The Econometrics of Shape Restrictions*

Denis Chetverikov[†]
Department of Economics
U.C. Los Angeles
chetverikov@econ.ucla.edu

Andres Santos[‡]
Department of Economics
U.C. Los Angeles
andres@econ.ucla.edu

Azeem M. Shaikh[§]
Department of Economics
University of Chicago
amshaikh@uchicago.edu

August 16, 2017

Abstract

We review recent developments in the econometrics of shapes 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.

KEYWORDS: shape restrictions, uniformity, irregular models

*We are grateful to Rosa Matzkin for helpful comments and Neng-Chieh Chang for research assistance.
[†]Research supported by NSF Grant SES-1628889.
[‡]Research supported by NSF Grant SES-1426882.
[§]Research supported by NSF Grants SES-1308260, SES-1227091, and SES-1530661.

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 as both 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 these 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 in the progress made in the last twenty 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 non-exhaustive 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 often also 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 intuition on 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: (i) the degree of sampling uncertainty and (ii) the region of the parameter space the underlying parameter lies in. For instance, when imposing a shape restriction such as monotonicity on an identified function θ_0, the finite-sample distribution of a constrained estimator depends on both the "steepness" of θ_0 and on how precisely θ_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 θ_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 on 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 respectively summarize recent developments in the theory of estimation and inference under shape restrictions. With regards to estimation, we discuss alternative methodologies for imposing shape restrictions and understanding the finite-sample properties of the resulting estimators. With regards 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 the reader does not attribute undue prominence to the selected examples, but instead finds their discussion a helpful starting point towards 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 non-exhaustive 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 due to Imbens and 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

$$D \equiv (1 - Z)D_0 + ZD_1 \qquad\qquad Y \equiv (1 - D)Y_0 + DY_1. \tag{1}$$

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

$$E[Y|Z=1] - E[Y|Z=0] = E[Y_1 - Y_0|D_1 - D_0 = 1]P(D_1 - D_0 = 1) \tag{2}$$
$$- E[Y_1 - Y_0|D_0 - D_1 = 1]P(D_0 - D_1 = 1). \tag{3}$$

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., (2)) and the average treatment effect for individuals induced *out* of treatment by the same change in Z (i.e., (3)). The conflation of these average treatment effects presents a fundamental impediment in identifying the causal effect of treatment.

To resolve this challenge, Imbens and Angrist (1994) impose 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 (3) equals zero and we obtain

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

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. See Heckman and Vytlacil (2005) and the references therein for further discussion. In particular, they employ this latent index structure to study the identification of what they refer to as policy relevant treatment effects. Heckman and 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 and 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 below.

Additional Examples. Shape restrictions motivated by economic theory have been extensively used for identification by Matzkin (1991, 1992). More recently, Allen and 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 non-additively separable manner, monotonicity is often employed to establish identification under both exogeneity (Matzkin, 2003) and endogeneity (Chernozhukov and Hansen, 2005). Similar arguments have also been successfully applied in nonseparable triangular models by Chesher (2003), Imbens and Newey (2009), Torgovitsky (2015), and D'Haultfœuille and Février (2015). Shi and Shum (2016) employ a generalization of monotonicity, termed cyclic monotonicity, to establish identification in multinomial choice models with fixed effects; see also Pakes and 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 due to Blundell et al. (2007b), who examine the evolution of wage inequality in the United Kingdom.

Concretely, letting W denote log-wages, $D \in \{0, 1\}$ a dummy variable indicating employment, and X 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

$$P(W \leq c|X) = P(W \leq c|X, D = 0)P(D = 0|X) + P(W \leq c|X, D = 1)P(D = 1|X), \tag{5}$$

which emphasizes the dependence of the conditional distribution of W given X on the unidentified distribution of wages of the unemployed. Equation (5) can be further used 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 "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 $c \in \mathbf{R}$

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

Restriction (6) can be combined with (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

$$P(W \leq c|X, Z') \leq P(W \leq c|X, Z) \tag{7}$$

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

Additional Examples. 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 such 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 US by exploiting the monotonicity restriction in Imbens and Angrist (1994). Finally, Kline and Tartari (2016) and Lee and Bhattacharya (2016) respectively employ revealed preference and Slutsky-type restrictions 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

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

must be strictly increasing in b, where G denotes the cdf 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 and Haile (2007). Lee et al. (2015) 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. The canonical examples of shape restrictions as testable implications belong to consumer theory (Samuelson, 1938). In this vein, McFadden and Richter (1990) characterize the empirical content of random utility models; see also Kitamura and 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. In relation to Section 2.1, we also note that the examined instrumental variables model generates restrictions on the distribution of the observed data; see, e.g., Imbens and Rubin (1997), Balke and Pearl (1997), Heckman and Vytlacil (2001), Machado et al. (2013), and Kitagawa (2015). Ellison and Ellison (2011) find 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 or not has strong economic implications. A central example here is whether goods are, loosely speaking, "complements" or "substitutes" – concepts that can often be formalized through the shape restrictions of supermodularity and submodularity (Milgrom and Roberts, 1995).

Supermodularity has particularly strong implications in matching markets. Following Shimer and 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 strictly increasing in X and Y. In this model, Shimer and Smith (2000) establish that supermodularity of F (and some of its derivatives) imply 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 datasets has motivated an important empirical literature; see Card et al. (2016) for a recent review. For example, following Abowd et al. (1999), a number of studies have estimated worker-specific and firm-specific fixed effects and found little correlation between them. However, as noted by Eeckhout and Kircher (2011), these fixed effects need not be connected to the underlying firm and worker types. Hagedorn et al. (2017) proposed an estimator of F, but its asymptotic properties are unknown. To our knowledge, no test of PAM or supermodularity of F is available.

Additional Examples. In related work, Athey and 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 was complementary to the scale of production. A novel model for studying whether goods are complements was introduced by Gentzkow (2007), who examined whether print and on-line media acted as complements or substitutes. The nonparametric identification of such a model was established by Fox and 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 Aït-Sahalia and Duarte (2003) in the nonparametric estimation of the state price density 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 state price density (SPD), the price $C(S_t, X, r)$ of the call option at time t is given by

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

Here, we have for simplicity omitted the dependence on the dividend yields of the asset and other state variables. Differentiating (9) with respect to X implies that

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

Exploiting (10), Aït-Sahalia and Lo (1998) constructed 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 (10), however, further implies that the call option pricing function must be nonincreasing and convex in the strike price. Building on this observation, Aït-Sahalia and Duarte (2003) build a nonparametric estimator of C that satisfies the constraints in (10), which they in turn differentiate to estimate the SPD. In estimating the S&P500 SPD, they find the constrained nonparametric estimator outperforms the constrained estimator.

Additional Examples. A related literature has noted that, in disagreement with theoretical expectations, estimates of the pricing kernel are often non-monotonic (Rosenberg and Engle, 2002). As a result, a series of papers has tested whether the violations from monotonicity are statistically significant; see, e.g., Beare and Schmidt (2016). Beare and Dossani (2017) imposed 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, were 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. Here, 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 $Y \in \mathbf{R}$

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

where $R \in \mathbf{R}$ and an individual is assigned to treatment whenever $R > 0$. In certain applications, a researcher may be confident maintaining that θ_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 $\{Y_i, R_i\}_{i=1}^n$, define the one-sided k-nearest neighbor estimators

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

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 θ_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 $\Delta\hat{\theta}_k(0) \equiv \hat{\theta}_{+,k}(0) - \hat{\theta}_{-,k}(0)$, c_α be the α 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 $k_{\min} \leq k_{\max}$ be given. The one-sided confidence interval

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

then possesses asymptotic coverage probability $1-\alpha$ despite k being chosen in (11) to make the interval as "short" as possible. Whenever the distribution of ϵ is known, as in Armstrong (2015), the resulting procedure is tuning parameter free in that we may set $k_{\min} = 1$ and $k_{\max} = n$. On the other hand, if the distribution of ϵ 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_α can be estimated using bootstrap methods such as those in Chetverikov (2012).

