

# Inference in Models Defined by Conditional Moment Inequalities with Continuous Covariates

Maria Ponomareva\*  
Northwestern University  
JOB MARKET PAPER

November 24, 2009

## Abstract

In this paper I present a novel approach to inference in models where the partially identified parameter is defined by a set of conditional moment inequalities with continuous covariates. This class of models covers many economic applications, including treatment response models and regression with missing or interval outcome data. Depending on the assumptions that a researcher is willing to make on conditional moment functions that define the inequalities, I propose inference procedure that is based on the distance between the set of conditional moment functions and the cone of non-positive (or non-negative) functions. If a researcher is reluctant to impose any assumptions about the shape of conditional moment functions except certain smoothness conditions, I offer a method that relies on bootstrapping of the simultaneous lower confidence bands for nonparametric estimators of conditional moments. In general, this inference procedure may lead to a conservative coverage. However, I show that under a particular set of shape restrictions on conditional moment functions one can construct confidence sets based on a Gaussian asymptotic approximation that is relatively easy to implement and attains accurate coverage in small samples. Finally, I conduct Monte Carlo simulations to illustrate both procedures.

---

\*I am indebted to Elie Tamer for his constant encouragement and support throughout this project. I am very grateful to Ivan Canay and Wenxin Jiang for valuable comments and suggestions. I want to thank Vasiliy Ponomarev and Daniil Manaenkov for helpful discussions.  
Original version: September 6, 2008.

# 1 Introduction

The majority of economic models focus on the *conditional* mean function. Additionally, the identification of the parameters of interest in many econometric settings often requires to impose fairly restrictive and sometimes unjustified assumptions. For example, in a simple linear regression model with missing data we need to impose additional assumptions about the mechanism that generates missing values in order to get point identification of the slope coefficients. Nonetheless, it is often possible to relax such assumptions and still get meaningful restrictions on the set of values that these parameters may take. In many cases this relaxation leads to models where the set of parameters compatible with the model is defined by a set of inequality constraints for conditional moment functions. Till this point, most of the literature focused either on constraints on unconditional moments or on the cases where the conditioning covariates are discrete. However, in real data many of the interesting applications deal with *continuous* covariates. One way to treat such a case is to transform unconditional moment inequalities into a finite set of conditional moment constraints and then apply the standard techniques developed in the literature. Unfortunately, this approach may lead to a loss of information, and as the result the researcher may end up with the confidence sets that are too large. In this paper I avoid a such a transformation and work directly with the (nonparametric) estimators of the conditional moment functions. This allows me to exploit all the information that the data contains. First, I consider the case where no shape restrictions are placed on the conditional moment functions and propose a consistent inference procedure that is based on the bootstrap approximation of the deviations of conditional moment functions (that are non-positive of a given parameter value belongs to the identified set) from their consistent nonparametric estimator. Second, I argue that a special case of shape restrictions on the conditional moments can be justified and for this case I propose an inference procedure that is based on the Gaussian asymptotic approximation and as such is relatively easy to implement. Finally, I conduct a Monte Carlo experiment that illustrates small sample performance of both methods.

This paper focuses on a particular class of partially identified models: models defined by a finite number of conditional moment inequalities with continuously distributed

covariates. Formally, the identified set  $\Theta_I$  is defined as

$$\Theta_I = \{\theta \in \Theta : E[m_j(Z, \theta)|X = x] \leq 0 \text{ for all } x \text{ and } j = 1, \dots, k\},$$

where  $\mathcal{X}$  is a compact (not necessarily proper) subset of the support of  $X$ . Here  $m(\cdot, \theta)$  is a vector-valued function that is known up to the value of the parameter of interest,  $\theta$ , and both  $Z$  and  $X$  are observable, possibly multidimensional, random variables. Random vector  $Z$  may or may not include vector of covariates,  $X$ . The objects of interest is a confidence set  $C_n$  that is *pointwise consistent in level* (the term is due to Romano and Shaikh, 2008), i.e.

$$\inf_{\theta \in \Theta_I} \lim_{n \rightarrow \infty} P\{\theta \in C_n\} \geq 1 - \alpha. \quad (1.1)$$

I construct confidence sets that satisfy (1.1) through pointwise testing procedure. That is, I construct  $(1 - \alpha)$  confidence set  $C_{n,1-\alpha}$  as a collection of all parameter value for which we fail to reject the null hypothesis  $H_0 : \theta \in \Theta_I$  against the alternative  $H_1 : \theta \notin \Theta_I$ . I choose the maximum deviation from zero of the estimator of conditional moment functions, properly normalized, as a statistic to test this null hypothesis. I use Nadaraya-Watson estimator of conditional moment functions. The limiting distribution of the statistic is different if different assumptions are made about the shape of conditional moment constraints. If we make no such assumptions besides some general smoothness conditions, then I propose a method based on bootstrap simultaneous confidence band (actually, half-band) for a conditional moment functions. However, this method may lead to conservative inference in small samples since it involves estimation of the set of zeros (or the argmax) of constraints, and without placing any restrictions this can be done only at the rate  $O_p((\log \log n/nh)^{1/2})$ , where  $h$  is the bandwidth choice.

I claim that in certain economic model it may be reasonable to assume that for the parameter values in the identified set conditional moment functions may have a unique binding point. That is, for any given parameter value of the boundary of the identified set, there is only one point in  $\mathcal{X}$  that some conditional moment inequality turns into an equality. In such a case one can learn the argmax set at the rate  $O((nh^3)^{-1/2})$ . In this setup, I show that one can construct an asymptotic approximation to the distribution of the chosen test statistic that behaves as a certain component-wise maximum of a multivariate Gaussian random variable. Basically, under these assumptions the

problem of constructing confidence set for the identified set reduces to estimation of moment conditions in a finite number of points, i.e. we are back to a finite number of inequalities. I propose an inference procedure analogous to the generalized moment selection approach of Andrews and Soares (2007), as well as give an approximation based on the wild bootstrap.

Finally, I consider an even more restrictive case when conditional moment functions are specified parametrically. If for each partially identified parameter value  $\theta$  we can estimate conditional moments  $\sqrt{n}$ -consistently, then I propose method based on the inference for the stochastic optimization problem studied in Shapiro (2000).

There is a number of techniques available in the literature that deal with the partially identified parameters. In the case of the scalar parameter of interest, the identification region is often an interval whose boundaries can be consistently estimated from the data. Inference procedures in such settings are developed in Horowitz and Manski (1998, 2000), Imbens and Manski (2004), Stoye (2007), and Fan and Park (2007). For a class of models whose identification region can be expressed as an expectation of a set valued random variable, Beresteanu and Molinari (2006) propose inference based on the Hausdorff distance between the identified set and its estimator. For a broad class of partially identified models, Chernozhukov, Hong, and Tamer (2007) extend the criterion function approach to set identified parameters, in particular those characterized by moment inequalities. They propose inference based on subsampling and asymptotic approximation to the distribution of sample criterion function. Building on this result, Romano and Shaikh (2006, 2007) suggest a different subsampling procedure that does not require an initial estimate of the identified set and that yields uniformly asymptotically valid confidence sets.

A number of papers focus explicitly on models with moment inequalities. Rosen (2006) constructs confidence sets based on pointwise testing and proposes a Wald-type statistic. Bugni (2007) and Canay (2007) introduce modifications of bootstrap procedure that are not subject to inconsistencies, while Andrews and Soares (2007) consider confidence sets based on generalized moment selection method and compare asymptotic power of various inference procedures. Andrews and Guggenberger (2007) address the issue of uniform inference and show uniform asymptotic validity for a broad class of confidence sets based on subsampling and “worst-case” asymptotic approximation. Those are the papers that focus on the univariate distribution of some criterion func-

tion. Another approach, introduced by Andrews, Berry, and Jia (2004), is to base inference directly on the distribution of sample analog of inequality constraints. The authors suggest a procedure that is based on a bootstrap approximation to the distribution of moment inequalities, while Pakes, Porter, Ho, and Ishii (2005) use an asymptotic approximation. Galichon and Henry (2006) extend this approach to more general settings and propose a dilation bootstrap approximation for their version of Kolmogorov-Smirnov test. Finally, Moon and Schorfheide (2007) and Liao and Jiang (2008) present a Bayesian approach to inference in partially identified model with unconditional moment inequalities. Liao and Jiang (2008) also show how in this setup one can consistently estimate the identified set for any real-valued continuous function of the parameter of interest.

The majority of the papers discussed above (with the exception of Chernozhukov, Hong, and Tamer (2007)) focus either on models with unconditional moment inequalities of models with discrete covariates. In a context of the randomly censored regression with continuous covariates, Khan and Tamer (2006) show that when the parameter of interest is point identified by conditional moment inequalities, it is possible to transform conditional moment constraints into unconditional ones and preserve point identification. The resulting estimator of the point identified parameter has a parametric convergence rate  $1/\sqrt{n}$ . Galichon and Henry (2006) suggest using the same transformation for partially identified model with continuous covariates. However, under this approach the number of unconditional moment constraints in the transformed model is a quadratic function of the sample size. In the same setup, Kim (2009) shows that the test statistic based on such transformation has a nonstandard limiting distribution that can be approximated by the infinite mixture of independent, centered chi-square distributions. Kim suggests subsampling as mean of constructing confidence sets. However, in the context of finite number of conditional moment inequalities Bugni (2007) shows that straightforward subsampling may lead to significant undercoverage in small samples. Menzel (2009) analyzes the case when the number of moment inequalities grows with the sample size. He also points out that when the the number of unconditional moments is large relative to the sample size, the subsampling procedure fails to achieve a desired confidence level in small samples. Andrews and Shi (2009) also study the above transformation of conditional moment inequalities into unconditional ones.

Another approach is to estimate conditional moment functions nonparametrically

and then to test whether they belong to the space of nonpositive functions. Ghosal, Sen, and van der Vaart (2000), Lee, Linton, and Whang (2007), and Birke and Dette (2007) use kernel estimator to test that a conditional mean function belongs to the space of (strictly) monotonic functions. Chernozhukov, Lee, and Rosen (2009) apply this approach to inference in models where the identified set is an interval with upper (lower) bound defined as a minimum (maximum) of a bound generating function that can be estimated parametrically or nonparametrically.

The rest of the paper is organized as follows. In section 2 I present several economic models where the identified set is defined by a number of conditional moment inequalities with continuous covariates. I also informally describe the methods proposed in the paper. In section 3 I formally describe the the setup and consider three cases distinguished by the restrictions placed on the conditional moment functions. For each of those cases I provide an inference method and justify it theoretically. Section 4 presents results of the Monte Carlo simulations, and section 5 concludes. All proofs of the results in this paper are collected in the Appendix.

## 2 Motivating Examples and Outline of the Inference Procedure

In this section I present three examples from the economic literature that deal with the partially identified parameters defined by the set of conditional moment inequalities and briefly describe the inference procedure.

**Example 1: Parametric Regression with Missing Data.** Sometimes the data available to the researcher may contain some missing values. If the researcher is interested in e.g. conditional mean of a certain variable that is not completely observed, then removal of the missing data from the sample may lead to a severe bias due to the fact that the data is not missing at random. The situation with missing outcome data can be treated as a special case of the regression with interval data, where some intervals (that correspond to the non-missing observations) are singletons. In particular, suppose that the conditional expectation of outcome  $Y$  given the vector of covariates  $X$  is  $E[Y|X = x] = g(x, \gamma)$ , where function  $g$  is known up to a vector of parameters  $\gamma$ . The

explanatory variable  $X$  is fully observed, however some observations of the outcome variable  $Y$  are missing. Let  $W$  be the indicator that the observation is non-missing, so that  $Y$  is observed if and only if  $W = 1$ . By the law of iterated expectations,

$$E[Y|X = x] = E[Y|X = x, W = 0]P\{W = 0|X = x\} + E[Y|X = x, W = 1]P\{W = 1|X = x\}.$$

If  $Y|X$  takes values in the bounded set  $[Y_0(X), Y_1(X)]$ , so that  $P\{Y \in [Y_0(x), Y_1(x)]|X = x\} = 1$  for any  $x \in \mathcal{X}$ , then the identified set for the parameter of interest,  $\gamma$ , is given by the following conditional moment constraints:

$$\Theta_I = \left\{ \begin{array}{l} \gamma \in \Theta \text{ such that } \forall x \in \mathcal{X} : \\ E[Y_0(X)|X = x, W = 0]p_0(x) + E[Y|X = x, W = 1]p_1(x) \\ \leq g(x, \gamma) \leq \\ E[Y_1(X)|X = x, W = 0]p_0(x) + E[Y|X = x, W = 1]p_1(x) \end{array} \right\},$$

where  $p_0(x) = P\{W = 0|X = x\}$  and  $p_1(x) = 1 - p_0(x)$ . Here both lower and upper bounds on  $g(x, \gamma)$  are the conditional expectations of the fully observed random variables and therefore can be consistently estimated.

**Example 2: Parametric Regression with Interval Data.** Another example is a parametric regression when the outcome is characterized by an interval rather than a point (see Manski and Tamer, 2002). For example, sometimes questionnaires elicit respondents' income brackets instead of the exact figure. Formally, let the conditional expectation of outcome  $Y$  given vector of covariates  $X$  be given by a parametric form  $E[Y|X = x] = f(x, \beta)$ , where function  $f$  is known up to the parameter  $\beta$ . However, instead of directly observing  $Y$ , we observe only upper and lower bounds on  $Y$ , so the data available to the researcher consists of  $(Y_0, Y_1, X)$ , where  $Pr\{Y_0 \leq Y \leq Y_1\} = 1$  and  $Y_0 \ll Y_1$ . The model implies that for any given point  $x$  that belongs to the support of the vector of covariates  $X$ , we have the following set of inequalities:

$$E[Y_0|X = x] \leq f(x, \beta) \leq E[Y_1|X = x].$$

Without imposing any additional assumptions about the distribution of  $Y|Y_0, Y_1, X$ , the identification set for  $\beta$  is given by  $\Theta_I =$

$\{\beta \in \Theta : E[Y_0|X = x] \leq f(x, \beta) \leq E[Y_1|X = x] \forall x \in \mathcal{X}\}$ . It is common to assume that  $f(x, \beta) = f(x\beta)$ . This setup includes linear, probit or logit model.

