Breaking the curse of dimensionality in conditional moment inequalities for discrete choice models

Le-Yu Chen
Sokbae Lee

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP51/17

Breaking the curse of dimensionality in conditional moment inequalities for discrete choice models*

Le-Yu Chen†

Institute of Economics, Academia Sinica

Sokbae Lee‡

Department of Economics, Columbia University

Centre for Microdata Methods and Practice, Institute for Fiscal Studies

19 October 2017

*We are grateful to the editor and two anonymous referees for constructive comments and suggestions. We also thank Hidehiko Ichimura, Kengo Kato, Shakeeb Khan, Toru Kitagawa, Dennis Kristensen, Adam Rosen, Kyungchul Song and participants at SNU Workshop on Advances in Microeconometrics and 2015 Econometric Society World Congress for helpful comments. This work was supported in part by the Ministry of Science and Technology, Taiwan (MOST105-2410-H-001-003-), Academia Sinica (Career Development Award research grant), the National Research Foundation of Korea (NRF-2015S1A5A2A01014041), the European Research Council (ERC-2014-CoG-646917-ROMIA), and the UK Economic and Social Research Council (ESRC) through research grant (ES/P008909/1) to the CeMMAP.

†E-mail: lychen@econ.sinica.edu.tw

‡E-mail: sl3841@columbia.edu

Abstract

This paper studies inference of preference parameters in semiparametric discrete choice models when these parameters are not point-identified and the identified set is characterized by a class of conditional moment inequalities. Exploring the semiparametric modeling restrictions, we show that the identified set can be equivalently formulated by moment inequalities conditional on only two continuous indexing variables. Such formulation holds regardless of the covariate dimension, thereby breaking the curse of dimensionality for nonparametric inference based on the underlying conditional moment inequalities. We further apply this dimension reducing characterization approach to the monotone single index model and to a variety of semiparametric models under which the sign of conditional expectation of a certain transformation of the outcome is the same as that of the indexing variable.

Keywords: *partial identification, conditional moment inequalities, discrete choice, monotone single index model, curse of dimensionality*
JEL codes: C14, C25.

1 Introduction

There has been substantial research carried out on partial identification since the seminal work of Manski. For example, see monographs by Manski (2003, 2007), a recent review by Tamer (2010), and references therein for extensive details. In its general form, identification results are typically expressed as nonparametric bounds via moment inequalities or other similar population quantities. When these unknown population quantities are high-dimensional (e.g. the dimension of covariates is high in conditional moment inequalities), there is a curse of dimensionality problem in that a very large sample is required to achieve good precision in estimation and inference (see, e.g. Chernozhukov, Lee, and Rosen (2013)). In this paper, we propose a method for inference that avoids the curse of dimensionality by exploiting the model structure. We illustrate our idea in the context of commonly used discrete choice models.

To explain this issue, suppose that one is interested in identifying a structural parameter in a binary choice model. In this model, it is quite common to assume that an individual's utility function is parametric while making weak assumptions regarding underlying unobserved heterogeneity. Specifically, consider the following model

$$Y = 1\{X'\beta \geq \varepsilon\}, \tag{1.1}$$

where Y is the binary outcome, X is an observed d dimensional covariate vector, ε is an unobserved random variable, $\beta \in \Gamma$ is a vector of unknown true parameters, and $\Gamma \subset \mathbb{R}^d$ is the parameter space for β.

Without sufficient exogenous variation from covariates, β is only partially identified. The resulting identification region is characterized by expressions involving nonparametric choice probabilities conditional on covariates. For example, under the assumption that the conditional median of ε is independent of X and other regularity conditions that will be given in Section 2, β is partially identified by

$$\Theta = \{b \in \Gamma : X'b\,[P(Y = 1|X) - 0.5] \geq 0 \text{ almost surely}\}. \tag{1.2}$$

Recently, Komarova (2013) and Blevins (2015) use this type of characterization to partially identify β. Both papers consider estimation and inference of the identified set Θ using a maximum score objective function; however, they do not develop inference methods for the parameter value β based on the conditional moment inequalities in

(1.2). Unlike theirs, we focus on inference for β as well as the issue of dimension reduction in the context of conditional moment inequalities.