Additional Examples. The theoretical literature studying inference under shape restrictions has seen a number of recent contributions including, among others, Freyberger and Horowitz (2015), Chernozhukov et al. (2015), Freyberger and Reeves (2017), Horowitz and 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 "non-standard" ways (Andrews, 1999, 2001). Before discussing estimation and inference, we therefore first develop intuition on the methodological complications that arise from imposing shape restrictions. Specifically, we focus on 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 conducted 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 or not, 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 will consider, for different values of $h \geq 0$, the specification

$$\Delta_L = \Delta_M + h \qquad \Delta_M = \frac{1}{2} \qquad \Delta_H = \Delta_M - h. \tag{12}$$

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 $\Delta_L \geq \Delta_M \geq \Delta_H$:

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

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 in (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 $h > 0$ in (12) so that the law of demand inequalities hold strictly. By consistency of the unconstrained estimators, it then follows that

$$\lim_{n \to \infty} P_h(\hat{\Delta}_L^U - \hat{\Delta}_M^U > 0 \text{ and } \hat{\Delta}_M^U - \hat{\Delta}_H^U > 0) = 1, \tag{14}$$

where we write P_h in place of P to emphasize the probability depends on h. However, if the *unconstrained* estimator satisfies the law of demand, then it must also solve the *constrained* optimization problem in (13). In other words, (14) implies 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 h is held fixed as n diverges to infinity. Somewhat negatively, they 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 (Aït-Sahalia and 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 (14) may be far from one. Whenever this is the case, pointwise asymptotics do not reflect the finite-sample situation and, as we will see in simulations below, approximations based on them 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 non-asymptotic (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 as well as Chetverikov and 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 due to Staiger and 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.

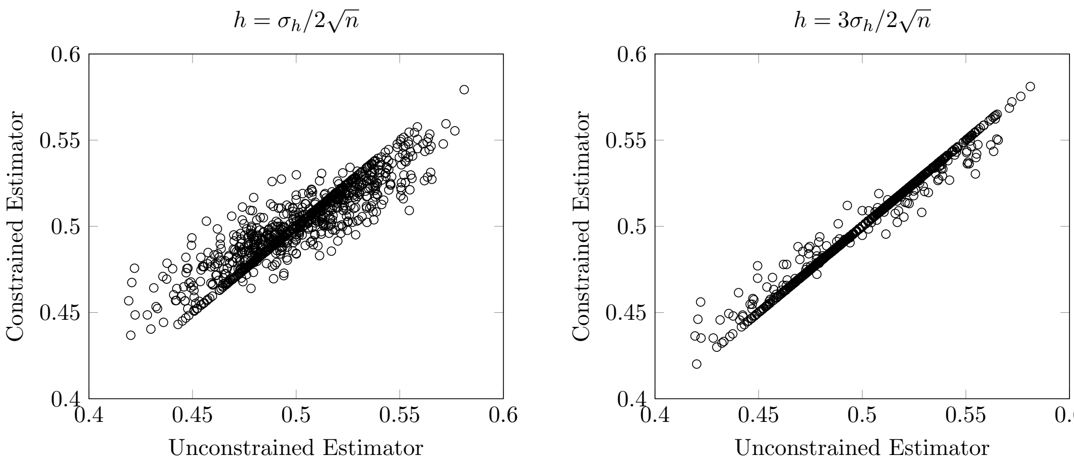
The first step in such an analysis is to develop an understanding of when we might expect shape restrictions to be informative. To this end, we return to our example and note the estimators $\sqrt{n}\{\hat{\Delta}_L^U - \hat{\Delta}_M^U\}$ and $\sqrt{n}\{\hat{\Delta}_M^U - \hat{\Delta}_H^U\}$ are approximately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

$$(\hat{\Delta}_L^U - \hat{\Delta}_M^U) \approx N(h, \frac{\sigma_h}{\sqrt{n}}) \qquad (\hat{\Delta}_M^U - \hat{\Delta}_H^U) \approx N(h, \frac{\sigma_h}{\sqrt{n}}), \tag{15}$$

where, in our design, the standard deviation is the same for both constraints. When h is "large" relative to $\sigma_h/\sqrt{n}$, the demand function is sufficiently elastic that the unconstrained estimator satisfies the law of demand with high probability. In contrast, when h is of the same order as $\sigma_h/\sqrt{n}$ (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 h) to the amount of sampling uncertainty (as measured by $\sigma_h/\sqrt{n}$).

Pointwise asymptotics (i.e., (14)) that rely on h being fixed as n diverges to infinity impose that $\sigma_h/\sqrt{n}$ is "small" relative to h. In this way, they mechanically model a finite-sample setting in which shape restrictions have no effect. In order to move away

Figure 1: Estimators of Δ_M and the Local Parameter



from this paradigm, we must consider an asymptotic framework in which h and $\sigma_h/\sqrt{n}$ remain of the same order regardless of the sample size. By (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 h is thus modelled as tending to zero with the sample size and hence is "local" to zero. We stress, however, that it is incorrect to think of a local analysis as merely modelling 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 vs. unconstrained estimators of Δ_M for different values of $\sqrt{n}h/\sigma_h$. As expected from the preceding discussion, we see that the differences between the constrained and unconstrained estimator decrease as $\sqrt{n}h/\sigma_h$ 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 $\sqrt{n}h/\sigma_h$ 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 for the constrained and unconstrained estimators for Δ_M (scaled by n). In accord with our discussion, we find that when sampling uncertainty (as measured $\sigma_h/\sqrt{n}$) is large relative to h, imposing the law of demand proves informative and the constrained estimator outperforms its unconstrained counterpart. On the other hand, as $\sqrt{n}h/\sigma_h$ 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

Table 1: Scaled Mean Squared Error

	$\sqrt{n}h/\sigma_h$								
	0.00	0.25	0.5	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

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 and 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 (Hall and Horowitz, 2005; Blundell et al., 2007a).

Remark 3.1. 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) and Romano and Shaikh (2012). Its importance in the analysis of "nonstandard" problems has been recently recognized in a variety of applications; see, among others, Leeb and Pötscher (2005), Mikusheva (2007), and Andrews and Cheng (2012). 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. ■

Remark 3.2. 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 (Lewbel, 1995; Haag et al., 2009), homogeneity (Keuzenkamp and Barten, 1995; Tripathi and Kim, 2003), or certain semi/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: (i) estimators that are built by imposing a shape restriction on an originally unconstrained estimator, and (ii) 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 what follows, we denote the parameter of interest by θ_0 and presume we have an estimator $\hat{\theta}_n$ available for it. It will be important to be explicit about the space in which θ_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 $\hat{\theta}_n^{\mathsf{2s}}$ that is obtained by imposing the relevant shape restriction on $\hat{\theta}_n$. Formally, $\hat{\theta}_n^{\mathsf{2s}}$ and $\hat{\theta}_n$ are therefore related by a known transformation $\phi\colon \mathbf{D} \to \mathbf{D}$ that maps the unconstrained estimator into a constrained version of it – i.e., $\hat{\theta}_n^{\mathsf{2s}} = \phi(\hat{\theta}_n)$.

In order to fix ideas, we introduce three examples of transformations ϕ.