**Example 3: Analysis of Treatment Response in the Presence of Instrumental Variables.** In general, unobservability of counterfactual outcomes does not allow for point identification of treatment effects without making strong assumptions such as random assignment to treatments. However, if there exist an instrument variable  $v$  such that treatment outcome is independent of  $v$  conditional on the vector of covariates  $x$ , then the sharp bounds on average treatment effect  $E[y(t)|x]$  are given by (see Manski, 2003):

$$E[y \cdot 1\{z = t\}|x, v] \leq E[y(t)|x] \leq E[y \cdot 1\{z = t\} + 1\{z \neq t\}|x, v],$$

where both upper and lower bounds are conditional expectations of the observed random variables, and therefore can be consistently estimated. For simplicity, assume that there are only two possibilities: either the subject receives treatment, and then  $t = 1$ , or he does not, and then  $t = 0$ . The object of interest is then  $\theta_x = E[y(1)|x]$  given the vector of person's characteristics  $x$ . Then the identified set for  $\theta_x$  is given by

$$\Theta_{x,I} = \{\theta \in \Theta : E[y \cdot 1\{z = t\}|x, v] \leq E[y(t)|x] \leq E[y \cdot 1\{z = t\} + 1\{z \neq t\}|x, v] \forall v \in \mathcal{V}\}.$$

**Outline of the Inference Procedure.** There is a direct connection between testing a given hypothesis and construction of a confidence set associated with this hypothesis, which we will use to construct confidence set  $C_{n,1-\alpha}$  for each element of the identified set. In particular, suppose that we want to test the null hypothesis that a candidate parameter  $\theta$  from the set of all possible parameter values  $\Theta$  belongs to the identified set  $\Theta_I$  against the alternative that it does not belong to the identified set. That is, we want to test  $H_0 : \theta \in \Theta_I$  vs  $H_1 : \theta \notin \Theta_I$  with asymptotic size less than or equal to some prespecified value  $\alpha$ . Now suppose we employ some statistic  $T_n(\theta)$  to test the null hypothesis, so that under the null hypothesis its properly scaled version converges in distribution to some nondegenerate random variable  $Z_\theta$ , i.e.  $a_n(T_n(\theta) - b_n) \xrightarrow{d} Z_\theta$ . Then we can construct a  $(1 - \alpha)$  confidence set  $C_{n,1-\alpha}$  as

$$C_{n,1-\alpha} = \{\theta \in \Theta : a_n T_n(\theta) < q(\theta, 1 - \alpha)\}, \quad (2.1)$$

where  $q(\theta, 1 - \alpha)$  is the  $(1 - \alpha)$  quantile of the distribution of  $Z_\theta$ . That is, we collect all values of  $\theta$  such that we fail to reject the null hypothesis that this parameter value belongs to the identified set. Lemma 6.4 in the Appendix shows that confidence sets constructed in this way satisfy (1.1).

In terms of the moment restrictions, our null and alternative hypotheses can be stated as:

$$\begin{aligned} H_0 &: E[m_j(Z, \theta)|X = x] \leq 0 \text{ for all } x \in \mathcal{X} \text{ and all } j \in \{1, \dots, k\} \\ H_1 &: E[m_{j'}(Z, \theta)|X = x'] > 0 \text{ for some } x' \in \mathcal{X} \text{ and some } j' \in \{1, \dots, k\}. \end{aligned} \tag{2.2}$$

Note that the null hypothesis can be equivalently stated as

$$\max_{x \in \mathcal{X}, j=1, \dots, k} E[m_j(Z, \theta)|X = x] \leq 0.$$

Therefore, I propose a test statistic based on a maximum of uniformly consistent and asymptotically unbiased estimators of the conditional moment functions  $\{E[m_j(Z, \theta)|X = x], x \in \mathcal{X}, j = 1, \dots, k\}$ :

$$T_n(\theta) = \max_{x \in \mathcal{X}, j=1, \dots, k} \frac{\widehat{E}_n[m_j(Z, \theta)|X = x]}{s_j(x; \theta)},$$

where  $s_j(x; \theta)$  is the standard error of  $\widehat{E}_n[m_j(Z, \theta)|X = x]$ . Let's consider the following random variable:  $\max_{j=1, \dots, k} \max_{x \in V_0^j(\theta)} \frac{\widehat{E}_n[m_j(Z, \theta)|X = x] - E[m_j(Z, \theta)|X = x]}{s_j(x; \theta)}$ , where  $V_0^j(\theta)$  is a set

of zeros of the  $j$ 'th conditional moment function, and if  $V_0^j(\theta) = \emptyset$ , then we define  $\max_{x \in V_0^j(\theta)} \frac{\widehat{E}_n[m_j(Z, \theta)|X = x] - E[m_j(Z, \theta)|X = x]}{s_j(x; \theta)} \equiv 0$ . If we knew some asymptotic approximation to the distribution of this random variable, then we could find its  $(1 - \alpha)$ -quantile  $q_{1-\alpha}(\theta)$  such that

$$\lim_{n \rightarrow \infty} P \left\{ \max_{j=1, \dots, k} \max_{x \in V_0^j(\theta)} \frac{\widehat{E}_n[m_j(Z, \theta)|X = x] - E[m_j(Z, \theta)|X = x]}{s_j(x; \theta)} \leq q_{1-\alpha}(\theta) \right\} = 1 - \alpha$$

and reject the null hypothesis if  $T_n(\theta)$  is greater than  $q_{1-\alpha}(\theta)$ . However, such approximation is not readily available except for the case  $k = 1$ , when the limiting distribution is an Extreme Value Type I distribution (see e.g. Johnston, 1982), and even in this case the convergence to the EVT I distribution is very slow. Therefore, I propose to

estimate  $q_{1-\alpha}(\theta)$  via bootstrap, which is shown in Hall (1993) to have a faster rate of convergence. Let  $\widehat{E}_n^*[m_j(Z, \theta)|X = x]$  denote the estimator of the conditional moment functions for the bootstrap sample, and  $s_j^*(x; \theta)$  denote the standard error of the estimator  $\widehat{E}_n^*[m_j(Z, \theta)|X = x]$ . Then we can estimate  $q_{1-\alpha}(\theta)$  with  $\widehat{q}_{1-\alpha}(\theta)$ , where the latter is chosen to satisfy:

$$P \left\{ \max_{j=1, \dots, k} \max_{x \in \widehat{V}_0^j(\theta)} \frac{\widehat{E}_n^*[m_j(Z, \theta)|X = x] - \widehat{E}_n[m_j(Z, \theta)|X = x]}{s_j^*(x; \theta)} \leq \widehat{q}_{1-\alpha}(\theta) \right\} = 1 - \alpha$$

where set  $\widehat{V}_0^j(\theta)$  asymptotically is a superset for  $V_0^j(\theta)$ . This method in general leads to the conservative coverage, since it uses a superset of  $V_0^j(\theta)$  rather than the set of zeros itself. This result is in parallel to those of Rosen (2006) for the model with unconditional moment inequalities, where an obvious upper bound on the number of binding inequalities is the total number of inequalities. However, if we know that  $V_0^j(\theta) = \mathcal{X}$ , then the method provides the exact pointwise coverage only if there exists some  $\theta_0 \in \Theta_I$  such that  $E[m_j(Z, \theta_0)|X = x] = 0$  for every  $x$  and  $j$ .

When is it also possible to obtain the exact coverage of the identified set? One possibility is to specify a parametric form of the conditional moment functions for any given  $\theta$ . For example, in the parametric regression with interval data one can also think of a parametric form for the conditional expectations of upper and lower interval points, i.e.  $E(Y_1|X) = g_1(x, \gamma_1)$  and  $E(Y_0|X) = g_0(x, \gamma_0)$ , where functions  $g_1$  and  $g_0$  are known up to the parameters  $\gamma_1$  and  $\gamma_0$ , correspondingly. Another possibility is to assume that for any  $\theta$  in the identified set for each  $j$  the set  $X_0^j(\theta) = \arg \max_{x \in \mathcal{X}} E[m_j(Z, \theta)|X = x]$  is either a singleton or an empty set. In other words, for any  $\theta$  on the boundary of the identified set conditional moment functions are not flat around zero. In both of this cases it is possible to approximate the distribution of the statistic  $T_n(\theta)$  with a certain functional of a multivariate Gaussian random variable  $Z(\theta)$  using a consistent estimator of the set of zeros rather than of its superset, and use its critical values to construct pointwise confidence sets.

Figures 1 and 2 illustrate the difference between flat and non-flat constraints for the model with interval outcome data, where  $f(x, \theta)$  is assumed to be linear, i.e.  $f(x, \theta) = \theta_0 + \theta_1 \cdot x$ . Next section presents the theoretical justification and the set of necessary assumptions for inference procedures in both of this cases.

### 3 Model and Inference

I start with formalizing the model. For the ease of presentation, assume that the model is characterized by  $k = 2$  conditional moment restrictions:

$$\begin{aligned} E[m_1(Z, \theta)|X = x] &\leq 0 \\ &\vdots \\ E[m_k(Z, \theta)|X = x] &\leq 0 \end{aligned} \tag{3.1}$$

where functions  $m_1(z, \theta), \dots, m_k(z, \theta)$  are known. Define  $J = \{1, \dots, k\}$ . We are interested in characterizing the set of parameter values  $\theta$  that satisfy all conditional moment restrictions for every  $x$  in some set  $\mathcal{X}^1$ . We call this set the *identified set* and denote it by  $\Theta_I$ . The researcher observes a random sample  $\{(X_i, Z_i), i = 1, \dots, n\}$  from population  $(\Omega, \mathcal{B}, \mathbf{P})$ , where  $\mathcal{B}$  is a  $\sigma$ -algebra of Borel sets on  $\Omega$ . Let  $S_X \subset \mathbb{R}^d$  and  $S_Z \subset \mathbb{R}^q$  denote the support of random vectors  $X$  and  $Z$ , respectively. We impose the following assumptions:

- (A1) (*Random sampling*) The data  $\{(X_i, Z_i), i = 1, \dots, n\}$  are i.i.d. sample from  $(\Omega, \mathcal{B}, \mathbf{P})$ .
- (A2) (*Compact parameter space*) Parameter of interest  $\theta$  belongs to the compact set  $\Theta \subset \mathbb{R}^m, 0 < m < +\infty$ .
- (A3) (*Conditioning over compact set*) The set  $\mathcal{X} \subseteq S_X$  is compact.

This gives the following definition of the identified set:

**Definition:** The identified set is given by

$$\Theta_I = \{\theta \in \Theta : (3.1) \text{ holds for all } x \in \mathcal{X}\} \tag{3.2}$$

The goal is to construct a confidence set  $C_{n,1-\alpha}$  for each element of  $\Theta_I$  that asymptotically consistent in level:

$$\inf_{\theta \in \Theta_I} \lim_{n \rightarrow \infty} P\{\theta \in C_{n,1-\alpha}\} \geq 1 - \alpha. \tag{3.3}$$

---

<sup>1</sup>In most of the cases we are interested in the case when moment conditions holds for any  $x$  in the support of  $X$ , so that  $\mathcal{X} = \text{supp}(X)$ .

As it was discussed in Section 2, the confidence sets in this paper are based on testing for each candidate value  $\theta \in \Theta$  the hypothesis that  $\theta \in \Theta_I$  against the alternative that  $\theta \notin \Theta_I$ , that is

$$\begin{aligned} H_0 &: E[m_j(Z, \theta)|X = x] \leq 0 \text{ for all } x \in \mathcal{X} \text{ and all } j \in \{1, 2\} \\ H_1 &: E[m_{j'}(Z, \theta)|X = x'] > 0 \text{ for some } x' \in \mathcal{X} \text{ and some } j' \in \{1, 2\}. \end{aligned} \tag{3.4}$$

I construct the test statistic for the null hypothesis in the following way. Let  $\widehat{E}_n[m_j(Z, \theta)|X = x]$  be an estimator (parametric or nonparametric) of the  $j$ 'th conditional moment function  $E[m_j(Z, \theta)|X = x]$  and let  $s_j(x, \theta)$  be the standard error of this estimator. Define a test statistic  $T_n(\theta)$  as

$$T_n(\theta) = \max \left\{ \max_{x \in \mathcal{X}} \frac{\widehat{E}_n[m_j(Z, \theta)|X = x]}{s_j(x, \theta)}; j \in J \right\} \tag{3.5}$$

If the null hypothesis is true for a candidate parameter  $\theta$ , we can expect to see small values of  $T_n(\theta)$ . Quantity  $[T_n(\theta)]_+^2$  can be viewed as a minimum (in the supremum norm) distance from the vector of estimators of the conditional moment functions to the space of non-positive functions.

The choice of the estimator of the conditional moment functions  $\widehat{E}_n[m_j(Z, \theta)|X = x]$  may be arbitrary as long as this estimator is uniformly consistent on  $\mathcal{X}$ . One can estimate conditional moment functions nonparametrically, or one can specify a parametric form for the conditional moments and use a  $\sqrt{n}$ -consistent estimator of this parameter to estimate those functions. In this paper I focus on inference based on the Nadaraya-Watson (kernel) estimator of conditional moment functions in the nonparametric case; and for the parametric case I provide results assuming availability of a parametric  $\sqrt{n}$ -consistent estimator of conditional moment functions for any  $\theta$  in the parameter space  $\Theta$ .

As it was noted in previous section, the asymptotic coverage results depend on the shape of the conditional moment functions. Therefore, I divide further discussion into three subsections. First subsection deals with the case when no restrictions are imposed on the conditional moment functions besides some necessary smoothness assumptions. This is the case that allows for the possibility of flat binding constraints. Second

---

<sup>2</sup>I define  $[u]_+ = \max\{u, 0\}$ .

subsection constructs confidence sets in the case where for any parameter  $\theta$  on the boundary of the identified set each of the binding conditional moment inequalities has a unique global maximum. That is, we rule out the possibility of flat binding constraints. In both subsections I use kernel estimators of conditional moment functions. Finally, third subsection treats the case of the parametric form of the conditional moment constraints.

