

# The Econometrics of Shape Restrictions

Denis Chetverikov,<sup>1</sup> Andres Santos,<sup>1</sup>  
and Azeem M. Shaikh<sup>2</sup>

<sup>1</sup>Department of Economics, University of California, Los Angeles, California 90095, USA;  
email: chetverikov@econ.ucla.edu, andres@econ.ucla.edu

<sup>2</sup>Department of Economics, University of Chicago, Chicago, Illinois 60637, USA;  
email: amshaikh@uchicago.edu

ANNUAL  
REVIEWS **CONNECT**

[www.annualreviews.org](http://www.annualreviews.org)

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Annu. Rev. Econ. 2018. 10:31–63

First published as a Review in Advance on  
March 14, 2018

The *Annual Review of Economics* is online at  
[economics.annualreviews.org](http://economics.annualreviews.org)

<https://doi.org/10.1146/annurev-economics-080217-053417>

Copyright © 2018 by Annual Reviews.  
All rights reserved

## Keywords

shape restrictions, uniformity, irregular models

## Abstract

We review recent developments in the econometrics of shape restrictions and their role in applied work. Our objectives are threefold. First, we aim to emphasize the diversity of applications in which shape restrictions have played a fruitful role. Second, we intend to provide practitioners with an intuitive understanding of how shape restrictions impact the distribution of estimators and test statistics. Third, we aim to provide an overview of new advances in the theory of estimation and inference under shape restrictions. Throughout the review, we outline open questions and interesting directions for future research.

## 1. INTRODUCTION

Shape restrictions have a long history in economics, with their crucial role being recognized as early as Slutsky (1915). Over a century later, we find their prominence increasing as breakthroughs across different literatures have widened their empirical applicability. In theoretical work, for instance, shape restrictions have continued to arise both as testable implications of models and as ways to obtain sharp counterfactual predictions. Meanwhile, econometric research has made important advances in developing suitable asymptotic approximations while continuing to find novel applications of shape restrictions for establishing point (or partial) identification. Finally, these developments in econometrics have been complemented by a growing literature in statistics focusing on shape-restricted estimation and inference.

In this article, we aim to provide an introduction to the complementary literatures in which shape restrictions have played a role. We take as our starting point an excellent earlier review by Matzkin (1994) and focus primarily on the progress made in the past 20 years. The breadth, scope, and sometimes technically challenging nature of the existing contributions make a detailed and comprehensive review impractical. As a result, we opt to instead structure our discussion around often simplified examples that nonetheless effectively illustrate important insights. We hope in this manner to provide the reader with not only an overview of recent advances, but also a helpful entry point into the different strands of the literature.

We begin, in Section 2, by discussing examples of the different roles that shape restrictions have played in empirical and theoretical work. Our selection of examples is necessarily nonexhaustive and intended primarily to illustrate the diversity of applications of shape restrictions. In identification analysis, for example, shape restrictions have often been imposed to achieve point identification or narrow the identified set of a partially identified parameter. Testing for the validity of shape restrictions is also often of interest, as their violation may provide evidence against particular economic theories, while their presence can have strong economic implications. Finally, shape restrictions that are deemed to hold can sometimes be employed to obtain more powerful tests and more accurate estimators—insights that have been applied to areas as diverse as state price density estimation and inference in regression discontinuity analysis.

In Section 3, we aim to provide insight into the methodological challenges that arise in estimation and inference under shape restrictions. Heuristically, the impact of shape restrictions on the finite-sample distribution of statistics depends on two main factors: (a) the degree of sampling uncertainty and (b) the region of the parameter space in which the underlying parameter lies. For instance, when imposing a shape restriction such as monotonicity on an identified function  $\theta_0$ , the finite-sample distribution of a constrained estimator depends on both the steepness of  $\theta_0$  and the statistical precision with which  $\theta_0$  can be estimated. Thus, shape restrictions can prove particularly helpful in applications in which the shape restrictions are close to binding or the model is hard to estimate—e.g., when the sample size is small, an unconstrained estimator for  $\theta_0$  has a slow rate of convergence, or the model is high dimensional. We emphasize, however, that it is precisely when shape restrictions are most informative that conventional asymptotic analysis may be unreliable. We illustrate these insights from the literature with a numerical example of the impact of imposing the law of demand in estimation. Fortunately, the econometrics literature has developed asymptotic approximations addressing this concern.

Finally, Sections 4 and 5 summarize recent developments in the theories of estimation and inference under shape restrictions, respectively. With regard to estimation, we discuss alternative methodologies for imposing shape restrictions and understanding the finite-sample properties of the resulting estimators. With regard to inference, we review different strategies for testing for shape restrictions and employing them to obtain sharper inference on an underlying parameter.

Throughout Sections 4 and 5, we again employ specific examples to guide our discussion. Our intent in this regard is to introduce the general insights of the broader literature by illustrating them through concrete statistical procedures. We thus hope that the reader does not attribute undue prominence to the selected examples, but instead finds the discussion of them to be a helpful starting point toward a more in-depth exploration of the literature.

## 2. THE ROLES OF SHAPE RESTRICTIONS

Shape restrictions can play a variety of roles in identification, estimation, and inference. In this section, we illustrate these uses by discussing different applications from the literature. Our examples are necessarily nonexhaustive and purposely selected with the aim of illustrating the diversity of applications of shape restrictions.

### 2.1. Establishing Point Identification

Imposing shape restrictions can be a powerful device for establishing identification of a parameter of interest. An influential example of this approach is developed by Imbens & Angrist (1994), who employ monotonicity to identify a local average treatment effect.

Consider a setting in which there are two potential outcomes  $(Y_0, Y_1)$ , a binary instrument  $Z \in \{0, 1\}$ , and two potential treatment decisions  $(D_0, D_1)$ . The observable variables are  $Z$ , the treatment decision  $D$ , and the outcome  $Y$ , which equal

$$D \equiv (1 - Z)D_0 + ZD_1, \quad Y \equiv (1 - D)Y_0 + DY_1. \quad 1.$$

Assuming that  $(Y_0, Y_1, D_0, D_1)$  are independent of  $Z$ , it then follows from Equation 1 that

$$E[Y|Z = 1] - E[Y|Z = 0] = E[Y_1 - Y_0|D_1 - D_0 = 1]P(D_1 - D_0 = 1) \quad 2.$$

$$- E[Y_1 - Y_0|D_0 - D_1 = 1]P(D_0 - D_1 = 1). \quad 3.$$

Heuristically, the above decomposition consists of the average treatment effect for individuals induced into treatment by a change of  $Z$  from zero to one (i.e., Equation 2) and the average treatment effect for individuals induced out of treatment by the same change in  $Z$  (i.e., Equation 3). The conflation of these average treatment effects presents a fundamental impediment to identifying the causal effect of treatment.

To resolve this challenge, Imbens & Angrist (1994) stipulate that the treatment be monotone in  $z$ —i.e., either  $D_1 \geq D_0$  almost surely or  $D_0 \geq D_1$  almost surely. Under this condition, assuming  $D_1 \geq D_0$ , the term in Equation 3 equals zero and we obtain

$$\frac{E[Y|Z = 1] - E[Y|Z = 0]}{P(D = 1|Z = 1) - P(D = 1|Z = 0)} = E[Y_1 - Y_0|D_1 - D_0 = 1]. \quad 4.$$

Thus, monotonicity enables us to identify the average treatment effect for individuals switched into treatment by the instrument.

Interestingly, the monotonicity restriction is equivalent to the existence of a latent index structure (Vytlacil 2002), which may also be viewed as a shape restriction. Heckman & Vytlacil (2005) and the references they cite provide further discussion of this. In particular, they employ this latent index structure to study the identification of what they refer to as policy-relevant treatment effects. Heckman & Pinto (2017) develop a more general notion of monotonicity, termed unordered monotonicity, that is motivated by choice-theoretic restrictions and applies to settings in which there is more than one treatment (see also Lee & Salanié 2017 for related results

concerning multiple treatments). Finally, we note that there is an extensive literature studying partial identification of treatment effects under shape restrictions, which we discuss in Section 2.2.

We conclude by noting that shape restrictions motivated by economic theory have been extensively used for identification by Matzkin (1991, 1992). More recently, Allen & Rehbeck (2016) employ a version of Slutsky symmetry to establish identification in a class of consumer choice models. In single equation models in which unobserved heterogeneity enters in a nonadditively separable manner, monotonicity is often employed to establish identification under both exogeneity (Matzkin 2003) and endogeneity (Chernozhukov & Hansen 2005). Similar arguments have also been successfully applied in nonseparable triangular models by Chesher (2003), Imbens & Newey (2009), Torgovitsky (2015), and D’Haultfoeuille & Février (2015). Shi & Shum (2016) employ a generalization of monotonicity, termed cyclic monotonicity, to establish identification in multinomial choice models with fixed effects (see also Pakes & Porter 2013).

## 2.2. Improving Partial Identification

In certain applications, shape restrictions may fail to deliver point identification but nonetheless provide informative bounds on the parameter of interest (Manski 1997). A particularly successful empirical application of this approach is developed by Blundell et al. (2007b), who examine the evolution of wage inequality in the United Kingdom.

Concretely, letting  $W$  denote log-wages,  $D \in \{0, 1\}$  denote a dummy variable indicating employment, and  $X$  denote a set of demographic characteristics, Blundell et al. (2007b) study how the interquartile range (IQR) of  $W$  conditional on  $X$  has evolved through time. The main challenge in their analysis is that the IQR is not (point) identified in the presence of selection into employment. The lack of identification follows from

$$P(W \leq c|X) = P(W \leq c|X, D = 0)P(D = 0|X) + P(W \leq c|X, D = 1)P(D = 1|X), \quad 5.$$

which emphasizes the dependence of the conditional distribution of  $W$  given  $X$  on the unidentified distribution of wages of the unemployed. One can further use Equation 5 to bound the conditional distribution of wages, and in turn the IQR, by noting that the unidentified distribution of wages of the unemployed must be bounded between zero and one. These so-called worst-case bounds were first studied by Manski (1989).

Blundell et al. (2007b) supplement the worst-case analysis by imposing additional shape restrictions that help narrow the bounds for the IQR. For example, in the presence of positive selection into employment, the distribution of  $W$  for workers first-order stochastically dominates the distribution of  $W$  for nonworkers—i.e., for all  $c \in \mathbf{R}$ ,

$$P(W \leq c|X, D = 1) \leq P(W \leq c|X, D = 0). \quad 6.$$

The restriction in Equation 6 can be combined with Equation 5 to improve on the worst-case bounds for the IQR. Alternatively, for  $Z$  equal to the unemployment benefits an individual is eligible for when unemployed, Blundell et al. (2007b) also examine the implications of imposing

$$P(W \leq c|X, Z') \leq P(W \leq c|X, Z) \quad 7.$$

whenever  $Z' \geq Z$  (see also Manski & Pepper 2000 for related restrictions). The constraints in both Equation 6 and Equation 7 prove to be informative, yielding empirically tighter bounds for the change in the IQR of log-wages of men between 1978 and 1998.

In related work, Kreider et al. (2012) apply shape restrictions to study the efficacy of the food stamps program. Bhattacharya et al. (2008) and Machado et al. (2013) find monotonicity restrictions that can be informative even if one is unwilling to assume the direction of the dependence

in  $Z$  (i.e., nondecreasing or nonincreasing). Lee (2009) bounds the average treatment effect of job training programs in the United States by exploiting the monotonicity restriction of Imbens & Angrist (1994). Finally, Kline & Tartari (2016) and Lee & Bhattacharya (2016) employ revealed preference and Slutsky-type restrictions, respectively, to sharpen their bounds.

### 2.3. Testing Model Implications

Economic theory sometimes yields testable implications that can be characterized through shape restrictions. An interesting example of this phenomenon arises in auction theory.

Consider a first price sealed bid auction with  $I$  bidders having independent and identically distributed valuations. The Bayesian Nash equilibrium in this auction is unique and symmetric, so that the resulting bids are independent and identically distributed as well. Since bids are observed and valuations are not, an interesting question is whether there exists a distribution of valuations such that the distribution of bids is compatible with bidders playing a Bayesian Nash equilibrium. Guerre et al. (2000) find that, for the distribution of bids to be compatible with a Bayesian Nash equilibrium, the function

$$\xi(b) \equiv b + \frac{G(b)}{(I-1)G'(b)} \quad 8.$$

must be strictly increasing in  $b$ , where  $G$  denotes the cumulative distribution function of the distribution of bids. Thus, monotonicity arises as a key testable implication of the model. An analogous result for affiliated private values has been established by Li et al. (2002) and Athey & Haile (2007). Lee et al. (2018) develop a general procedure that may be applied to test these monotonicity restrictions, while Jun et al. (2010) construct a nonparametric test of affiliation in auction models.

Additional examples of shape restrictions as testable implications are present in consumer theory (Samuelson 1938). In this vein, McFadden & Richter (1990) characterize the empirical content of random utility models (see also Kitamura & Stoye 2013 for a formal test). More recently, Bhattacharya (2017) characterizes the empirical content of discrete choice models as shape restrictions on the conditional choice probabilities. We also note that the instrumental variables model examined in Section 2.1 generates restrictions on the distribution of the observed data (see, e.g., Balke & Pearl 1997, Heckman & Vytlacil 2001, Imbens & Rubin 1997, Kitagawa 2015, Machado et al. 2013). Ellison & Ellison (2011) find that a test for monotonicity can be employed to detect strategic investments by firms that aim to deter entrance into their markets.

### 2.4. Delivering Economic Implications

In certain applications, whether shape restrictions are satisfied has strong economic implications. A central example is whether goods are, loosely speaking, complements or substitutes—concepts that can often be formalized through the shape restrictions of supermodularity and submodularity (Milgrom & Roberts 1995).

Supermodularity has particularly strong implications in matching markets. Following Shimer & Smith (2000), consider a two-sided market where workers are matched with firms. Unmatched workers of type  $X \in [0, 1]$  engage in a random search and, upon meeting a firm of type  $Y \in [0, 1]$ , can generate output  $V$ , given by

$$V = F(X, Y),$$

where the production function  $F$  is assumed to be strictly increasing in  $X$  and  $Y$ . In this model, Shimer & Smith (2000) establish that supermodularity of  $F$  (and some of its derivatives) implies

positive assortative matching (PAM) in equilibrium, i.e., higher-type workers are employed by higher-type firms. Thus, higher-type workers receive higher salaries both due to their type and by virtue of being matched to higher-type firms. As a result, supermodularity of  $F$  can translate into higher dispersion in wages.

The implications of PAM for the wage distribution and the increasing availability of employer–employee matched data sets have motivated an important empirical literature (see Card et al. 2016 for a recent review). For example, following Abowd et al. (1999), several studies have estimated worker-specific and firm-specific fixed effects and found little correlation between them. However, as noted by Eeckhout & Kircher (2011), these fixed effects need not be connected to the underlying firm and worker types. Hagedorn et al. (2017) propose an estimator of  $F$ , but its asymptotic properties are unknown. To our knowledge, no test of PAM or supermodularity of  $F$  is available.

In related work, Athey & Stern (1998) employ supermodularity to define whether different firm organizational practices are complements or substitutes. Kretschmer et al. (2012) apply their approach to determine whether the adoption of a new software application is complementary to the scale of production. A novel model for studying whether goods are complements is introduced by Gentzkow (2007), who examines whether print and online media act as complements or substitutes. The nonparametric identification of such a model is established by Fox & Lazzati (2013) (see Chernozhukov et al. 2015 for a test of complementarity).

## 2.5. Informing Estimation

When shape restrictions implied by economic theory are deemed to hold, they can be employed in applications to improve estimation of a parameter of interest. This approach has been pursued, for example, by Ait-Sahalia & Duarte (2003) in the nonparametric estimation of the state price density (SPD) function.

Consider a call option on an asset with strike price  $X$  expiring at time  $T$ . For  $S_t$ , the price of the underlying asset at time  $t$ ;  $r$ , the deterministic risk free rate; and  $p^*$ , the SPD, the price  $C(S_t, X, r)$  of the call option at time  $t$  is given by

$$C(S_t, X, r) = e^{-(T-t)r} \int_0^\infty \max\{S_T - X, 0\} p^*(S_T) dS_T. \quad 9.$$

In this case, we have, for simplicity, omitted the dependence on the dividend yields of the asset and other state variables. Differentiating Equation 9 with respect to  $X$  implies that

$$-e^{-(T-t)r} \leq \frac{\partial}{\partial X} C(S_t, X, r) \leq 0 \leq \frac{\partial^2}{\partial X^2} C(S_t, X, r) = p^*(X) e^{(T-t)r}. \quad 10.$$

Exploiting Equation 10, Ait-Sahalia & Lo (1998) construct an unconstrained nonparametric estimator of the SPD by estimating the second derivative of the pricing function  $C$  with respect to the strike price  $X$ . The derivation in Equation 10, however, further implies that the call option pricing function must be nonincreasing and convex in the strike price. Building on this observation, Ait-Sahalia & Duarte (2003) build a nonparametric estimator of  $C$  that satisfies the constraints in Equation 10, which they in turn differentiate to estimate the SPD. In estimating the S&P 500 Index SPD, they find that the constrained nonparametric estimator outperforms the constrained estimator.