Example 4.1. 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 and 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

$$(\hat{\beta}_n(\tau), \hat{\theta}_n(\tau)) \equiv \arg\min_{(\beta,\theta)} \frac{1}{n}\sum_{i=1}^n \rho_\tau(Y_i - X_i'\beta - D_i\theta)^2, \tag{16}$$

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 τ. The estimation procedure in (16), however, does not guarantee $\hat{\theta}_n$ to be monotone. Chernozhukov et al. (2010) therefore suggest instead employing an estimator $\hat{\theta}_n^{\mathsf{2s}} \equiv \phi(\hat{\theta}_n)$, where

$$\phi(\theta)(\tau) \equiv \inf\left\{c : \int_0^1 1\{\theta(\tilde{u}) \le c\} d\tilde{u} \ge \tau\right\}. \tag{17}$$

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

Example 4.2. Building on Example 4.1, an alternative to employing the monotone rearrangement of $\hat{\theta}_n$ is to instead let $\hat{\theta}_n^{\mathsf{2s}}$ be the "closest" monotone function to $\hat{\theta}_n$, e.g.,

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

In practice, (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 $\hat{\theta}_n$ is in a space $\mathbf{D}$ with norm $\|\cdot\|_{\mathbf{D}}$, and note we may think of the set of parameters satisfying a shape restriction as a subset $C \subset \mathbf{D}$; e.g., in (18), $\|\theta\|_{\mathbf{D}}^2 = \int \theta(u)^2 du$ and C is the set of nondecreasing functions. We may then let $\hat{\theta}_n^{\mathsf{2s}}$ be the "closest" parameter to $\hat{\theta}_n$ satisfying the desired shape restriction by defining $\phi \colon \mathbf{D} \to \mathbf{D}$ to equal

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

and setting $\hat{\theta}_n^{\mathsf{2s}} = \phi(\hat{\theta}_n)$. Applying this approach, Fang and 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 4.3. An alternative approach to (19) for imposing concavity is to employ the least concave majorant (lcm) of a function. Specifically, for a bounded function θ defined on, e.g., $[0, 1]$, the lcm of θ is the function $\phi(\theta)$ defined pointwise as

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

Intuitively, the lcm of θ is the "smallest" concave function that is "larger" than θ. Thus, letting $\hat{\theta}_n^{\mathsf{2s}} \equiv \phi(\hat{\theta}_n)$ we obtain a concave function $\hat{\theta}_n^{\mathsf{2s}}$ as a transformation of $\hat{\theta}_n$. The lcm has been widely studied in statistics; see Robertson et al. (1988) and Section 4.2. Within econometrics, the lcm has been employed by Delgado and Escanciano (2012) in testing stochastic monotonicity, Beare and Schmidt (2016) in examining the monotonicity of the pricing kernel, and Luo and Wan (2017) in studying auctions. ■

4.1.1 Local Analysis via 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, $\hat{\theta}_n^{\mathsf{2s}}$ is a deterministic transformation of an original estimator $\hat{\theta}_n$. Second, $\hat{\theta}_n$ is unconstrained and hence its asymptotic distribution is often readily available. These two aspects of the problem make it amenable to the Delta Method.

In what follows, 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 4.1, $\theta_{0,n}$ corresponds to the quantile coefficient function when the data is distributed according to P_n. It is in addition convenient to impose

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

where θ_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 $\hat{\theta}_n$ satisfies

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

where the limit $\mathbb{G}_0$ does not depend on λ. Intuitively, (22) demands that $\hat{\theta}_n$ be robust to local perturbations of the underlying distribution – notice, e.g., that in Table 1 the mean squared error of the unconstrained estimator does not depend on $\sqrt{n}h/\sigma_h$.

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

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

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

Definition 4.1. *Let $\mathbf{D}, \mathbf{E}$ be Banach spaces with norms $\|\cdot\|_{\mathbf{D}}$ and $\|\cdot\|_{\mathbf{E}}$ and $\phi\colon \mathbf{D} \to \mathbf{E}$.*

(i) ϕ is Hadamard differentiable at θ if there is a continuous linear map $\phi'_\theta\colon \mathbf{D} \to \mathbf{E}$ such that for all sequences $\{h_n\} \subset \mathbf{E}$ and $\{t_n\} \subset \mathbf{R}$ with $h_n \to h$ and $t_n \to 0$

$$\lim_{n\to\infty} \|\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h)\|_{\mathbf{E}} = 0. \tag{24}$$

(ii) ϕ is Hadamard directionally differentiable at θ if there is a continuous map $\phi'_\theta\colon \mathbf{D} \to \mathbf{E}$ such that for all sequences $\{h_n\} \subset \mathbf{D}$ and $\{t_n\} \subset \mathbf{R}_+$ with $h_n \to h$ and $t_n \downarrow 0$

$$\lim_{n\to\infty} \|\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h)\|_{\mathbf{E}} = 0. \tag{25}$$

A map ϕ is (fully) Hadamard differentiable at θ if it can be locally approximated by a linear map ϕ'_θ. In turn, ϕ is Hadamard directionally differentiable at θ if a similar approximation requirement holds for a map ϕ'_θ 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 $\theta > 0$, then ϕ is Hadamard differentiable and $\phi'_\theta(h) = h$ for all $h \in \mathbf{R}$. On the other hand, if $\theta = 0$, then ϕ is Hadamard directionally differentiable with $\phi'_\theta(h) = \max\{h, 0\}$ for all $h \in \mathbf{R}$. 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 the Delta Method continues to apply when ϕ is Hadamard directionally (but not fully) differentiable. In particular, the local analysis in Dümbgen (1993) together with (21), (22), and (23) establishes that

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

Crucially, if ϕ is (fully) Hadamard differentiable, then ϕ'_{θ_0} is linear and (26) implies the asymptotic distribution of $\hat{\theta}_n^{2s}$ does not depend on λ. In applications involving shape restrictions, however, ϕ'_{θ_0} is often nonlinear, reflecting that ϕ is Hadamard directionally (but not fully) differentiable. In such instances, the limiting distribution in (26) depends on λ, 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 λ implies that, whenever $\mathbb{G}_0$ is Gaussian, a "naive" plug-in bootstrap is inconsistent (Fang and Santos, 2014). Nonetheless, result (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 and Li (2014). Finally, we note that (26) can also be used to study the risk and optimality (or lack thereof) of estimators (Fang, 2014).

Returning to our examples, we note Chernozhukov et al. (2010) establishes the (full) Hadamard differentiability of the monotone rearrangement operator (i.e., ϕ as in (17)) at any strictly increasing θ. Whenever θ is not strictly increasing, ϕ remains (fully) Hadamard differentiable if the domain of θ is restricted to areas in which the derivative of θ is bounded away from zero. Whether ϕ 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., ϕ as in (19)) was shown in 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., ϕ as in (20)) was proven by Beare and Moon (2015) and Beare and Fang (2016).

4.1.2 Finite-Sample Improvements

Chernozhukov et al. (2009) propose imposing monotonicity to improve confidence intervals for monotone functions. Here, we apply their ideas to general shape restrictions.