When X contains several continuous covariates yet their support is not rich enough to ensure point identification, we can, for instance, construct a confidence region for β by inverting the test of Chernozhukov, Lee, and Rosen (2013, henceforth CLR), who plug in nonparametric (kernel or series based) estimators to form one-sided Kolmogorov-Smirnov type statistic for testing the conditional moment inequalities. In order to conduct inference based on the CLR method, we need to estimate conditional expectation $E(Y|X) = P(Y = 1|X)$ nonparametrically. In this context, it is difficult to carry out inference in a fully nonparametric fashion when d is large. One may attempt to use parametric models to fit the choice probabilities. However, that can lead to misspecification which may invalidate the whole partial identification approach. Hence, it is important to develop dimension reduction methods that avoid misspecification but improve the precision of inference, compared to fully nonparametric methods.

In this paper, we establish an alternative characterization of Θ that is free from the curse of dimensionality. One of the main results of this paper (Lemma 1 in Section 2) is that $\Theta = \widetilde{\Theta}$, where

$$\widetilde{\Theta} \equiv \{b \in \Gamma : X'b\,[P(Y = 1|X'b, X'\gamma) - 0.5] \geq 0 \text{ almost surely for all } \gamma \in \Gamma\}. \quad (1.3)$$

This characterization of the identified set Θ enables us to break the curse of dimensionality since we now need to deal with the choice probability conditional on only two indexing variables. The benefit of using the characterization in $\widetilde{\Theta}$, as opposed to Θ, is most clear when we estimate the conditional expectation functions directly. The local power of a Kolmogorov-Smirnov type test decreases as the dimension of conditional variables gets large (for example, see CLR and Armstrong (2014, 2015, 2016)). If the method of CLR is utilized with (1.2), the dimension of nonparametric smoothing is d. Whereas, if the same method is combined with (1.3), note that the dimension of nonparametric smoothing is always 2. This is true even if d is large. Therefore, the latter method is free from the curse of dimensionality.

The remainder of the paper is organized as follows. In Section 2, we provide a formal statement about the binary choice model (1.1). In Section 3, we show that our approach can be extended to the class of semiparametric models under which the sign of conditional expectation of a certain transformation of the outcome is the same as that of the indexing variable. This extension covers a variety of discrete choice models in the literature. Section 4 describes how to construct a confidence set based on CLR and Section 5 presents some results of Monte Carlo simulation experiments that illustrate finite-sample advantage of using the dimension reducing approach. In Section 6, we discuss how to apply our dimension reducing approach to the monotone single index model, which admits related yet different sign restrictions from those studied in Section 3. We conclude the paper in Section 7. Proofs and further results are collated in Appendix A.

2 Conditional moment inequalities for a binary choice model

To convey the main idea of this paper in a simple form, we start with a binary choice model. Recall that in the binary choice model (1.1), we have that $Y = 1\{X'\beta \geq \varepsilon\}$, where the distribution of ε conditional on X is unknown. Let Γ_X denote the support of X. Write $X = (X_1, \widetilde{X})$ where $\widetilde{X}$ is the subvector of X excluding its first element. Let Γ be the parameter space that contains the true parameter vector value β. Let b denote a generic element of Γ. Let $Q_\tau(U|V)$ denote the τ quantile of the distribution of a random variable U conditional on a random vector V. We study inference of the model under the following assumptions.

Condition 1. *(i) $|b_1| = 1$ for all $b \in \Gamma$. (ii) The distribution of X_1 conditional on $\widetilde{X} = \widetilde{x}$ is absolutely continuous with respect to Lebesgue measure for almost every realization $\widetilde{x}$.*

Condition 2. *(i) For some $\tau \in (0,1)$ and for all $x \in \Gamma_X$, $Q_\tau(\varepsilon|X = x) = 0$. (ii) For all $x \in \Gamma_X$, there is an open interval containing zero such that the distribution of ε conditional on $X = x$ has a Lebesgue density that is everywhere positive on this interval.*