A related literature has noted that, in disagreement with theoretical expectations, estimates of the pricing kernel are often nonmonotonic (Rosenberg & Engle 2002). As a result, a series of studies has tested whether the violations from monotonicity are statistically significant (see, e.g., Beare & Schmidt 2016). Beare & Dossani (2018) impose monotonicity of the pricing kernel to inform forecasts. Within economics, monotonicity constraints have been imposed by Henderson

et al. (2012) in the empirical study of auctions. Restrictions from consumer theory, such as Slutsky inequalities, are imposed in estimation under exogeneity of prices by Blundell et al. (2012) and under endogeneity by Blundell et al. (2013).

## 2.6. Informing Inference

Finally, shape restrictions may help conduct inference on parameters of interest. In this section, we present an example of this way of using shape restrictions from Armstrong (2015).

Consider a sharp regression discontinuity (RD) model in which, for an outcome  $Y \in \mathbf{R}$ ,

$$Y = \theta_0(R) + \epsilon, \quad E[\epsilon | R] = 0,$$

where  $R \in \mathbf{R}$ , and an individual is assigned to treatment whenever  $R > 0$ . In certain applications, a researcher may be confident in maintaining that  $\theta_0$  is nondecreasing near (but not necessarily at) the discontinuity point zero. Armstrong (2015) demonstrates that such knowledge can be exploited in the construction of one-sided confidence intervals for the average treatment effect at zero, which equals  $\lim_{r \downarrow 0} \theta_0(r) - \lim_{r \uparrow 0} \theta_0(r)$  (see Hahn et al. 2001 for explanations in terms of the potential outcome framework). In particular, given a sample  $\{Y_i, R_i\}_{i=1}^n$ , one can define the one-sided  $k$ -nearest-neighbor estimators as

$$\hat{\theta}_{+,k}(0) \equiv \frac{1}{k} \sum_{i \in A_+(k)} Y_i, \quad \hat{\theta}_{-,k}(0) \equiv \frac{1}{k} \sum_{i \in A_-(k)} Y_i,$$

where  $A_+(k) \equiv \{i : \sum_{j=1}^n 1\{0 < R_j \leq R_i\} \leq k\}$  and  $A_-(k) \equiv \{i : \sum_{j=1}^n 1\{R_j \leq R_i \leq 0\} \leq k\}$ . The monotonicity of  $\theta_0$  ensures directional control of the bias, which greatly facilitates the choice of  $k$  in an optimal (minimax) way (see Section 5.2 for detailed related arguments). Concretely, let  $\Delta \hat{\theta}_k(0) \equiv \hat{\theta}_{+,k}(0) - \hat{\theta}_{-,k}(0)$ , let  $c_\alpha$  be the  $\alpha$  quantile of

$$\min_{k_{\min} \leq k \leq k_{\max}} \sqrt{k} \{ \Delta \hat{\theta}_k(0) - E[\Delta \hat{\theta}_k(0) | \{R_i\}_{i=1}^n] \}$$

conditional on  $\{R_i\}_{i=1}^n$ , and let  $k_{\min} \leq k_{\max}$  be given. The one-sided confidence interval

$$\left( -\infty, \min_{k_{\min} \leq k \leq k_{\max}} \left\{ \Delta \hat{\theta}_k(0) - \frac{c_\alpha}{\sqrt{k}} \right\} \right] \quad 11.$$

then possesses asymptotic coverage probability  $1 - \alpha$  despite  $k$  being chosen in Equation 11 to make the interval as short as possible. Whenever the distribution of  $\epsilon$  is known, as in the work of Armstrong (2015), the resulting procedure is tuning parameter free in that we may set  $k_{\min} = 1$  and  $k_{\max} = n$ . In contrast, if the distribution of  $\epsilon$  is unknown, then  $k_{\min}$  and  $k_{\max}$  may be set to equal  $k_{\min} = \sqrt{n}$  and  $k_{\max} = n / \log(n)$ , and  $c_\alpha$  can be estimated using bootstrap methods such as those of Chetverikov (2012).

The work of Armstrong (2015) is one of numerous recent advances in the theoretical literature studying inference under shape restrictions. Other recent contributions include those of Freyberger & Horowitz (2015), Chernozhukov et al. (2015), Freyberger & Reeves (2017), Horowitz & Lee (2017), and Mogstad et al. (2017) from econometrics and Dümbgen (2003) and Cai et al. (2013) from statistics. We review this literature in Section 5.

## 3. INTUITION FOR ASYMPTOTICS

A common feature of the examples in Section 2 is that shape restrictions can affect the distribution of statistics in nonstandard ways (Andrews 1999, 2001). Before discussing estimation and inference,

we therefore first develop insight into the methodological complications that arise from imposing shape restrictions. Specifically, we focus on situations when we might expect shape restrictions to matter and on the appropriateness of different asymptotic frameworks.

### 3.1. Basic Model

We base our exposition on a simple example inspired by Dupas (2014), who conducts a randomized pricing experiment of malaria nets. Consider a sample of  $n$  individuals, each of whom is independently assigned a price  $X_i \in \{L, M, H\}$  with probabilities

$$P(X_i = L) = P(X_i = M) = P(X_i = H) = \frac{1}{3}.$$

Upon observing the price, individual  $i$  decides whether to purchase the net, and we let  $Y_i$  be a binary variable indicating purchase. The parameters of interest are

$$\Delta_j \equiv P(Y_i = 1 | X_i = j)$$

for  $j \in \{L, M, H\}$ . We study, for different values of  $b \geq 0$ , the specification

$$\Delta_L = \Delta_M + b, \quad \Delta_M = \frac{1}{2}, \quad \Delta_H = \Delta_M - b. \quad 12.$$

We consider two different estimators for  $\Delta \equiv (\Delta_L, \Delta_M, \Delta_H)$ . First, we examine a constrained estimator that imposes the law of demand  $\Delta_L \geq \Delta_M \geq \Delta_H$ :

$$(\hat{\Delta}_L^C, \hat{\Delta}_M^C, \hat{\Delta}_H^C) \equiv \arg \min_{\delta_L \geq \delta_M \geq \delta_H} \frac{1}{n} \sum_{i=1}^n \left( Y_i - \sum_{j \in \{L, M, H\}} \delta_j 1\{X_i = j\} \right)^2. \quad 13.$$

Second, we examine an unconstrained estimator  $\hat{\Delta}^U \equiv (\hat{\Delta}_L^U, \hat{\Delta}_M^U, \hat{\Delta}_H^U)$  that minimizes the same criterion as that in Equation 13 but without imposing the constraint  $\delta_L \geq \delta_M \geq \delta_H$ .

### 3.2. Pointwise Asymptotics

Early research on shape restrictions made the observation that, if the restrictions hold strictly, then the unconstrained estimator will asymptotically satisfy the constraints. To illustrate this logic, suppose  $b > 0$  in Equation 12 so that the law of demand inequalities holds strictly. Due to the consistency of the unconstrained estimators, it then follows that

$$\lim_{n \rightarrow \infty} P_b(\hat{\Delta}_L^U - \hat{\Delta}_M^U > 0 \text{ and } \hat{\Delta}_M^U - \hat{\Delta}_H^U > 0) = 1, \quad 14.$$

where we write  $P_b$  in place of  $P$  to emphasize that the probability depends on  $b$ . However, if the unconstrained estimator satisfies the law of demand, then it must also solve the constrained optimization problem in Equation 13. In other words, Equation 14 implies that the constrained and unconstrained estimators equal each other with probability tending to one.

The preceding arguments rely on pointwise asymptotics—the name reflecting the fact that  $b$  is held fixed as  $n$  diverges to infinity. Somewhat negatively, these arguments seem to imply that imposing shape restrictions has no effect. Yet such a theoretical conclusion clashes with empirical studies that have found imposing shape restrictions to be informative in a variety of contexts (Ait-Sahalia & Duarte 2003, Blundell et al. 2012). This apparent tension may be reconciled by noting that, for a given sample size  $n$ , the probability on the left-hand side of Equation 14 may be far from one. Whenever this is the case, pointwise asymptotics do not reflect the finite-sample



situation, and as we see in simulations below, approximations based on these asymptotics can be very misleading.

In the next section, we describe an alternative asymptotic framework that better reflects a finite-sample setting in which shape restrictions are informative. Before proceeding, however, we note that, in some cases, nonasymptotic (i.e., finite-sample) bounds on the error of estimators subject to shape restrictions are also available (see Section 4.2.2 for a discussion of some such results; see also Chetverikov & Wilhelm 2017).

### 3.3. Local Asymptotics

A local asymptotic analysis is one way to improve on a pointwise asymptotic approximation. A prominent example of such an approach is that developed by Staiger & Stock (1997), who use weak-instrument asymptotics to model a finite-sample situation in which the first stage  $F$  statistic is small. For our purposes, we desire a local asymptotic analysis that reflects a finite-sample setting in which imposing shape restrictions proves informative.

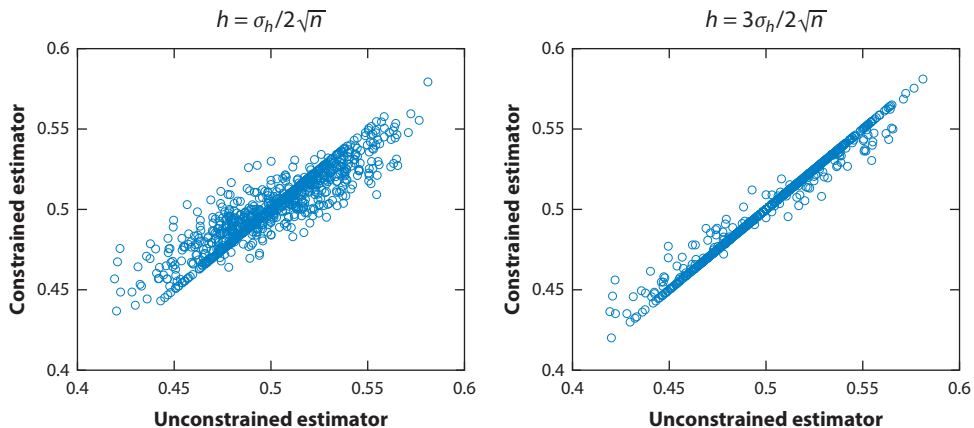
The first step in such an analysis is to develop an understanding of when we might expect shape restrictions to be informative. To this end, we return to our example and note that the estimators  $\sqrt{n}\{\hat{\Delta}_L^U - \hat{\Delta}_M^U\}$  and  $\sqrt{n}\{\hat{\Delta}_M^U - \hat{\Delta}_H^U\}$  are approximately normal with

$$(\hat{\Delta}_L^U - \hat{\Delta}_M^U) \approx N\left(b, \frac{\sigma_b}{\sqrt{n}}\right), \quad (\hat{\Delta}_M^U - \hat{\Delta}_H^U) \approx N\left(b, \frac{\sigma_b}{\sqrt{n}}\right), \quad 15.$$

where, in our design, the standard deviation is the same for both constraints. When  $b$  is large relative to  $\sigma_b/\sqrt{n}$ , the demand function is sufficiently elastic that the unconstrained estimator satisfies the law of demand with high probability. In contrast, when  $b$  is of the same order as  $\sigma_b/\sqrt{n}$  (or smaller), the amount of sampling uncertainty is such that a priori knowledge of the law of demand is informative. Therefore, whether imposing the law of demand affects estimation and inference depends on the ratio of the elasticity of demand (as measured by  $b$ ) to the amount of sampling uncertainty (as measured by  $\sigma_b/\sqrt{n}$ ).

Pointwise asymptotics (i.e., Equation 14) that rely on  $b$  being fixed as  $n$  diverges to infinity require that  $\sigma_b/\sqrt{n}$  be small relative to  $b$ . In this way, they mechanically model a finite-sample setting in which shape restrictions have no effect. To move away from this paradigm, we must consider an asymptotic framework in which  $b$  and  $\sigma_b/\sqrt{n}$  remain of the same order regardless of the sample size. As a consequence of Equation 15, such a framework ensures that the unconstrained estimators violate the law of demand with positive probability even as  $n$  diverges to infinity—i.e., shape restrictions remain informative asymptotically. The resulting analysis is termed local in that  $b$  is thus modeled as tending to zero with the sample size and is thus local to zero. We stress, however, that it is incorrect to think of a local analysis as merely modeling inelastic demand curves. Rather, local asymptotics are simply a device for approximating finite-sample settings in which the amount of sampling uncertainty renders imposing the law of demand informative.

**Figure 1** depicts scatter plots of the constrained versus unconstrained estimators of  $\Delta_M$  for different values of  $\sqrt{nb}/\sigma_b$ . As expected from the preceding discussion, we see that the differences between the constrained and unconstrained estimator decrease as  $\sqrt{nb}/\sigma_b$  increases. While **Figure 1** is based on simulations with  $n$  equal to a thousand, the results are qualitatively similar for different values of  $n$ . **Figure 1** hides, however, that the value of  $\sqrt{nb}/\sigma_b$  affects the distribution of the constrained estimator but not the distribution of the unconstrained estimator. This contrast is illustrated in **Table 1**, which summarizes the mean squared error (MSE) for the constrained and unconstrained estimators for  $\Delta_M$  (scaled by  $n$ ). In accord with our discussion, we find that, when sampling uncertainty (as measured  $\sigma_b/\sqrt{n}$ ) is large relative to  $b$ , imposing the law of demand



**Figure 1**

Estimators of  $\Delta_M$  and the local parameter.

proves informative, and the constrained estimator outperforms its unconstrained counterpart. In contrast, as  $\sqrt{nb}/\sigma_b$  increases, we find that the improvements in estimation obtained by the constrained estimator diminish.

We conclude this section with a few important takeaways. First, the finite-sample distribution of statistics can be significantly impacted by the presence of shape restrictions. As a result, it is imperative to employ asymptotic frameworks that reflect this phenomenon, such as the local approximations discussed above. Second, the higher the degree of sampling uncertainty, the more informative shape restrictions may be. This importance of sampling uncertainty is dramatically exemplified by Chetverikov & Wilhelm (2017), who study the impact of imposing monotonicity in nonparametric instrumental variable regression—a setting in which the rate of convergence can be as slow as logarithmic in  $n$  (Blundell et al. 2007a, Hall & Horowitz 2005).

**Remark 1.** Local asymptotic analysis arises naturally in establishing the uniform asymptotic validity of statistical procedures, such as tests and confidence regions. This more demanding notion of validity often leads to procedures that have desirable properties in finite samples (see, e.g., the discussion in Andrews et al. 2011, Romano & Shaikh 2012). Its importance in the analysis of nonstandard problems has been recently recognized in a variety of applications (see, e.g., Andrews & Cheng 2012, Leeb & Pötscher 2005, Mikusheva 2007). In the case of shape restrictions, such a notion of validity would, in particular, ensure that a test has approximately the right size or that a confidence region has approximately the right coverage probability in large samples regardless of the informativeness of the shape restrictions.

**Table 1** Scaled mean squared error

	$\sqrt{nb}/\sigma_b$								
	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00
Constrained estimator	0.32	0.39	0.45	0.52	0.59	0.64	0.69	0.72	0.74
Unconstrained estimator	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77

**Remark 2.** Multiple shape restrictions, such as concavity, monotonicity, and supermodularity, may be intuitively thought of as inequality restrictions. In contrast, other shape restrictions, such as symmetry (Haag et al. 2009, Lewbel 1995), homogeneity (Keuzenkamp & Barten 1995, Tripathi & Kim 2003), or certain semi- or nonparametric specifications (Blundell et al. 2007a), can be thought of as equality restrictions. It is worth noting that pointwise asymptotic approximations are often more reliable under equality restrictions than under inequality restrictions.

## 4. ESTIMATION

In this section, we discuss methods for estimating parameters that satisfy a conjectured shape restriction. We organize our discussion around two approaches: (a) estimators that are built by imposing a shape restriction on an originally unconstrained estimator and (b) estimators that are obtained as constrained optimizers to a criterion function.

### 4.1. Building on Unconstrained Estimators

In many applications, an unconstrained estimator for a parameter of interest is readily available. Such an estimator may then be transformed to satisfy a desired shape restriction in a variety of ways. Because unconstrained estimators are often easy to compute and analyze, these two-step approaches can be computationally and theoretically straightforward.

In this section, we denote the parameter of interest by  $\theta_0$  and presume that we have an estimator  $\hat{\theta}_n$  available for it. It is important to be explicit about the space in which  $\theta_0$  and  $\hat{\theta}_n$  reside, and we therefore let  $\theta_0, \hat{\theta}_n \in \mathbf{D}$ , where  $\mathbf{D}$  is a complete vector space with norm  $\|\cdot\|_{\mathbf{D}}$ —i.e.,  $\mathbf{D}$  is a Banach space. Our objective is to understand the properties of an estimator  $\hat{\theta}_n^{2s}$  that is obtained by imposing the relevant shape restriction on  $\hat{\theta}_n$ . Formally,  $\hat{\theta}_n^{2s}$  and  $\hat{\theta}_n$  are therefore related by a known transformation  $\phi: \mathbf{D} \rightarrow \mathbf{D}$  that maps the unconstrained estimator into a constrained version of it—i.e.,  $\hat{\theta}_n^{2s} = \phi(\hat{\theta}_n)$ .