For simplicity, we assume θ_0 is a scalar-valued bounded function with domain $[0, 1]$. In many applications, it is possible to construct a confidence interval for θ_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 and Christensen (2017) for constructions for nonparametric regression without and with endogeneity. These confidence intervals employ functions $\hat{l}_n$ and $\hat{u}_n$ satisfying

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

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

Whenever θ_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 imposing monotonicity on $\hat{l}_n$ and $\hat{u}_n$ can yield finite-sample improvements on confidence intervals for a monotone function θ_0. Specifically, suppose $\phi\colon \mathbf{D} \to \mathbf{D}$ assigns to any function $\theta \in \mathbf{D}$ another function $\phi(\theta) \in \mathbf{D}$ satisfying the desired shape restriction. Moreover, assume: (i) $\phi(\theta) = \theta$ whenever θ satisfies the shape restriction, (ii) ϕ satisfies

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

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

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

For a map ϕ satisfying these requirements, Chernozhukov et al. (2009) propose employing $\hat{l}_n^{\mathsf{2s}} \equiv \phi(\hat{l}_n)$ and $\hat{u}_n^{\mathsf{2s}} \equiv \phi(\hat{u}_n)$ to obtain a transformed confidence region for θ_0. By construction, $\hat{l}_n^{\mathsf{2s}}$ and $\hat{u}_n^{\mathsf{2s}}$ now satisfy the shape restriction and

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

by (28) and $\phi(\theta_0) = \theta_0$. Hence, the transformed confidence region still has confidence level at least $1-\alpha$ by (27). Also, by condition (29), we can conclude $\|\hat{l}_n^{\mathsf{2s}} - \hat{u}_n^{\mathsf{2s}}\|_{\mathbf{D}} \le \|\hat{l}_n - \hat{u}_n\|_{\mathbf{D}}$, and thus the new confidence region is in this sense no larger than the original.

Returning to our examples, we note Chernozhukov et al. (2009) establish the monotone rearrangement operator (i.e., ϕ as in (17)) satisfies (28) and (29). Here, we also observe the projection operator (i.e., ϕ as in (18)) satisfies the desired properties whenever $\mathbf{D}$ is a Hilbert space and C is closed, convex, and the pointwise minimum and maximum of any $\theta_1, \theta_2 \in C$ also belongs to C.[1] The lcm operator (i.e., ϕ as in (20)) satisfying (28) is immediate from its definition, while the fact that the lcm map satisfies requirement (29) follows from Theorem 5.11 in Eggermont and LaRiccia (2001).

Finally, we mention a recent proposal by Freyberger and 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.

[1] The fact that projection operators satisfy (28) follows form Lemma 2.4 in Nishimura and Ok (2012), while condition (29) is well known to be satisfied; see, e.g., Lemma 46.5.4 in Zeidler (1984).

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 and 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 i.i.d. sample $\{X_i\}_{i=1}^n$ from a distribution on $\mathbf{R}_+$ with density f_0, the Grenander estimator equals

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

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 4.3); see also Prakasa Rao (1969) for 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 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 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

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

where W is a standard two-sided Brownian motion with $W(0) = 0$. Groeneboom and 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 (31) may be employed for inference given an estimator of $|f_0(x_0)f_0'(x_0)|$.[2] 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, on the other hand, unfortunately inconsistent (Kosorok, 2008). We emphasize, however, these inferential procedures are justified under pointwise asymptotics, and they may be inaccurate whenever f_0 is not sufficiently steep at

[2] To this end, note $\hat{f}_n(x_0)$ is consistent for $f_0(x_0)$ but $\hat{f}_n'(x_0)$ *is not* consistent for $f_0'(x_0)$.

x_0 (relative to the sample size). In particular, the discussion in Groeneboom (1985) implies that the asymptotic distribution in (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 (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 in 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 Groeneboom (1985), who shows that if f_0 is nonincreasing, has compact support, and a continuous first derivative, then it follows that

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

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 binwidth histogram, where the length of the binwidth 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 binwidth 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

$$f_0(x_0) = f_0(x_0|X \in A)P(X \in A) \tag{33}$$

which suggests an immediate estimator for $f_0(x_0)$. Specifically, we may estimate $P(X \in A)$ by its sample analogue 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 (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, studied estimation of a density that is known to be nondecreasing/nonincreasing to the left/right of an unknown point μ. In turn, Rufibach (2007) proposes computing a nonparametric maximum likelihood estimator under the assumption that f_0 is log-concave; see also Dümbgen and Rufibach (2009) and Balabdaoui et al. (2009) for its asymptotic properties and Koenker and Mizera (2010) for computational aspects. Finally, Balabdaoui and Wellner (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 (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 and Jongbloed (2014). Here, 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 what follows we let $Y \in \mathbf{R}$, $X \in \mathbf{R}$ be continuously distributed, and suppose

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

for some unobservable $\epsilon \in \mathbf{R}$ and unknown regression function θ_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 θ_0 is given by

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

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 (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 in 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 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 θ_0 over all observations between the u^{th} and v^{th} lowest (i.e., between $X_{(u)}$ and $X_{(v)}$). Exploiting θ_0 is nonincreasing and martingale arguments like in Zhang (2002), it is then possible to show for any $0 \leq m \leq \min(i_0 - 1, n - i_0)$ that

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

Result (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 θ_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. Hence, setting $m \asymp n^{2/3}$ in (36) implies via Markov's inequality that

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

On the other hand, if θ_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. Hence, setting $m \asymp n$ gives

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

Thus, like in the case of the Grenander density estimator, $\hat{\theta}_n(x_0)$ typically has a $n^{-1/3}$ rate of convergence, but if θ_0 is flat around x_0, then the estimator is able to adapt to this situation and its convergence improves to a $n^{-1/2}$ rate.

The finite-sample bound obtained in (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 θ_0 is in. For this reason, recent studies of the risk of constrained estimators have focused on finite-sample bounds such as (36). Chatterjee and Lafferty (2015), for example, derive finite-sample bounds for nonparametric regression estimators constrained to be nondecreasing/nonincreasing to the left/right of an unknown point in the support of X. They find a $n^{-1/3}$ rate of convergence under a particular norm, with improvements as θ_0 approaches the boundary of the constraint set. In turn, Guntuboyina and Sen (2015) shows nonparametric regression estimators constrained to be convex converge at a $n^{-2/5}$ rate (up to log factors), with improvements near the boundary of the constraint set. For related additional results, see Chatterjee et al. (2014) and Bellec (2016).

Finally, we note the fact that $\hat{\theta}_n$ (as in (35)) and the Grenander estimator $\hat{f}_n$ (as in (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 θ_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 pdf 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 (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. See Anevski and Hössjer (2006) for a study of asymptotic distributions and Durot et al. (2012) for 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 (34)) with θ_0 twice differentiable and $\theta_0'(x_0) < 0$, the isotonic estimator $\hat{\theta}_n(x_0)$ in (35) converges at a $n^{-1/3}$ rate while a kernel or series estimator can attain a $n^{-2/5}$ rate (Horowtiz, 2009; Belloni et al., 2015). On the other hand, the constrained estimators of Section 4.2 can posses a faster rate of converge 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 and Huang (2001) develop a clever method for combining kernel and constrained estimators. Here, 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 price, and Q_i 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

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

where h 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

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

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

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

Intuitively, the weights $(\xi_1, \ldots, \xi_n)$ ensure $\hat{\theta}_{n,C}$ satisfies the Slutsky restrictions while being as close as possible to the empirical distribution weights $(1/n, \ldots, 1/n)$. 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 (37). Indeed, the original proposal in Hall and 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 enable the use of shape restrictions in a rich class of settings (Chen, 2007). Here, we illustrate such an approach through the nonparametric instrumental variable (NPIV) model of Newey and Powell (2003). Specifically, suppose that for some unknown θ_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 and Wilhelm (2017) study the problem of estimating θ_0 under the assumption that it is nonincreasing. Specifically, let $p(u) = (p_1(u), \ldots, p_k(u))'$ be a vector of functions such as splines, wavelets, or polynomials. The simplest version of the constrained estimator studied in Chetverikov and Wilhelm (2017) is then

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

where $\hat{\beta}_n$ are the 2SLS 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 θ_0 can suffer from a very slow, potentially logarithmic, rate of convergence (Hall and Horowitz, 2005; Blundell et al., 2007a). 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

θ_0 is rather "steep". It is less clear, however, why the improvements from imposing the constraint are as substantial as those found in simulations. Towards answering this question, Chetverikov and Wilhelm (2017) show that when the function θ_0 is constant, under certain conditions, the constrained estimator $\hat{\theta}_{n,CW}$ does not suffer from the ill-posedness of the model (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 and Wilhelm (2017) derive a finite-sample risk bound that reveals $\hat{\theta}_{n,CW}$ has superior estimation properties when θ_0 is in a neighborhood of a constant function. Crucially, this neighborhood can be rather large depending on the degree of ill-posedness.