The event $X'\beta \geq \varepsilon$ determining the choice is invariant with respect to an arbitrary positive scalar multiplying both sides of the inequality. Therefore, the scale of β is not identified; following the literature (e.g., Horowitz (1992)), we assume Condition 1 (i) for scale normalization. Condition 1 (i) and (ii) together imply that the model admits at least one continuous covariate. Condition 2 (i), due to Manski (1985, 1988), is a quantile independence assumption and allows for nonparametric specification of the preference shock with a general form of heteroskedasticity. Condition 2 (ii) implies that, for all $x \in \Gamma_X$, $P(\varepsilon \leq t|X = x)$ is strictly increasing in t around the neighborhood of the point $t = 0$. This is a fairly weak restriction which is not confined to the case where the distribution of ε conditional on X has a Lebesgue density that is everywhere positive on $\mathbb{R}$.

Under Condition 2, Manski (1988, Proposition 2) established that the necessary and sufficient condition for point identification of β is that, for $b \neq \beta$,

$$P\left(X'b < 0 \leq X'\beta \text{ or } X'b \geq 0 > X'\beta\right) > 0. \tag{2.1}$$

Given the scale normalizing assumption, the condition (2.1) effectively requires that the covariates X should be observed with sufficient variation. Hence, lack of adequate support of the distribution of X may result in non-identification of β. For example, Manski (1988) and Horowitz (1998, Section 3.2.2) constructed non-identification cases for which all covariates take discrete values. Admitting continuous covariates does not guarantee identification either. As indicated by Manski (1985, Lemma 1), non-identification also arises when the covariates are distributed over a bounded support such that one of the choices is observed with probability well below τ for almost every realized value of X. In empirical applications of the discrete choice model, it is quite common to include continuous variables in the covariate specification. Therefore, the present paper addresses and develops the method for inference of β in the presence of continuous covariates for the model where the support of data may not be rich enough to fulfill the point-identifying condition (2.1).

Though Conditions 1 and 2 do not suffice for point identification of β, it still induces restrictions on possible values of preference parameters, which results in set identification of β. To see this, note that Condition 2 implies that for all $x \in \Gamma_X$,

$$P(Y = 1|X = x) > \tau \Leftrightarrow x'\beta > 0, \tag{2.2}$$

$$P(Y = 1|X = x) = \tau \Leftrightarrow x'\beta = 0, \tag{2.3}$$

$$P(Y = 1|X = x) < \tau \Leftrightarrow x'\beta < 0. \tag{2.4}$$

Given Condition 1, $X'b$ is continuous for any $b \in \Gamma$. Thus $P(Y = 1|X) = \tau$ occurs with zero probability. The set of observationally equivalent preference parameter values that conform with Condition 2 can hence be characterized by

$$\Theta = \{b \in \Gamma : X'b\,[P(Y = 1|X) - \tau] \geq 0 \text{ almost surely}\}. \tag{2.5}$$

Given (2.2), (2.3) and (2.4), we also have that

$$\Theta = \{b \in \Gamma : b'XX'\beta \geq 0 \text{ almost surely}\}. \tag{2.6}$$

Namely, the vector b is observationally equivalent to β if and only if the indexing variables $X'b$ and $X'\beta$ are of the same sign almost surely.

Operationally, one could make inference on β by pointwise inverting a test of the conditional moment inequalities given in (2.5). However, as discussed in Section 1, there is the curse of dimensionality in nonparametric inference of the conditional expectation when the dimension of continuous covariates is high. By exploiting the restrictions implies by Conditions 1 and 2, we now present below a novel set of conditional moment inequalities that can equivalently characterize the set Θ yet enable inference to be performed free from the curse of dimensionality.

