Journal of Econometrics 201 (2017) 108–126

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Nonparametric estimation and inference under shape restrictions

Joel L. Horowitz [a,*], Sokbae Lee [b,c]

[a] Northwestern University, Evanston, IL 60208, USA
[b] Institute for Fiscal Studies, London, WC1E 7AE, UK
[c] Department of Economics, Columbia University, New York, NY 10027, USA

A R T I C L E I N F O

Article history:
Received 16 July 2016
Received in revised form 22 March 2017
Accepted 21 June 2017
Available online 13 July 2017

JEL classification:
C13
C14
C21

Keywords:
Conditional mean function
Constrained estimation
Monotonic
Convex
Slutsky condition

A B S T R A C T

Economic theory often provides shape restrictions on functions of interest in applications, such as monotonicity, convexity, non-increasing (non-decreasing) returns to scale, or the Slutsky inequality of consumer theory; but economic theory does not provide finite-dimensional parametric models. This motivates nonparametric estimation under shape restrictions. Nonparametric estimates are often very noisy. Shape restrictions stabilize nonparametric estimates without imposing arbitrary restrictions, such as additivity or a single-index structure, that may be inconsistent with economic theory and the data. This paper explains how to estimate and obtain an asymptotic uniform confidence band for a conditional mean function under possibly nonlinear shape restrictions, such as the Slutsky inequality. The results of Monte Carlo experiments illustrate the finite-sample performance of the method, and an empirical example illustrates its use in an application.

1. Introduction

Let Y be a scalar random variable and X be a scalar random variable or vector. This paper presents a method for nonparametrically estimating and carrying out inference about the conditional mean function

$$g(x) \equiv E(Y|X = x)$$

under a shape restriction on g such as monotonicity, convexity, non-increasing (non-decreasing) returns to scale, or the Slutsky inequality of consumer theory. Economic theory often provides shape restrictions but does not provide finite-dimensional parametric models. For example, cost functions are monotone increasing, concave in input prices, and may exhibit non-increasing or non-decreasing returns to scale. Demand functions satisfy the Slutsky inequality, which is nonlinear. This motivates nonparametric estimation under shape restrictions. This paper explains how to estimate and form a uniform confidence band for g under shape restrictions that are more complicated than monotonicity or convexity and may be nonlinear.

It is well known that g can be estimated consistently and with the optimal rate of convergence without imposing shape restrictions. Fan and Gijbels (1996) and Härdle (1990), among many others, describe nonparametric estimation and rates of convergence without shape restrictions. Mammen (1991a, b), Mammen and Thomas-Agnan (1999), and Wang and Shen (2013) discuss rates of convergence with shape restrictions. However, fully nonparametric estimates can be noisy and inconsistent with economic theory due to random sampling errors. For example, Blundell et al. (2012, 2017) found fully nonparametric estimates of demand functions to be wiggly and non-monotonic. Blundell et al. (2012, 2017) also found that imposing the Slutsky restriction reduced random noise and led to well-behaved nonparametric estimates without the need for arbitrary and possibly incorrect parametric or semiparametric assumptions.

Many methods are available for carrying out consistent nonparametric estimation under shape restrictions. See, for example, Hall and Huang, 2001, 2002; Hall et al., 2001; Hall and Presnell, 1999; Matzkin, 1994 and the references cited in the foregoing paragraph. Asymptotic inference is not difficult if the values of x at which the shape restriction binds or does not bind in the sampled population are known. Liew (1976) illustrates this in the context of inequality constrained estimation of a linear model. Du et al. (2013) carry out kernel nonparametric estimation. In applications, however, it is not known where in the sampled population the shape restriction does or does not bind. This greatly complicates inference, because random sampling errors can cause the shape restriction to bind or not bind the estimated and true g at different

* Corresponding author.
E-mail address: joel-horowitz@northwestern.edu (J.L. Horowitz).

http://dx.doi.org/10.1016/j.jeconom.2017.06.019

values of x. A similar problem arises in inference about a finite-dimensional parameter that may be on the boundary of the parameter set (Andrews, 1999). Existing results on inference about a shape-restricted, nonparametrically estimated conditional mean function are limited to functions that are assumed to be monotonic or convex. The literature on inference under monotonicity or convexity restrictions is vast. See, among many others, Banerjee and Wellner, 2001; Birke and Dette, 2006; Chernozhukov et al., 2009; Dette et al., 2006; Dumbgen, 2003; Groeneboom and Jongbloed, 2015; Groeneboom et al., 2001; Pal and Woodroofe, 2007; and the references therein. Existing results do not treat shape restrictions such as increasing or decreasing returns to scale and the Slutsky inequality that are of particular importance in economics. There is also a large literature on testing the hypothesis that a shape restriction holds. See, for example, Andrews and Shi, 2013; Chernozhukov et al., 2013; Hall and Yatchew, 2005; Lee et al., 2013; Romano et al., 2014 and the references therein.

This paper is concerned with inference under shape restrictions, such as the Slutsky restriction, that may be nonlinear in a sense that is defined in Section 5. We formulate the estimation problem as minimization of a local quadratic objective function subject to constraints that implement the shape restriction. In general, the shape restriction generates a continuum of constraints. We reduce the number of constraints to a finite value by imposing the shape restriction and estimating g only on a discrete grid of points x in the support of X. We obtain a confidence band that is uniform over points in the grid. The grid becomes finer as the sample size, n, increases, thereby ensuring that, asymptotically, the shape restriction holds everywhere in the support of X. This enables us to obtain a confidence band for g that, asymptotically, is uniform over the support of X and satisfies the shape restriction. In practice, a confidence band can be computed only on a grid, so there is little practical difference between a band that is uniform over grid points and one that is uniform over a continuum.