We note, however, that the results in Chetverikov and Wilhelm (2017) rely upon a monotone IV assumption, which requires the conditional distribution of X given W to be nondecreasing in W (in the sense of first-order stochastic dominance). While 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 their results 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. Here, 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. Here, we discuss a simplified version of the test in Lee et al. (2015).

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 τ^{th} quantile of the bid distribution conditional on X and the auction receiving I bids. Under appropriate restrictions, Bayesian Nash equilibrium bidding behavior implies

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

almost surely in X whenever $I_1 < I_2$. Lee et al. (2015) 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 (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. (2015) consider the test statistic

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

where $n = n_1 + n_2$, F is a weighting measure chosen by the researcher, and $h \downarrow 0$ 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 the bandwidth h is chosen appropriately, it is possible to show that

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

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. Exploiting such an expansion, it then follows for any distribution satisfying the null hypothesis in (40) that

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

Moreover, since $E[\psi_n(B_i, X_i|\tau, x, I_j)] = 0$ for all (τ, x), the quantiles of the upper bound in (41) are easily estimated by the bootstrap. Concretely, for $\hat{q}_n^*(\tau|x, I_j)$ the bootstrap analogue to $\hat{q}_n(\tau|x, I_j)$ for $j \in \{1, 2\}$, Lee et al. (2015) show that the $1 - \alpha$ quantile of

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

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. (2015) further provide alternative critical values that, loosely speaking, attempt to determine at what values of (τ, x) equation (40) holds with equality, which can improve the power of the test against certain alternatives. Finally, we note that the general construction in Lee et al. (2015) more broadly applies to testing whether an unknown function θ_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 θ_0 (as in (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 a Bahadur representation is valid imposes restrictive conditions on the choice of bandwidth h. 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 h. 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 θ_0, and unobservable $\epsilon \in \mathbf{R}$

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

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 θ_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 hypothesis based on the process (indexed by $x \in [0,1]$)

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

where $K\colon \mathbf{R} \to \mathbf{R}_+$ is a kernel function and $h > 0$ is a bandwidth. Intuitively, $U_{n,h}(x)$ is a local measure of association between Y and X similar to Kendall's τ statistic. In particular, the limiting expectation of $U_{n,h}(x)$ as $h \downarrow 0$ is negative if θ_0 is nonincreasing at x, but positive otherwise. Thus, Ghosal et al. (2000) consider the test statistic

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

where $\hat{\sigma}^2_{n,h}(x)$ is an appropriate variance normalization. They establish that the asymptotic distribution of $T_{n,h}^{GSV}$ 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 α. Crucially, the construction of $T_{n,h}^{GSV}$ avoids estimating θ_0, so that a Bahadur representation is unnecessary. As a result, asymptotic size control is achieved under weaker conditions on the bandwidth h than those required by Lee et al. (2015).

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

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

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 Chernozhukov et al. (2013, 2017) to develop several bootstrap methods that yield 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 α.

The test of Chetverikov (2012) is minimax rate-optimal against certain Hölder classes. However, it may potentially be improved by using the arguments in Dümbgen and Spokoiny (2001). Intuitively, for small values of h, the statistic $T_{n,h}^{GSV}$ 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 h in $\mathcal{H}_n$ can significantly increase the quantiles of $T_n^C \equiv \sup_{h \in \mathcal{H}_n} T_{n,h}^{GSV}$ and hence 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 be best detected by large values of h, revealing a sensitivity of the procedure to whether small values of h are included in $\mathcal{H}_n$ or not. In the related Gaussian white noise model, Dümbgen and Spokoiny (2001) solve this problem by employing h-dependent critical values. Within our context, such a test would reject the null hypothesis that θ_0 is nonincreasing whenever, for appropriate choices of $c_{\alpha,n}(h)$, we find that

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

The analysis in Dümbgen and Spokoiny (2001) of the Gaussian white noise model suggests that the modification in (43) should substantially increases the power against alternatives that are best detected by large values of h with almost no effect on the power against alternatives that are best detected by small values of h. It would be of interest to extend the analysis in Dümbgen and Spokoiny (2001) to cover the standard mean regression model by studying the properties of the test in (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

$$Y = \theta_0(X) + \epsilon \quad E[\epsilon|X] = 0, \tag{44}$$

for some unknown function θ_0, unobservable $\epsilon \in \mathbf{R}$, and where for notational simplicity we let $X \in [0,1]$. Suppose that we observe an i.i.d. 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 θ_0: the smoother the function θ_0 is, the better $\theta_0(x_0)$ can be estimated. In most applications, however, the smoothness of θ_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 θ_0 are of particular interest in such settings. These confidence intervals should be shorter the smoother θ_0 is. Regrettably, a fundamental result due to Low (1997) says that adaptive confidence intervals for $\theta_0(x_0)$ typically do not exist. For example, suppose we know that θ_0 is Lipschitz-continuous, i.e., $\theta_0 \in \Lambda(M)$ where $\Lambda(M)$ is given by

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

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

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

where we write P_{θ_0} in place of P to emphasize that the probability depends on θ_0. It then follows from the results in Low (1997) that for all θ_0 that are Lipschitz-continuous with Lipschitz constant $M' < M$, we will have for some constant $K > 0$ that

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

which corresponds to the precision of estimating a Lipschitz-continuous function. For instance, when θ_0 is a constant function we would hope for the confidence region to shrink at a $n^{-1/2} \ll n^{-1/3}$ 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 θ_0 because it is constrained to control size as in (46) – i.e., the confidence region fails to *adapt* to the smoothness of θ_0.[3]

Adaptive confidence intervals for $\theta_0(x_0)$ exist, however, if we assume that θ_0 is either nondecreasing/nonincreasing or convex/concave (Dümbgen, 2003; Cai et al., 2013). Here, we discuss the construction in Cai et al. (2013) for nondecreasing θ_0, and refer

[3] Low (1997) establishes the result for density estimation, but the extension to regression models is immediate; see also Cai and Low (2004).

the reader to the original paper for the other cases. In addition, since Cai et al. (2013) work with Gaussian ϵ, we slightly modify their procedure to allow for non-Gaussian ϵ.

In order 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 $X_{(1)}, \ldots, X_{(n_1)}$ so that $x_0 < X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n_1)}$, 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 $X_{(-1)}, \ldots, X_{(-n_2)}$ so that $x_0 \geq X_{(-1)} \geq X_{(-2)} \geq \cdots \geq X_{(-n_2)}$, where $n_2 \equiv n - n_1$ is the number of observations i with $X_i \leq x_0$. In addition, let $Y_{(1)}, \ldots, Y_{(n_1)}$ be the Y_i corresponding to $X_{(1)}, \ldots, X_{(n_1)}$, and $Y_{(-1)}, \ldots, Y_{(-n_2)}$ be the Y_i corresponding to $X_{(-1)}, \ldots, X_{(-n_2)}$. 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

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

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

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

for any $1 \leq j \leq J$, which are one-sided nearest neighbor estimators of $\theta_0(x_0)$. Moreover, we note that since θ_0 is nondecreasing, the biases of $\delta_{j,R}$ and $\delta_{j,L}$ can be signed:

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

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

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

where $\sigma^2 \equiv E[\epsilon^2 | X = x_0]$. Letting c_α 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 (50) and (51) that

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

It is worth emphasizing the fundamental role that the monotonicity of θ_0 plays in ensuring the constructed confidence intervals are valid for all $1 \leq j \leq J$ (as in (52)). Without monotonicity, (50) may not hold and it is possible to find a θ_0 for which the (now uncontrolled) biases of $\delta_{j,L}$ and $\delta_{j,R}$ cause the coverage in (52) to fail. In contrast, since thanks to monotonicity of θ_0 the coverage in (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

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

where the first term on the right-hand side is nonincreasing in j and the second one is nondecreasing in j. Hence, in order 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

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

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 the confidence interval

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

covers $\theta_0(x_0)$ with asymptotic probability at least $1-\alpha$ uniformly over all nondecreasing functions θ_0. Moreover, CI_α^* adapts to θ_0 in the sense that its expected length (under θ_0) is bounded from above up to a constant by that of the "best" confidence interval, which minimizes the expected length under θ_0 subject to the constraint of guaranteeing coverage uniformly over all monotonic functions.