Since I will be using a  $L_\infty$  distance from the conditional moment function to the set of non-positive functions as the test statistic, the problem of testing the non-positivity of conditional moment functions amounts to estimating the distance from a regression function to a cone of non-positive functions. The latter problem is studied in Juditsky and Nemirovski (2002). They show in the context of a white-noise model that the minimax risk of estimating the  $L_r$ -distance,  $1 \leq r < \infty$ , from the (unobserved) regression function to the cone of non-positive functions is essentially the same (up to the  $\ln n$ ) as the minimax risk of estimating the regression function itself. For example, for the class of bounded and  $k = 2$  times continuously differentiable functions, the minimax risk  $R^*(n)$  of estimating the distance  $\Phi_r$  from a regression function to the cone of non-positive functions on  $[0, 1]$ , defined as  $R^*(n) = \inf_{\hat{f}_n} \sup_{f \in C^2[0,1]} E\{|\Phi_r(\hat{f}_n) - \Phi_r(f)|\}$ , where the infimum is taken over the set of all nonparametric estimates, is bounded by

$$O(n^{-4/5}(\ln n)^{-\tau}) \leq R^*(n) \leq O((n \ln n)^{-4/5}). \quad (3.6)$$

For a given function  $f$ , the distance  $\Phi_r(f)$  from a given function  $f$  to the cone of non-positive functions  $\mathcal{M}$  is defined as  $\Phi_r(f) = \inf\{\|f - g\|_r : g \in \mathcal{M}\}$ . Note that the convergence rate  $n^{-4/5}$  in (3.6) corresponds to the minimal AMISE of the kernel regression estimator in the case of twice continuously differentiable functions (see Ullah and Pagan, 1999). Likewise, Ghosal et al (2000) and Dümbgen and Spokoiny (2001) provide results for the tests of qualitative hypotheses based on the estimation of the  $L_\infty$ -distance to the cone of non-positive, monotone or convex functions and obtain results similar to Juditsky and Nemirovski (2002). These findings suggest that unlike the models comprised of the unconditional moment inequalities or conditional moment inequalities where covariates have finite support, in the present setup of the conditional moment restrictions with continuous covariates we cannot expect to achieve a parametric convergence rate  $1/\sqrt{n}$  for the boundary of the identified set without imposing

additional parametric assumptions about conditional moment functions that define the identified set. Therefore, finite sample performance of any inferential procedure based on nonparametric estimators of conditional moment constraints is going to be inferior to the finite sample performance of methods proposed in the literature that deal with unconditional constraints (that can be estimated at parametric rate).

### 3.1 No shape restrictions

I first start with the case where we do not wish to impose any shape restrictions on the conditional moment inequalities. Critical values for the test of the null hypothesis that a candidate parameter value  $\theta$  belongs to the identified set based on the statistic  $T_n(\theta)$  defined in (3.5) are constructed based on the bootstrap approximation to the distribution of

$$\max \left\{ \max_{x \in V_0^j(\theta)} \frac{\widehat{E}_n[m_j(Z, \theta)|X = x] - E[m_j(Z, \theta)|X = x]}{s_j(x; \theta)}, j \in J \right\} \quad (3.7)$$

where  $V_0^j(\theta) = \{x \in \mathcal{X} : E[m_j(Z; \theta)|X = x] \geq 0\}$ <sup>3</sup> and  $s_j(x; \theta)$  is standard error of an estimator  $\widehat{E}_n[m_j(Z, \theta)|X = x]$ . A multivariate analog of a strong approximation result for kernel estimators of regression functions can be used to show that under certain conditions, distribution of (3.7) may be approximated by the distribution of

$$\max \left\{ \max_{x \in V_0^j(\theta)} \Gamma_n(x), j \in J \right\}$$

where  $\{\Gamma_n(\cdot), n = 1, \dots\}$  is a sequence of multivariate zero mean Gaussian processes indexed by  $\mathcal{X}$  with continuous sample paths and known covariance function. This implies that the random variable in (3.7) converges in distribution to some stable law no matter whether or not  $\theta$  belongs to  $\Theta_I$ . However, if  $\theta$  does not belong to the identified set, then the statistic  $T_n(\theta)$  defined in (3.5) asymptotically goes to plus infinity with probability one, while the limiting distribution of (3.7) still exists.

Although quantiles of the distribution of (3.7) are not readily available, and in

---

<sup>3</sup>With this definition,  $V_0^j(\theta)$  is indeed a set of zeros of  $j$ 'th conditional moment function if  $\theta$  belongs to the boundary of the identified set  $\Theta_I$ . If  $\theta$  belongs to the interior of  $\Theta_I$ , then  $V_0^j(\theta)$  is empty. Finally, if  $\theta$  does not belong to  $\Theta_I$ , then  $V_0^j(\theta)$  is a compact set with a non-empty interior.

general case sets  $V_0^j(\theta)$  are unknown<sup>4</sup>, I start with the lemma that supports the validity of confidence set for each element of the identified set that is based on quantiles of this distribution if we substitute  $V^j$ -a superset for  $V_0^j(\theta)$

**Lemma 3.1** *Suppose that assumptions (A1)-(A3) are satisfied, and the identified set if given by (3.2). Let  $\widehat{E}_n[m_j(Z, \theta)|X = x]$  be a uniformly consistent estimator of  $E[m_j(Z, \theta)|X = x]$  on  $\mathcal{X}$  for each  $j \in J$ , with  $s_j(x; \theta)$  being the standard error of this estimator. Let  $(1 - \alpha)$  be the desired confidence level,  $\alpha \in (0, 1)$ . Let  $V_0^j(\theta) \subset V^j$ . Suppose that  $q_{1-\alpha}(\theta)$  is defined as*

$$\lim_{n \rightarrow \infty} P \left\{ \max_{j \in J} \left\{ \max_{x \in V^j} \frac{\widehat{E}_n[m_j(Z, \theta)|X = x] - E[m_j(Z, \theta)|X = x]}{s_j(x; \theta)} \right\} \leq q_{1-\alpha}(\theta) \right\} = 1 - \alpha \quad (3.8)$$

Define  $C_{n,1-\alpha} = \{\theta \in \Theta : T_n(\theta) \leq q_{1-\alpha}(\theta)\}$ , where  $T_n(\theta)$  is given by (3.5). Then  $\inf_{\theta \in \Theta_I} \lim_{n \rightarrow \infty} P \{\theta \in C_{n,1-\alpha}\} \geq 1 - \alpha$

The intuition behind this lemma is as follows: given a candidate parameter  $\theta$ , equation (3.8) gives one-sided (lower) simultaneous confidence bound for true (unobserved) conditional moment curves that correspond to this parameter value. Then the confidence set  $C_{n,1-\alpha}$  is constructed as the collection of all parameter values such that the constant zero function  $c(x) \equiv 0$  lies above this lower bound for all conditional moment functions simultaneously.

In general, an asymptotic approximation of the distribution of (3.7) is not readily available. One exclusion in the case  $k = 1$  and  $V_0^j(\theta) = \mathcal{X}$ , when Johnson (1982) shows that under certain conditions a studentized version of the Nadaraya-Watson kernel estimator allows a strong approximation by a sequence of Gaussian processes. Then one can apply extremal types theorem for stationary processes in Leadbetter and Rootzen (1988) to obtain an approximation of the distribution of the supremum of the Gaussian process with Extreme Value Type I distribution. This approximation is known to have a slow (logarithmic in  $n$ ) rate of convergence (see Hall 1979, 1991), but one can use the result due to Konakov and Piterbarg (1984) that provides a refinement to this

---

<sup>4</sup>Although the set  $\{x \in \mathcal{X} : E_n[m_j(Z; \theta)|X = x] \geq 0\}$  is a consistent estimator of the set  $V_0^j(\theta)$  when the latter is nonempty, we cannot simply plug-in this estimator since the limiting distribution of (3.7) is discontinuous in  $\{V_0^j(\theta), j \in J\}$ . Moreover, if  $V_0^j(\theta)$  is a singleton, in small samples its estimator will always be a compact set with a non-empty interior.

asymptotic approximation that converges at polynomial rate. For the comparison of two approximations in the context of testing for monotonicity, see Ghosal et al (2000), Lee, Linton and Whang (2007).

I do not want to rule out the possibility of having more than one conditional moment constraint that defines the identified set. Therefore, I resort to the approach due to Hall (1993) and estimate quantiles  $q_{1-\alpha}(\theta)$  by the quantile of bootstrap version of (3.7) for a kernel estimator of conditional moment functions. I start with introducing some notation. Let  $f(x)$  denote the density of  $X$ , and let  $\sigma_j^2(x; \theta)$  denote the conditional variance of the error term

$$w_i^j = m_j(Z_i, \theta) - E[m_j(Z_i, \theta) | X_i = x]$$

I employ a kernel estimator of the conditional moment functions to construct the test statistic  $T_n(\theta)$ . Given the choice of bandwidth  $h$  and kernel function  $K$ , the kernel estimator for the  $j$ 'th conditional moment constraint is defined by

$$\widehat{E}_n[m_j(Z_i, \theta) | X_i = x] = \frac{\sum_{i=1}^n m_j(Z_i, \theta) K((x - X_i)/h)}{\sum_{i=1}^n K((x - X_i)/h)} \quad (3.9)$$

Denote  $c_K = \int K^2(u) du$  and define

$$\begin{aligned} K_h(u) &= \frac{1}{h} K(u/h) \\ \hat{f}(x) &= n^{-1} \sum_{i=1}^n K_h(x - X_i) \\ \hat{\sigma}_j^2(x; \theta) &= n^{-1} \sum_{i=1}^n (m_j(Z_i, \theta))^2 K_h(x - X_i) / \hat{f}(x) - \left( \widehat{E}_n[m_j(Z_i, \theta) | X_i = x] \right)^2 \end{aligned}$$

If we choose the bandwidth and the kernel in such a way that the estimator is asymptotically unbiased, we can estimate its variance at any given  $x$  with

$$s_j(x, \theta) = \frac{c_K \hat{\sigma}_j^2(x; \theta)}{nh \hat{f}(x)} \quad (3.10)$$

In order to obtain an asymptotic approximation of the test statistic (3.5) for a kernel

estimator  $\widehat{E}_n[m_j(Z, \theta)|X = x]$  we need to impose some regularity conditions that deal with the smoothness of conditional moment functions and moments of random variable  $m_j(Z, \theta)|X$ .

**(R1)** Density  $f(x)$  is bounded away from zero on  $\mathcal{X}$ .

**(R2)** For any  $\theta \in \Theta$  and any  $j \in J$ , functions  $E[m_j(Z; \theta)|X = x]$ ,  $f(x)$  and  $\sigma_j^2(x; \theta)$  are twice continuously differentiable on  $\mathcal{X}$ .

**(R2)** For any  $\theta \in \Theta$  and  $j \in J$ ,  $E[|m_j(Z; \theta)|^{2+p}|X = x] \leq A_p < \infty$  for some  $p > 0$ .

To estimate quantiles of distribution from Lemma 3.1, I propose the following bootstrap procedure:

1. Estimate conditional moment functions with (3.9). Choose  $\tau_n = (2 \log \log n)^{1/2}$ .
2. Choose a small  $\varepsilon > 0^5$  and for each  $j$ , construct an estimator of the superset of  $V_0^j(\theta)$  as

$$\widehat{V}_0^j(\theta) = \left\{ x \in \mathcal{X} : \widehat{E}_n[m_j(Z_i, \theta)|X_i = x] \geq -\varepsilon - \tau_n s_j(x, \theta) \right\} \quad (3.11)$$

3. Construct  $R$  bootstrap samples of size  $n$  by sampling randomly with replacements from the data  $(Z_i, X_i)$ .
4. For each bootstrap sample, estimate conditional moment functions with (3.9) and denote those estimates  $\widehat{E}_n^*[m_j(Z_i, \theta)|X_i = x]$ . Similarly, calculate  $s_j^*(x, \theta)$  for each bootstrap sample. Finally, compute bootstrap analog of (3.7):

$$B_n^*(\theta) = \max_{j \in J} \max_{x \in \widehat{V}_0^j(\theta)} \frac{\widehat{E}_n^*[m_j(Z, \theta)|X = x] - \widehat{E}_n[m_j(Z, \theta)|X = x]}{s_j^*(x; \theta)} \quad (3.12)$$

5. Let  $\hat{q}_{n, 1-\alpha}^B(\theta)$  be  $(1 - \alpha)$  quantile of the empirical distribution of  $B_n^*(\theta)$ .

**Theorem 3.2** *Let assumptions (A1)-(A3) and (R1)-(R3) hold. Suppose that*

1.  $nh^5 \log n \rightarrow 0$  and  $nh^3 \rightarrow \infty$  as  $n \rightarrow \infty$ ;

---

<sup>5</sup>I propose to choose  $\varepsilon > 0$  so that first even if  $V_0^j(\theta)$  is a singleton, its superset has a nonempty interior as well as the estimator of the superset. And second, this way we can set  $C = 1$  in  $\tau_n = C(\log \log n)^{1/2}$  and thus avoid choosing a particular value for this constant.

2.  $K$  is a differentiable kernel that vanishes outside the interval  $[-1, 1]$ .

Define  $C_{n,1-\alpha}^B = \{\theta \in \Theta : T_n(\theta) \leq \hat{q}_{1-\alpha}^B(\theta)\}$ , where  $T_n(\theta)$  is given by (3.5). Then  $\inf_{\theta \in \Theta_I} \lim_{n \rightarrow \infty} P\{\theta \in C_{n,1-\alpha}^B\} \geq 1 - \alpha$ .

Condition  $nh^5 \log n \rightarrow 0$  ensures that kernel estimator is asymptotically unbiased, and therefore we do not need to correct for the bias when constructing simultaneous lower confidence bound for conditional moment functions. Choice of a sequence  $\{\tau_n, n \in \mathbb{N}\}$  follows from a law of iterated logarithm for Nadaraya-Watson regression estimator (see Härdle, 1984). I choose to undersmooth and set the bias asymptotically to zero by requiring  $nh^5 \log n \rightarrow 0$  rather than estimating the bias because Hall (1993) shows that inaccuracy of bias-correction procedure may affect the accuracy of the bootstrap approximation. Finally, we choose to estimate an  $\varepsilon$ -superset  $\{x \in \mathcal{X} : |E[m_j(Z, \theta)|X = x]| < \varepsilon\}$  of  $V_0^j(\theta)$  rather than the set of zeros itself because statistic  $T_n(\theta)$  has different limiting behavior depending on whether the set of zeros of conditional moment functions for a given  $\theta$  is finite or a continuum.