Note that the restrictions (2.2), (2.3) and (2.4) imply that

$$Q_{1-\tau}(Y|X) = 1\{X'\beta > 0\} = Q_{1-\tau}(Y|X'\beta) \text{ almost surely.}$$

In other words, we have that with probability 1,

$$\text{sgn}[P(Y = 1|X) - \tau] = \text{sgn}[P(Y = 1|X'\beta) - \tau] = \text{sgn}[X'\beta], \tag{2.7}$$

where $\text{sgn}(\cdot)$ is the sign function such that $\text{sgn}(u) = 1$ if $u > 0$; $\text{sgn}(u) = 0$ if $u = 0$; $\text{sgn}(u) = -1$ if $u < 0$. The sign equivalence (2.7) motivates use of indexing variables instead of the full set of covariates as the conditioning variables in nonparametric estimation of the conditional expectation, thereby breaking the curse of dimensionality as raised in the discussion above. To be precise, let

$$\widetilde{\Theta} \equiv \{b \in \Gamma : X'b\,[P(Y = 1|X'b, X'\gamma) - \tau] \geq 0 \text{ almost surely for all } \gamma \in \Gamma\}.$$

The first key result of our approach is the following lemma showing that the identified set Θ can be equivalently characterized by $\widetilde{\Theta}$, which is based on the choice probabilities conditional on two indexing variables.

Lemma 1. *Under Conditions 1 and 2, we have that* $\Theta = \widetilde{\Theta}$.

To explain the characterization result of Lemma 1, note that the model (1.1) under Condition 2 implies that for any $\gamma \in \Gamma$,

$$\text{sgn}[P(Y = 1|X'\beta, X'\gamma) - \tau] = \text{sgn}[X'\beta] \text{ almost surely.} \tag{2.8}$$

Thus, intuitively speaking, for any b that is observationally equivalent to β, equation (2.8) should also hold for b in place of β in the statement. Define

$$\underline{\Theta} \equiv \{b \in \Gamma : X'b\,[P(Y = 1|X'\gamma) - \tau] \geq 0 \text{ almost surely for all } \gamma \in \Gamma\}, \tag{2.9}$$
$$\overline{\Theta} \equiv \{b \in \Gamma : X'b\,[P(Y = 1|X'b) - \tau] \geq 0 \text{ almost surely}\}. \tag{2.10}$$

In contrast with the set $\widetilde{\Theta}$, the sets $\underline{\Theta}$ and $\overline{\Theta}$ are based on moment inequalities conditional on a single indexing variable. The next lemma, which extends the result of Lemma 1, establishes the relation between the sets Θ, $\widetilde{\Theta}$, $\underline{\Theta}$ and $\overline{\Theta}$.

Lemma 2. *Under Conditions 1 and 2, we have that*

$$\underline{\Theta} \subset \Theta = \widetilde{\Theta} \subset \overline{\Theta}. \tag{2.11}$$

It is interesting to note that the set inclusion in (2.11) can be strict as demonstrated in the examples of Appendix A.2. Namely, the set $\underline{\Theta}$ is too restrictive and a test of the inequalities given by (2.9) may inadequately reject the true parameter value β with probability approaching unity. Moreover, the set $\overline{\Theta}$ is not sharp and thus a test of inequalities given by (2.10) would not be consistent against some b values that are incompatible with the inequality restrictions given by (2.5).

The identifying relationship in (2.11) can be viewed as a conditional moment inequality analog of well-known index restrictions in semiparametric binary response models (e.g., Cosslett (1983), Powell, Stock, and Stoker (1989), Han (1987), Ichimura (1993), Klein and Spady (1993), Coppejans (2001)). The main difference between our setup and those models is that we allow for partial identification as well as a general form of heteroskedasticity. It is also noted that to ensure equivalent characterization of the set Θ, we need two indices unlike ones in the point-identified cases.

3 General results for a class of semiparametric models under sign restrictions

In this section, we extend the dimension reducing characterization approach of the previous section to a variety of semiparametric discrete choice models under which the sign of conditional expectation of a certain transformation of the outcome is the same as that of the indexing variable. We treat univariate and multivariate outcome models in a unified abstract setting given as follows.

Let (Y, X) be the data vector of an individual observation where Y is a vector of outcomes and X is a vector of covariates. The econometric model specifying the distribution of Y conditional on X depends on a finite dimensional parameter vector β and is characterized by the following sign restrictions.