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

$$\hat{\sigma}^2 \equiv \frac{1}{2k_{J,n}} \sum_{i=1}^{k_{J,n}} \left(Y_{(i)}^2 + Y_{(-i)}^2 \right) - \left(\frac{1}{2k_{J,n}} \sum_{i=1}^{k_{J,n}} (Y_{(i)} + Y_{(-i)}) \right)^2. \tag{56}$$

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 extensions by Self and Liang (1987), Shapiro (1989), and King and Rockafellar (1993), and in studies of linear and nonlinear models by Gourieroux et al. (1981, 1982) and Wolak (1989).

Intuitively, inequality restrictions on a vector may be thought of as the finite-dimensional analogue 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 what follows, we illustrate a solution to these challenges through a special case 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

$$E[\rho(X, \theta_0)|Z] = 0, \tag{57}$$

where $\rho\colon \mathbf{R}^{d_x} \times \Theta \to \mathbf{R}$ is a known function assumed to be scalar valued for simplicity. Inference in this model has been extensively studied under the assumption that θ_0 is in the "interior" of the parameter space; see Hansen (1985), Ai and Chen (2003), and Chen and Pouzo (2015) for parametric, semiparametric, and nonparametric specifications.

Testing for and/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 in Chernoff (1954). Here, we focus on Chernozhukov et al. (2015) who examine hypothesis tests with the structure

$$H_0 : \theta_0 \in R \qquad\qquad H_1 : \theta_0 \notin R, \tag{58}$$

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

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

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

For illustrative purposes, we consider an example in which $X = (V, W)$ with $V \in [0, 1]$, θ_0 is twice continuously differentiable function of V, and we are interested in building a confidence region for a functional $g\colon \Theta \to \mathbf{R}$ of θ_0 while imposing concavity. In such an application, we would let Θ 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 $\mathbf{G}$ the set of continuous functions on $[0, 1]$. The set R then becomes

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

and we may obtain a confidence region for $g(\theta_0)$ that imposes concavity on θ_0 by conducting test inversion of (58) for R as in (60) over different values of $\lambda \in \mathbf{R}$.

5.3.2 Statistic and Critical Values

Since θ_0 satisfies the conditional moment restriction in (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, for some k_n increasing with the sample size let $q^{k_n}(Z_i) \equiv (q_1(Z_i), \ldots, q_{k_n}(Z_i))'$, and define the test statistic

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

where Θ_n is a finite-dimensional approximation to Θ; i.e., Θ_n is a "sieve" such as polynomials, splines, or wavelets, whose size increases with the sample size (Chen, 2007). Heuristically, if θ_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. On the other hand, if θ_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 θ_0 is. To elucidate this relation it is convenient to define

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

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

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

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

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

These derivations yield two important observations. First, the distribution of T_n depends on "where" θ_0 is in the parameter space through the restriction $\theta_0 + h/\sqrt{n} \in$

$\Theta_n \cap R$ in (64). For instance, returning to our example in (60), if we impose that θ_0 be concave, then the set of functions h such that $\theta_0 + h/\sqrt{n}$ is concave depends on θ_0. Second, (64) emphasizes that the distribution of T_n only depends on three unknowns: the distribution of $\mathbb{G}_n(\theta_0)$, the expectation $E[\nabla_\theta \rho(X_i, \theta_0)[h] q^{k_n}(Z_i)]$, and the (unknown) set of h that satisfy $\theta_0 + h/\sqrt{n} \in \Theta_n \cap R$. 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

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

where $\hat{\theta}_n$ is the minimizer of (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 (65) is simply a computationally convenient method for simulating a Gaussian vector whose covariance matrix is the sample analogue of the covariance matrix of $\mathbb{G}_n(\theta_0)$.

The set of h satisfying the constraint $\theta_0 + h/\sqrt{n} \in \Theta_n \cap R$ 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 (60) reduces to restricting h to the set[4]

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

Here, 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 analogue T_n^* to the statistic T_n that is given by

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

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 α. 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 (66) for each draw of $\{\omega_i\}_{i=1}^n$, and obtaining the $1-\alpha$ quantile across simulations of the corresponding T_n^*.

[4]In a more general setting with Υ_G linear, $\hat{C}_n \equiv \{\theta \in \Theta_n : \Upsilon_F(\hat{\theta}_n + h) = 0$ and $\Upsilon_G(h) \leq (-(\Upsilon_G(\hat{\theta}_n) + r_n \mathbf{1_G})) \vee 0\}$ for "$\vee$" the least upper bound and "$\mathbf{1_G}$" the "one" element in $\mathbf{G}$ (i.e., the order unit).

6 Conclusion

In this review, we have discussed 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; see, however, a literature studying the canonical limiting experiment under a tangent cone assumption (van der Vaart, 1989; Chen and Santos, 2015). 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.

References

Abadie, A., Angrist, J. and Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, **70** 91–117.

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, **67** 251–333.

Ai, C. and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, **71** 1795–1843.

Aït-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, **116** 9–47.

Aït-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance*, **53** 499–547.

Allen, R. and Rehbeck, J. (2016). Complementarity in perturbed utility models. Tech. rep., University of California San Diego.

Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica*, **67** 1341–1383.

Andrews, D. W. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, **69** 683–734.

Andrews, D. W. and Cheng, X. (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica*, **80** 2153–2211.

Andrews, D. W., Cheng, X. and Guggenberger, P. (2011). Generic results for establishing the asymptotic size of confidence sets and tests.

Anevski, D. and Hössjer, O. (2006). A general asymptotic scheme for inference under order restrictions. *The Annals of Statistics*, **34** 1874–1930.

Armstrong, T. (2015). Adaptive testing on a regression function at a point. *The Annals of Statistics*, **43** 2086–2101.

Athey, S. and Haile, P. A. (2007). Nonparametric approaches to auctions. *Handbook of econometrics*, **6** 3847–3965.

Athey, S. and Stern, S. (1998). An empirical framework for testing theories about complimentarity in organizational design. Tech. rep., National Bureau of Economic Research.

Balabdaoui, F., Rufibach, K. and Wellner, J. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Annals of statistics*, **37** 1299–1331.

Balabdaoui, F. and Wellner, J. A. (2007). Estimation of a k-monotone density: limit distribution theory and the spline connection. *The Annals of Statistics* 2536–2564.

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, **92** 1171–1176.

Beare, B. K. and Dossani, A. (2017). Option augmented density forecasts of market returns with monotone pricing kernel. Tech. rep., University of California, San Diego.

Beare, B. K. and Fang, Z. (2016). Global limit theory for the grenander estimator under nonstrict concavity. Tech. rep., University of California San Diego.

Beare, B. K. and Moon, J.-M. (2015). Nonparametric tests of density ratio ordering. *Econometric Theory*, **31** 471–492.

Beare, B. K. and Schmidt, L. D. (2016). An empirical test of pricing kernel monotonicity. *Journal of Applied Econometrics*, **31** 338–356.