The use of a discrete grid of points x enables us to overcome the problem of not knowing which constraints are binding in the sampled population. Let $\overline{\mathcal{S}}_n$ be the set of constraints that bind in the population or nearly bind in a sense that is defined in Section 4. This set is unknown. We find a data-based set $\hat{\mathcal{S}}_n$ of "possibly binding" constraints and carry out estimation under the (possibly false) assumption that $\hat{\mathcal{S}}_n = \overline{\mathcal{S}}_n$. We show that $\hat{\mathcal{S}}_n = \overline{\mathcal{S}}_n$ with probability approaching 1 as $n \to \infty$. Consequently $\overline{\mathcal{S}}_n$ can be treated as known asymptotically, and asymptotic inference can be carried out as if $\overline{\mathcal{S}}_n$ were known and $\hat{\mathcal{S}}_n = \overline{\mathcal{S}}_n$.

Let $g_0(x)$ and $\hat{g}(x)$, respectively, denote the true conditional mean function and the shape-restricted nonparametric estimator. We show that with suitable scaling, $\hat{g}(x) - g_0(x)$ is asymptotically jointly normally distributed with mean 0 over grid points. Asymptotic normality makes it possible to obtain a confidence band for g_0 that is uniform over grid points. As $n \to \infty$ and the distance between grid points approaches 0, the uniform confidence band over grid points converges to a uniform confidence band over all values of x.

Estimation of $g(x)$ at points x that are not in the grid is unnecessary for forming an asymptotic uniform confidence band for g but may be of interest for other reasons. Estimation of $g(x_{new})$ at a point x_{new} that is not in the grid can be carried out using the methods of this paper by shifting the location of the grid so that x_{new} is a point of the shifted grid. Alternatively, $g(x_{new})$ can be estimated using any of a variety of methods for interpolating $g(x)$ between grid points subject to the shape restrictions. The choice among interpolation methods is arbitrary and, except in special cases, does not yield an estimator that converges in probability as rapidly as an estimator based on the shifted grid.

Section 2 outlines the main steps involved in implementing our method. Section 3 presents the unconstrained and constrained nonparametric estimators of g and defines the grid. Section 4 describes the method for finding the set $\hat{\mathcal{S}}_n$ of possibly binding constraints. Section 5 explains how to carry out inference about g and form a uniform confidence band for g under shape restrictions. The confidence band obtained in Section 5 is uniform over the support of X and also over a class of functions g that includes nearly binding constraints. To minimize notational complexity, the discussion in Sections 2–5 assumes that X is a scalar random variable. The extension to higher dimensions is outlined in Section 6. Section 7 presents the results of Monte Carlo experiments and an empirical example that illustrates the numerical performance of our methods. Section 8 presents concluding comments. The proofs of theorems are in the Appendix.

2. A guide to implementation

This section outlines the main steps of our method for estimating and obtaining a uniform confidence band for g. We assume here that X is a scalar random variable whose support is [0, 1]. The extension to a multidimensional X is presented in Section 6.

1. Define a grid $0 < x_1 < x_2 < \cdots < x_J < 1$ of J equally spaced points on (0, 1). A data-based method for choosing J in applications is presented in Section 7.
2. Estimate $g(x_j)$ $(j = 1, \ldots, J)$ nonparametrically by using local quadratic estimation with bandwidth h. Let $\tilde{g}(x_j)$ denote the resulting estimate. A method for choosing h in applications is presented in Section 3.1.
3. Use the estimates $\tilde{g}(x_j)$ to find the set $\hat{\mathcal{S}}_n$ of possibly binding shape constraints. $\hat{\mathcal{S}}_n$ is given by Eq. (4.6).
4. Re-estimate $g(x_j)$ $(j = 1, \ldots, J)$ nonparametrically using constrained local quadratic estimation under the restriction that the shape constraints in $\hat{\mathcal{S}}_n$ are binding (that is, they are equalities) and ignoring all other shape constraints.
5. Form a uniform confidence band for g using either the method of Eqs. (5.8) and (5.9) or the method of Section 5.3.

3. The estimators of g

This section describes our methods for estimating g with and without shape restrictions. The unrestricted estimator is used to estimate the set of possibly binding constraints. The shape-restricted estimator is an extension of the unrestricted estimator. Section 3.1 presents the unrestricted estimator. Section 3.2 presents grid and the shape-restricted estimator.

3.1. The unrestricted estimator

This section presents the unrestricted nonparametric estimator of g that is used throughout the remainder of this paper. Let $\{Y_i, X_i : i = 1, \ldots, n\}$ denote an independent random sample from the distribution of (Y, X). Assume for now that X is a scalar random variable. The extension to a multidimensional X is presented in Section 6. Also assume that the support of X is a compact interval. Without further loss of generality, let this interval be [0, 1].

We use local quadratic estimation with bandwidth $h \propto n^{-1/5}$ to obtain the unrestricted nonparametric estimator of g. In applications, the bandwidth can be chosen by using cross-validation or plug-in methods for local constant or local linear estimation. Under our assumptions, local quadratic estimation with $h \propto n^{-1/5}$ provides an estimator of g that is free of asymptotic bias, and the bandwidth can be selected by standard methods. Local constant, local linear, and series estimation methods with a bandwidth selected by cross-validation or plug-in methods do not have this property. They require undersmoothing or explicit bias correction