To fix ideas, we introduce three examples of transformations  $\phi$ .

**Example 1.** When estimating quantile functions, we face the possibility that our estimators are not monotonic in the quantile. This quantile crossing can manifest itself, for example, when employing quantile regression or quantile instrumental variable methods (Abadie et al. 2002, Chernozhukov & Hansen 2005). Suppose we observe  $\{Y_i, X_i, D_i\}_{i=1}^n$  with  $Y_i, D_i \in \mathbf{R}$ ,  $X_i \in \mathbf{R}^{d_x}$ , and we estimate

$$[\hat{\beta}_n(\tau), \hat{\theta}_n(\tau)] \equiv \arg \min_{(\beta, \theta)} \frac{1}{n} \sum_{i=1}^n \rho_{\tau}(Y_i - X_i' \beta - D_i \theta), \quad 16.$$

where  $\rho_{\tau}(u)$  is the check function  $\rho_{\tau}(u) \equiv u(\tau - 1\{u < 0\})$ . We are interested in the quantile regression function  $\tau \mapsto \hat{\theta}_n(\tau)$ , which should be monotonic in  $\tau$ . The estimation procedure in Equation 16, however, does not guarantee that  $\hat{\theta}_n$  will be monotone. Chernozhukov et al. (2010) therefore suggest instead employing an estimator  $\hat{\theta}_n^{2s} \equiv \phi(\hat{\theta}_n)$ , where

$$\phi(\theta)(\tau) \equiv \inf \left\{ c : \int_0^1 1\{\theta(\bar{u}) \leq c\} d\bar{u} \geq \tau \right\}. \quad 17.$$

The resulting estimator  $\hat{\theta}_n^{2s}$  is called the monotone rearrangement of  $\hat{\theta}_n$ . Intuitively,  $\hat{\theta}_n^{2s}(\tau)$  is simply the  $\tau$ th quantile of  $\{\hat{\theta}_n(u) : u \in [0, 1]\}$ , and therefore  $\hat{\theta}_n^{2s}$  is monotonic.

**Example 2.** Building on Example 1, an alternative to employing the monotone re-arrangement of  $\hat{\theta}_n$  is to instead let  $\hat{\theta}_n^{2s}$  be the closest monotone function to  $\hat{\theta}_n$ , e.g.,

$$\hat{\theta}_n^{2s} \equiv \arg \min_{f: [0,1] \rightarrow \mathbf{R}} \int_0^1 [\hat{\theta}_n(u) - f(u)]^2 du \quad \text{s.t. } f \text{ nondecreasing.} \quad 18.$$

In practice, Equation 18 may be solved over a grid of  $[0, 1]$ . Importantly, this approach can be easily generalized to shape restrictions beyond monotonicity. To this end, recall that  $\hat{\theta}_n$  is in a space  $\mathbf{D}$  with norm  $\|\cdot\|_{\mathbf{D}}$  and note that we may think of the set of parameters satisfying a shape restriction as a subset  $C \subset \mathbf{D}$ ; for example, in Equation 18,  $\|\theta\|_{\mathbf{D}}^2 = \int \theta(u)^2 du$  and  $C$  is the set of nondecreasing functions. We may then let  $\hat{\theta}_n^{2s}$  be the closest parameter to  $\hat{\theta}_n$  satisfying the desired shape restriction by defining  $\phi: \mathbf{D} \rightarrow \mathbf{D}$  to equal

$$\phi(\theta) \equiv \arg \min_{f \in C} \|f - \theta\|_{\mathbf{D}} \quad 19.$$

and setting  $\hat{\theta}_n^{2s} = \phi(\hat{\theta}_n)$ . Applying this approach, Fang & Santos (2014) compare an unconstrained trend in the dispersion of residual wage inequality to the closest concave trend to examine whether skill-biased technical change has decelerated.

**Example 3.** An alternative approach to Equation 19 for imposing concavity is to employ the least concave majorant (LCM) of a function. Specifically, for a bounded function  $\theta$  defined on, e.g.,  $[0, 1]$ , the LCM of  $\theta$  is the function  $\phi(\theta)$  defined pointwise as

$$\phi(\theta)(u) \equiv \inf\{g(u) : g \text{ is concave and } g(u) \geq \theta(u) \text{ for all } u \in [0, 1]\}. \quad 20.$$

Intuitively, the LCM of  $\theta$  is the smallest concave function that is larger than  $\theta$ . Thus, letting  $\hat{\theta}_n^{2s} \equiv \phi(\hat{\theta}_n)$ , we obtain a concave function  $\hat{\theta}_n^{2s}$  as a transformation of  $\hat{\theta}_n$ . The LCM has been widely studied in statistics (see Robertson et al. 1988). Within econometrics, the LCM has been employed by Delgado & Escanciano (2012) in testing stochastic monotonicity, Beare & Schmidt (2016) in examining the monotonicity of the pricing kernel, and Luo & Wan (2018) in studying auctions.

**4.1.1. Local analysis via the delta method.** As emphasized in Section 3, it is important to employ asymptotic approximations that accurately reflect the impact of shape restrictions on the finite-sample distribution of statistics. Two features of the present context make developing a local approximation particularly tractable. First,  $\hat{\theta}_n^{2s}$  is a deterministic transformation of an original estimator  $\hat{\theta}_n$ . Second,  $\hat{\theta}_n$  is unconstrained, and thus its asymptotic distribution is often readily available. These two aspects of the problem make it amenable to the delta method.

In this section, we keep the exposition informal for conciseness but refer the reader to the cited material for additional details. Since we are interested in a local approximation, we let the distribution of the data depend on the sample size  $n$  and denote it by  $P_n$ . The parameter of interest therefore also depends on  $n$ , and we denote it by  $\theta_{0,n}$ . For instance, in Example 1,  $\theta_{0,n}$  corresponds to the quantile coefficient function when the data are distributed according to  $P_n$ . It is, in addition, convenient to impose

$$\theta_{0,n} = \theta_0 + \frac{\lambda}{\sqrt{n}}, \quad 21.$$

where  $\theta_0$  may be understood as the limiting value of  $\theta_{0,n}$  along  $P_n$ , and  $\lambda \in \mathbf{D}$  is often referred to as the local parameter. Letting  $\xrightarrow{L_n}$  denote convergence in distribution along  $P_n$ , we assume that

$\hat{\theta}_n$  satisfies

$$\sqrt{n}\{\hat{\theta}_n - \theta_{0,n}\} \xrightarrow{L_n} \mathbb{G}_0, \quad 22.$$

where the limit  $\mathbb{G}_0$  does not depend on  $\lambda$ . Intuitively, Equation 22 demands that  $\hat{\theta}_n$  be robust to local perturbations of the underlying distribution—notice, e.g., that, in **Table 1**, the MSE of the unconstrained estimator does not depend on  $\sqrt{nh}/\sigma_b$ .

To complete our setup, we presume that  $\phi: \mathbf{D} \rightarrow \mathbf{D}$  maps any function satisfying the desired shape restriction into itself. Since  $\hat{\theta}_n^{2s} \equiv \phi(\hat{\theta}_n)$ , we may then write

$$\sqrt{n}\{\hat{\theta}_n^{2s} - \theta_{0,n}\} = \sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta_{0,n})\}, \quad 23.$$

where we exploit  $\phi(\theta_{0,n}) = \theta_{0,n}$  due to  $\theta_{0,n}$  satisfying the shape restriction. The equality in Equation 23, together with Equation 22, reveals the potential applicability of the delta method. However, one last obstacle remains: In our problems, the map  $\phi$  often fails to be (fully) differentiable. Fortunately, a remarkable extension of the delta method developed by Shapiro (1991) and Dümbgen (1993) continues to apply provided  $\phi$  is directionally differentiable instead. The relevant concepts of full and directional differentiability are as follows.

**Definition 1.** Let  $\mathbf{D}, \mathbf{E}$  be Banach spaces with norms  $\|\cdot\|_{\mathbf{D}}, \|\cdot\|_{\mathbf{E}}$ , and  $\phi: \mathbf{D} \rightarrow \mathbf{E}$ .

1.  $\phi$  is Hadamard differentiable at  $\theta$  if there is a continuous linear map  $\phi'_\theta: \mathbf{D} \rightarrow \mathbf{E}$  such that, for all sequences  $\{b_n\} \subset \mathbf{E}$  and  $\{t_n\} \subset \mathbf{R}$  with  $b_n \rightarrow b$  and  $t_n \rightarrow 0$ ,

$$\lim_{n \rightarrow \infty} \left\| \frac{\phi(\theta + t_n b_n) - \phi(\theta)}{t_n} - \phi'_\theta(b) \right\|_{\mathbf{E}} = 0. \quad 24.$$

2.  $\phi$  is Hadamard directionally differentiable at  $\theta$  if there is a continuous map  $\phi'_\theta: \mathbf{D} \rightarrow \mathbf{E}$  such that, for all sequences  $\{b_n\} \subset \mathbf{D}$  and  $\{t_n\} \subset \mathbf{R}_+$  with  $b_n \rightarrow b$  and  $t_n \downarrow 0$ ,

$$\lim_{n \rightarrow \infty} \left\| \frac{\phi(\theta + t_n b_n) - \phi(\theta)}{t_n} - \phi'_\theta(b) \right\|_{\mathbf{E}} = 0. \quad 25.$$

A map  $\phi$  is (fully) Hadamard differentiable at  $\theta$  if it can be locally approximated by a linear map  $\phi'_\theta$ . In turn,  $\phi$  is Hadamard directionally differentiable at  $\theta$  if a similar approximation requirement holds for a map  $\phi'_\theta$  that may no longer be linear. As an illustrative example, suppose  $\mathbf{D} = \mathbf{E} = \mathbf{R}$  and  $\phi(\theta) = \max\{\theta, 0\}$ . It is then straightforward to verify that, if we have  $\theta > 0$ , then  $\phi$  is Hadamard differentiable and we have  $\phi'_\theta(b) = b$  for all  $b \in \mathbf{R}$ . In contrast, if we have  $\theta = 0$ , then  $\phi$  is Hadamard directionally differentiable with  $\phi'_\theta(b) = \max\{b, 0\}$  for all  $b \in \mathbf{R}$ . We further note that, in some applications, a more general concept called tangential Hadamard (directional) differentiability is required.

Shapiro (1991) and Dümbgen (1993) originally noted that the delta method continues to apply when  $\phi$  is Hadamard directionally (but not fully) differentiable. In particular, the local analysis in Dümbgen (1993), together with Equations 21–23, establishes that

$$\sqrt{n}\{\hat{\theta}_n^{2s} - \theta_{0,n}\} \xrightarrow{L_n} \phi'_{\theta_0}(\mathbb{G}_0 + \lambda) - \phi'_{\theta_0}(\lambda). \quad 26.$$

Crucially, if  $\phi$  is (fully) Hadamard differentiable, then  $\phi'_{\theta_0}$  is linear, and Equation 26 implies that the asymptotic distribution of  $\hat{\theta}_n^{2s}$  does not depend on  $\lambda$ . In applications involving shape restrictions, however,  $\phi'_{\theta_0}$  is often nonlinear, reflecting that  $\phi$  is Hadamard directionally (but not fully) differentiable. In such instances, the limiting distribution in Equation 26 depends on  $\lambda$ , entailing an impact of shape restrictions on the finite-sample distribution (see, e.g., the MSE of the constrained estimator in **Table 1**). This dependence on the local parameter  $\lambda$  implies that, whenever

$\mathbb{G}_0$  is Gaussian, a naive plug-in bootstrap is inconsistent (Fang & Santos 2014). Nonetheless, the result in Equation 26 can be employed to study the validity of alternative resampling schemes such as the rescaled bootstrap (Dümbgen 1993),  $m$  out of  $n$  bootstrap (Shao 1994), or subsampling (Politis et al. 1999) (see, e.g., Hong & Li 2014). Finally, we note that Equation 26 can also be used to study the risk and optimality (or lack thereof) of estimators (Fang 2014).

Returning to our examples, Chernozhukov et al. (2010) establish the (full) Hadamard differentiability of the monotone rearrangement operator (i.e.,  $\phi$  as in Equation 17) at any strictly increasing  $\theta$ . Whenever  $\theta$  is not strictly increasing,  $\phi$  remains (fully) Hadamard differentiable if the domain of  $\theta$  is restricted to areas in which the derivative of  $\theta$  is bounded away from zero. Whether  $\phi$  remains Hadamard directionally differentiable without such domain restrictions appears to be an open question. We further note that the Hadamard directional differentiability of the projection operator (i.e.,  $\phi$  as in Equation 19) is shown by Zarantonello (1971) whenever  $C$  is closed and convex and  $\mathbf{D}$  is a Hilbert space. Finally, the Hadamard directional differentiability of the LCM operator (i.e.,  $\phi$  as in Equation 20) is proven by Beare & Moon (2015) and Beare & Fang (2018).

**4.1.2. Finite-sample improvements.** Chernozhukov et al. (2009) propose imposing monotonicity to improve confidence intervals for monotone functions. In this section, we apply their ideas to general shape restrictions.

For simplicity, we assume  $\theta_0$  is a scalar-valued bounded function with domain  $[0, 1]$ . In many applications, it is possible to construct a confidence interval for  $\theta_0$  over a subset  $A \subseteq [0, 1]$  by employing an unconstrained estimator  $\hat{\theta}_n$  (see, e.g., Belloni et al. 2015 and Chen & Christensen 2018 for constructions for nonparametric regression without and with endogeneity, respectively). These confidence intervals employ functions  $\hat{l}_n$  and  $\hat{u}_n$ , satisfying

$$\liminf_{n \rightarrow \infty} P(\hat{l}_n(u) \leq \theta_0(u) \leq \hat{u}_n(u) \text{ for all } u \in A) \geq 1 - \alpha \quad 27.$$

for some prespecified confidence level  $1 - \alpha$ . Moreover, the asymptotic coverage can often be shown to hold uniformly in a suitable class of underlying distributions.

Whenever  $\theta_0$  is known to satisfy a particular shape restriction, it may be desirable for  $\hat{l}_n$  and  $\hat{u}_n$  to satisfy it as well. Chernozhukov et al. (2009), for example, observe that imposing monotonicity on  $\hat{l}_n$  and  $\hat{u}_n$  can yield finite-sample improvements on confidence intervals for a monotone function  $\theta_0$ . Specifically, suppose  $\phi : \mathbf{D} \rightarrow \mathbf{D}$  assigns to any function  $\theta \in \mathbf{D}$  another function  $\phi(\theta) \in \mathbf{D}$  satisfying the desired shape restriction. Moreover, assume that (a)  $\phi(\theta) = \theta$  whenever  $\theta$  satisfies the shape restriction; (b)  $\phi$  satisfies

$$\phi(\theta_1)(u) \leq \phi(\theta_2)(u) \text{ for all } u \in [0, 1] \quad 28.$$

whenever  $\theta_1(u) \leq \theta_2(u)$  for all  $u \in [0, 1]$ ; and (c) for any  $\theta_1, \theta_2 \in \mathbf{D}$ , we have

$$\|\phi(\theta_1) - \phi(\theta_2)\|_{\mathbf{D}} \leq \|\theta_1 - \theta_2\|_{\mathbf{D}}. \quad 29.$$

For a map  $\phi$  satisfying these requirements, Chernozhukov et al. (2009) propose employing  $\hat{l}_n^{2s} \equiv \phi(\hat{l}_n)$  and  $\hat{u}_n^{2s} \equiv \phi(\hat{u}_n)$  to obtain a transformed confidence region for  $\theta_0$ . By construction,  $\hat{l}_n^{2s}$  and  $\hat{u}_n^{2s}$  now satisfy the shape restriction and

$$P(\hat{l}_n^{2s}(u) \leq \theta_0(u) \leq \hat{u}_n^{2s}(u) \text{ for all } u \in A) \geq P(\hat{l}_n(u) \leq \theta_0(u) \leq \hat{u}_n(u) \text{ for all } u \in A)$$

as a consequence of Equation 28 and  $\phi(\theta_0) = \theta_0$ . Thus, the transformed confidence region still has a confidence level of at least  $1 - \alpha$  as a consequence of Equation 27. Also, using the condition

in Equation 29, we can conclude that  $\|\hat{l}_n^{2s} - \hat{u}_n^{2s}\|_{\mathbf{D}} \leq \|\hat{l}_n - \hat{u}_n\|_{\mathbf{D}}$ , and thus that the new confidence region is, in this sense, no larger than the original.