Bellec, P. (2016). Sharp oracle inequalities for least squares estimators in shape restricted regression. *arXiv: 1510.08029.*

Belloni, A., Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, **186** 345–366.

Bhattacharya, D. (2017). The empirical content of discrete choice models. Tech. rep., University of Cambridge.

Bhattacharya, J., Shaikh, A. M. and Vytlacil, E. (2008). Treatment effect bounds under monotonicity assumptions: An application to swan-ganz catheterization. *The American Economic Review*, **98** 351–356.

Birge, L. (1989). The grenander estimator: a nonasymptotic approach. *The Annals of Statistics*, **17** 1532–1549.

Birge, L. (1997). Estimation of unimodal densities without smoothness assumptions. *The Annals of Statistics*, **25** 970–981.

Blundell, R., Chen, X. and Kristensen, D. (2007a). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, **75** 1613–1669.

Blundell, R., Gosling, A., Ichimura, H. and Meghir, C. (2007b). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica*, **75** 323–363.

Blundell, R., Horowitz, J. and Parey, M. (2013). Nonparametric estimation of a heterogeneous demand function under the slutsky inequality restriction. Tech. rep., Centre for Microdata Methods and Practice.

Blundell, R., Horowitz, J. L. and Parey, M. (2012). Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation. *Quantitative Economics*, **3** 29–51.

Brunk, H. (1970). Estimation of isotonic regression. *Nonparametric techniques in statistical inference* 177–195.

Cai, T. and Low, M. (2004). An adaptation theory for nonparametric confidence intervals. *The Annals of Statistics*, **32** 18051840.

Cai, T., Low, M. and Xia, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *The Annals of Statistics*, **41** 722750.

Card, D., Cardoso, A. R., Heining, J. and Kline, P. (2016). Firms and labor market inequality: Evidence and some theory. Tech. rep., National Bureau of Economic Research.

Cattaneo, M. D., Jansson, M. and Nagasawa, K. (2017). Bootstrap-based inference for cube root consistent estimators. *arXiv preprint arXiv:1704.08066.*

Chatterjee, S., Guntuboyina, A. and Sen, B. (2014). On risk bounds in isotonic and other shape restricted regression problems. *The Annals of Statistics*, **43** 1774–1800.

Chatterjee, S. and Lafferty, J. (2015). Adaptive risk bounds in unimodal regression. *arXiv: 1512.0295.*

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of econometrics*, **6** 5549–5632.

Chen, X. and Christensen, T. (2017). Optimal sup-norm rates, adaptivity and inference in nonparametric instrumental variables estimation. *Quntitative Economics* Forthcoming.

Chen, X. and Pouzo, D. (2015). Sieve wald and qlr inferences on semi/nonparametric conditional moment models. *Econometrica*, **83** 1013–1079.

Chen, X. and Santos, A. (2015). Overidentification in regular models. Tech. rep., Yale University.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics* 573–578.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, **41** 2786–2819.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability, forthcoming.*

Chernozhukov, V., Fernandez-Val, I. and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, **96** 559–575.

Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, **78** 1093–1125.

Chernozhukov, V. and Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, **73** 245–261.

Chernozhukov, V., Newey, W. K. and Santos, A. (2015). Constrained conditional moment restriction models. *arXiv preprint arXiv:1509.06311.*

Chesher, A. (2003). Identification in nonseparable models. *Econometrica*, **71** 1405–1441.

Chetverikov, D. (2012). Testing regression monotonicity in econometric models. *arXiv preprint arXiv:1212.6757.*

Chetverikov, D. and Wilhelm, D. (2017). Nonparametric instrumental variable estimation under monotonicity. *Econometrica*, **85** 1303–1320.

Delgado, M. A. and Escanciano, J. C. (2012). Distribution-free tests of stochastic monotonicity. *Journal of Econometrics*, **170** 68–75.

D'Haultfœuille, X. and Février, P. (2015). Identification of nonseparable triangular models with discrete instruments. *Econometrica*, **83** 1199–1210.

Dümbgen, L. (1993). On nondifferentiable functions and the bootstrap. *Probability Theory and Related Fields*, **95** 125–140.

Dümbgen, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli*, **9** 423449.

Dümbgen, L. and 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. and Spokoiny, V. G. (2001). Multiscale testing of qualitative hypotheses. *Annals of Statistics* 124–152.

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, V. N. and Lopuhaä, H. P. (2012). The limit distribution of the l^∞-error of grenander-type estimators. *The Annals of Statistics*, **40** 1578–1608.

Eeckhout, J. and Kircher, P. (2011). Identifying sortingin theory. *The Review of Economic Studies*, **78** 872–906.

Eggermont, P. P. B. and LaRiccia, V. N. (2001). *Maximum penalized likelihood estimation*, vol. 1. Springer.

Ellison, G. and Ellison, S. F. (2011). Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration. *American Economic Journal: Microeconomics*, **3** 1–36.

Fang, Z. (2014). Optimal plug-in estimators of directionally differentiable functionals. Tech. rep., Texas A&M.

Fang, Z. and Santos, A. (2014). Inference on directionally differentiable functions. *arXiv preprint arXiv:1404.3763*.

Fox, J. T. and Lazzati, N. (2013). Identification of discrete choice models for bundles and binary games. Tech. rep., Centre for Microdata Methods and Practice.

Freyberger, J. and Horowitz, J. L. (2015). Identification and shape restrictions in nonparametric instrumental variables estimation. *Journal of Econometrics*, **189** 41–53.

Freyberger, J. and Reeves, B. (2017). Inference under shape restrictions. Tech. rep., University of Wisconsin, Madison.

Gentzkow, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. *The American Economic Review*, **97** 713–744.

Ghosal, S., Sen, A. and Van Der Vaart, A. W. (2000). Testing monotonicity of regression. *Annals of statistics* 1054–1082.

Gourieroux, C., Holly, A. and Monfort, A. (1981). Kuhn-tucker, likelihood ratio and wald tests for nonlinear models with inequality constraints on the parameters. *Journal of Econometrics*, **16** 166.

Gourieroux, C., Holly, A. and Monfort, A. (1982). Likelihood ratio test, wald test, and kuhn-tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* 63–80.

Grenander, U. (1956). On the theory of mortality measurement: part ii. *Scandinavian Actuarial Journal*, **1956** 125–153.

Groeneboom, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, **2** 539–555.

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric estimation under shape constraints*, vol. 38. Cambridge University Press.

Groeneboom, P. and Wellner, J. (2001). Computing chernoff's distribution. *Journal of Computational and Graphical Statistics*, **10** 338–400.

Guerre, E., Perrigne, I. and Vuong, Q. (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica*, **68** 525–574.

Guntuboyina, A. and Sen, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probability Theory and Related Fields*, **163** 379–411.

Haag, B. R., Hoderlein, S. and Pendakur, K. (2009). Testing and imposing slutsky symmetry in nonparametric demand systems. *Journal of Econometrics*, **153** 33–50.

Hagedorn, M., Law, T. H. and Manovskii, I. (2017). Identifying equilibrium models of labor market sorting. *Econometrica*, **85** 29–65.

Hahn, J., Todd, P. and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, **69** 201–209.

Hall, P. and Horowitz, J. L. (2005). Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, **33** 2904–2929.

Hall, P. and Huang, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics* 624–647.

Hansen, L. P. (1985). A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics*, **30** 203–238.

Heckman, J. J. and Pinto, R. (2017). Unordered monotonicity. Tech. rep., National Bureau of Economic Research.

Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation1. *Econometrica*, **73** 669–738.

Heckman, J. J. and Vytlacil, E. J. (2001). Instrumental variables, selection models, and tight bounds on the average treatment effect. *Econometric Evaluation of Labour Market Policies* 1–15.