The method proposed in Theorem 3.2 will lead to conservative coverage in small samples if no  $\theta$  exists such that for this parameter value all binding conditional moment functions are identical zeros for a continuum of points in the support. Consider a case when  $V_0^j(\theta)$  is a singleton for some  $j$  (this case is a subject of the discussion in the next section). Due to continuity of conditional moment functions, for any finite  $n$  estimator  $\hat{V}_0^j(\theta)$  will be a segment or a collection of segments, but not a singleton. Since the situation when no such parameter value exists is not implausible, next subsection attempt to improve the coverage under the assumption that for each  $\theta$  all binding conditional moments (there may be none) have unique maximum (which is equivalent to the assumption that all binding conditional moment functions have unique zero point).

### 3.2 Conditional moment functions with unique maximum

Suppose that the researcher has reasons to believe that for any  $\theta$  in the identified set each of the binding conditional moment functions (there might be none if  $\theta$  belongs to the interior of the identified set) equals zero only at a single point in  $\mathcal{X}$ . Given some smoothness assumptions about conditional moments, this belief is equivalent to the statement that for any  $\theta$  on the boundary of the identified set each of the binding

constraints for this parameter value has a unique global maximum, and the size of this maximum is equal to zero.

As an illustration of the economic model where one might expect the binding conditional moments to have a unique maximum, consider a game-theoretic model with multiple equilibria, such as e.g. entry-exit games. Tamer (2003) shows that in the absence of equilibrium selection mechanism, there may exist a set of parameter values that are consistent with more than one outcome. For example, let  $y_i$  denote the observed outcome of an entry-exit game in market  $i$ , and let  $x_i$  be a vector of profit shifters for all players in market  $i$ . Denote by  $\mathcal{Y}$  and  $\mathcal{X}$  the supports of  $y_i$  and  $x_i$ , respectively. Then given a parametric model for the probability of outcome  $y$ ,  $P(y|x;\theta)$ , the possibility of multiple equilibria implies that the identified set for  $\theta$  is characterized by

$$P_0(y|x) - P(y|x;\theta) \leq 0, \quad \forall (y, x) \in (\mathcal{Y}, \mathcal{X}), \quad (3.13)$$

where  $P_0(y|x)$  is the true (observed) probability of outcome  $y$ . For the detailed presentation of this model, see Andrews, Berry, and Jia (2004). Because the parametric model provides only an approximation for the true more complicated data generating process, one cannot always expect to find  $\theta$  such that (3.13) holds with equality for all  $(y, x) \in (\mathcal{Y}, \mathcal{X})$ .

Note that the set  $V_0^j(\theta) = \{x \in \mathcal{X} : E[m_j(Z, \theta)|x] \geq 0\}$  is also the argmax set for the  $j$ 'th conditional moment functions given a candidate value  $\theta$  belongs to the set  $\Theta_I$ . The following list of assumptions summarizes shape restrictions under consideration in this subsection, restrictions on location of the argmax set, and some additional smoothness requirements for the behavior of conditional moment functions in the neighborhood of the argmax sets.

**(UM1)** (*Unique mode*) For any parameter  $\theta$  in the identified set and any  $j \in J$  the argmax set  $V_0^j(\theta)$  is a singleton:  $V_0^j(\theta) = \{x_0^j(\theta)\}$ .

**(UM2)** (*Additional smoothness in the neighborhood of argmax*) There exist some  $\varepsilon > 0$  such that each conditional moment function is at least 3 times continuously differentiable in each interval  $[x_0^j(\theta) - \varepsilon, x_0^j(\theta) + \varepsilon]$ . Also,  $f(x)$  is at least 3 times continuously differentiable on each interval  $[x_0^j(\theta) - \varepsilon, x_0^j(\theta) + \varepsilon]$ . Finally, the Jacobian matrix of  $j$ 'th conditional moment function, evaluated at the argmax,

is negative definite: for any vector  $u \neq 0$ ,

$$u' \left( \frac{\partial^2 E[m_j(Z; \theta) | X = x]}{\partial x \partial x'} \Big|_{x=x_0^j(\theta)} \right) u < 0.$$

**(UM3)** (*Location of the argmax*) For any parameter  $\theta$  in the identified set,  $x_0^j(\theta)$  belongs to the interior of  $\mathcal{X}$  for  $j \in J$ .

Assumption (UM1) formally states that for any parameter on the boundary of the identified set there is a unique value of covariate vector  $X$  for which some of the moment inequalities turn into equalities. Assumption (UM2) ensures the appropriate behavior of conditional moment functions in some neighborhood of the argmax set that allows to construct an asymptotic Gaussian approximation of for the distribution of test statistic  $T_n(\theta)$ . Finally, assumption (UM3) rules out the case when the maximum is attained on the boundary of the set  $\mathcal{X}$  in order to avoid possible boundary effects in the case when the compact set  $\mathcal{X}$  (over which we want the conditional moment constraints to hold) coincides with the support of  $X$ . This assumption can be removed if  $\mathcal{X}$  is a proper subset of a continuous support of  $X$ .

In this setting we assume that all binding conditional moment functions have unique maximum, and therefore I will focus on asymptotic approximation of the behavior of the conditional moment functions locally (i.e. around the location of the argmax) rather than globally, as it was done in previous subsection. Therefore, instead of considering the maximum of the studentized version of kernel estimator (given by  $T_n(\theta)$  in (3.5)), I will analyze the studentized version of the maximum, given by

$$\tilde{T}_n(\theta) = \max \left\{ \frac{\left[ \max_{x \in \mathcal{X}} \hat{E}_n[m_j(Z_i, \theta) | X_i = x] \right]_+}{s_j(\hat{x}_0(\theta), \theta)}; j \in J \right\} \quad (3.14)$$

where  $\hat{x}_0(\theta) = \arg \max_{x \in \mathcal{X}} \hat{E}_n[m_j(Z_i, \theta) | X_i = x]$ .

Next theorem provides this asymptotic approximation, and inferential method presented in this section is based on this result.

**Theorem 3.3** *Let assumptions (A1)-(A3), (R1)-(R3) and (UM1)-(UM3) hold. Suppose that*

1.  $nh^7 \rightarrow \lambda^2 \geq 0$  and  $nh^5 \rightarrow \infty$  as  $n \rightarrow \infty$ ;
2.  $K$  is a symmetric, twice continuously differentiable third order kernel that vanishes outside the interval  $[-1, 1]$ .

Then for any  $\theta \in \Theta$

$$\sqrt{nh}(\max_{x \in \mathcal{X}} \widehat{E}_n[m(Z_i, \theta)|X_i = x] - \max_{x \in \mathcal{X}} E[m(Z_i, \theta)|X_i = x]) \xrightarrow{d} N(0, \Sigma(\theta)), \quad (3.15)$$

where

$$(\Sigma(\theta))_{j,l} = \begin{cases} 0 & \text{if } x_0^j(\theta) \neq x_0^l(\theta) \\ c_K \frac{E[W_{jl}(Z, X, \theta)|X=x_0^j(\theta)]}{f(x_0^j(\theta))} & \text{otherwise} \end{cases} \quad (3.16)$$

and

$$W_{jl}(Z, X, \theta) = (m_j(Z, \theta) - E[m_j(Z, \theta|X)])(m_l(Z, \theta) - E[m_l(Z, \theta|X)])$$

This result is an extension of Theorem 4.1 in Ziegler (2003). Theorem 3.3 claims that under certain conditions on the choice of a bandwidth and kernel function, the estimator of the size of the maximum of conditional moment constraints is asymptotically normally distributed with zero mean. Here we need  $nh^5 \rightarrow \infty$  to deliver consistency of the estimates of the derivatives up to second order; we also need  $nh^7 n \rightarrow \lambda^2 \geq 0$  to achieve a necessary rate of convergence of  $\hat{x}_0^j(\theta) = \arg \max_{x \in \mathcal{X}} \widehat{E}_n[m(Z_i, \theta)|X_i = x]$  to  $x_0^j(\theta)$ . Assumption (UM2) that requires conditional moment functions and density of  $X$  to be three times continuously differentiable in some neighborhood of the maxima is crucial here, since it allows us to eliminate the asymptotic bias of the kernel estimator of a maximum by using a third order kernel while allowing  $nh^5 \rightarrow \infty$ . Finally note that if conditional moment functions have different locations of the global maximum, the variance-covariance matrix (3.16) becomes diagonal. Appendix provides a consistent estimator of  $\Sigma(\theta)$ .

Under the null hypothesis that  $\theta \in \Theta_I$ , we have:

- $\max_{x \in \mathcal{X}} E[m_j(Z_i, \theta)|X_i = x] < 0$  for  $j \in J$  if  $\theta$  belongs to the interior of  $\Theta_I$ ;
- $\max_{x \in \mathcal{X}} E[m_j(Z_i, \theta)|X_i = x] \leq 0$  for  $j \in J$ ,  
and for some  $j' \in \{1, 2\}$   $\max_{x \in \mathcal{X}} E[m_{j'}(Z_i, \theta)|X_i = x] = 0$  if  $\theta$  belongs to the boundary of  $\Theta_I$ .

This observation together with the asymptotic approximation of Theorem 3.3 allow us to test the null hypothesis that a candidate parameter value  $\theta$  belongs to the identified set using statistic  $\tilde{T}_n(\theta)$  introduced in (3.14). Basically, this result reduces the continuum of moment inequalities to a finite number of constraints evaluated at certain points, and inferential theory for this case is readily available in the literature. The only difference is in the estimator of moment conditions. It can be shown that the limiting distribution of  $\tilde{T}_n(\theta)$  depends on the number of binding conditional moment constraints for the value of  $\theta$ . In particular, if  $\theta$  belongs to the interior of the identified set, then  $\tilde{T}_n(\theta)$  is  $o_p(1)$ . If  $\theta$  belongs to the boundary of the identified set, then  $\tilde{T}_n(\theta)$  is  $O_p(1)$  and is distributed as the maximum of the  $[Z_{b(\theta)}(\theta)]_+$ , where  $b(\theta)$  is the number of binding constraints for this  $\theta$  and  $Z_b(\theta)$  is  $b$ -dimensional Gaussian random variable. We cannot directly plug in any consistent estimator of  $b(\theta)$  since the limiting distribution is discontinuous in  $b(\theta)$ . Therefore, I propose an asymptotic approximation analogous to the one proposed in Bugni (2007) for the case of unconditional moment inequalities. In this case the quantiles of the approximating distribution need to be obtained by simulations. The procedure is as follows: for a given candidate  $\theta$

1. Choose  $\tau_n = (2 \log \log n)^{1/2}$ . Let  $\hat{\Sigma}(\theta)$  be a consistent estimator of variance-covariance matrix in Theorem 3.3.
2. Draw a samples of size  $R$  from  $N(0, \hat{\Sigma}(\theta))$ . Let  $Z^r(\theta)$  be the  $r$ 'th element of this sample. Compute

$$G_{r,n}(\theta) = \max \left\{ \frac{[Z_j^r(\theta)]_+}{\left(\hat{\Sigma}(\theta)\right)_{jj}} \cdot \mathbf{1} \left\{ \frac{\hat{E}_n[m(Z_i, \theta) | X_i = \hat{x}_0^j(\theta)]}{s_j(\hat{x}_0^j(\theta), \theta)} \geq -\tau_n \right\}, j \in J \right\}$$

3. Define  $q_{n,1-\alpha}^G(\theta)$  as the  $(1 - \alpha)$  quantile of the empirical distribution of  $G_r(\theta)$ .

The choice of sequence  $\{\tau_n, n \in \mathbb{N}\}$  comes from the a law of iterated logarithm for kernel estimators introduced in Härdle (1984). This procedure is analogous to the generalized moment selection procedure introduced in Andrews and Soares (2007) and Chernozhukov, Hong and Tamer (2007).

**Theorem 3.4** *Let the assumptions of Theorem 3.3 hold.*

*Define  $C_{n,1-\alpha}^G = \left\{ \theta \in \Theta : \tilde{T}_n(\theta) \leq \hat{q}_{1-\alpha}^G(\theta) \right\}$ , where  $\tilde{T}_n(\theta)$  is given by (3.14).*

Then  $\inf_{\theta \in \Theta_I} \lim_{n \rightarrow \infty} P \{ \theta \in C_{n,1-\alpha}^G \} = 1 - \alpha$ .

It seems possible to avoid estimating variance-covariance matrix and use wild bootstrap to approximate distribution of  $\tilde{T}_n(\theta)$  under the null hypothesis. The procedure is somewhat analogous to the method of testing strict monotonicity of a function proposed in Bowman, Jones and Gijbels (1998). It relies on the wild bootstrap approximation (see e.g. Wu, 1986) and is given here as a conjecture. For a given candidate  $\theta$

1. Choose  $\tau_n = (2 \log \log n)^{1/2}$ . Estimate residuals by  $\hat{u}_{ji} = m_j(Z_i, \theta) - \hat{E}_n[m(Z_i, \theta) | X_i]$ .
2. Draw a random sample  $\{C_i, i = 1, \dots, n\}$  from the distribution that puts mass  $(\sqrt{5} + 1)/2\sqrt{5}$  on  $(1 - \sqrt{5})/2$  and mass  $(\sqrt{5} - 1)/2\sqrt{5}$  on  $(1 + \sqrt{5})/2$ .
3. Generate bootstrap sample  $\{(Y_i^*(\theta), X_i) | i = 1, \dots, n\}$  as

$$Y_{ij}^*(\theta) = \hat{E}_n[m(Z_i, \theta) | X_i] - \hat{E}_n[m(Z_i, \theta) | X_i = \hat{x}_0^j(\theta)] \cdot p_j(\theta, \tau_n) + C_i \hat{u}_{ji}$$

where

$$p_j(\theta, \tau_n) = \mathbf{1} \left\{ \frac{\hat{E}_n[m(Z_i, \theta) | X_i = \hat{x}_0^j(\theta)]}{s_j(\hat{x}_0^j(\theta), \theta)} \geq -\tau_n \right\}$$

$$4. \text{ Compute } \tilde{T}_{r,n}^*(\theta) = \max \left\{ \frac{\left[ \max_{x \in \mathcal{X}} \hat{E}_n[Y_{ij}^*(\theta) | X_i = x] \right]_+}{s_j^*(\hat{x}_0^j(\theta), \theta)}; j \in J \right\}$$

5. Repeat steps 1-3  $R$  times. Calculate  $\hat{q}_{n,1-\alpha}^{wb}$  as  $(1 - \alpha)$  quantile of empirical distribution of  $\{\tilde{T}_{r,n}^*(\theta), r = 1, \dots, R\}$ . Define  $C_{n,1-\alpha}^{wb} = \left\{ \theta \in \Theta : \tilde{T}_n(\theta) \leq \hat{q}_{1-\alpha}^{wb}(\theta) \right\}$

Unlike the procedure based on the asymptotic approximation, the wild bootstrap procedure does not require to estimate the whole matrix  $\Sigma(\theta)$ , just its diagonal elements. The covariance structure is preserved by sampling the whole vector of residuals. To sum up, in this section I considered the case when set of zeros  $V_0^j(\theta)$  is a singleton and this allowed us to estimate it directly, and not the superset of it. Next section presents another example where we can estimate the set of zeros directly.