Returning to our examples, Chernozhukov et al. (2009) establish that the monotone rearrangement operator (i.e.,  $\phi$  as in Equation 17) satisfies Equations 28 and 29. In this review, we also observe that the projection operator (i.e.,  $\phi$  as in Equation 18) satisfies the desired properties whenever  $\mathbf{D}$  is a Hilbert space,  $C$  is closed and convex, and the pointwise minimum and maximum of any  $\theta_1, \theta_2 \in C$  also belong to  $C$ .<sup>1</sup> The LCM operator (i.e.,  $\phi$  as in Equation 20) satisfying Equation 28 is immediate from its definition, while the fact that the LCM map satisfies the requirement in Equation 29 follows from theorem 5.11 of Eggermont & LaRiccia (2001).

Finally, we mention a recent proposal by Freyberger & Reeves (2017), who obtain confidence bands for certain parameters via test inversion. Their construction applies to a rich class of problems in which constrained estimators are equal to the projection of the unconstrained estimator. While computationally intensive, the resulting confidence bands are shown to be valid uniformly in the underlying distribution of the data.

## 4.2. Constrained Estimation: Bandwidth Free

A recent literature in statistics has found multiple applications in which nonparametric estimation under shape restrictions may be carried out without the need to select a smoothing parameter. We illustrate these results by reviewing select examples and refer the reader to Groeneboom & Jongbloed (2014) for a broader review of the literature.

**4.2.1. Density estimation.** Motivated by the study of mortality, Grenander (1956) proposes a density estimator based on a nonparametric maximum likelihood procedure subject to the constraint that the density be nonincreasing. Specifically, given an independent and identically distributed sample  $\{X_i\}_{i=1}^n$  from a distribution on  $\mathbf{R}_+$  with density  $f_0$ , the Grenander estimator equals

$$\hat{f}_n \equiv \arg \max_{f: \mathbf{R}_+ \rightarrow \mathbf{R}_+} \prod_{i=1}^n f(X_i) \text{ s.t. } f \text{ nonincreasing and } \int_{\mathbf{R}_+} f(x) dx = 1. \quad 30.$$

The Grenander estimator is straightforward to compute, as it in fact equals the left derivative of the least concave majorant of the empirical distribution function (recall Example 3); Prakasa Rao (1969) provides a closed form expression for  $\hat{f}_n$ .

Especially notable of the Grenander estimator is that it requires no smoothing parameter akin to the bandwidth of a kernel estimator. This remarkable feature has led to a significant literature examining the statistical properties of  $\hat{f}_n$ . In particular, Prakasa Rao (1969) establishes that, for any  $x_0$  in the interior of the support of  $X_i$ ,  $\hat{f}_n(x_0)$  is consistent for the true density  $f_0(x_0)$  provided that  $f_0$  is indeed nonincreasing and continuous. Under the additional requirements that  $f_0$  be differentiable at  $x_0$  and  $f_0'(x_0) \neq 0$ , Prakasa Rao (1969) further finds the asymptotic distribution of  $\hat{f}_n(x_0)$  to equal

$$n^{1/3} [\hat{f}_n(x_0) - f_0(x_0)] \xrightarrow{L} |4f_0(x_0)f_0'(x_0)|^{1/3} \times \arg \max_{u \in \mathbf{R}} [W(u) - u^2], \quad 31.$$

where  $W$  is a standard two-sided Brownian motion with  $W(0) = 0$ . Groeneboom & Wellner (2001) tabulate the quantiles of  $\arg \max_{u \in \mathbf{R}} [W(u) - u^2]$ , which is said to have Chernoff's distribution, and thus Equation 31 may be employed for inference given an estimator of

<sup>1</sup>The fact that projection operators satisfy Equation 28 follows from lemma 2.4 of Nishimura & Ok (2012), while the condition in Equation 29 is well known to be satisfied (see, e.g., lemma 46.5.4 in Zeidler 1984).

$|f_0(x_0)f_0'(x_0)|$ .<sup>2</sup> Alternatively, the quantiles of the limiting distribution of the Grenander estimator may be estimated by subsampling (Politis et al. 1999), the  $m$  out of  $n$  bootstrap (Sen et al. 2010), or a procedure proposed by Cattaneo et al. (2017). The nonparametric bootstrap is, however, unfortunately inconsistent (Kosorok 2008). We emphasize, though, that these inferential procedures are justified under pointwise asymptotics, and they may be inaccurate whenever  $f_0$  is not sufficiently steep at  $x_0$  (relative to the sample size). In particular, the discussion by Groeneboom (1985) implies that the asymptotic distribution in Equation 31 can be a poor approximation for the finite-sample distribution of  $n^{1/3}[\hat{f}_n(x_0) - f_0(x_0)]$  whenever  $f_0'(x_0)$  is close to zero.

The asymptotic distribution in Equation 31 reveals an interesting feature of the Grenander estimator: The closer  $f_0$  is to the boundary of the constraint set in the neighborhood of  $x_0$  [i.e., the smaller  $|f_0'(x_0)|$  is], the more accurate the estimator  $\hat{f}_n(x_0)$  is. In fact, even though the rate of convergence of  $\hat{f}_n(x_0)$  is  $n^{-1/3}$  whenever  $f_0'(x_0) \neq 0$ , the rate improves to  $n^{-1/2}$  whenever  $f_0$  is flat in the neighborhood of  $x_0$  (Groeneboom 1985).

While the analysis of Prakasa Rao (1969) concerns the asymptotic behavior of  $\hat{f}_n$  at a point, other studies have examined the properties of  $\hat{f}_n$  as a global estimator of  $f_0$ . We highlight work by Groeneboom (1985), who shows that if  $f_0$  is nonincreasing, has compact support, and has a continuous first derivative, then it follows that

$$\lim_{n \rightarrow \infty} n^{1/3} E \left[ \int_{\mathbf{R}_+} |\hat{f}_n(x) - f_0(x)| dx \right] = 0.82 \int_{\mathbf{R}_+} |f_0(x)f_0'(x)/2|^{1/3} dx. \quad 32.$$

Birge (1989) derives a finite-sample estimation error bound for  $\hat{f}_n$  in the  $L^1$  norm that holds uniformly over all nonincreasing  $f_0$ . One of the main takeaways from his analysis is that  $\hat{f}_n$  may be interpreted as a variable bin width histogram, where the length of the bin width at each point  $x \in \mathbf{R}_+$  is selected in an (almost) optimal way. Thus, even though computing  $\hat{f}_n$  does not require choosing a smoothing parameter,  $\hat{f}_n$  may nonetheless be viewed as the estimator corresponding to an (almost) optimal choice of an underlying smoothing parameter (i.e., the bin width length).

Although the assumption of a monotone density may be difficult to justify in economic applications, the described results are useful because they provide a good benchmark for analyses under weaker assumptions. For example, for a point  $x_0$  in the support of  $X_i$ , we may instead assume that the density  $f_0$  of  $X_i$  is nonincreasing in a set  $A$  containing  $x_0$ . Letting  $f_0(\cdot|X \in A)$  be the density of  $X$  conditional on  $X \in A$ , we obtain

$$f_0(x_0) = f_0(x_0|X \in A)P(X \in A), \quad 33.$$

which suggests an immediate estimator for  $f_0(x_0)$ . Specifically, we may estimate  $P(X \in A)$  by its sample analog and  $f_0(x_0|X \in A)$  by computing the Grenander estimator on the subsample  $\{X_i : X_i \in A\}$ . The asymptotic distribution of this local Grenander estimator is immediate from Equation 31, since estimating  $P(X \in A)$  has no asymptotic impact.

We conclude by mentioning a number of shape restrictions beyond monotonicity that have been shown to enable bandwidth-free nonparametric estimation. Birge (1997), for instance, studies estimation of a density that is known to be nondecreasing or nonincreasing to the left or right of an unknown point  $\mu$ . In turn, Rufibach (2007) proposes computing a nonparametric maximum likelihood estimator under the assumption that  $f_0$  is log-concave (see also Dümbgen & Rufibach 2009 and Balabdaoui et al. 2009 for the asymptotic properties of the maximum likelihood estimator and Koener & Mizera 2010 for computational aspects). Finally, Balabdaoui & Wellner

<sup>2</sup>To this end, note that  $\hat{f}_n(x_0)$  is consistent for  $f_0(x_0)$  but that  $\hat{f}_n'(x_0)$  is not consistent for  $f_0'(x_0)$ .



(2007) study the estimation of  $k$ -monotone densities, which include monotonicity and convexity restrictions as special cases. As with the Grenander estimator, these shape restrictions may be applied locally by exploiting Equation 33.

**4.2.2. Regression estimation.** The insights gained from studying the shape-restricted maximum likelihood density estimator have been successfully applied to other settings, including hazard rate estimation, censored models, and deconvolution problems (see Groeneboom & Jongbloed 2014). In this section, we review recent advances in the study of shape-restricted nonparametric regression. In particular, we focus on theoretical insights characterizing the impact of shape restrictions on the finite-sample performance of estimators.

In this section, we let  $Y \in \mathbf{R}$ ,  $X \in \mathbf{R}$  be continuously distributed, and suppose

$$Y = \theta_0(X) + \epsilon, \quad E[\epsilon|X] = 0, \quad 34.$$

for some unobservable  $\epsilon \in \mathbf{R}$  and unknown regression function  $\theta_0$  that is assumed to be nonincreasing. For simplicity, we further suppose  $X$  has support  $[0, 1]$ , in which case the shape-constrained nonparametric estimator of  $\theta_0$  is given by

$$\hat{\theta}_n \in \arg \min_{\theta: [0,1] \rightarrow \mathbf{R}} \frac{1}{n} \sum_{i=1}^n [Y_i - \theta(X_i)]^2 \quad \text{s.t. } \theta \text{ is nonincreasing.} \quad 35.$$

Thus, computing  $\hat{\theta}_n$  at points in the sample  $\{X_i\}_{i=1}^n$  only requires solving a quadratic optimization problem subject to linear constraints. Since  $\hat{\theta}_n$  is not uniquely determined by Equation 35 at points  $x_0$  outside the sample,  $\{X_i\}_{i=1}^n$ ,  $\hat{\theta}_n$  is often additionally required to be left continuous and piecewise constant between observations. The resulting  $\hat{\theta}_n$  then equals the left derivative of the least concave majorant of a cumulative sum diagram—a characterization that reveals a close connection between  $\hat{\theta}_n$  and Grenander's estimator.

Let  $x_0 \in (0, 1)$  and suppose  $\theta'_0(x_0)$  exists and  $E[\epsilon^2|X] \leq \sigma^2$  almost surely holds true for some  $\sigma^2 > 0$ . Also, let  $X_{(j)}$  denote the  $j$ th lowest value in  $\{X_i\}_{i=1}^n$  and set  $1 \leq i_0 \leq n$  to be the smallest integer such that  $X_{(i_0)} \geq x_0$ . For any  $1 \leq u \leq v \leq n$ , further define  $\bar{\theta}_0^{u,v} \equiv (v - u + 1)^{-1} \sum_{j=u}^v \theta_0[X_{(j)}]$ , which is simply the sample average of the function  $\theta_0$  over all observations between the  $u$ th and  $v$ th lowest [i.e., between  $X_{(u)}$  and  $X_{(v)}$ ]. Exploiting  $\theta_0$  being nonincreasing and martingale arguments like those of Zhang (2002), it is then possible to show for any  $0 \leq m \leq \min(i_0 - 1, n - i_0)$  that

$$\begin{aligned} E[|\hat{\theta}_n(x_0) - \theta_0(x_0)| | \{X_i\}_{i=1}^n] \\ \leq \bar{\theta}_0^{i_0-m, i_0} - \bar{\theta}_0^{i_0, i_0+m} + \frac{2\sigma}{\sqrt{m+1}} + \theta_0(X_{(i_0-1)}) - \theta_0(X_{(i_0)}). \end{aligned} \quad 36.$$

The result in Equation 36 is important because it can be used to understand how the finite-sample accuracy of  $\hat{\theta}_n(x_0)$  depends on the flatness of  $\theta_0$  around the point  $x_0$ . For instance, note that  $\theta_0(X_{(i_0-1)}) - \theta_0(X_{(i_0)}) = O_p(n^{-1})$  and  $\bar{\theta}_0^{i_0-m, i_0} - \bar{\theta}_0^{i_0, i_0+m} = O_p(m/n)$  since  $\theta'_0(x_0)$  exists. Thus, setting  $m \asymp n^{2/3}$  in Equation 36 implies, via Markov's inequality, that

$$|\hat{\theta}_n(x_0) - \theta_0(x_0)| = O_p(n^{-1/3}).$$

However, if  $\theta_0$  is constant in a neighborhood of  $x_0$ , then  $\bar{\theta}_0^{i_0-m, i_0} = \bar{\theta}_0^{i_0, i_0+m}$  for  $m$  up to order  $n$ . Thus, setting  $m \asymp n$  gives

$$|\hat{\theta}_n(x_0) - \theta_0(x_0)| = O_p(n^{-1/2}).$$

Thus, as in the case of the Grenander density estimator,  $\hat{\theta}_n(x_0)$  typically has an  $n^{-1/3}$  rate of convergence, but if  $\theta_0$  is flat around  $x_0$ , then the estimator is able to adapt to this situation, and its convergence improves to an  $n^{-1/2}$  rate.

The finite-sample bound obtained in Equation 36 emphasizes that studying the rate of convergence of shape-constrained estimators is a nuanced problem. In particular, as discussed in Section 3, the finite-sample impact of imposing a shape restriction in estimation depends on both the sampling uncertainty and the region of the parameter space that  $\theta_0$  is in. For this reason, recent studies of the risk of constrained estimators have focused on finite-sample bounds such as Equation 36. Chatterjee & Lafferty (2015), for example, derive finite-sample bounds for nonparametric regression estimators constrained to be nondecreasing or nonincreasing to the left or right of an unknown point in the support of  $X$ . They find an  $n^{-1/3}$  rate of convergence under a particular norm, with improvements as  $\theta_0$  approaches the boundary of the constraint set. In turn, Guntuboyina & Sen (2015) show that nonparametric regression estimators constrained to be convex converge at an  $n^{-2/5}$  rate (up to log factors), with improvements near the boundary of the constraint set (for related additional results, see Bellec 2016, Chatterjee et al. 2014).

Finally, we note that the fact that  $\hat{\theta}_n$  (as in Equation 35) and the Grenander estimator  $\hat{f}_n$  (as in Equation 30) equal the left derivative of a least concave majorant leads to similarities in their analysis. Brunk (1970), for instance, obtains an asymptotic distribution by showing, under mild assumptions, that if  $\theta_0$  is differentiable and  $\theta'_0(x_0) \neq 0$ , then

$$n^{1/3}[\hat{\theta}_n(x_0) - \theta_0(x_0)] \xrightarrow{L} 2 \left| \frac{\sigma_0^2 \theta'_0(x_0)}{2f_X(x_0)} \right|^{1/3} \times \arg \max_{u \in \mathbf{R}} [W(u) - u^2],$$

where  $f_X$  is the probability density function of  $X$ ,  $\sigma_0^2 \equiv E[\epsilon^2 | X = x_0]$ , and  $W$  is a standard two-sided Brownian motion with  $W(0) = 0$  (compare to Equation 31). The common structure present in both  $\hat{\theta}_n$  and  $\hat{f}_n$  has led to a more general literature studying the properties of left derivatives of least concave majorants of stochastic processes. Anevski & Hössjer (2006) provide a study of asymptotic distributions, and Durot et al. (2012) provide a study of uniform confidence bands.

### 4.3. Constrained Estimators with Smoothing

An advantage of the estimators discussed in Section 4.2 is that they do not require selecting smoothing parameters. However, if the function to be estimated is sufficiently smooth, then unconstrained kernel or series estimators can outperform the procedures of Section 4.2. For example, in the mean regression model (as in Equation 34), with  $\theta_0$  twice differentiable and  $\theta'_0(x_0) < 0$ , the isotonic estimator  $\hat{\theta}_n(x_0)$  in Equation 35 converges at an  $n^{-1/3}$  rate, while a kernel or series estimator can attain an  $n^{-2/5}$  rate (Belloni et al. 2015, Horowitz 2009). However, the constrained estimators of Section 4.2 can possess a faster rate of convergence than their kernel or series counterparts near the boundary of the constraint set. These observations motivate the study of shape-constrained kernel or series estimators as a way to combine the advantages of both approaches.