Assumption 1. *For some set C and some known functions G and H, and for all $c \in C$, the following statements hold with probability 1. That is, with probability 1,*

$$G(X, c, \beta) > 0 \Longleftrightarrow E\left(H(Y, c)|X\right) > 0, \tag{3.1}$$

$$G(X, c, \beta) = 0 \Longleftrightarrow E\left(H(Y, c)|X\right) = 0, \tag{3.2}$$

$$G(X, c, \beta) < 0 \Longleftrightarrow E\left(H(Y, c)|X\right) < 0. \tag{3.3}$$

Let β be the true data generating parameter vector. Assume $\beta \in \Gamma$ where Γ denotes the parameter space. Let b be a generic element of Γ. Note that the functions G and H in Assumption 1 are determined by the specification of the given model. For example, for the binary choice model of Section 2, Assumption 1 is fulfilled by taking $G(X, c, b) = X'b$ and $H(Y, c) = Y - \tau$, both being independent of c. Other examples satisfying Assumption 1 are presented below.

Define

$$\Theta_0 = \{b \in \Gamma : (3.1), (3.2) \text{ and } (3.3) \text{ hold with } b \text{ in place of } \beta \text{ almost surely for all } c \in C\}.$$

Note that Θ_0 consists of observationally equivalent parameter values that conform with the sign restrictions of Assumption 1. We impose the following continuity assumption.

Assumption 2. *For all $c \in C$ and for all $b \in \Gamma$, the event that $G(X, c, b) = 0$ occurs with zero probability.*

Under Assumptions 1 and 2, we can reformulate the identified set Θ_0 using weak conditional moment inequalities given by the set

$$\Theta \equiv \{b \in \Gamma : G(X, c, b)E\left(H(Y, c)|X\right) \geq 0 \text{ almost surely for all } c \in C\}. \tag{3.4}$$

We now derive the equivalent characterization of the set Θ using indexing variables. Define

$$\begin{aligned}
\widetilde{\Theta} &\equiv \{b \in \Gamma : G(X, c, b)E(H(Y, c)|G(X, c, b), G(X, c, \gamma)) \geq 0 \text{ almost surely for all } (\gamma, c) \in \Gamma \times C\}, \\
\underline{\Theta} &\equiv \{b \in \Gamma : G(X, c, b)E(H(Y, c)|G(X, c, \gamma)) \geq 0 \text{ almost surely for all } (\gamma, c) \in \Gamma \times C\}, \\
\overline{\Theta} &\equiv \{b \in \Gamma : G(X, c, b)E(H(Y, c)|G(X, c, b)) \geq 0 \text{ almost surely for all } c \in C\}.
\end{aligned}$$

The following theorem generalizes the results of Lemmas 1 and 2.

Theorem 1. *Given Assumptions 1 and 2, we have that*

$$\underline{\Theta} \subset \Theta_0 = \Theta = \widetilde{\Theta} \subset \overline{\Theta}. \tag{3.5}$$

In the following subsections, we discuss examples of semiparametric models that fit within the setting of sign restrictions of Assumption 1. In addition, the general framework in this section can be applied to monotone transformation models (e.g., see Abrevaya (1999, 2000), Chen (2010) and Pakes and Porter (2016, Section 2)).

Example 1: Ordered choice model under quantile independence restriction

Consider an ordered response model with $K+1$ choices. Let $\{1, ..., K+1\}$ denote the choice index set. The agent chooses alternative c if and only if

$$\lambda_{c-1} < X'\theta + \varepsilon \leq \lambda_c \tag{3.6}$$

where $\lambda_0 = -\infty < \lambda_1 < .... < \lambda_K < \lambda_{K+1} = \infty$. Let $\lambda \equiv (\lambda_1, ..., \lambda_K)$ be the vector of unknown threshold parameters. We assume that X does not contain a constant component because the coefficient (intercept) associated with a constant covariate cannot be separately identified from the threshold parameters. Let Y be the observed choice, which is given by

$$Y = \sum_{c=1}^{K+1} c1\{\lambda_{c-1} < X'\theta + \varepsilon \leq \lambda_c\}. \tag{3.7}$$

