. 11)), which attests to the popularity of the methods among students and researchers alike.

Though kernel methods are popular, they are but one of many approaches toward the construction of flexible models. Approaches to flexible modeling include spline, nearest neighbor, neural network, and a variety of flexible series methods, to name but a few. In this article, however, we shall restrict attention to the class of nonparametric kernel methods, and will also touch on semiparametric kernel methods as well. We shall also focus on more practical aspects of the methods and direct the interested reader to Li and Racine (2007a) and the references listed above for details on the theoretical underpinnings in order to keep this review down to a manageable size.

It bears mentioning that there are two often heard complaints regarding the state of nonparametric kernel methods, namely, (1) the lack of software, and (2) the numerical burden associated with these methods. We are of course sympathetic to both complaints. The latter may unavoidable and simply be “the nature of the beast” as they say, though see Computational Considerations for a discussion of the issues. However, the former is changing and recent developments
hold the promise for computational breakthroughs. Many statistical software packages now contain some elementary nonparametric methods (one-dimensional density estimation, one-dimensional regression) though they often use rule-of-thumb methods for bandwidth selection which, though computationally appealing, may not be robust choices in all applications. Recently, an R (R Development Core Team (2007)) package “np” has been created that provides an easy to use and open platform for kernel estimation, and we direct the interested reader to Hayfield and Racine (2007) for details. All examples in this review were generated using the np package, and code to replicate these results is available upon request.


References


References


References

References