In the context of kernel estimation of conditional means, Hall & Huang (2001) develop a clever method for combining kernel and constrained estimators. In this section, we illustrate their approach, as applied by Blundell et al. (2012) to impose the Slutsky restrictions. Specifically, let  $\{Y_i, P_i, Q_i\}_{i=1}^n$  be a random sample with  $Y_i$  denoting income,  $P_i$  denoting price, and  $Q_i$  denoting quantity demanded. The classical Nadaraya-Watson kernel estimator of the conditional mean of

$Q_i$  given  $(P_i, Y_i)$  at a point  $(p_0, y_0)$  is given by

$$\hat{\theta}_n(p_0, y_0) \equiv \frac{\sum_{i=1}^n Q_i K((P_i - p_0)/b, (Y_i - y_0)/b)}{\sum_{i=1}^n K((P_i - p_0)/b, (Y_i - y_0)/b)},$$

where  $b$  is a bandwidth, and  $K$  is a bivariate kernel function. The estimator  $\hat{\theta}_n$ , however, need not satisfy the Slutsky restrictions implied by economic theory. Therefore, Blundell et al. (2012) propose instead employing the estimator

$$\hat{\theta}_{n,C}(p_0, y_0) \equiv \frac{\sum_{i=1}^n \xi_i Q_i K((P_i - p_0)/b, (Y_i - y_0)/b)}{n^{-1} \sum_{i=1}^n K((P_i - p_0)/b, (Y_i - y_0)/b)},$$

where  $\{\xi_i\}_{i=1}^n$  are weights chosen to impose the Slutsky restrictions on  $\hat{\theta}_{n,C}$ . In particular, for a prespecified set  $\{(p_j, y_j)\}_{j=1}^J$ , a suitable way to select  $\{\xi_i\}_{i=1}^n$  is to let

$$\begin{aligned} \{\xi_i\}_{i=1}^n &\equiv \arg \min_{\{w_i\}_{i=1}^n} \left\{ n - \sum_{i=1}^n (nw_i)^{1/2} \right\} \quad \text{s.t. } w_i \geq 0 \text{ for all } i, \sum_{i=1}^n w_i = 1, \\ \text{and } \max_{1 \leq j \leq J} &\left\{ \frac{\partial \hat{\theta}_{n,C}(p_j, y_j)}{\partial p} + \hat{\theta}_{n,C}(p_j, y_j) \frac{\partial \hat{\theta}_{n,C}(p_j, y_j)}{\partial y} \right\} \leq 0. \end{aligned} \quad 37.$$

Intuitively, the weights  $(\xi_1, \dots, \xi_n)$  ensure that  $\hat{\theta}_{n,C}$  satisfies the Slutsky restrictions while being as close as possible to the empirical distribution weights  $(1/n, \dots, 1/n)$ . Note that the Slutsky restrictions are only imposed on a subset of points rather than on the entire support. This approach produces satisfactory results as long as the spacing between the subset of points is sufficiently small. We also observe that  $\hat{\theta}_{n,C}$  can be potentially modified to allow for other shape restrictions by simply changing the constraints in Equation 37. Indeed, the original proposal of Hall & Huang (2001) concerns estimation of monotonic conditional means.

Imposing shape restrictions on series (or sieve) estimators is also straightforward. Moreover, the wide applicability of sieve estimators enables the use of shape restrictions in a rich class of settings (Chen 2007). We illustrate such an approach through the nonparametric instrumental variable (NPIV) model of Newey & Powell (2003). Specifically, suppose that for some unknown  $\theta_0$  we have

$$Y = \theta_0(X) + \epsilon, \quad E[\epsilon|W] = 0, \quad 38.$$

where  $Y \in \mathbf{R}$ ,  $X \in \mathbf{R}$  is endogenous, and  $W \in \mathbf{R}$  is an instrument. In this context, Chetverikov & Wilhelm (2017) study the problem of estimating  $\theta_0$  under the assumption that it is nonincreasing. Specifically, let  $p(u) = [p_1(u), \dots, p_k(u)]'$  be a vector of functions such as splines, wavelets, or polynomials. The simplest version of the constrained estimator studied in Chetverikov & Wilhelm (2017) is then

$$\hat{\theta}_{n,CW}(x) \equiv p(x)' \hat{\beta}_n, \quad 39.$$

where  $\hat{\beta}_n$  are the two-stage least squares coefficients obtained from regressing  $Y$  on the vector  $p(X)$ , employing  $p(W)$  as instruments subject to the constraint  $\hat{\theta}'_{n,CW}(x) \leq 0$  for all  $x$  in a grid  $\{x_j\}_{j=1}^J$ . For series estimators, we note that properly selecting  $\{x_j\}_{j=1}^J$  may ensure  $\hat{\theta}'_{n,CW}(x) \leq 0$  at all points, not just for  $x \in \{x_j\}_{j=1}^J$  (see, e.g., Mogstad et al. 2017).

It is, by now, well known that the NPIV model is ill posed and that, as a result, the unconstrained estimator of  $\theta_0$  can suffer from a very slow, potentially logarithmic rate of convergence (Blundell et al. 2007a, Hall & Horowitz 2005). Given our discussion in Section 3, it is therefore intuitively clear that the constrained estimator  $\hat{\theta}_{n,CW}$  can outperform its unconstrained counterpart even in large samples and when  $\theta_0$  is rather steep. It is less clear, however, why the improvements from

imposing the constraint are as substantial as those found in simulations. In an effort to answer this question, Chetverikov & Wilhelm (2017) show that, when the function  $\theta_0$  is constant, under certain conditions, the constrained estimator  $\hat{\theta}_{n,CW}$  does not suffer from the ill-posedness of the model in Equation 38 and has a fast rate of convergence in a (truncated)  $L^2$  norm:  $(k^2 \log n/n)^{1/2}$  if  $p$  consists of polynomials and  $(k \log n/n)^{1/2}$  if  $p$  consists of splines. Moreover, Chetverikov & Wilhelm (2017) derive a finite-sample risk bound that reveals that  $\hat{\theta}_{n,CW}$  has superior estimation properties when  $\theta_0$  is in a neighborhood of a constant function. Crucially, this neighborhood can be rather large depending on the degree of ill-posedness.

However, the results in Chetverikov & Wilhelm (2017) rely upon a monotone instrumental variables assumption, which requires the conditional distribution of  $X$  given  $W$  to be nondecreasing in  $W$  (in the sense of first-order stochastic dominance). Although this is plausible in many applications, it is unclear whether this assumption is necessary for their results to hold. In addition, their estimation error bounds apply only in a truncated  $L^2$  norm, which is defined as the usual  $L^2$  norm but with integration being over a strict subset of the support of  $X$ . It would be of interest to investigate under what conditions the results of Chetverikov & Wilhelm (2017) can be extended to the usual  $L^2$  (or other stronger) norms (see, however, Scaillet 2016 for important challenges in this regard).

## 5. INFERENCE

We next examine recent contributions to inference under shape restrictions. For conciseness, we focus on three specific areas. First, we review tests of whether shape restrictions are satisfied by a parameter of interest. Second, we illustrate the role shape restrictions can play in informing inference by delivering adaptive confidence intervals. Third, we discuss inference methods based on constrained minimization of criterion functions.

### 5.1. Testing Shape Restrictions

There are multiple ways to test whether a parameter of interest satisfies a shape restriction. In this section, we discuss an approach based on unconstrained estimators and an alternative that avoids parameter estimation altogether. A third construction based on the constrained minimization of criterion functions is examined in Section 5.3.

**5.1.1. Using unconstrained estimators.** Unconstrained estimators may be used to test for shape restrictions by assessing whether violations of the conjectured restrictions are statistically significant. In this section, we discuss a simplified version of the test used by Lee et al. (2018).

We consider, as in Section 2.3, first price sealed bid auctions in which we observe bids and an auction characteristic  $X \in \mathbf{R}$  such as appraisal value. Let  $q(\tau|X, I)$  denote the  $\tau$ th quantile of the bid distribution conditional on  $X$  and the auction receiving  $I$  bids. Under appropriate restrictions, Bayesian Nash equilibrium bidding behavior implies

$$q(\tau|X, I_2) - q(\tau|X, I_1) \leq 0 \text{ for all } \tau \in (0, 1) \quad 40.$$

almost surely in  $X$  whenever  $I_1 < I_2$ . Lee et al. (2018) construct a test of this implication of equilibrium behavior as an application of their general procedure. In particular, suppose we observe two samples  $\{B_i, X_i\}_{i=1}^{n_1}$  and  $\{B_i, X_i\}_{i=1}^{n_2}$  of auctions of size  $I_1$  and  $I_2$ , where  $B_i$  is the vector of submitted bids at auction  $i$ . We may then test whether Equation 40 holds by employing local quantile regression estimators  $\hat{q}_n(\tau|x, I_j)$  of  $q(\tau|x, I_j)$  for  $j \in \{1, 2\}$ . Specifically, Lee et al. (2018)

consider the test statistic

$$T_n \equiv \int \max\{0, \sqrt{nb}[\hat{q}_n(\tau|x, I_2) - \hat{q}_n(\tau|x, I_1)]\}dF(\tau, x),$$

where  $n = n_1 + n_2$ ,  $F$  is a weighting measure chosen by the researcher, and  $b \downarrow 0$  is the bandwidth employed in computing the local quantile regression estimators [we assume for simplicity that the same bandwidth is employed to estimate  $q(\tau|x, I_1)$  and  $q(\tau|x, I_2)$ ].

Provided that the bandwidth  $b$  is chosen appropriately, it is possible to show that

$$\sqrt{n_j b}[\hat{q}_n(\tau|x, I_j) - q(\tau|x, I_j)] = \frac{1}{\sqrt{n_j b}} \sum_{i=1}^{n_j} \psi_n(B_i, X_i|\tau, x, I_j) + o_p(1)$$

for  $j \in \{1, 2\}$  and some functions  $\psi_n(\cdot, \cdot|\tau, x, I_j)$  satisfying  $E[\psi_n(B_i, X_i|\tau, x, I_j)] = 0$ . Expansions of this type are known as Bahadur representations. If we exploit such an expansion, it then follows for any distribution satisfying the null hypothesis in Equation 40 that

$$T_n \leq \int \max \left\{ 0, \sum_{j=1}^2 \frac{(-1)^j \sqrt{n}}{n_j \sqrt{b}} \sum_{i=1}^{n_j} \psi_n(B_i, X_i|\tau, x, I_j) \right\} dF(\tau, x) + o_p(1). \quad 41.$$

Moreover, since  $E[\psi_n(B_i, X_i|\tau, x, I_j)] = 0$  holds true for all  $(\tau, x)$ , the quantiles of the upper bound in Equation 41 are easily estimated by the bootstrap. Concretely, for  $\hat{q}_n^*(\tau|x, I_j)$ , the bootstrap analog to  $\hat{q}_n(\tau|x, I_j)$  for  $j \in \{1, 2\}$ , Lee et al. (2018) show that the  $1 - \alpha$  quantile of

$$\int \max\{0, \sqrt{nb}\{\hat{q}_n^*(\tau|x, I_2) - \hat{q}_n^*(\tau|x, I_1) - [\hat{q}_n(\tau|x, I_2) - \hat{q}_n(\tau|x, I_1)]\}\}dF(\tau, x)$$

conditional on the data provides a valid critical value for the test statistic  $T_n$ . Such a critical value is often called least favorable in that it corresponds to the largest (pointwise) asymptotic distribution possible under the null hypothesis.

Lee et al. (2018) further provide alternative critical values that, loosely speaking, attempt to determine at what values of  $(\tau, x)$  Equation 40 holds with equality, which can improve the power of the test against certain alternatives. Finally, we note that the general construction of Lee et al. (2018) more broadly applies to testing whether an unknown function  $\theta_0$  of  $X$  satisfies  $\theta_0(X) \leq 0$  almost surely. As in our discussion, their proposed test statistic is based on the positive part of a kernel-based estimator  $\hat{\theta}_n$  for  $\theta_0$  (as in Equation 41), and critical values are obtained by the bootstrap. The procedure is applicable in many settings, including testing for monotonicity, convexity, and supermodularity in both mean and quantile regression models.

**5.1.2. Avoiding parameter estimation.** A challenge of the tests discussed in Section 5.1.1 is that ensuring that a Bahadur representation is valid imposes restrictive conditions on the choice of bandwidth  $b$ . In certain applications, it may be possible to avoid estimation of the underlying parameter and obtain a valid test under weaker restrictions on the choice of  $b$ . We illustrate such an approach in the context of testing for monotonicity in the mean regression model.

Suppose that, for observable  $Y, X \in \mathbf{R}$ , unknown function  $\theta_0$ , and unobservable  $\epsilon \in \mathbf{R}$ , we have

$$Y = \theta_0(X) + \epsilon, \quad E[\epsilon|X] = 0.$$

We also let  $X$  have support  $[0, 1]$  and  $\{Y_i, X_i\}_{i=1}^n$  be a random sample. The null hypothesis to be tested is that  $\theta_0$  is nonincreasing on  $[0, 1]$ , and the alternative is that there exist  $x_1, x_2 \in [0, 1]$  such that  $x_1 < x_2$  but  $\theta_0(x_1) > \theta_0(x_2)$ . Ghosal et al. (2000) propose a test of such a hypothesis based on

the process (indexed by  $x \in [0, 1]$ )

$$U_{n,b}(x) \equiv \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \text{sign}(Y_i - Y_j) \text{sign}(X_i - X_j) K\left(\frac{X_i - x}{b}\right) K\left(\frac{X_j - x}{b}\right),$$

where  $K : \mathbf{R} \rightarrow \mathbf{R}_+$  is a kernel function and  $b > 0$  is a bandwidth. Intuitively,  $U_{n,b}(x)$  is a local measure of association between  $Y$  and  $X$  similar to Kendall's  $\tau$  statistic. In particular, the limiting expectation of  $U_{n,b}(x)$  as  $b \downarrow 0$  is negative if  $\theta_0$  is nonincreasing at  $x$  but positive otherwise. Thus, Ghosal et al. (2000) consider the test statistic

$$T_{n,b}^{GSV} \equiv \sup_{x \in [0,1]} \frac{\sqrt{n}U_{n,b}(x)}{\hat{\sigma}_{n,b}(x)}, \quad 42.$$

where  $\hat{\sigma}_{n,b}^2(x)$  is an appropriate variance normalization. They establish that the asymptotic distribution of  $T_{n,b}^{GSV}$  is bounded from above by a Gumbel distribution, and in this manner obtain analytical critical values that ensure the resulting test is of asymptotic level  $\alpha$ . Crucially, the construction of  $T_{n,b}^{GSV}$  avoids estimating  $\theta_0$ , so that a Bahadur representation is unnecessary. As a result, asymptotic size control is achieved under weaker conditions on the bandwidth  $b$  than those required by Lee et al. (2018).

While the test of Ghosal et al. (2000) is easy to implement and has asymptotic size control under weak conditions on  $b$ , it has good power only if  $b$  is carefully selected. To address this drawback, Chetverikov (2012) suggests taking the supremum in Equation 42 over both  $x \in [0, 1]$  and  $b \in \mathcal{H}_n$ , where  $\mathcal{H}_n$  is a growing set of possible bandwidth values. Concretely, Chetverikov (2012) considers the test statistic

$$T_n^C \equiv \sup_{b \in \mathcal{H}_n} T_{n,b}^{GSV} = \sup_{x \in [0,1], b \in \mathcal{H}_n} \frac{\sqrt{n}U_{n,b}(x)}{\hat{\sigma}_{n,b}(x)}.$$

This modification substantially complicates the derivation of the limiting distribution of the test statistic, since the extreme value theory arguments employed by Ghosal et al. (2000) are no longer applicable. Instead, Chetverikov (2012) relies on the work of Chernozhukov et al. (2013, 2017) to develop several bootstrap methods that yield the critical value  $c_{\alpha,n}^C$  for which the test that rejects whenever  $T_n^C$  exceeds  $c_{\alpha,n}^C$  also has asymptotic level  $\alpha$ .

The test of Chetverikov (2012) is minimax rate optimal against certain Hölder classes. However, it may potentially be improved by using the arguments of Dümbgen & Spokoiny (2001). Intuitively, for small values of  $b$ , the statistic  $T_{n,b}^{GSV}$  can take large values even under the null, since it contains the maximum over many asymptotically independent random variables. As a result, including small values of  $b$  in  $\mathcal{H}_n$  can significantly increase the quantiles of  $T_n^C \equiv \sup_{b \in \mathcal{H}_n} T_{n,b}^{GSV}$  and thus also the corresponding critical value  $c_{\alpha,n}^C$ . In turn, the resulting larger critical values  $c_{\alpha,n}^C$  undermine the power of the test based on the pair  $(T_n^C, c_{\alpha,n}^C)$  against alternatives that can best be detected by large values of  $b$ , revealing a sensitivity of the procedure to whether small values of  $b$  are included in  $\mathcal{H}_n$ . In the related Gaussian white noise model, Dümbgen & Spokoiny (2001) solve this problem by employing  $b$ -dependent critical values. Within our context, such a test would reject the null hypothesis that  $\theta_0$  is nonincreasing whenever, for appropriate choices of  $c_{\alpha,n}(b)$ , we find that

$$\sup_{x \in [0,1]} \frac{\sqrt{n}U_{n,b}(x)}{\hat{\sigma}_{n,b}(x)} > c_{\alpha,n}(b) \quad \text{for at least for one } b \in \mathcal{H}_n. \quad 43.$$

The analysis in Dümbgen & Spokoiny (2001) of the Gaussian white noise model suggests that the modification in Equation 43 should substantially increase the power against alternatives that are best detected by large values of  $b$  with almost no effect on the power against alternatives that are best detected by small values of  $b$ . It would be of interest to extend the analysis of Dümbgen &

Spokoiny (2001) to cover the standard mean regression model by studying the properties of the test in Equation 43.