We are interested in inference of $\beta \equiv (\theta, \lambda)$. Lee (1992) and Komarova (2013) studied inference of the ordered response model under quantile independence restriction. Assume the distribution of ε conditional on X satisfies Condition 2. Using this restriction, we see that Assumption 1 holds with $C = \{1, ..., K\}$, $H(Y, c) = 1\{Y \leq c\} - \tau$ and $G(X, c, \beta) = \widetilde{X}_c'\beta$ where $\widetilde{X}_c \equiv (-X', l_c')'$ with l_c being the K dimensional vector $(l_{c,1}, ..., l_{c,K})$ such that $l_{c,j} = 1$ if $j = c$ and $l_{c,j} = 0$ otherwise.

Example 2: Multinomial choice model

Consider a multinomial choice model with K alternatives. Let $\{1, ..., K\}$ denote the choice index set. The utility from choosing alternative j is

$$U_j = X_j{}'\beta + \varepsilon_j \tag{3.8}$$

where $X_j \in \mathbb{R}^q$ is a vector of observed choicewise covariates and ε_j is a choicewise preference shock. The agent chooses alternative k if $U_k > U_j$ for all $j \neq k$. Let X denote the vector $(X_1, ..., X_K)$ and Y denote the observed choice. We assume that the unobservables $\varepsilon \equiv (\varepsilon_1, ..., \varepsilon_K)$ should satisfy the following rank ordering property.

Condition 3. *For any pair (s, t) of choices, we have that with probability 1,*

$$X_s{}'\beta > X_t'\beta \iff P(Y = s|X) > P(Y = t|X). \tag{3.9}$$

Manski (1975), Matzkin (1993) and Fox (2007) used Condition 3 as an identifying restriction in the multinomial choice model to allow for nonparametric unobservables with unknown form of heteroskedasticity. Goeree, Holt, and Palfrey (2005, Proposition 5) showed that it suffices for Condition 3 to assume that the joint distribution of ε conditional on X for almost every realization of X is exchangeable and has a joint density that is everywhere positive on $\mathbb{R}^K$.

Under Condition 3, Assumption 1 holds for this example by taking $C \equiv \{(s, t) \in \{1, ..., K\}^2 : s < t\}$, $G(X, s, t, \beta) = (X_s - X_t)'\beta$ and $H(Y, s, t) = 1\{Y = s\} - 1\{Y = t\}$.

Example 3: Binary choice panel data with fixed effect

Consider the following binary choice panel data model

$$Y_t = 1\{X_t'\beta + v \geq \varepsilon_t\},\ t \in \{1, ..., T\} \tag{3.10}$$

where $X_t \in \mathbb{R}^q$ is a vector of per-period covariates and v is an unobserved fixed effect. Let X be the vector $(X_1, ..., X_T)$. Let $Y = (Y_1, ..., Y_T)$ denote the vector of outcomes. Manski (1987) imposed the following restrictions on the transitory shocks ε_t.

Condition 4. *The distribution of ε_t conditional on (X, v) is time invariant and has a Lebesgue density that is everywhere positive on $\mathbb{R}$ for almost every realization of (X, v).*

Under Condition 4 and by Lemma 1 of Manski (1987), Assumption 1 holds for this example by taking $C \equiv \{(s,t) \in \{1, ..., T\}^2 : s < t\}$, $G(X, s, t, \beta) = (X_s - X_t)'\beta$ and $H(Y, s, t) = Y_s - Y_t$.

Example 4: Ordered choice panel data with fixed effect

This example is concerned with the ordered choice model of Example 1 in the panel data context. Let $\{1, ..., K + 1\}$ denote the choice index set. For each period $t \in \{1, ..., T\}$, we observe the agent's ordered response outcome Y_t that is generated by

$$Y_t = \sum_{j=1}^{K+1} j1\{\lambda_{j-1} < X_t'\beta + v + \varepsilon_t \leq \lambda_j\}, \tag{3.11}$$