### 3.3 Parametric conditional moment restrictions

Here I consider the case when the researcher is willing to assume a parametric form of conditional moment constraints:  $E(m_j(Z_i, \theta)|X_i = x) = M_j(x, \theta, \beta)$ . Many of the parametric specifications for conditional moment functions that can arise in econometric models are *separated in the parameters*  $\theta$  and  $\beta$ , i.e.

$$M_j(x, \theta, \beta) = h_j(x, \theta) + q_j(x, \beta) \quad (3.17)$$

and here I will focus only on such models. For example, in the context of the interval outcome data, one can assume that:

$$\begin{aligned} E(Y_0|X = x) &= q_0(x\beta_0); \\ E(Y_1|X = x) &= q_1(x\beta_1); \\ E(Y|X = x) &= x\theta \end{aligned}$$

and the identified set is defined as  $\Theta_I = \{\theta \in \Theta : q_0(x\beta_0) - x\theta \leq 0 \text{ and } x\theta - q_1(x\beta_1) \leq 0 \text{ for all } x \in \mathcal{X}\}$ . Both  $q_0$  and  $q_1$  are known functions, and under some regularity assumptions  $\beta_0$  and  $\beta_1$  can be estimated at parametric rate.

As in the previous section, the inference method will be based on the approximation to the distribution of the statistic

$$T_n(\theta) = \max \left\{ \max_{x \in \mathcal{X}} \frac{[\widehat{E}_n[m_j(Z_i, \theta)|X_i = x]]_+}{s_j(x, \theta)}; j \in J \right\} \quad (3.18)$$

only in this case estimators of conditional moment constraints are  $\sqrt{n}$  uniformly consistent on  $\mathcal{X}$ . The following theorem provides a basis for an asymptotic approximation to the distribution of  $\widetilde{T}_n(\theta)$ , with the second claim following directly from Theorem 3.1 in Shapiro (2000).

**Theorem 3.5** *Let  $\sqrt{n}(\widehat{\beta}_n - \beta) \xrightarrow{d} N(0, \Sigma_\beta)$ . Suppose that  $B$  is a compact set such that  $\beta \in B$ , and for any  $b \in B$ , function  $f(x, b)$  is twice continuously differentiable on  $\mathcal{X}$ . Then stochastic process  $Z_n(x) = \sqrt{n}(f(x, \widehat{\beta}) - f(x, \beta))$  converges in the space of all continuous functions on  $\mathcal{X}$  to a continuous Gaussian process  $Z_\infty(x)$  with a known*

covariance function:

$$Z_n(x) = Z_\infty(x) + o_p(1)$$

Moreover, if  $V(\beta) = \arg \max_{x \in \mathcal{X}} f(x, \beta)$ , then

$$\sqrt{n}(\max_{x \in \mathcal{X}} f(x, \hat{\beta}_n) - \max_{x \in \mathcal{X}} f(x, \beta)) \xrightarrow{d} \max_{x \in V(\beta)} Z_\infty(x)$$

Similar result is also obtained in Chernozhukov, Lee, and Rosen (2009). Based on this result, I propose the following procedure given  $\hat{\beta}$  - a consistent estimator of  $\beta$  in (3.17) such that  $\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \Sigma_\beta)$ .

1. Let  $\tau_n = (2 \log \log n)^{1/2}$ .
2. Draw a sample  $\{Z_r, r = 1, \dots, R\}$  from distribution  $N(0, \Sigma_\beta)$ . Define  $w(x) = \left( \frac{\partial q_1(x, \hat{\beta})}{\partial \beta}, \frac{\partial q_2(x, \hat{\beta})}{\partial \beta} \right)'$  and  $\nu_r(x) = w(x)Z_r$ .
3. Let  $\hat{V}_0^j(\theta) = \{x \in \mathcal{X} : h_j(x, \theta) + q_j(x, \hat{\beta}) > -\tau_n \max_{x \in \mathcal{X}} s_j(x)\}$ , where  $s_j(x) = (1/n)^{1/2} (d(x)\Sigma_\beta d(x)')_{jj}$ .
4. Calculate  $\Lambda_r(\theta) = \max_{j \in J} \max_{x \in \hat{V}_0^j(\theta)} \nu_{r,j}(x)$
5. Define  $\hat{q}_{1-\alpha}^G(\theta)$  as  $(1 - \alpha)$  quantile of the empirical distribution of  $\Lambda_r(\theta)$ .
6. Calculate  $C_{n,1-\alpha} = \{\theta \in \Theta : T_n(\theta) \leq \hat{q}_{1-\alpha}^G(\theta)\}$ .

### 3.4 Further Issues

**Uniform validity:** Although uniform validity of the inference procedure is beyond the scope of this paper, it seems plausible that under some restrictions on the family of data generating processes (DGP)  $\mathcal{P}$  the proposed bootstrap procedure is uniformly consistent. For example, Kim (2009) also requires to impose certain restrictions on the family of the DGP for the subsampling to be asymptotically valid. In the context of shape restrictions, one can apply result if Andrews and Guggenberger (2007) to show that the asymptotic approximation is uniformly valid if we restrict the family of DGP to have the Jacobian matrices of conditional moment functions evaluated at the argmax to be bounded away from zero.

**Which procedure to choose:** In general, it might be advisable to use both inference procedures. While the bootstrap procedure may lead to a conservative coverage in small samples if the conditional moment functions are non-flat, it is always consistent. On the other hand, the procedure based on the shape restrictions of conditional moment functions leads to undercoverage if these restrictions are violated, it performs accurately in small samples when shape restrictions are indeed true. Doing both procedures will provide the researcher with a benchmark on how wide might be the confidence set for the data at hand.

**Finite vs infinite number of moment inequalities:** As it was discussed above, if a researcher is not willing to lose any information by transforming a model into a finite set of unconditional moment inequalities, then he cannot hope to achieve a parametric convergence rate that he has in the case of a finite set of unconditional moment inequalities. In other words, confidence sets based on a finite number of moments may be wide due to the loss of information, and the confidence sets based on an infinite number of conditional moment inequalities may be wide due to the slow rate of convergence of the nonparametric estimator. Menzel (2009) also points out to this trade-off between the convergence rate and the number of moment inequalities. It seems that in the cases when conditional moment functions are highly correlated across  $j$ , one might lose a lot of information by converting the model into a finite number of conditional moment inequalities. On the other hand, if the degree of dependence among the conditional moment functions across  $j$  is low, it might be preferable to transform the model into a finite number of inequalities and use the advantage of the fast parametric convergence rate in this case.

## 4 Monte Carlo Simulations

In this section, I use the setup of Example 1 (parametric regression with interval outcome) to evaluate finite sample performance of the inferential methods proposed in the paper. In particular, I assume that  $X \in \mathbb{R}$ , so that  $d = 1$ , and the conditional mean of  $Y|X = x$  is a linear function of  $x$ :

$$E[Y|X = x] = \beta_0 + \beta_1 x.$$

The parameter vector  $\beta = (\beta_0, \beta_1)$  is partially identified since only  $([Y_0, Y_1], X)$  are observed, with  $Pr\{Y \in [Y_0, Y_1]\} = 1$  and  $Y_0$  is strictly smaller than  $Y_1$ . To illustrate the differences in inference with curved constraints (unique global maximum) vs flat constraints, I use the following designs for conditional distributions of interval outcome,  $[Y_0, Y_1]$ .

**Design 1:**  $X$  is uniformly distributed on  $[-0.28, 1.12]$ . Given  $X = x$ ,  $Y_0$  and  $Y_1$  are generated as  $Y_0 = 5x - 4(x - 1/2)^2 + \varepsilon$  and  $Y_1 = 2 + 5x + 4(x - 1/2)^2 + \varepsilon$ , where  $\varepsilon \sim N(0, 1)$  and is independent of  $X$ . The identified set is given by

$$\Theta_I = \{\beta \in \Theta : E[Y_0|X = x] \leq \beta_0 + \beta_1 x \leq E[Y_1|X = x] \text{ for any } x \in [0, 1]\}. \quad (4.1)$$

Figure 1.

**Design 2:**  $X$  is uniformly distributed on  $[0, 1]$ . Given  $X = x$ ,  $Y_0$  and  $Y_1$  are generated as  $Y_0 = 1/2 + 3x + \varepsilon$  and  $Y_1 = 2 + 7x + \varepsilon$ , where  $\varepsilon \sim N(0, 1)$  and is independent of  $X$ . The identified set in this case is also given by (4.1). Figure 2.

The confidence sets proposed in the paper are based on pointwise testing procedure. Therefore, to evaluate the performance of the inference procedures proposed in the paper, I first need to choose a number of points in the parameter space. Corollaries 3.4, 3.6 and 3.10 suggest that any point in the identified set will be covered with a probability that is no less than the desired coverage level  $1 - \alpha$ . In particular, for any point in the interior of the identified set, the coverage probability converges to one, while for some (or all) points on the boundary of the identified set, the coverage probability converges to  $1 - \alpha$ . Finally, the coverage probability for points outside the identified set converges to zero. Therefore, for each design I present simulation results for the set of parameter values that include points on the boundary of the identified set, points in the interior of the identified set, and points that do not belong to the identified set. In both designs, only one out of two constraints can be binding, so we do not need to estimate the number of binding constraints.

Finally, for both designs, I consider the number of observation  $n = \{100, 500, 1000\}$  and confidence level  $1 - \alpha = \{0.75, 0.85, 0.95\}$ . I take independent draws of  $x_i$

from the corresponding uniform distribution, and  $\varepsilon_i \sim N(0, 1)$  to simulated the data  $\{([y_{0,i}, y_{1,i}], x_i)\}_{i=1}^n$ . The number of Monte Carlo replications is 1000, number of bootstrap samples is 500.

## 4.1 Design 1

In the setup of the Design 1, corresponding conditional moment inequalities have a unique global maximum. Therefore, I use inferential procedure described in Section 3.1, which is based on the Normal limiting distribution. In order to estimate conditional moment constraints, the bandwidth  $h = 1.7n^{-1/6.5}$  and a second order kernel  $K(u) = (15/32)(7u^4 - 10u^2 + 3)I(|u| \leq 1)$  are used. Table 4.1 presents empirical coverage probabilities for different points in the parameter space. Here  $\beta = (-3/4, 7)$  belongs to the boundary of the identified set, while  $\beta = (-1/2, 7)$  is a point in the interior of the identified set. Points  $\beta = (-1, 7)$  and  $\beta = (-5/4, 5)$  do not belong to the identified set, but while the line  $y = -1 + 7x$  crosses the lower conditional moment constraint  $E[Y_0|x] = 5x - 4(x - 1/2)^2$ , the line  $y = -5/4 + 5x$  lies completely outside the area between upper and lower conditional moment constraints. The parameter value  $\beta = (-1, 9)$  belongs to the boundary of the identified set. However, in a contrast to assumption (S1), the distance between the line  $y = -1 + 9x$  and the lower conditional moment constraint is minimized (and equal to zero) at the boundary of the set  $[0, 1]$ . Still, as one can see from Table 4.1, the proposed inferential procedure still provides satisfactory results in this case.

**Table 4.1:** Coverage probability for  $\beta$

Coverage	0.75			0.85			0.95		
Sample size	100	500	1000	100	500	1000	100	500	1000
Asymptotic approximation									
$\beta = (-3/4, 7)$	0.8870	0.8255	0.7640	0.9305	0.8965	0.8425	0.9700	0.9625	0.9435
$\beta = (-1/2, 7)$	0.9965	1.0000	1.0000	0.9985	1.0000	1.0000	1.0000	1.0000	1.0000
$\beta = (-1, 7)$	0.2870	0.0080	0.0000	0.4120	0.0145	0.0005	0.6185	0.0580	0.0010
$\beta = (-1, 9)$	0.5705	0.6645	0.6985	0.6775	0.7865	0.8280	0.8175	0.9215	0.9405
$\beta = (-5/4, 5)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Subsampling approximation, $\beta = (-3/4, 7)$									
$a_n = n/5$	0.4605	0.3290	0.5415	0.6150	0.3840	0.7035	0.6515	0.4460	0.8650
$a_n = n/7$	0.5495	0.2350	0.4005	0.5910	0.1965	0.5095	0.5985	0.3235	0.6260
$a_n = n/10$	0.5370	0.2120	0.2335	0.5345	0.1925	0.2295	0.5290	0.5260	0.2990

## 4.2 Design 2

In the setup of Design 2, conditional moment constraints are linear functions of  $x$ , and therefore there is a continuum of points that maximize conditional moment inequalities. Hence, the inference procedure for this case is based on bootstrap approximation. To estimate conditional moment constraints, I use the bandwidth  $h = n^{-1/4.5}$  and kernel function  $K(u) = (15/16)(1 - u^2)I(|u| \leq 1)$ . Table 4.2 shows the empirical coverage for a set of point in the parameter space. In particular, points  $\beta = (2, 7)$  and  $\beta = (1/2, 3)$  belong to the boundary of the identified set with constraints are binding everywhere on  $[0, 1]$ . Point  $\beta = (2, 5)$  also belongs to the boundary of the identified set, but constraints are binding in a single point on the boundary of  $[0, 1]$ . Finally, points  $\beta = (-1/2, 8)$  and  $\beta = (2, 15/2)$  do not belong to the identified set. The difference is that the line with  $y = -1/2 + 8x$  crosses the area between upper and lower constraints, while the line  $y = 2 + (15/2)x$  lies completely outside this area.