Henderson, D. J., List, J. A., Millimet, D. L., Parmeter, C. F. and Price, M. K. (2012). Empirical implementation of nonparametric first-price auction models. *Journal of Econometrics*, **168** 17–28.

Hong, H. and Li, J. (2014). The numerical delta method and bootstrap. Tech. rep., Stanford University.

Horowitz, J. L. and Lee, S. (2017). Nonparametric estimation and inference under shape restrictions. *Journal of Econometrics* Forthcoming.

Horowtiz, J. (2009). *Semiparametric and nonparametric methods in econometrics*. Springer.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62** 467–475.

Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, **77** 1481–1512.

Imbens, G. W. and Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies*, **64** 555–574.

Jun, S. J., Pinkse, J. and Wan, Y. (2010). A consistent nonparametric test of affiliation in auction models. *Journal of Econometrics*, **159** 46–54.

Keuzenkamp, H. A. and Barten, A. P. (1995). Rejection without falsification on the history of testing the homogeneity condition in the theory of consumer demand. *Journal of Econometrics*, **67** 103–127.

King, A. J. and Rockafellar, R. T. (1993). Asymptotic theory for solutions in statistical estimation and stochastic programming. *Mathematics of Operations Research*, **18** 148–162.

Kitagawa, T. (2015). A test for instrument validity. *Econometrica*, **83** 2043–2063.

Kitamura, Y. and Stoye, J. (2013). Nonparametric analysis of random utility models: testing. Tech. rep., Yale University.

Kline, P. and Tartari, M. (2016). Bounding the labor supply responses to a randomized welfare experiment: A revealed preference approach. *The American Economic Review*, **106** 971–1013.

Koenker, R. and Mizera, I. (2010). Quasi-concave density estimation. *The Annals of Statistics* 2998–3027.

Kosorok, M. (2008). Bootstrapping the grenander estimator. *IMS Collections. Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, **1** 282–292.

Kreider, B., Pepper, J. V., Gundersen, C. and Jolliffe, D. (2012). Identifying the effects of snap (food stamps) on child health outcomes when participation is endogenous and misreported. *Journal of the American Statistical Association*, **107** 958–975.

Kretschmer, T., Miravete, E. J. and Pernías, J. C. (2012). Competitive pressure and the adoption of complementary innovations. *The American Economic Review*, **102** 1540–1570.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, **76** 1071–1102.

Lee, S. and Salanié, B. (2017). Identifying effects of multivalued treatments. Tech. rep., Columbia University.

Lee, S., Song, K. and Whang, Y.-J. (2015). Testing for a general class of functional inequalities. *Econometric Theory, forthcoming.*

Lee, Y.-Y. and Bhattacharya, D. (2016). Applied welfare analysis for discrete choice with interval-data on income. Tech. rep., University of California - Irvine.

Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, **21** 21–59.

Lewbel, A. (1995). Consistent nonparametric hypothesis tests with an application to slutsky symmetry. *Journal of Econometrics*, **67** 379–401.

Li, T., Perrigne, I. and Vuong, Q. (2002). Structural estimation of the affiliated private value auction model. *RAND Journal of Economics* 171–193.

Low, M. (1997). On nonparametric confidence intervals. *The Annals of Statistics*, **25** 25472554.

Luo, Y. and Wan, Y. (2017). Integrated-quantile-based estimation for first-price auction models. *Journal of Business & Economic Statistics* 1–8.

Machado, C., Shaikh, A. and Vytlacil, E. (2013). Instrumental variables and the sign of the average treatment effect. Tech. rep.

Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human resources* 343–360.

Manski, C. F. (1997). Monotone treatment response. *Econometrica* 1311–1334.

Manski, C. F. and Pepper, J. V. (2000). Monotone instrumental variables: with an application to the returns to schooling. *Econometrica*, **68** 997–1010.

Matzkin, R. L. (1991). Semiparametric estimation of monotone and concave utility functions for polychotomous choice models. *Econometrica* 1315–1327.

Matzkin, R. L. (1992). Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica* 239–270.

Matzkin, R. L. (1994). Restrictions of economic theory in nonparametric methods. *Handbook of econometrics*, **4** 2523–2558.

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica*, **71** 1339–1375.

McFadden, D. and Richter, M. K. (1990). Stochastic rationality and revealed stochastic preference. *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO* 161–186.

Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica*, **75** 1411–1452.

Milgrom, P. and Roberts, J. (1995). Complementarities and fit strategy, structure, and organizational change in manufacturing. *Journal of accounting and economics*, **19** 179–208.

Mogstad, M., Santos, A. and Torgovitsky, A. (2017). Using instrumental variables for inference about policy relevant treatment parameters. Tech. rep., National Bureau of Economic Research.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, **71** 1565–1578.

Nishimura, H. and Ok, E. A. (2012). Solvability of variational inequalities on hilbert lattices. *Mathematics of Operations Research*, **37** 608–625.

Pakes, A. and Porter, J. (2013). Moment inequalities for semiparametric multinomial choice with fixed effects. Tech. rep.

Politis, D. N., Romano, J. and Wolf, M. (1999). *Subsampling*. Springer, New York.

Prakasa Rao, B. (1969). Estimation of a unimodal density. *Sankhya: The Indian Journal of Statistics, Series A* 23–36.

Robertson, T., Wright, F. and Dykstra, R. (1988). *Order restricted statistical inference*. Wiley, New York.

Romano, J. P. and Shaikh, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *The Annals of Statistics*, **40** 2798–2822.

Rosenberg, J. V. and Engle, R. F. (2002). Empirical pricing kernels. *Journal of Financial Economics*, **64** 341–372.

Rufibach, K. (2007). Computing maximum likelihood estimators of a log-concave density function. *Journal of Statistical Computation and Simulation*, **77** 561–574.

Samuelson, P. A. (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. *The Econometrics Journal*, **19** 232–236.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82** 605–610.

Sen, B., Banerjee, M. and Woodroofe, M. (2010). Inconsistency of bootstrap: the grenander estimator. *The Annals of Statistics*, **38** 1953–1977.

Shao, J. (1994). Bootstrap sample size in nonregular cases. *Proceedings of the American Mathematical Society*, **122** 1251–1262.

Shapiro, A. (1989). Asymptotic properties of statistical estimators in stochastic programming. *The Annals of Statistics* 841–858.

Shapiro, A. (1991). Asymptotic analysis of stochastic programs. *Annals of Operations Research*, **30** 169–186.

Shi, X. and Shum, M. (2016). Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. Tech. rep., University of Wisconsin - Madison.

Shimer, R. and Smith, L. (2000). Assortative matching and search. *Econometrica*, **68** 343–369.

Slutsky, E. (1915). Sulla teoria del bilancio del consumatore. *Giornale degli economisti e rivista di statistica* 1–26.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, **65** 557–586.

Torgovitsky, A. (2015). Identification of nonseparable models using instruments with small support. *Econometrica*, **83** 1185–1197.

Tripathi, G. and Kim, W. (2003). Nonparametric estimation of homogeneous functions. *Econometric Theory*, **19** 640–663.

van der Vaart, A. (1989). On the asymptotic information bound. *The Annals of Statistics* 1487–1500.

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, **70** 331–341.

Wolak, F. A. (1989). Testing inequality constraints in linear econometric models. *Journal of econometrics*, **41** 205–235.

Zarantonello, E. H. (1971). *Projections on convex sets in Hilbert space and spectral theory.* University of Wisconsin.

Zeidler, E. (1984). *Nonlinear Functional Analysis and its Applications III.* Springer-Verlag, New York.

Zhang, C.-H. (2002). Risk bounds in isotonic regression. *The Annals of Statistics*, **30** 528–555.