## 5.2. Adaptive Confidence Intervals via Shape Restrictions

We consider a standard mean regression model in which  $Y \in \mathbf{R}$ ,  $X \in \mathbf{R}$ , and

$$Y = \theta_0(X) + \epsilon \quad E[\epsilon|X] = 0 \quad 44.$$

for some unknown function  $\theta_0$  and unobservable  $\epsilon \in \mathbf{R}$ , and where, for notational simplicity, we let  $X \in [0, 1]$ . Suppose that we observe an independent and identically distributed sample  $\{Y_i, X_i\}_{i=1}^n$  and are interested in estimating  $\theta_0(x_0)$  for some  $x_0 \in (0, 1)$ . It is well known that the precision with which  $\theta_0(x_0)$  can be estimated depends on the smoothness of  $\theta_0$ : The smoother the function  $\theta_0$  is, the better  $\theta_0(x_0)$  can be estimated. In most applications, however, the smoothness of  $\theta_0$  is unknown, and it is therefore unclear how well  $\theta_0(x_0)$  can be estimated. Adaptive confidence intervals that are as precise as possible given the unknown smoothness of  $\theta_0$  are of particular interest in such settings. These confidence intervals should be shorter the smoother  $\theta_0$  is. Regrettably, a fundamental result developed by Low (1997) is that adaptive confidence intervals for  $\theta_0(x_0)$  typically do not exist. For example, suppose we know that  $\theta_0$  is Lipschitz continuous, i.e.,  $\theta_0 \in \Lambda(M)$ , where  $\Lambda(M)$  is given by

$$\Lambda(M) \equiv \{\theta: [0, 1] \rightarrow \mathbf{R} \text{ s.t. } |\theta(a) - \theta(b)| \leq M|a - b| \text{ for all } a, b \in [0, 1]\}. \quad 45.$$

In addition, suppose  $[c_{L,\alpha}, c_{R,\alpha}]$  is a confidence region with confidence level  $1 - \alpha$ , so that

$$\inf_{\theta_0 \in \Lambda(M)} P_{\theta_0}(c_{L,\alpha} \leq \theta_0(x_0) \leq c_{R,\alpha}) \geq 1 - \alpha, \quad 46.$$

where we write  $P_{\theta_0}$  in place of  $P$  to emphasize that the probability depends on  $\theta_0$ . It then follows from the results of Low (1997) that, for all  $\theta_0$  that are Lipschitz continuous with Lipschitz constant  $M' < M$ , we will find for some constant  $K > 0$  that

$$E[c_{R,\alpha} - c_{L,\alpha}] \geq \frac{K}{n^{1/3}}, \quad 47.$$

which corresponds to the precision of estimating a Lipschitz-continuous function. For instance, when  $\theta_0$  is a constant function, we would hope for the confidence region to shrink at an  $n^{-1/2} \ll n^{-1/3}$  rate, since  $\theta_0(x_0)$  can then be estimated by the sample mean of  $\{Y_i\}_{i=1}^n$ . However, the confidence interval  $[c_{L,\alpha}, c_{R,\alpha}]$  will not be able to take advantage of the smoothness of a constant  $\theta_0$  because it is constrained to control size, as in Equation 46—i.e., the confidence region fails to adapt to the smoothness of  $\theta_0$ .<sup>3</sup>

Adaptive confidence intervals for  $\theta_0(x_0)$  exist, however, if we assume that  $\theta_0$  is either non-decreasing/nonincreasing or convex/concave (Cai et al. 2013, Dümbgen 2003). We discuss the construction of Cai et al. (2013) for nondecreasing  $\theta_0$  and refer the reader to their original work for the other cases. In addition, since Cai et al. (2013) work with Gaussian  $\epsilon$ , we slightly modify their procedure to allow for non-Gaussian  $\epsilon$ .

To construct an adaptive confidence interval for  $\theta_0(x_0)$ , we first order the data according to the regressors  $\{X_i\}_{i=1}^n$ . Specifically, consider all  $X_i$  such that  $X_i > x_0$  and order them into  $X_{(1)}, \dots, X_{(n_1)}$  so that  $x_0 < X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n_1)}$ , where  $n_1$  is the number of observations  $i$  with  $X_i > x_0$ . Similarly, consider all  $X_i$  such that  $X_i \leq x_0$  and order them into  $X_{(-1)}, \dots, X_{(-n_2)}$  so that  $x_0 \geq X_{(-1)} \geq X_{(-2)} \geq \dots \geq X_{(-n_2)}$ , where  $n_2 \equiv n - n_1$  is the number of observations  $i$  with  $X_i \leq x_0$ .

<sup>3</sup>Low (1997) establishes the result for density estimation, but the extension to regression models is immediate (see also Cai & Low 2004).

In addition, let  $Y_{(1)}, \dots, Y_{(n_1)}$  be the  $Y_i$  corresponding to  $X_{(1)}, \dots, X_{(n_1)}$  and  $Y_{(-1)}, \dots, Y_{(-n_2)}$  be the  $Y_i$  corresponding to  $X_{(-1)}, \dots, X_{(-n_2)}$ . Finally, for any  $a \in \mathbf{R}$ , let  $\lfloor a \rfloor$  denote the largest integer smaller than or equal to  $a$ , set the integers  $k_{0,n}$  and  $k_{j,n}$  to be given by

$$k_{j,n} \equiv \left\lfloor \frac{k_{0,n}}{2^j} \right\rfloor \quad \text{and} \quad k_{0,n} \equiv \min \left\{ n_1, n_2, \left\lfloor \frac{n}{\log(n)} \right\rfloor \right\}, \quad 48.$$

and let  $J$  be the largest integer such that  $k_{0,n}/2^J \geq \sqrt{n}$ . Given this notation, we define

$$\delta_{j,L} \equiv \frac{1}{k_{j,n}} \sum_{i=1}^{k_{j,n}} Y_{(-i)} \quad \text{and} \quad \delta_{j,R} \equiv \frac{1}{k_{j,n}} \sum_{i=1}^{k_{j,n}} Y_{(i)} \quad 49.$$

for any  $1 \leq j \leq J$ , which are one-sided nearest-neighbor estimators of  $\theta_0(x_0)$ . Moreover, we note that, since  $\theta_0$  is nondecreasing, the biases of  $\delta_{j,R}$  and  $\delta_{j,L}$  can be signed:

$$E[\delta_{j,L} | \{X_i\}_{i=1}^n] \leq \theta_0(x_0) \leq E[\delta_{j,R} | \{X_i\}_{i=1}^n]. \quad 50.$$

Under mild regularity conditions, the variances of  $\delta_{j,L}$  and  $\delta_{j,R}$  are approximately

$$\text{Var}\{\delta_{j,R} | \{X_i\}_{i=1}^n\} \approx \text{Var}\{\delta_{j,L} | \{X_i\}_{i=1}^n\} \approx \frac{\sigma^2}{k_{j,n}}, \quad 51.$$

where  $\sigma^2 \equiv E[\epsilon^2 | X = x_0]$ . Letting  $c_\alpha$  denote the  $\sqrt{1 - \alpha}$  quantile of a standard normal distribution, these derivations suggest, for each  $1 \leq j \leq J$ , building the confidence region  $[c_{j,L,\alpha}, c_{j,R,\alpha}] \equiv [\delta_{j,L} - c_\alpha \sigma / \sqrt{k_{j,n}}, \delta_{j,R} + c_\alpha \sigma / \sqrt{k_{j,n}}]$ . Indeed, notice that, by independence of  $\delta_{j,L}$  and  $\delta_{j,R}$  conditional on  $\{X_i\}_{i=1}^n$ , we obtain from Equations 50 and 51

$$\begin{aligned} & P \left( \delta_{j,L} - \frac{\sigma}{\sqrt{k_{j,n}}} c_\alpha \leq \theta_0(x_0) \leq \delta_{j,R} + \frac{\sigma}{\sqrt{k_{j,n}}} c_\alpha \right) \\ & \geq P \left( \frac{\sqrt{k_{j,n}}}{\sigma} \{\delta_{j,L} - E[\delta_{j,L}]\} \leq c_\alpha \right) P \left( -c_\alpha \leq \frac{\sqrt{k_{j,n}}}{\sigma} \{\delta_{j,R} - E[\delta_{j,R}]\} \right) \approx 1 - \alpha. \quad 52. \end{aligned}$$

It is worth emphasizing the fundamental role that the monotonicity of  $\theta_0$  plays in ensuring that the constructed confidence intervals are valid for all  $1 \leq j \leq J$  (as in Equation 52). Without monotonicity, Equation 50 may not hold, and it is possible to find a  $\theta_0$  for which the (now uncontrolled) biases of  $\delta_{j,L}$  and  $\delta_{j,R}$  cause the coverage in Equation 52 to fail. In contrast, since, thanks to the monotonicity of  $\theta_0$ , the coverage in Equation 52 holds for all  $1 \leq j \leq J$ , we are now free to search for the best  $j$  in a data-dependent way. Specifically, we note that

$$E[c_{j,R,\alpha} - c_{j,L,\alpha} | \{X_i\}_{i=1}^n] = E[\delta_{j,R} - \delta_{j,L} | \{X_i\}_{i=1}^n] + \frac{2\sigma}{\sqrt{k_{j,n}}} c_\alpha, \quad 53.$$

where the first term on the right-hand side is nonincreasing in  $j$  and the second term is nondecreasing in  $j$ . Thus, to minimize the expected length of the confidence interval, we would like to set  $j$  to make these two terms equal. However, this choice is not feasible, since  $E[\delta_{j,R} - \delta_{j,L} | \{X_i\}_{i=1}^n]$  is unknown. Instead, Cai et al. (2013) define

$$\xi_j \equiv \frac{1}{k_{j-1,n}} \sum_{i=k_{j,n}+1}^{k_{j-1,n}} [Y_{(i)} - Y_{(-i)}] \quad 54.$$

and set  $\hat{j}$  to be the smallest  $j$  such that  $\xi_j \leq 3c_\alpha \sigma / (2k_j)$ —if  $\hat{j} > J$  or  $\hat{j}$  does not exist, then let  $\hat{j} = J$ . The arguments in Cai et al. (2013) then imply that the confidence interval

$$CI_\alpha^* \equiv \left[ \delta_{\hat{j},L} - \frac{\sigma}{\sqrt{k_{\hat{j},n}}} c_\alpha, \delta_{\hat{j},R} + \frac{\sigma}{\sqrt{k_{\hat{j},n}}} c_\alpha \right] \quad 55.$$



covers  $\theta_0(x_0)$  with asymptotic probability of at least  $1 - \alpha$  uniformly over all nondecreasing functions  $\theta_0$ . Moreover,  $CI_\alpha^*$  adapts to  $\theta_0$  in the sense that its expected length (under  $\theta_0$ ) is bounded from above up to a constant by that of the best confidence interval, which minimizes the expected length under  $\theta_0$  subject to the constraint of guaranteeing coverage uniformly over all monotonic functions.

Finally, we note that, while we have assumed that  $\sigma^2 \equiv E[\epsilon|X = x_0]$  is known for simplicity, the construction of a feasible confidence region requires a suitable consistent estimator for  $\sigma^2$ . One possible such estimator  $\hat{\sigma}^2$  is given by

$$\hat{\sigma}^2 \equiv \frac{1}{2k_{J_n}} \sum_{i=1}^{k_{J_n}} \left[ Y_{(i)}^2 + Y_{(-i)}^2 \right] - \left\{ \frac{1}{2k_{J_n}} \sum_{i=1}^{k_{J_n}} [Y_{(i)} + Y_{(-i)}] \right\}^2. \quad 56.$$

### 5.3. Criterion-Based Tests

The classical analysis of criterion-based tests, such as the likelihood ratio test, assumes that the parameter of interest is in the interior of the parameter space. As early as Chernoff (1954), however, it was found that imposing inequality restrictions on the parameter of interest leads to nonstandard (pointwise) limiting distributions. Subsequently, related conclusions were found by a variety of authors, including in extensions by Self & Liang (1987), Shapiro (1989), and King & Rockafellar (1993) and in studies of linear and nonlinear models by Gouriéroux et al. (1981, 1982) and Wolak (1989).

Intuitively, inequality restrictions on a vector may be thought of as the finite-dimensional analog of shape restrictions on nonparametric parameters. As a result, it is to be expected that similar complications will arise when employing criterion-based tests to conduct inference under shape restrictions. In this section, we illustrate a solution to these challenges through a special case of the analysis of Chernozhukov et al. (2015).

**5.3.1. Testing problem.** Suppose that, for some observable  $X \in \mathbf{R}^{d_x}$  and  $Z \in \mathbf{R}^{d_z}$ , the parameter of interest  $\theta_0 \in \Theta$  is identified by the conditional moment restriction

$$E[\rho(X, \theta_0)|Z] = 0, \quad 57.$$

where  $\rho: \mathbf{R}^{d_x} \times \Theta \rightarrow \mathbf{R}$  is a known function assumed to be scalar valued for simplicity. Inference in this model has been extensively studied under the assumption that  $\theta_0$  is in the interior of the parameter space (see Hansen 1985, Ai & Chen 2003, and Chen & Pouzo 2015 for parametric, semiparametric, and nonparametric specifications, respectively).

Testing for or imposing shape restrictions, however, often requires studying the behavior of test statistics in regions near the boundary of the parameter space. Intuitively, numerous shape restrictions can be thought of as inequality constraints that generate similar challenges to those originally found by Chernoff (1954). We focus on the work of Chernozhukov et al. (2015), who examine hypothesis tests with the structure

$$H_0 : \theta_0 \in R, \quad H_1 : \theta_0 \notin R, \quad 58.$$

where the set  $R$  represents the restrictions that we are interested in. Specifically, Chernozhukov et al. (2015) allow for equality and inequality constraints by introducing maps  $\Upsilon_G : \Theta \rightarrow \mathbf{G}$  and  $\Upsilon_F : \Theta \rightarrow \mathbf{F}$  (for spaces  $\mathbf{G}$  and  $\mathbf{F}$ ) and setting  $R$  to equal

$$R \equiv \{\theta \in \Theta : \Upsilon_F(\theta) = 0 \text{ and } \Upsilon_G(\theta) \leq 0\}. \quad 59.$$

The spaces  $\mathbf{G}$  and  $\mathbf{F}$  must be sufficiently general to encompass a diverse set of constraints such as homogeneity, monotonicity, supermodularity, or Slutsky restrictions (see Chernozhukov et al. 2015 for technical details).

For illustrative purposes, we consider an example in which  $X = (V, W)$  with  $V \in [0, 1]$ ,  $\theta_0$  is a twice continuously differentiable function of  $V$ , and we are interested in building a confidence region for a functional  $g : \Theta \rightarrow \mathbf{R}$  of  $\theta_0$  while imposing concavity. In such an application, we would let  $\Theta$  be the space of twice continuously differentiable functions, set  $\Upsilon_F(\theta) = g(\theta) - \lambda$  for a  $\lambda \in \mathbf{R}$ , and let  $\Upsilon_G(\theta) = \nabla^2 \theta$  with  $\mathbf{G}$  being the set of continuous functions on  $[0, 1]$ . The set  $R$  then becomes

$$R = \{\theta \in \Theta : g(\theta) = \lambda \text{ and } \nabla^2 \theta(v) \leq 0 \text{ for all } v \in [0, 1]\}, \quad 60.$$

and we may obtain a confidence region for  $g(\theta_0)$  that imposes concavity on  $\theta_0$  by conducting a test inversion of Equation 58 for  $R$ , as in Equation 60, over different values of  $\lambda \in \mathbf{R}$ .

**5.3.2. Statistical and critical values.** Since  $\theta_0$  satisfies the conditional moment restriction in Equation 57, a possible approach for conducting inference is to construct an overidentification test. To this end, let  $\{q_j\}_{j=1}^\infty$  be a set of functions of  $Z_i$ ; for some  $k_n$  increasing with the sample size, let  $q^{k_n}(Z_i) \equiv [q_1(Z_i), \dots, q_{k_n}(Z_i)]'$ ; and define the test statistic

$$T_n \equiv \inf_{\theta \in \Theta_n \cap R} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \rho(X_i, \theta) q^{k_n}(Z_i) \right\|, \quad 61.$$

where  $\Theta_n$  is a finite-dimensional approximation to  $\Theta$ , i.e.,  $\Theta_n$  is a sieve, such as a polynomial, spline, or wavelet, whose size increases with the sample size (Chen 2007). Heuristically, if  $\theta_0$  indeed satisfies the conjectured restrictions (i.e.,  $\theta_0 \in R$ ), then the unconditional population moments equal zero for some  $\theta \in \Theta$  and  $T_n$  should converge in distribution. In contrast, if  $\theta_0$  does not satisfy the restrictions (i.e.,  $\theta_0 \notin R$ ), then it will not be possible to zero the moment conditions, and  $T_n$  should diverge to infinity.