**Table 4.2:** Coverage probability for  $\beta$

Coverage	0.75			0.85			0.95		
Sample size	100	500	1000	100	500	1000	100	500	1000
Bootstrap approximation									
$\beta = (2, 7)$	0.6645	0.6150	0.6160	0.8230	0.7640	0.7615	0.9630	0.9440	0.9445
$\beta = (1/2, 3)$	0.5385	0.5940	0.6310	0.7235	0.7575	0.7915	0.9430	0.9410	0.9485
$\beta = (2, 5)$	0.9915	0.9990	1.0000	0.9970	1.0000	1.0000	0.9995	1.0000	1.0000
$\beta = (1, 5)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\beta = (-1/2, 8)$	0.2220	0.0000	0.0000	0.3695	0.0003	0.0000	0.6925	0.0050	0.0000
$\beta = (2, 15/2)$	0.0160	0.0000	0.0000	0.0460	0.0000	0.0000	0.2315	0.0000	0.0000
Subsampling approximation, $\beta = (2, 7)$									
$m = n/5$	0.8310	0.7480	0.7380	0.9470	0.8920	0.8795	0.9950	0.9860	0.9830
$m = n/7$	0.8440	0.7365	0.7165	0.9525	0.8735	0.8620	0.9970	0.9800	0.9765
$m = n/10$	0.8650	0.7255	0.6625	0.9715	0.8655	0.8365	1.0000	0.9835	0.9685
$\beta = (5/4, 2)$	0.9215	0.9575	0.9820	0.9665	0.9760	0.9800	0.9960	0.9970	0.9970

Finally, Table 4.3 illustrates “misspecified” case, when inference procedure that is based on the assumption that the conditional moment constraints can have flat parts is applied to the data generated by the model with non-flat constraints (Design 1), and when inference procedure that is based on the assumption that the conditional moment constraints can have unique binding points is applied to the data generated by the model with flat constraints (Design 2). Here we see that bootstrap approximation over-covers in small samples if constraints are non-flat, while the asymptotic approximation

severely undercovers when constraints are flat. Moreover, in the case of asymptotic approximation with flat constraints leads to inconsistent inference asymptotically.

**Table 4.3:** "Misspecified" models

Coverage	0.75			0.85			0.95		
Sample size	100	500	1000	100	500	1000	100	500	1000
Design 2, but normal approximation is used.									
$\beta = (2, 7)$	0.4120	0.3700	0.3220	0.5885	0.5395	0.5165	0.8205	0.7975	0.7780
Design 1, but bootstrap approximation is used									
$\beta = (-3/4, 7)$	0.9360	0.9640	0.9820	0.9890	0.9960	0.9940	0.9960	0.9970	0.9970

## 5 Conclusion

In this paper I present an approach to constructing confidence sets for each element of partially identified parameters defined by the finite number of conditional moment inequalities. The paper goes beyond the assumption that the conditioning covariates are discrete, and introduces inference procedure applicable in the case where the identified set of the parameter values is defined by a number of conditional moment inequalities with continuously distributed covariates. The inference procedure is based on the supremum statistic for a Nadaraya-Watson estimator of the conditional moment functions that define the set of constraints.

I consider three main scenarios and provide a way to construct confidence sets in each of the scenarios: no restrictions on the conditional moment functions; conditional moment functions have unique maximums; conditional moment functions are specified parametrically. Using Monte Carlo simulations I show that in small samples ignoring the shape of the constraints may lead to a conservative coverage for each element of the identified set. I also show that in small samples confidence sets based on subsampling approximation may result in significant undercoverage when conditional moment functions on the boundary of the identified set have peaks. If the researcher is willing to assume that conditional moment functions have unique maximums for the parameter values on the boundary of the identified set, I show that the problem essentially reduces to the problem with finite number of inequalities, evaluated at certain point (estimates of the location of the maximum) and one can use an analog of the Generalized Moment Selection method of Andrews and Soares (2007) to construct confidence sets. Finally, using Monte Carlo simulations, I show that in small samples there is a trade-off between

coverage and assumptions about the shape of the constraints.

The findings of this paper leave open several questions. Our Monte Carlo exercise shows that in small samples the coverage of confidence sets based on a subsampling approximation is greatly affected by the shape of conditional moment functions that define the constraints. This finding can be explained by an observation that if conditional moment functions have peaks, the behavior of the supremum statistic will be mostly determined by the observations in the small neighborhood of the peak. Therefore, with small sample sizes subsampling approximation may fail to capture this behavior due to insufficient number of observation around the location of the peak. This observation is also relevant for the confidence sets proposed in Kim (2008). It would be of interest to see to what extent the approach proposed by Kim suffers the same drawback.

Another interesting question would be to investigate to what extent transforming conditional inequalities into a finite number of unconditional inequalities affects the trade-off between the size of the identified set in this case and the size of the confidence sets. In general, when transforming conditional inequalities into unconditional ones we lose some information, which leads to a larger identified set. However, in small samples confidence set for the “unconditional” identified set based on parametric estimation of unconditional moments can be expected to be tighter than confidence sets for the “conditional” identified set based on nonparametric estimation of conditional moment functions.

## 6 Appendix

Notation: let  $\{(Y_{1i}, Y_{2i}, X_i), i = 1, \dots, n\}$  be an i.i.d. sample. Define

$f(x)$  marginal density of  $X$

$\mu_j(x) = E(Y_j|X = x), j \in J$

$\hat{\mu}_j(x)$  estimator of  $\mu_j(x)$

$\sigma_j^2(x) = E((Y_j - \mu_j(x))^2|X = x)$

$\hat{\sigma}_j^2(x)$  estimator of  $\sigma_j^2(x)$

$x_0^j = \arg \max_{x \in \mathcal{X}} \mu_j(x)$

$$\hat{x}_0^j = \arg \max_{x \in \mathcal{X}} \hat{\mu}_j(x)$$

$$c_K = \int K^2(u) du$$

$$K_h(u) = \frac{1}{h} K(u/h)$$

**Proof of Lemma 3.1.** For any fixed  $\theta \in \Theta_I$ , any  $j \in J$  and any constant  $q$ , the event  $A_1 = \left\{ \max_{x \in V^j} \frac{\hat{E}_n[m_j(Z, \theta)|X=x] - E[m_j(Z, \theta)|X=x]}{s_j(x, \theta)} \leq c \right\}$  implies  $A_2 = \left\{ \max_{x \in \mathcal{X}} \frac{\hat{E}_n[m_j(Z, \theta)|X=x]}{s_j(x, \theta)} \leq c \right\}$  since if  $\theta \in \Theta_I$ , then  $E[m_j(Z, \theta)|X = x] \leq 0$ . Therefore, for any  $q$  the event  $A'_1 = \left\{ \max_{j \in J} \max_{x \in V^j} \frac{\hat{E}_n[m_j(Z, \theta)|X=x] - E[m_j(Z, \theta)|X=x]}{s_j(x; \theta)} \leq q \right\}$  implies event  $A'_2 = \{T_n(\theta) \leq q\}$ . So, for any  $\theta \in \Theta_I$ ,

$$P\{T_n(\theta) \leq q_{n, 1-\alpha}\} \geq P\left\{ \max_{j \in J} \max_{x \in V^j} \frac{\hat{E}_n[m_j(Z, \theta)|X = x] - E[m_j(Z, \theta)|X = x]}{s_j(x; \theta)} \leq q_{n, 1-\alpha} \right\},$$

where the latter probability is equal to  $(1 - \alpha)$  by definition of the quantity  $q_{n, 1-\alpha}$ . This concludes the proof.  $\square$

**Proof of Theorem 3.2.** Suppose first that the collection of sets  $\{\hat{V}_0^j(\theta), j \in J\}$  is known. For any  $k$ -dimensional vector  $t$ , consider  $W_n(x, \theta) = \sum_{j \in J} t_j \frac{\hat{E}_n[m_j(Z_i; \theta)|X_i=x] - E[m_j(Z_i; \theta)|X_i=x]}{s_j(x; \theta)}$ . We can use lemmas A.2-A.6 in Johnston (1982) to show that under the assumptions of Theorem 3.2 there exists a sequence of stationary mean zero Gaussian processes with continuous sample paths  $\{\Gamma_n(x, \theta), n \in \mathbb{N}\}$  such that  $\sup_{x \in \mathcal{X}} |W_n(x, \theta) - \Gamma_n(x, \theta)| = o_p((\log n)^{-1/2})$ . Since the choice of vector  $t$  was arbitrary, there exists a sequence of  $k$ -dimensional stationary mean-zero Gaussian processes  $\{\tilde{\Gamma}_n(x, \theta), n \in \mathbb{N}\}$  such that

$$\sup_{x \in \mathcal{X}, j \in J} \left\| \frac{\hat{E}_n[m_j(Z_i; \theta)|X_i = x] - E[m_j(Z_i; \theta)|X_i = x]}{s_j(x; \theta)} - \tilde{\Gamma}_{j,n}(x, \theta) \right\| = o_p((\log n)^{-1/2})$$

Observe that the functional  $M(g) = \max_{x \in V^j} g_j(x), j \in J$  is Lipschitz continuous on  $\mathcal{C}(\mathcal{X})$ . Therefore, we can approximate the distribution of

$$Q_n(\theta) = \max_{j \in J} \max_{x \in V_0^j(\theta)} \frac{\hat{E}_n[m_j(Z, \theta)|X = x] - E[m_j(Z, \theta)|X = x]}{s_j(x; \theta)}$$

by the distribution of  $M\left(\tilde{\Gamma}_n(\cdot, \theta)\right) = \max\left\{\max_{x \in V_0^j(\theta)} \Gamma_{j,n}(x, \theta), j \in J\right\}$ . In the case  $k = 1$  the distribution of this random variable is known to converge to the extreme value type I distribution, although the rate of convergence is slow (see Hall 1979, 1991). We can use result in Hall (1993) to show that one can approximate distribution of  $M(\Gamma_n(\cdot, \theta))$  by the bootstrap distribution. I will now prove that  $P\left\{\liminf_{n \rightarrow \infty} \left\{V_\varepsilon^j(\theta) \subseteq \widehat{V}_0^j(\theta)\right\}\right\} = 1$ ,

where  $V_\varepsilon^j(\theta) = \{x \in \mathcal{X} : |E[m_j(Z, \theta)|X = x]| \leq \varepsilon\}$ .

The event  $\left\{\max_{x \in V_0^j(\theta)} \frac{|\widehat{E}_n[m_j(Z, \theta)|X = x] - 0|}{s_j(x, \theta)} < \frac{\varepsilon}{\min_{x \in V_0^j(\theta)} s_j(x, \theta)} + \tau_n\right\}$  implies the event  $\left\{\max_{x \in V_0^j(\theta)} \frac{|\widehat{E}_n[m_j(Z, \theta)|X = x] - \varepsilon|}{s_j(x, \theta)} < \tau_n\right\}$ , which in turn implies the event  $\left\{V_0^j(\theta) \subset \widehat{V}_\varepsilon^j(\theta)\right\}$ .

By a law of iterated logarithms for kernel regression estimator, we have

$$P\left\{\liminf_{n \rightarrow \infty} \left\{\max_{x \in V_0^j(\theta)} \frac{|\widehat{E}_n[m_j(Z, \theta)|X = x] - 0|}{s_j(x, \theta)} < \frac{\varepsilon}{\min_{x \in V_0^j(\theta)} s_j(x, \theta)} + \tau_n\right\}\right\} = 1$$

Therefore,  $P\left\{\liminf_{n \rightarrow \infty} \left\{V_\varepsilon^j(\theta) \subseteq \widehat{V}_0^j(\theta)\right\}\right\} = 1$  and  $V_0^j(\theta) \subset V_\varepsilon^j(\theta)$ . Therefore, we can use  $\widehat{V}_0^j(\theta)$  instead of  $V_0^j(\theta)$  in the bootstrap procedure, and because  $V_0^j(\theta) \subset V_\varepsilon^j(\theta)$ , by Lemma 3.1 bootstrap distribution asymptotically stochastically dominates distribution of  $M(\Gamma_n(\theta))$ .  $\square$

**Proof of Theorem 3.3.** To prove this theorem, I will first introduce two auxiliary lemmas.

**Lemma 6.1** *Suppose that*

1.  $nh_n^5 \log n \rightarrow 0$  and  $nh_n \rightarrow \infty$  as  $n \rightarrow \infty$ ;
2.  $K$  is a symmetric, twice continuously differentiable kernel that vanishes outside the interval  $[-1, 1]$ ;
3.  $\mu_1(x)$ ,  $\mu_2(x)$ ,  $\sigma_1^2(x)$ ,  $\sigma_2^2(x)$  and  $f(x)$  are twice continuously differentiable everywhere on  $\mathcal{X}$
4.  $E[|Y_j - \mu_j(x)|^{2+p}|X = x] \leq C_j < \infty$  for some  $p > 0$

5.  $f(x) > 0$  everywhere on  $\mathcal{X}$

Let  $x', x''$  be two distinct points in  $\mathcal{X}$ . Then

$$\sqrt{nh_n} \begin{pmatrix} \hat{\mu}_1(x') - \mu_1(x') \\ \hat{\mu}_2(x') - \mu_2(x') \\ \hat{\mu}_1(x'') - \mu_1(x'') \\ \hat{\mu}_2(x'') - \mu_2(x'') \end{pmatrix} \xrightarrow{d} N(0, \Sigma),$$

where  $\Sigma$  is a symmetric  $4 \times 4$  matrix with

$$\Sigma_{11} = c_K \sigma_1^2(x')/f(x'), \quad \Sigma_{22} = c_K \sigma_2^2(x')/f(x');$$

$$\Sigma_{33} = c_K \sigma_1^2(x'')/f(x''), \quad \Sigma_{44} = c_K \sigma_2^2(x'')/f(x'');$$

$$\Sigma_{12} = c_K (E[Y_1 \cdot Y_2 | X = x'] - \mu_1(x')\mu_2(x'))/f(x');$$

$$\Sigma_{13} = \Sigma_{14} = \Sigma_{23} = 0;$$

$$\Sigma_{34} = c_K (E[Y_1 \cdot Y_2 | X = x''] - \mu_1(x'')\mu_2(x''))/f(x'').$$

Finally, if  $K$  is a third order kernel, then we can obtain the above asymptotic result under  $nh_n^7 \rightarrow \lambda^2 \geq 0$  and  $nh_n^5 \rightarrow \infty$  as  $n \rightarrow \infty$ .

**Proof:** To prove joint asymptotic normality, I will employ the Cramér-Wold device. Therefore, let  $t = (t_1, t_2, t_3, t_4)$  be an arbitrary vector. I will show that random variable

$$\begin{aligned} \xi_n = & t_1(\hat{\mu}_1(x') - \mu_1(x')) + t_2(\hat{\mu}_2(x') - \mu_2(x')) \\ & + t_3(\hat{\mu}_1(x'') - \mu_1(x'')) + t_4(\hat{\mu}_2(x'') - \mu_2(x'')) \end{aligned} \quad (6.1)$$

is asymptotically normally distributed with zero mean.

Following Härdle (1992), for a fixed  $x$  define  $\hat{r}_j(x) = \hat{\mu}_j(x)\hat{f}(x)$ . It can be written as  $\hat{r}_j(x) = n^{-1} \sum_{i=1}^n Y_{ji}K_{h_n}(x - X_i)$ . For any  $x$ , we have

$$(\hat{\mu}_j(x) - \mu_j(x)) - \frac{\hat{r}_j(x) - \mu_j(x)\hat{f}(x)}{f(x)} = o_p((nh_n)^{-1/2})$$

Define random variable

$$\begin{aligned}\zeta_n = & t_1 \frac{\hat{r}_1(x') - \mu_1(x') \hat{f}'(x)}{f(x')} + t_2 \frac{\hat{r}_2(x'') - \mu_2(x'') \hat{f}(x')}{f(x')} \\ & + t_3 \frac{\hat{r}_1(x'') - \mu_1(x'') \hat{f}(x'')}{f(x'')} + t_4 \frac{\hat{r}_2(x'') - \mu_2(x'') \hat{f}(x'')}{f(x'')} \end{aligned} \quad (6.2)$$

Then  $|(nh_n)^{1/2}(\xi_n - \zeta_n)| = o_p(1)$ .

For each  $x$  and  $j \in J$  have the following expression for  $\frac{\hat{r}_j(x) - \mu_j(x) \hat{f}(x)}{f(x)}$ :

$$\frac{\hat{r}_j(x) - \mu_j(x) \hat{f}(x)}{f(x)} = \frac{1}{f(x)} \frac{1}{n} \sum_{i=1}^n K_{h_n}(x - X_i) (Y_{ji} - \mu_j(x))$$

For each  $x \in \{x', x''\}$  and  $j \in J$ , the bias  $E\left(\frac{\hat{r}_j(x) - \mu_j(x) \hat{f}(x)}{f(x)}\right) = O(h_n^2)$ , so that  $E((nh_n)^{1/2} \zeta_n) = O(nh_n^5)$ . Therefore,  $E((nh_n)^{1/2} \zeta_n) = o(1)$  by condition (1) of the lemma. Condition (4) of the lemma ensures that Lindeberg condition is satisfied for the triangular array  $\{(nh_n)^{1/2} \zeta_n\}$ , which allows to apply Liapunov's central limit theorem to  $\{(nh_n)^{1/2} \zeta_n\}$ . This implies, together with zero asymptotic bias of  $\{(nh_n)^{1/2} \zeta_n\}$ , that  $(nh_n)^{1/2} \xi_n \xrightarrow{d} N(0, \sigma_\xi(t))$  for any vector  $t$ . Thus, the Cramér-Wold device warrants joint normal distribution of random vector  $(nh_n)^{1/2}(\hat{\mu}_1(x') - \mu_1(x'), \hat{\mu}_2(x') - \mu_2(x'), \hat{\mu}_1(x'') - \mu_1(x''), \hat{\mu}_2(x'') - \mu_2(x''))$ . It only remains to derive the asymptotic variance-covariance matrix. I will do this only for  $\Sigma_{12}$ , since the proof for any other element of matrix  $\Sigma$  is similar.

Given joint asymptotic normality,  $\Sigma_{12}$  is equal to

$$\begin{aligned}\Sigma_{12} &= \lim_{n \rightarrow \infty} nh_n E [(\hat{\mu}_1(x') - \mu_1(x'))(\hat{\mu}_2(x') - \mu_2(x'))] \\ &= \lim_{n \rightarrow \infty} \frac{nh_n}{f(x')^2} \left( E \left[ \frac{1}{n} \sum_{i=1}^n K_{h_n}(x' - X_i) (Y_{1i} - \mu(x')) \times \right. \right. \\ &\quad \left. \left. \times \frac{1}{n} \sum_{i=1}^n K_{h_n}(x' - X_i) (Y_{2i} - \mu(x')) \right] + o((nh)^{-1}) \right) \end{aligned} \quad (6.3)$$

Finally,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{nh_n} E \left[ \sum_{i=1}^n K \left( \frac{x' - X_i}{h_n} \right) (Y_{1i} - \mu(x')) \sum_{i=1}^n K \left( \frac{x' - X_i}{h_n} \right) (Y_{2i} - \mu(x')) \right] \\ & = c_K f(x') [E(Y_1 Y_2 | X = x') - \mu_1(x') \mu_2(x')] \end{aligned}$$

The last statement of the lemma follows from the expression of the bias for the higher order kernels. This concludes the proof of the lemma.  $\square$

Next lemma shows consistency and asymptotic normality of kernel estimator of a unique maximum and provides a simplified version of the proof in Ziegler (2000), since here we are not interested in asymptotic normality of the estimator of the location of the maximum. This allows to relax assumptions about the choice of bandwidth. Here for the ease of the presentation I omit index  $j$  in  $\mu_j$ ,  $\sigma_j$  and  $x_0^j$ .

**Lemma 6.2** *Let the equivalent of conditions (A1)-(A3), (R1)-(R3) and (UM1)-(UM3) hold for i.i.d. sample  $\{(Y_i, X_i), i = 1, \dots, n\}$  and  $\mu(x) = E(Y|X = x)$ . Let  $x_0 = \arg \max \mu(x)$  and  $\hat{x}_0 = \arg \max \hat{\mu}(x)$ . Suppose that*

1.  $nh^7 \rightarrow \lambda^2 \geq 0$  and  $nh^5 \rightarrow \infty$  as  $n \rightarrow \infty$ ;
2.  $K$  is a symmetric, twice continuously differentiable third order kernel that vanishes outside the interval  $[-1, 1]$ .

Then

$$\sqrt{nh}(\hat{\mu}(\hat{x}_0) - \mu(x_0)) \xrightarrow{d} N(0, \omega)$$

where  $\omega = c_K(E(Y^2|X = x_0) - \mu^2(x_0))/f(x_0)$ .

**Proof:** We can write

$$\hat{\mu}(\hat{x}_0) - \mu(x_0) = (\hat{\mu}(\hat{x}_0) - \hat{\mu}(x_0)) + (\hat{\mu}(x_0) - \mu(x_0))$$

Lemma 6.1 implies that under our assumptions,  $(nh)^{1/2}(\hat{\mu}(x_0) - \mu(x_0)) \xrightarrow{d} N(0, \omega)$ . It remains to show that  $(nh)^{1/2}(\hat{\mu}(\hat{x}_0) - \hat{\mu}(x_0)) = o_p(1)$ .

Since  $\mathcal{X}$  is compact and  $\mu(\cdot)$  is continuous on  $\mathcal{X}$ , then by Heine-Cantor theorem  $\mu(\cdot)$  is also uniformly continuous on  $\mathcal{X}$ . Together with the assumptions of the lemma

it implies is uniform consistency of kernel estimator  $\hat{\mu}(x)$  on  $\mathcal{X}$  (see Bierens, 1987). Therefore

$$\sup_{x \in \mathcal{X}} |\hat{\mu}(x) - \mu(x)| \xrightarrow{p} 0 \text{ as } n \rightarrow \infty$$

This in turn implies that  $\|\hat{x}_0 - x_0\| \xrightarrow{p} 0$ . So,  $\hat{x}_0$  is a consistent estimator for  $x_0$ .

By Taylor series expansion of  $\hat{\mu}(x_0)$  around  $x = \hat{x}_0$  we have

$$\hat{\mu}(\hat{x}_0) - \hat{\mu}(x_0) = \frac{\partial \hat{\mu}(\hat{x}_0)}{\partial x} (x_0 - \hat{x}_0) + O_p(\|x_0 - \hat{x}_0\|^2),$$

where the last term follows from the fact that Hessian of  $\hat{\mu}(x_0)$  is bounded with probability 1, since as  $nh^5 \rightarrow \infty$  the second derivative of  $\hat{\mu}(x)$  evaluated at  $x = x_0$  is a consistent estimator of the second derivative of  $\mu(x)$  evaluated at  $x = x_0$ . By the construction of  $\hat{x}_0$ , we have  $\frac{\partial \hat{\mu}(\hat{x}_0)}{\partial x} = 0$ , so that  $\hat{\mu}(\hat{x}_0) - \hat{\mu}(x_0) = O_p(\|x_0 - \hat{x}_0\|^2)$ . Therefore, in order to evaluate the term  $(nh)^{1/2}(\hat{\mu}(\hat{x}_0) - \hat{\mu}(x_0))$  we need to find a rate of convergence of  $\hat{x}_0$ .

Exact Taylor series expansion of first order conditions for  $\hat{x}_0$  around  $x = x_0$  gives us

$$0 = \frac{\partial \hat{\mu}(x_n)}{\partial x} = \frac{\partial \hat{\mu}(x_0)}{\partial x} + \frac{\partial^2 \hat{\mu}(\tilde{x})}{\partial x \partial x'} (x_0 - x_n),$$

where  $\tilde{x}$  lies between  $x_0$  and  $\hat{x}_0$  and therefore  $\|\tilde{x} - x_0\| = o_p(1)$ . Provided that  $nh^5 \rightarrow \infty$ , we have

$$\frac{\partial^2 \hat{\mu}(x_0)}{\partial x \partial x'} - \frac{\partial^2 \mu(x_0)}{\partial x \partial x'} = o_p(1)$$

Continuity of the second derivative of  $\mu(\cdot)$  in the neighborhood of  $x_0$  (assumption (UM2)) implies that also

$$\frac{\partial^2 \hat{\mu}(\tilde{x})}{\partial x \partial x'} - \frac{\partial^2 \mu(x_0)}{\partial x \partial x'} = o_p(1)$$

By assumption (UM2),  $\frac{\partial^2 \mu(x_0)}{\partial x \partial x'}$  is negative definite, therefore its inverse exists and we can write

$$(\hat{x}_0 - x_0) = \left( \left[ \frac{\partial^2 \mu(x_0)}{\partial x \partial x'} \right]^{-1} + o_p(1) \right) \frac{\partial \hat{\mu}(x_0)}{\partial x}$$

Since  $nh^3 \rightarrow 0$ , then  $\frac{\partial \hat{\mu}(x_0)}{\partial x}$  is a consistent estimator of  $\frac{\partial \mu(x_0)}{\partial x}$ . Moreover, since  $nh^7 \rightarrow \lambda^2$ , then  $\left\| \frac{\partial \hat{\mu}(x_0)}{\partial x} - \frac{\partial \mu(x_0)}{\partial x} \right\| = O_p((nh^3)^{-1})$ . By definition,  $x_0$  is the argmax of

$\mu(x)$  and it belongs to the interior of  $\mathcal{X}$ , so by assumption (UM3) we have

$$\frac{\partial \mu(x_0)}{\partial x} = 0 \quad (6.4)$$

This implies that  $\|\hat{x}_0 - x_0\| = O_p((nh^3)^{-1})$ . Therefore,  $(nh)^{1/2}(\hat{\mu}(\hat{x}_0) - \hat{\mu}(x_0)) = O_p\left(\frac{(nh)^{1/2}}{n^2h^6}\right) = O_p(n^{-3/2}h^{-11/2}) = o_p(1)$  since  $nh^5 \rightarrow \infty$ . This completes the proof of the lemma.  $\square$

Theorem 3.2 immediately follows from Lemmas 6.1 and 6.2 and the fact that  $K$  is a third order kernel.  $\square$

Lemma 6.1 also provides a consistent estimator of variance-covariance matrix  $\Sigma(\theta)$ .

**Lemma 6.3** (Estimator of  $\Sigma(\theta)$ ) *For any  $j, l \in \{1, 2\}$ ,  $(\widehat{\Sigma}(\theta))_{jl}$  given by*

$$\begin{aligned} (\widehat{\Sigma}(\theta))_{jl} &= \frac{nh}{\hat{f}(\hat{x}_0^j)\hat{f}(\hat{x}_0^l)} \left( \frac{1}{n} \sum_{i=1}^n K_h(\hat{x}_0^j - X_i)(m_j(Z_i, \theta) - \widehat{E}_n[m_j(Z_i, \theta)|X = \hat{x}_0^j]) \times \right. \\ &\quad \left. \times \left( \frac{1}{n} \sum_{i=1}^n K_h(\hat{x}_0^l - X_i)(m_l(Z_i, \theta) - \widehat{E}_n[m_l(Z_i, \theta)|X = \hat{x}_0^l]) \right) \right) \end{aligned}$$

*is a consistent estimator of  $(\Sigma(\theta))_{jl}$ .*

**Proof of Theorem 3.4.** We need to show that  $G_{r,n}(\theta)$  and  $\widetilde{T}_n(\theta)$  have the same limiting distribution if  $\theta$  belongs to the identified set  $\Theta_I$ . Asymptotic distribution of  $\widetilde{T}_n(\theta)$  follows from Theorem 3.3, so it remains to show that  $G_{r,n}(\theta)$  converges in distribution to the same functional of multivariate Gaussian random variable with variance-covariance matrix  $\Sigma(\theta)$ . This follows from consistency of estimators  $\widehat{E}_n[m_j(Z, \theta)|X = x]$ ,  $\hat{x}_0^j$ ,  $\widehat{\Sigma}(\theta)$ , and a law of iterated logarithm for kernel estimator (see Härdle, 1984), which implies that for any  $\theta \in \Theta_I$  and our choice of a sequence  $\{\tau_n, n = 1, \dots\}$  we have

$$P \left\{ \lim_{n \rightarrow \infty} \mathbf{1} \left\{ \frac{\widehat{E}_n[m(Z_i, \theta)|X_i = \hat{x}_0^j(\theta)]}{s_j(\hat{x}_0^j(\theta), \theta)} \geq -\tau_n \right\} = \mathbf{1} \{E[m_j(Z_i, \theta)|X_i = x_0^j(\theta)] = 0\} \right\} = 1$$