As expected from Section 3, the finite-sample distribution of  $T_n$  depends on where on the parameter space  $\theta_0$  is. To elucidate this relationship, it is convenient to define

$$\mathbb{G}_n(\theta) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\rho(X_i, \theta) q^{k_n}(Z_i) - E[\rho(X_i, \theta) q^{k_n}(Z_i)]\}, \quad 62.$$

which we note should be approximately normally distributed for any  $\theta \in \Theta$ . It is further convenient, but not necessary, to assume that  $\rho(X, \cdot)$  is differentiable in  $\theta$ , and we let  $\nabla_\theta \rho(X_i, \theta)[b] \equiv \frac{\partial}{\partial \tau} \rho(X_i, \theta_0 + \tau b)|_{\tau=0}$ . Under appropriate conditions, we then obtain

$$T_n = \inf_{b: \theta_0 + \frac{b}{\sqrt{n}} \in \Theta_n \cap R} \left\| \mathbb{G}_n \left( \theta_0 + \frac{b}{\sqrt{n}} \right) + \sqrt{n} E \left[ \rho \left( X_i, \theta_0 + \frac{b}{\sqrt{n}} \right) q^{k_n}(Z_i) \right] \right\| \quad 63.$$

$$= \inf_{b: \theta_0 + \frac{b}{\sqrt{n}} \in \Theta_n \cap R} \left\| \mathbb{G}_n(\theta_0) + E[\nabla_\theta \rho(X_i, \theta_0)[b] q^{k_n}(Z_i)] \right\| + o_p(1), \quad 64.$$

where Equation 63 follows by parameterizing  $b = \sqrt{n}\{\theta - \theta_0\}$ , and Equation 64 follows by arguing through consistency that the value  $\hat{b}_n$  minimizing in Equation 63 must be such that  $\hat{b}_n/\sqrt{n} = o_p(1)$  holds true.

These derivations yield two important observations. First, the distribution of  $T_n$  depends on where  $\theta_0$  is in the parameter space through the restriction  $\theta_0 + b/\sqrt{n} \in \Theta_n \cap R$  in Equation 64. For instance, returning to our example in Equation 60, if we stipulate that  $\theta_0$  be concave, then the set of functions  $b$  such that  $\theta_0 + b/\sqrt{n}$  is concave depends on  $\theta_0$ . Second, Equation 64 emphasizes that the distribution of  $T_n$  only depends on three unknowns: the distribution of  $\mathbb{G}_n(\theta_0)$ , the expectation

$E[\nabla_{\theta} \rho(X_i, \theta_0)[b]q^{k_n}(Z_i)]$ , and the (unknown) set of  $b$  that satisfies  $\theta_0 + b/\sqrt{n} \in \Theta_n \cap R$ . Critical values for  $T_n$  may therefore be obtained by employing suitable substitutes for these three unknowns.

In particular, the distribution of  $\mathbb{G}_n(\theta_0)$  may be approximated via simulation or the bootstrap. Chernozhukov et al. (2015) propose, for example, employing

$$\hat{\mathbb{G}}_n(\hat{\theta}_n) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \omega_i \left\{ \rho(X_i, \hat{\theta}_n)q^{k_n}(Z_i) - \frac{1}{n} \sum_{i=1}^n \rho(X_i, \hat{\theta}_n)q^{k_n}(Z_i) \right\}, \quad 65.$$

where  $\hat{\theta}_n$  is the minimizer of Equation 61, and  $\{\omega_i\}_{i=1}^n$  are drawn by the researcher from a standard normal distribution independently of  $\{X_i, Z_i\}_{i=1}^n$ . Notice that, conditional on the data,  $\hat{\mathbb{G}}_n(\hat{\theta}_n)$  follows a normal distribution, and thus Equation 65 is simply a computationally convenient method for simulating a Gaussian vector whose covariance matrix is the sample analog of the covariance matrix of  $\mathbb{G}_n(\theta_0)$ .

The set of  $b$  satisfying the constraint  $\theta_0 + b/\sqrt{n} \in \Theta_n \cap R$  cannot be uniformly consistently estimated. As a result, Chernozhukov et al. (2015) propose a construction that, when applied to the set  $R$  as defined in Equation 60, reduces to restricting  $b$  to the set<sup>4</sup>

$$\hat{C}_n \equiv \left\{ b : g \left( \hat{\theta}_n + \frac{b}{\sqrt{n}} \right) = \lambda \text{ and } \frac{\nabla^2 b(v)}{\sqrt{n}} \leq \max\{0, -[\nabla^2 \hat{\theta}_n(v) + r_n]\} \text{ for all } v \in [0, 1] \right\}.$$

In this case,  $r_n$  is a bandwidth selected by the researcher that is meant to reflect the sampling uncertainty present in  $\nabla^2 \hat{\theta}_n$  as an estimator for  $\nabla^2 \theta_0$ . Combining these constructions then leads to a bootstrap analog  $T_n^*$  to the statistic  $T_n$  that is given by

$$T_n^* \equiv \inf_{b \in \hat{C}_n} \left\| \hat{\mathbb{G}}_n(\hat{\theta}_n) + \frac{1}{n} \sum_{i=1}^n \nabla_{\theta} \rho(X_i, \hat{\theta}_n)[b]q^{k_n}(Z_i) \right\|. \quad 66.$$

The  $1 - \alpha$  quantile of  $T_n^*$  conditional on the data (but unconditional on  $\{\omega_i\}_{i=1}^n$ ) then provides a valid critical value for  $T_n$ . Specifically, a test that rejects the null hypothesis whenever  $T_n$  is larger than such a critical value has asymptotic level  $\alpha$ . We note that, from a computational perspective, obtaining the desired quantile requires simulating a sample  $\{\omega_i\}_{i=1}^n$  multiple times, solving the optimization problem in Equation 66 for each draw of  $\{\omega_i\}_{i=1}^n$ , and obtaining the  $1 - \alpha$  quantile across simulations of the corresponding  $T_n^*$ .

## 6. CONCLUSION

In this review, we discuss recent developments in the econometrics of shape restrictions. While important advances have been made, particularly in estimation and inference, there undoubtedly remain multiple exciting areas for future research. Optimality results have often been limited to the nonparametric white noise Gaussian model, and their extension to richer economic models is needed. Along these lines, our understanding of efficient semiparametric estimation under shape restrictions remains limited; however, there is a literature studying the canonical limiting experiment under a tangent cone assumption (Chen & Santos 2015, van der Vaart 1989). Finally, we note that we find the possibility of extending the bandwidth-free nonparametric estimation methods of Section 4.2 to a richer class of models particularly exciting.

<sup>4</sup>In a more general setting, with  $\Upsilon_G$  being linear, we find  $\hat{C}_n \equiv \{\theta \in \Theta_n : \Upsilon_F(\hat{\theta}_n + b) = 0 \text{ and } \Upsilon_G(b) \leq \{-[\Upsilon_G(\hat{\theta}_n) + r_n \mathbf{I}_G]\} \vee 0\}$  for  $\vee$  being the least upper bound and  $\mathbf{I}_G$  being the one element in  $\mathbf{G}$  (i.e., the order unit).

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

D.C. is supported by National Science Foundation (NSF) grant SES-1628889. A.S. is supported by NSF grant SES-1426882. A.M.S. is supported by NSF grants SES-1308260, SES-1227091, and SES-1530661.

## LITERATURE CITED

- Abadie A, Angrist J, Imbens G. 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica* 70:91–117
- Abowd JM, Kramarz F, Margolis DN. 1999. High wage workers and high wage firms. *Econometrica* 67:251–333
- Ai C, Chen X. 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71:1795–843
- Ait-Sahalia Y, Duarte J. 2003. Nonparametric option pricing under shape restrictions. *J. Econom.* 116:9–47
- Ait-Sahalia Y, Lo AW. 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. *J. Finance* 53:499–547
- Allen R, Rehbeck J. 2016. *Complementarity in perturbed utility models*. Tech. Rep., Univ. Calif., San Diego
- Andrews DW. 1999. Estimation when a parameter is on a boundary. *Econometrica* 67:1341–83
- Andrews DW. 2001. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 69:683–734
- Andrews DW, Cheng X. 2012. Estimation and inference with weak, semi-strong, and strong identification. *Econometrica* 80:2153–211
- Andrews DW, Cheng X, Guggenberger P. 2011. *Generic results for establishing the asymptotic size of confidence sets and tests*. Discuss. Pap., Cowles Found. Res. Econ., Yale Univ., New Haven, CT
- Anevski D, Hössjer O. 2006. A general asymptotic scheme for inference under order restrictions. *Ann. Stat.* 34:1874–930
- Armstrong T. 2015. Adaptive testing on a regression function at a point. *Ann. Stat.* 43:2086–101
- Athey S, Haile PA. 2007. Nonparametric approaches to auctions. In *Handbook of Econometrics*, Vol. 6A, ed. JJ Heckman, EE Learner, pp. 3847–965. Amsterdam: Elsevier
- Athey S, Stern S. 1998. *An empirical framework for testing theories about complementarity in organizational design*. NBER Work. Pap. 6600
- Balabdaoui F, Rufibach K, Wellner J. 2009. Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Stat.* 37:1299–331
- Balabdaoui F, Wellner JA. 2007. Estimation of a k-monotone density: limit distribution theory and the spline connection. *Ann. Stat.* 2536–64
- Balke A, Pearl J. 1997. Bounds on treatment effects from studies with imperfect compliance. *J. Am. Stat. Assoc.* 92:1171–76
- Beare BK, Dossani A. 2018. Option augmented density forecasts of market returns with monotone pricing kernel. *Quant. Finance* 18(4):623–35
- Beare BK, Fang Z. 2018. Global limit theory for the Grenander estimator under nonstrict concavity. *Electron. J. Stat.* 11(2):3841–70
- Beare BK, Moon J-M. 2015. Nonparametric tests of density ratio ordering. *Econom. Theory* 31:471–92
- Beare BK, Schmidt LD. 2016. An empirical test of pricing kernel monotonicity. *J. Appl. Econom.* 31:338–56
- Bellec P. 2016. Sharp oracle inequalities for Least Squares estimators in shape restricted regression. arXiv:1510.08029 [math.ST]
- Belloni A, Chernozhukov V, Chetverikov D, Kato K. 2015. Some new asymptotic theory for least squares series: pointwise and uniform results. *J. Econom.* 186:345–66

- Bhattacharya D. 2017. *The empirical content of discrete choice models*. Tech. Rep., Univ. Cambridge, UK
- Bhattacharya J, Shaikh AM, Vytlacil E. 2008. Treatment effect bounds under monotonicity assumptions: an application to Swan–Ganz catheterization. *Am. Econ. Rev.* 98:351–56
- Birge L. 1989. The Grenander estimator: a nonasymptotic approach. *Ann. Stat.* 17:1532–49
- Birge L. 1997. Estimation of unimodal densities without smoothness assumptions. *Ann. Stat.* 25:970–81
- Blundell R, Chen X, Kristensen D. 2007a. Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica* 75:1613–69
- Blundell R, Gosling A, Ichimura H, Meghir C. 2007b. Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica* 75:323–63
- Blundell R, Horowitz J, Parey M. 2013. *Nonparametric estimation of a heterogeneous demand function under the Slutsky inequality restriction*. Tech. Rep., Cent. Microdata Methods Pract., London
- Blundell R, Horowitz JL, Parey M. 2012. Measuring the price responsiveness of gasoline demand: economic shape restrictions and nonparametric demand estimation. *Quant. Econ.* 3:29–51
- Brunk H. 1970. Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference*, ed. ML Puri, pp. 177–95. Cambridge, UK: Cambridge Univ. Press
- Cai T, Low M. 2004. An adaptation theory for nonparametric confidence intervals. *Ann. Stat.* 32:1805–40
- Cai T, Low M, Xia Y. 2013. Adaptive confidence intervals for regression functions under shape constraints. *Ann. Stat.* 41:722–50
- Card D, Cardoso AR, Heining J, Kline P. 2016. *Firms and labor market inequality: evidence and some theory*. NBER Work. Pap. 22850
- Cattaneo MD, Jansson M, Nagasawa K. 2017. Bootstrap-based inference for cube root consistent estimators. arXiv:1704.08066 [math.ST]
- Chatterjee S, Guntuboyina A, Sen B. 2014. On risk bounds in isotonic and other shape restricted regression problems. *Ann. Stat.* 43:1774–800
- Chatterjee S, Lafferty J. 2015. Adaptive risk bounds in unimodal regression. arXiv:1512.02956 [math.ST]
- Chen X. 2007. Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics*, Vol. 6B, ed. JJ Heckman, EE Learner, pp. 5549–632. Amsterdam: Elsevier
- Chen X, Christensen TM. 2018. Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. *Quant. Econ.* 9:39–84
- Chen X, Pouzo D. 2015. Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica* 83:1013–79
- Chen X, Santos A. 2015. *Overidentification in regular models*. Tech. Rep., Yale Univ., New Haven, CT
- Chernoff H. 1954. On the distribution of the likelihood ratio. *Ann. Math. Stat.* 25:573–78
- Chernozhukov V, Chetverikov D, Kato K. 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Stat.* 41:2786–819
- Chernozhukov V, Chetverikov D, Kato K. 2017. Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* 45(5):2309–52
- Chernozhukov V, Fernández-Val I, Galichon A. 2009. Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* 96:559–75
- Chernozhukov V, Fernández-Val I, Galichon A. 2010. Quantile and probability curves without crossing. *Econometrica* 78:1093–125
- Chernozhukov V, Hansen C. 2005. An IV model of quantile treatment effects. *Econometrica* 73:245–61
- Chernozhukov V, Newey WK, Santos A. 2015. Constrained conditional moment restriction models. arXiv:1509.06311 [math.ST]
- Chesher A. 2003. Identification in nonseparable models. *Econometrica* 71:1405–41
- Chetverikov D. 2012. Testing regression monotonicity in econometric models. arXiv:1212.6757 [math.ST]
- Chetverikov D, Wilhelm D. 2017. Nonparametric instrumental variable estimation under monotonicity. *Econometrica* 85:1303–20
- Delgado MA, Escanciano JC. 2012. Distribution-free tests of stochastic monotonicity. *J. Econom.* 170:68–75
- D’Haultfœuille X, Février P. 2015. Identification of nonseparable triangular models with discrete instruments. *Econometrica* 83:1199–210
- Dümbgen L. 1993. On nondifferentiable functions and the bootstrap. *Probab. Theory Relat. Fields* 95:125–40

- Dümbgen L. 2003. Optimal confidence bands for shape-restricted curves. *Bernoulli* 9:423–49
- Dümbgen L, Rufibach K. 2009. Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency. *Bernoulli* 15:40–68
- Dümbgen L, Spokoiny VG. 2001. Multiscale testing of qualitative hypotheses. *Ann. Stat.* 29:124–52
- Dupas P. 2014. Short-run subsidies and long-run adoption of new health products: evidence from a field experiment. *Econometrica* 82:197–228
- Durot C, Kulikov VN, Lopuhaä HP. 2012. The limit distribution of the  $F^\infty$ -error of Grenander-type estimators. *Ann. Stat.* 40:1578–608
- Eeckhout J, Kircher P. 2011. Identifying sorting in theory. *Rev. Econ. Stud.* 78:872–906
- Eggermont PPB, LaRiccia VN. 2001. *Maximum Penalized Likelihood Estimation*, Vol. 1. Berlin: Springer
- Ellison G, Ellison SF. 2011. Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration. *Am. Econ. J. Microecon.* 3(1):1–36
- Fang Z. 2014. *Optimal plug-in estimators of directionally differentiable functionals*. Tech. Rep., Texas A & M Univ., College Station, TX
- Fang Z, Santos A. 2014. Inference on directionally differentiable functions. arXiv:1404.3763 [math.ST]
- Fox JT, Lazzati N. 2013. *Identification of discrete choice models for bundles and binary games*. Tech. Rep., Cent. Microdata Methods Pract., London
- Freyberger J, Horowitz JL. 2015. Identification and shape restrictions in nonparametric instrumental variables estimation. *J. Econom.* 189:41–53
- Freyberger J, Reeves B. 2017. *Inference under shape restrictions*. Tech. Rep., Univ. Wis., Madison
- Gentzkow M. 2007. Valuing new goods in a model with complementarity: online newspapers. *Am. Econ. Rev.* 97:713–44
- Ghosal S, Sen A, van der Vaart AW. 2000. Testing monotonicity of regression. *Ann. Stat.* 28:1054–82
- Gouriéroux C, Holly A, Monfort A. 1981. Kuhn-Tucker, likelihood ratio and Wald tests for nonlinear models with inequality constraints on the parameters. *J. Econom.* 16:166
- Gouriéroux C, Holly A, Monfort A. 1982. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* 50:63–80
- Grenander U. 1956. On the theory of mortality measurement: part II. *Scand. Actuar. J.* 1956:125–53
- Groeneboom P. 1985. Estimating a monotone density. *Proc. Berkeley Conf. Honor Jerzy Neyman Jack Kiefer*, Vol. II, pp. 539–55. Belmont, CA: Wadsworth
- Groeneboom P, Jongbloed G. 2014. *Nonparametric Estimation Under Shape Constraints: Estimators, Algorithms and Asymptotics*. Cambridge, UK: Cambridge Univ. Press
- Groeneboom P, Wellner J. 2001. Computing Chernoff’s distribution. *J. Comput. Graph. Stat.* 10:338–400
- Guerre E, Perrigne I, Vuong Q. 2000. Optimal nonparametric estimation of first-price auctions. *Econometrica* 68:525–74
- Guntuboyina A, Sen B. 2015. Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Relat. Fields* 163:379–411
- Haag BR, Hoderlein S, Pendakur K. 2009. Testing and imposing Slutsky symmetry in nonparametric demand systems. *J. Econom.* 153:33–50
- Hagedorn M, Law TH, Manovskii I. 2017. Identifying equilibrium models of labor market sorting. *Econometrica* 85:29–65
- Hahn J, Todd P, Van der Klaauw W. 2001. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69:201–9
- Hall P, Horowitz JL. 2005. Nonparametric methods for inference in the presence of instrumental variables. *Ann. Stat.* 33:2904–29
- Hall P, Huang L-S. 2001. Nonparametric kernel regression subject to monotonicity constraints. *Ann. Stat.* 29:624–47
- Hansen LP. 1985. A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *J. Econom.* 30:203–38
- Heckman JJ, Pinto R. 2017. *Unordered monotonicity*. NBER Work. Pap. 23497
- Heckman JJ, Vytlacil EJ. 2001. Instrumental variables, selection models, and tight bounds on the average treatment effect. In *Econometric Evaluation of Labour Market Policies*, ed. M Lechner, F Pfeiffer, pp. 1–15. Berlin: Springer

- Heckman JJ, Vytlacil EJ. 2005. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73:669–738
- Henderson DJ, List JA, Millimet DL, Parmeter CF, Price MK. 2012. Empirical implementation of nonparametric first-price auction models. *J. Econom.* 168:17–28
- Hong H, Li J. 2014. *The numerical delta method and bootstrap*. Tech. Rep., Stanford Univ., Stanford, CA
- Horowitz JL. 2009. *Semiparametric and Nonparametric Methods in Econometrics*. Berlin: Springer
- Horowitz JL, Lee S. 2017. Nonparametric estimation and inference under shape restrictions. *J. Econom.* 201(1):108–26
- Imbens GW, Angrist JD. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62:467–75
- Imbens GW, Newey WK. 2009. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* 77:1481–512
- Imbens GW, Rubin DB. 1997. Estimating outcome distributions for compliers in instrumental variables models. *Rev. Econ. Stud.* 64:555–74
- Jun SJ, Pinkse J, Wan Y. 2010. A consistent nonparametric test of affiliation in auction models. *J. Econom.* 159:46–54
- Keuzenkamp HA, Barten AP. 1995. Rejection without falsification on the history of testing the homogeneity condition in the theory of consumer demand. *J. Econom.* 67:103–27
- King AJ, Rockafellar RT. 1993. Asymptotic theory for solutions in statistical estimation and stochastic programming. *Math. Oper. Res.* 18:148–62
- Kitagawa T. 2015. A test for instrument validity. *Econometrica* 83:2043–63
- Kitamura Y, Stoye J. 2013. *Nonparametric analysis of random utility models: testing*. Tech. Rep., Yale Univ., New Haven, CT
- Kline P, Tartari M. 2016. Bounding the labor supply responses to a randomized welfare experiment: a revealed preference approach. *Am. Econ. Rev.* 106:971–1013
- Koenker R, Mizera I. 2010. Quasi-concave density estimation. *Ann. Stat.* 38:2998–3027
- Kosorok M. 2008. Bootstrapping the Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, ed. N Balakrishnan, EA Peña, MJ Silvapulle, pp. 282–92. Inst. Math. Stat. Collect. 1. Beachwood, OH: Inst. Math. Stat.
- Kreider B, Pepper JV, Gundersen C, Jolliffe D. 2012. Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported. *J. Am. Stat. Assoc.* 107:958–75
- Kretschmer T, Miravete EJ, Pernas JC. 2012. Competitive pressure and the adoption of complementary innovations. *Am. Econ. Rev.* 102:1540–70
- Lee DS. 2009. Training, wages, and sample selection: estimating sharp bounds on treatment effects. *Rev. Econ. Stud.* 76:1071–102
- Lee S, Salanié B. 2017. *Identifying effects of multivalued treatments*. Tech. Rep., Columbia Univ., New York
- Lee S, Song K, Whang Y-J. 2018. Testing for a general class of functional inequalities. *Econom. Theory*. In press
- Lee Y-Y, Bhattacharya D. 2016. *Applied welfare analysis for discrete choice with interval-data on income*. Tech. Rep., Univ. Calif., Irvine
- Leeb H, Pötscher BM. 2005. Model selection and inference: facts and fiction. *Econom. Theory* 21:21–59
- Lewbel A. 1995. Consistent nonparametric hypothesis tests with an application to Slutsky symmetry. *J. Econom.* 67:379–401
- Li T, Perrigne I, Vuong Q. 2002. Structural estimation of the affiliated private value auction model. *RAND J. Econ.* 33:171–93
- Low M. 1997. On nonparametric confidence intervals. *Ann. Stat.* 25:2547–54
- Luo Y, Wan Y. 2018. Integrated-quantile-based estimation for first-price auction models. *J. Bus. Econ. Stat.* 36(1):173–80
- Machado C, Shaikh A, Vytlacil E. 2013. *Instrumental variables and the sign of the average treatment effect*. Tech. Rep., Univ. Chicago
- Manski CF. 1989. Anatomy of the selection problem. *J. Hum. Resour.* 24:343–60
- Manski CF. 1997. Monotone treatment response. *Econometrica* 65:1311–34

- Manski CF, Pepper JV. 2000. Monotone instrumental variables: with an application to the returns to schooling. *Econometrica* 68:997–1010
- Matzkin RL. 1991. Semiparametric estimation of monotone and concave utility functions for polychotomous choice models. *Econometrica* 59:1315–27
- Matzkin RL. 1992. Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica* 60:239–70
- Matzkin RL. 1994. Restrictions of economic theory in nonparametric methods. In *Handbook of Econometrics*, Vol. 4, ed. R Engle, D McFadden, pp. 2523–58. Amsterdam: Elsevier
- Matzkin RL. 2003. Nonparametric estimation of nonadditive random functions. *Econometrica* 71:1339–75
- McFadden D, Richter MK. 1990. Stochastic rationality and revealed stochastic preference. In *Preferences, Uncertainty, and Optimality: Essays in Honor of Leonid Hurwicz*, ed. JS Chipman, pp. 161–86. Boulder, CO: Westview Press
- Mikusheva A. 2007. Uniform inference in autoregressive models. *Econometrica* 75:1411–52
- Milgrom P, Roberts J. 1995. Complementarities and fit strategy, structure, and organizational change in manufacturing. *J. Account. Econ.* 19:179–208
- Mogstad M, Santos A, Torgovitsky A. 2017. *Using instrumental variables for inference about policy relevant treatment parameters*. NBER Work. Pap. 23568
- Newey WK, Powell JL. 2003. Instrumental variable estimation of nonparametric models. *Econometrica* 71:1565–78
- Nishimura H, Ok EA. 2012. Solvability of variational inequalities on Hilbert lattices. *Math. Oper. Res.* 37:608–25
- Pakes A, Porter J. 2013. *Moment inequalities for semiparametric multinomial choice with fixed effects*. Tech. Rep., Harvard Univ., Cambridge, MA
- Politis DN, Romano J, Wolf M. 1999. *Subsampling*. Berlin: Springer
- Prakasa Rao B. 1969. Estimation of a unimodal density. *Sankhya Indian J. Stat. A* 31:23–36
- Robertson T, Wright F, Dykstra R. 1988. *Order Restricted Statistical Inference*. New York: Wiley
- Romano JP, Shaikh AM. 2012. On the uniform asymptotic validity of subsampling and the bootstrap. *Ann. Stat.* 40:2798–822
- Rosenberg JV, Engle RF. 2002. Empirical pricing kernels. *J. Financ. Econ.* 64:341–72
- Rufibach K. 2007. Computing maximum likelihood estimators of a log-concave density function. *J. Stat. Comput. Simul.* 77:561–74
- Samuelson PA. 1938. A note on the pure theory of consumer’s behaviour. *Economica* 5:61–71
- Scaillet O. 2016. On ill-posedness of nonparametric instrumental variable regression with convexity constraints. *Econom. J.* 19:232–36
- Self SG, Liang K-Y. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.* 82:605–10
- Sen B, Banerjee M, Woodrooffe M. 2010. Inconsistency of bootstrap: the Grenander estimator. *Ann. Stat.* 38:1953–77
- Shao J. 1994. Bootstrap sample size in nonregular cases. *Proc. Am. Math. Soc.* 122:1251–62
- Shapiro A. 1989. Asymptotic properties of statistical estimators in stochastic programming. *Ann. Stat.* 17:841–58
- Shapiro A. 1991. Asymptotic analysis of stochastic programs. *Ann. Oper. Res.* 30:169–86
- Shi X, Shum M. 2016. *Estimating semi-parametric panel multinomial choice models using cyclic monotonicity*. Tech. Rep., Univ. Wis., Madison
- Shimer R, Smith L. 2000. Assortative matching and search. *Econometrica* 68:343–69
- Slutsky E. 1915. Sulla teoria del bilancio dei consumatore. *G. Econ. Riv. Stat.* 51:1–26
- Staiger D, Stock JH. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65:557–86
- Torgovitsky A. 2015. Identification of nonseparable models using instruments with small support. *Econometrica* 83:1185–97
- Tripathi G, Kim W. 2003. Nonparametric estimation of homogeneous functions. *Econom. Theory* 19:640–63



- van der Vaart A. 1989. On the asymptotic information bound. *Ann. Stat.* 17:1487–500
- Vytlacil E. 2002. Independence, monotonicity, and latent index models: an equivalence result. *Econometrica* 70:331–41
- Wolak FA. 1989. Testing inequality constraints in linear econometric models. *J. Econom.* 41:205–35
- Zarantonello EH. 1971. *Projections on Convex Sets in Hilbert Space and Spectral Theory*. Madison, WI: Univ. Wis
- Zeidler E. 1984. *Nonlinear Functional Analysis and Its Applications*, Vol. III. Berlin: Springer
- Zhang C-H. 2002. Risk bounds in isotonic regression. *Ann. Stat.* 30:528–55



## New From Annual Reviews:

### ***Annual Review of Criminology***

criminol.annualreviews.org • Volume 1 • January 2018

**ONLINE NOW!**

**Co-Editors: Joan Petersilia, *Stanford University* and Robert J. Sampson, *Harvard University***

The *Annual Review of Criminology* provides comprehensive reviews of significant developments in the multidisciplinary field of criminology, defined as the study of both the nature of criminal behavior and societal reactions to crime. International in scope, the journal examines variations in crime and punishment across time (e.g., why crime increases or decreases) and among individuals, communities, and societies (e.g., why certain individuals, groups, or nations are more likely than others to have high crime or victimization rates). The societal effects of crime and crime control, and why certain individuals or groups are more likely to be arrested, convicted, and sentenced to prison, will also be covered via topics relating to criminal justice agencies (e.g., police, courts, and corrections) and criminal law.

Annu. Rev. Econ. 2018.10:31-63. Downloaded from www.annualreviews.org. Access provided by 73.75.44.232 on 08/07/18. For personal use only.

#### TABLE OF CONTENTS FOR VOLUME 1:

##### THE DISCIPLINE

- *Reflections on Disciplines and Fields, Problems, Policies, and Life*, James F. Short
- *Replication in Criminology and the Social Sciences*, William Alex Pridemore, Matthew C. Makel, Jonathan A. Plucker

##### CRIME AND VIOLENCE

- *Bringing Crime Trends Back into Criminology: A Critical Assessment of the Literature and a Blueprint for Future Inquiry*, Eric P. Baumer, María B. Vélez, Richard Rosenfeld
- *Immigration and Crime: Assessing a Contentious Issue*, Graham C. Ousey, Charis E. Kubrin
- *The Long Reach of Violence: A Broader Perspective on Data, Theory, and Evidence on the Prevalence and Consequences of Exposure to Violence*, Patrick Sharkey
- *Victimization Trends and Correlates: Macro- and Microinfluences and New Directions for Research*, Janet L. Lauritsen, Maribeth L. Rezey
- *Situational Opportunity Theories of Crime*, Pamela Wilcox, Francis T. Cullen
- *Schools and Crime*, Paul J. Hirschfield

##### PUNISHMENT AND POLICY

- *Collateral Consequences of Punishment: A Critical Review and Path Forward*, David S. Kirk, Sara Wakefield
- *Understanding the Determinants of Penal Policy: Crime, Culture, and Comparative Political Economy*, Nicola Lacey, David Soskice, David Hope

- *Varieties of Mass Incarceration: What We Learn from State Histories*, Michael C. Campbell
- *The Politics, Promise, and Peril of Criminal Justice Reform in the Context of Mass Incarceration*, Katherine Beckett

##### THE PRISON

- *Inmate Society in the Era of Mass Incarceration*, Derek A. Kreager, Candace Kruttschnitt
- *Restricting the Use of Solitary Confinement*, Craig Haney

##### DEVELOPMENTAL AND LIFE-COURSE CRIMINOLOGY

- *Desistance from Offending in the Twenty-First Century*, Bianca E. Bersani, Elaine Eggleston Doherty
- *On the Measurement and Identification of Turning Points in Criminology*, Holly Nguyen, Thomas A. Loughran

##### ECONOMICS OF CRIME

- *Gun Markets*, Philip J. Cook
- *Offender Decision-Making in Criminology: Contributions from Behavioral Economics*, Greg Pogarsky, Sean Patrick Roche, Justin T. Pickett

##### POLICE AND COURTS

- *Policing in the Era of Big Data*, Greg Ridgeway
- *Reducing Fatal Police Shootings as System Crashes: Research, Theory, and Practice*, Lawrence W. Sherman
- *The Problems With Prosecutors*, David Alan Sklansky
- *Monetary Sanctions: Legal Financial Obligations in US Systems of Justice*, Karin D. Martin, Bryan L. Sykes, Sarah Shannon, Frank Edwards, Alexes Harris
- *Forensic DNA Typing*, Erin Murphy



# Contents

Sorting in the Labor Market <i>Jan Eeckhout</i> .....	1
The Econometrics of Shape Restrictions <i>Denis Chetverikov, Andres Santos, and Azeem M. Shaikh</i> .....	31
Networks and Trade <i>Andrew B. Bernard and Andreas Moxnes</i> .....	65
Economics of Child Protection: Maltreatment, Foster Care, and Intimate Partner Violence <i>Joseph J. Doyle, Jr. and Anna Aizer</i> .....	87
Fixed Effects Estimation of Large- $T$ Panel Data Models <i>Iván Fernández-Val and Martin Weidner</i> .....	109
Radical Decentralization: Does Community-Driven Development Work? <i>Katherine Casey</i> .....	139
The Cyclical Job Ladder <i>Giuseppe Moscarini and Fabien Postel-Vinay</i> .....	165
The Consequences of Uncertainty: Climate Sensitivity and Economic Sensitivity to the Climate <i>John Hassler, Per Krusell, and Conny Olovsson</i> .....	189
Measuring Global Value Chains <i>Robert C. Johnson</i> .....	207
Implications of High-Frequency Trading for Security Markets <i>Oliver Linton and Sobeil Mahmoodzadeh</i> .....	237
What Does (Formal) Health Insurance Do, and for Whom? <i>Amy Finkelstein, Neale Mahoney, and Matthew J. Notowidigdo</i> .....	261
The Development of the African System of Cities <i>J. Vernon Henderson and Sebastian Kriticos</i> .....	287
Idea Flows and Economic Growth <i>Francisco J. Buera and Robert E. Lucas, Jr.</i> .....	315
Welfare Reform and the Labor Market <i>Marc K. Chan and Robert Moffitt</i> .....	347

Spatial Patterns of Development: A Meso Approach <i>Stelios Michalopoulos and Elias Papaioannou</i> .....	383
Prosocial Motivation and Incentives <i>Timothy Besley and Maitreesh Ghatak</i> .....	411
Social Incentives in Organizations <i>Nava Ashraf and Oriana Bandiera</i> .....	439
Econometric Methods for Program Evaluation <i>Alberto Abadie and Matias D. Cattaneo</i> .....	465
The Macroeconomics of Rational Bubbles: A User's Guide <i>Alberto Martin and Jaume Ventura</i> .....	505
Progress and Perspectives in the Study of Political Selection <i>Ernesto Dal Bó and Frederico Finan</i> .....	541
Identification and Extrapolation of Causal Effects with Instrumental Variables <i>Magne Mogstad and Alexander Torgovitsky</i> .....	577
Macroeconomic Nowcasting and Forecasting with Big Data <i>Brandyn Bok, Daniele Caratelli, Domenico Giannone, Argia M. Sbordone, and Andrea Tambalotti</i> .....	615

## Indexes

Cumulative Index of Contributing Authors, Volumes 6–10 .....	645
Cumulative Index of Article Titles, Volumes 6–10 .....	648

## Errata

An online log of corrections to *Annual Review of Economics* articles may be found at  
<http://www.annualreviews.org/errata/economics>