This completes the proof.  $\square$

**Proof of Theorem 3.5.** The second part of the claim follows directly from Theorem 3.1 in Shapiro (2000). It only remains to show that if function  $f(x, \beta)$  satisfies the

conditions of Theorem 3.5 and  $\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \Sigma_\beta)$ , then stochastic process  $Z_n(x) = \sqrt{n}(f(x, \hat{\beta}) - f(x, \beta))$  can be approximated by a stationary Gaussian process  $Z_\infty(x)$  with continuous sample paths. For a fixed  $x$ , Delta method implies that

$$\sqrt{n}(f(x, \hat{\beta}) - f(x, \beta)) = \frac{\partial f(x, b)}{\partial b} \Big|_{b=\beta} N(0, \Sigma_\beta) + o_p(1) \quad (6.5)$$

Define  $Z_\infty(x) = \frac{\partial f(x, b)}{\partial b} \Big|_{b=\beta} N(0, \Sigma_\beta)$ . Then  $Z_\infty(x)$  has continuous sample paths since  $\frac{\partial f(x, b)}{\partial b} \Big|_{b=\beta}$  is continuous on  $\mathcal{X}$  and  $\sup_{x \in \mathcal{X}} |Z_n(x) - Z_\infty(x)| = o_p(1)$ . This concludes the proof.  $\square$

## References

- [1] Andrews, D.W.K., S. Berry, and P. Jia (2004). “Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location,” working paper, Yale University.
- [2] Andrews, D.W.K. and P. Guggenberger (2007). “Validity of Subsampling and Plug-in Asymptotic Inference for Parameters Defined by Moment Inequalities,” working paper, Yale University.
- [3] Andrews, D.W.K. and X. Shi (2009). “Inference for Parameters Defined by Conditional Moment Inequalities,” working paper, Yale University.
- [4] Andrews, D.W.K. and G. Soares (2007). “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” working paper, Yale University.
- [5] Beresteanu, A. and F. Molinari (2007). “Asymptotic Properties for a Class of Partially Identified Models,” working paper, Cornell University
- [6] Berman, S.M. (1980). “A Compound Poisson Limit for Stationary Sums, and Sojourns of Gaussian Processes,” *The Annals of Probability*, 8(3), 511-538.
- [7] Bierens, H.J. (1987). “Kernel Estimators of Regression Functions,” Ch. 3 in T.F. Bewley (ed.), *Advances in Econometrics*, Cambridge University Press, Vol. I, 99-144.

- [8] Birke, M., and H. Dette (2007). “Testing Strict Monotonicity in Nonparametric Regression,” *Mathematical Methods of Statistics*, 16(2), 110-123.
- [9] Bowman, A.V., M. C. Jones, and I. Gijbels (1998). “Testing monotonicity of regression,” *Journal of Computational and Graphical Statistics*, 7, 489-500.
- [10] Bugni, F. (2007). “Bootstrap Inference for Partially Identified Models,” Working paper, Northwestern University.
- [11] Canay, I. (2007). “EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity,” Working paper, University of Wisconsin - Madison.
- [12] Chernozhukov, V., H. Hong, and E. Tamer (2007). “Parameter Set Inference in a Class of Econometric Models. *Econometrica*, 75(5), 1243-1284.
- [13] Chernozhukov, V., S. Lee, and A.M. Rosen (2009). “Intersection Bounds: Estimation and Inference,” CEMMAP working paper CWP19/09.
- [14] Dümbgen, L. and V.G. Spokoiny. “Multiscale Testing of Qualitative Hypotheses,” *The Annals of Statistics*, 29(1), 124-152.
- [15] Fan, Y. and S. Park (2007). “Confidence Sets for Some Partially Identified Parameters,” working paper, Vanderbilt University.
- [16] Galichon, A. and M. Henry (2006). “Dilation Bootstrap: A Natural Approach to Inference in Incomplete Models,” working paper.
- [17] Ghosal, A., A. Sen, and A.W. van der Vaart (2000). “Testing Monotonicity of Regression,” *The Annals of Statistics*, 28(4), 1054-1082.
- [18] Hall, P. (1979) “On the Rate of Convergence of Normal Extremes,” *Journal of Applied Probability*, 16, 433-439.
- [19] Hall, P. (1991). “On Convergence Rates of Suprema,” *Probability Theory Related Fields*, 89, 447-455.
- [20] Hall, P. (1993). “On Edgeworth Expansion and Bootstrap Confidence Bands in Nonparametric Curve Estimation,” *Journal of the Royal Statistical Society, Series B (Methodological)*, 55(1), 291-304.
- [21] Härdle, W. (1984). “A Law of the Iterated Logarithm for Nonparametric Regression Function Estimators,” *The Annals of Statistics*, 12(2), 624-635.

- [22] Härdle, W. (1992). *Applied Nonparametric Regression*. New York: Cambridge University Press.
- [23] Horowitz, J.L. and C.F. Manski (2000). “Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data,” *Journal of the American Statistical Association*, 95(449), 77-84.
- [24] Imbens, G.W. and C.F. Manski (2004). “Confidence Intervals for Partially Identified Parameters,” *Econometrica*, 72(6), 1845-1857.
- [25] Johnston, G.J. (1982). “Probabilities of Maximal Deviations for Nonparametric Regression Function Estimates,” *Journal of Multivariate Analysis*, 12, 402-414.
- [26] Juditsky, A. and A. Nemirovski (2002). “On Nonparametric Tests of Positivity/Monotonicity/Convexity,” *The Annals of Statistics*, 30(2), 498-527.
- [27] Khan, S. and E. Tamer (2007). “Inference on Randomly Censored Regression Models Using Conditional Moment Inequalities,” working paper.
- [28] Kim, K. (2009). “Set Estimation and Inference with Models Characterized by Conditional Moment Inequalities,” working paper, University of Minnesota.
- [29] Konakov, V.D. and V.I. Piterbarg (1984). “On the Convergence Rate of Maximal Deviation Distribution for Kernel Regression Estimates,” *Journal of Multivariate Analysis*, 15, 279-294.
- [30] Lee, S., O. Linton, and Y. Whang (2007). “Testing for Stochastic Monotonicity,” working paper.
- [31] Liao, Y., and W. Jiang (2008). “Bayesian Analysis in Moment Inequality Models,” working paper, Northwestern University.
- [32] Manski, C.F. and E. Tamer (2002). “Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica*, 70(2), 519-546.
- [33] Manski, C.F. (2003). *Partial Identification of Probability Distributions*. Springer-Verlag, New York.
- [34] Menzel, K. (2009) “Estimation and Inference with Many Moment Inequalities,” working paper.
- [35] Moon, H.R. and F. Schorfheide (2007). “A Bayesian Look at Partially-Identified Models,” working paper, University of Pennsylvania.

- [36] Pakes, A., J. Porter, K. Ho, and J. Ishii (2005). “Moment Inequalities and Their Application,” working paper.
- [37] Romano, J.P. and A.M. Shaikh (2006). “Inference for the Identified Set in Partially Identified Econometric Models,” working paper.
- [38] Romano, J.P. and A.M. Shaikh (2007). “Inference for Identifiable Parameters in Partially Identified Econometric Models,” working paper.
- [39] Rosen, A.M. (2006). “Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities,” working paper, University College London.
- [40] Shapiro, A. (2000). “Statistical Inference of Stochastic Optimization Problems,” in S. P. Uryasev (ed.), *Probabilistic Constrained Optimization: Methodology and Applications*, Kluwer Academic Publishers, 282-304.
- [41] Stoye, J. (2007). “More on Confidence Intervals for Partially Identified Parameters,” working paper, NYU.
- [42] Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria, *Review of Economic Studies*, 70(1), 147-165.
- [43] Ullah, A. and A. Pagan (1999). *Nonparametric econometrics*. Cambridge University Press.
- [44] Wu, C.F.Y. (1986). “Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis,” *Annals of Statistics*, 14, 1261-1350.
- [45] Ziegler, K. (2000). “Nonparametric Estimation of Location and Size of Maxima of Regression Functions in the Random Design Case Based on the Nadaraya-Watson Estimator with Data-Dependent Bandwidths,” Habilitation Thesis.
- [46] Ziegler, K. (2003). “On the Asymptotic Normality of Kernel Regression Estimators of the Mode in the Random Design Model,” *Journal of Statistical Planning and Inference*, 115, 123-144.

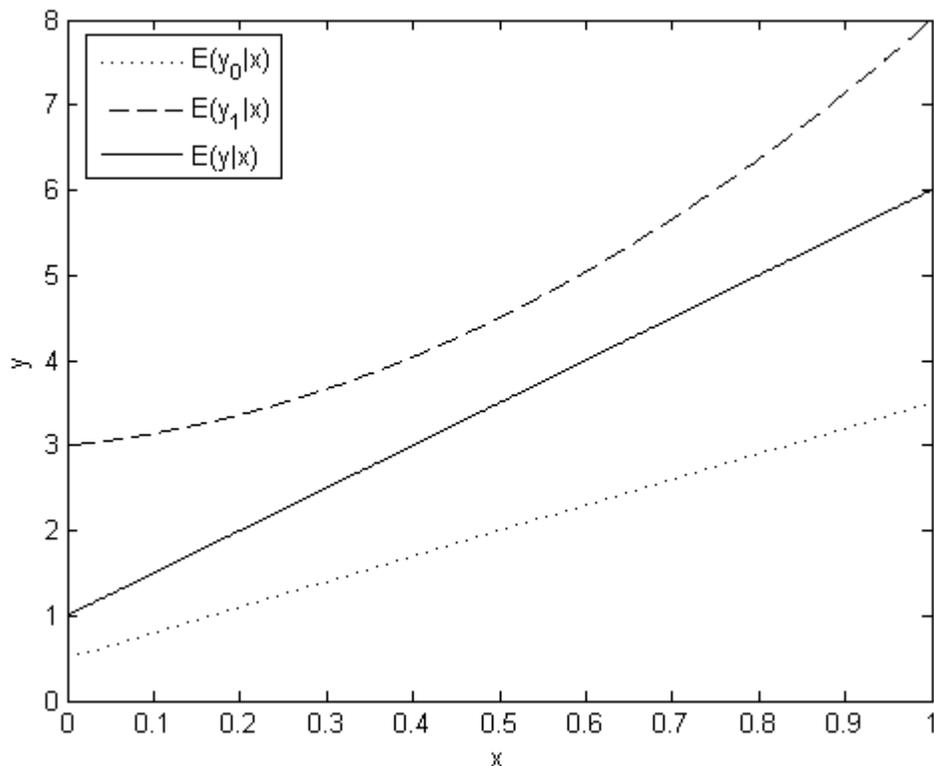


Figure 1: **Both flat and non-flat constraints.** This figure illustrates the case where  $E(Y_1|x)$  is linear in  $x$ , while  $E(Y_0|x)$  is nonlinear. Therefore, there exists some  $\theta'$  in the identified set  $\Theta_I$  such that  $\theta'_0 + \theta'_1 \cdot x - E(Y_1|x) < 0$  and  $E(Y_0|x) - \theta'_0 - \theta'_1 \cdot x = 0$  for all  $x \in [0, 1]$ . There also exist some  $\theta''$  in the identified set  $\Theta_I$  such that  $E(Y_0|x) - \theta''_0 - \theta''_1 \cdot x < 0$  and  $\theta''_0 + \theta''_1 \cdot x - E(Y_1|x) \leq 0$  for any  $x \in [0, 1]$ , and  $\theta''_0 + \theta''_1 \cdot x_0 - E(Y_1|x_0) = 0$  for some  $x_0 \in [0, 1]$ . In this case the bootstrap method will give conservative coverage for  $\theta''$  even though it belongs to the boundary of the identified set.

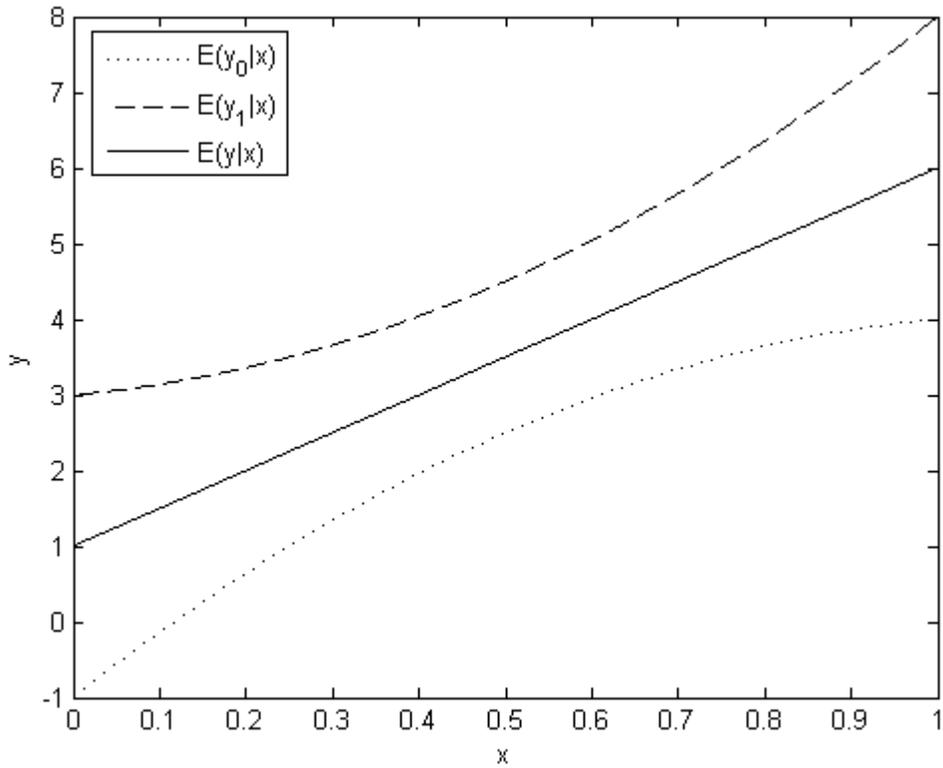


Figure 2: **Both constrains are non-flat.** This figure illustrates the case where both  $E(Y_1|x)$  and  $E(Y_0|x)$  are nonlinear. In this case method based on critical values obtained from the bootstrap approximation gives conservative coverage for any point on the boundary of the identified set. In contrast, method based on the Gaussian approximation provides the coverage that has asymptotically exact size.