where v is an unobserved fixed effect and $\lambda_0 = -\infty < \lambda_1 < .... < \lambda_K < \lambda_{K+1} = \infty$. Let X and Y denote the covariate vector $(X_1, ..., X_T)$ and outcome vector $(Y_1, ..., Y_T)$, respectively. Suppose the shocks ε_t also satisfy Manski (1987)'s stationarity assumption given by Condition 4. Under this restriction and by applying the law of iterated expectations, we see that Assumption 1 holds for this example by taking $C = \{(k, s, t) : k \in \{1, ..., K\}, (s, t) \in \{1, ..., T\}^2$ such that $s < t\}$, $G(X, k, s, t, \beta) = (X_t - X_s)'\beta$ and $H(Y, k, s, t) = 1\{Y_s \leq k\} - 1\{Y_t \leq k\}$.

4 The $(1-\alpha)$ level confidence set

This section describes how to construct a confidence set for the true value β based on the conditional moment inequalities that define the set $\widetilde{\Theta}$. Let $v \equiv (x, \gamma, c)$ and $\mathcal{V} \equiv \{(x, \gamma, c) : x \in \Gamma_X, \gamma \in \Gamma, c \in C\}$. Assume the set $\mathcal{V}$ is nonempty and compact. Define

$$\begin{aligned} m_b(v) \quad &\equiv \quad E\left(G(X,c,b)H(Y,c)|G(X,c,b) = G(x,c,b), G(X,c,\gamma) = G(x,c,\gamma)\right) \\ &\quad \times f_{b,c,\gamma}\left(G(x,c,b), G(x,c,\gamma)\right), \end{aligned}$$

where the function $f_{b,c,\gamma}$ denotes the joint density function of the indexing variables $(G(X,c,b), G(X,c,\gamma))$. Under the assumption that $f_{b,c,\gamma}\left(G(x,c,b), G(x,c,\gamma)\right) > 0$, note that for almost every $v \in \mathcal{V}$,

$$\begin{aligned} &m_b(v) \geq 0 \\ &\Longleftrightarrow E\left(G(X,c,b)H(Y,c)|G(X,c,b) = G(x,c,b), G(X,c,\gamma) = G(x,c,\gamma)\right) \geq 0. \end{aligned}$$

Thus we have that

$$\widetilde{\Theta} = \{b \in \Gamma : m_b(v) \geq 0 \text{ for almost every } v \in \mathcal{V}\}. \tag{4.1}$$

Assume that we observe a random sample of individual outcomes and covariates $(Y_i, X_i)_{i=1,\ldots,n}$. For inference on the true parameter value β, we aim to construct a set estimator $\widehat{\Theta}$ at the $(1-\alpha)$ confidence level such that

$$\liminf_{n \longrightarrow \infty} P(\beta \in \widehat{\Theta}) \geq 1 - \alpha. \tag{4.2}$$

We now delineate an implementation of the set estimator $\widehat{\Theta}$ based on a kernel version of CLR. To estimate the function m_b, we consider the following kernel type estimator:

$$\widehat{m}_b(v) \equiv \{nh_n(c,b)h_n(c,\gamma)\}^{-1} \sum_{i=1}^{n} G(X_i,c,b)H(Y_i,c)K_n(X_i,v,b), \tag{4.3}$$

where

$$K_n(X_i, v, b) \equiv K\left(\frac{G(x,c,b) - G(X_i,c,b)}{h_n(c,b)}, \frac{G(x,c,\gamma) - G(X_i,c,\gamma)}{h_n(c,\gamma)}\right), \tag{4.4}$$

$K(\cdot,\cdot)$ is a bivariate kernel function, and $h_n(c,\gamma)$ is a sequence of bandwidths for each (c,γ). Note that $h_n(c,\gamma)$ is a function of (c,γ) and thus it can be different from $h_n(c,b)$. Define

$$T(b) \equiv \inf_{v\in\mathcal{V}} \frac{\widehat{m}_b(v)}{\widehat{\sigma}_b(v)}, \tag{4.5}$$

where

$$\begin{aligned}
\widehat{\sigma}_b^2(v) \quad &\equiv \quad n^{-2}[h_n(c,b)]^{-2}[h_n(c,\gamma)]^{-2}\sum_{i=1}^n \widehat{u}_i^2(b,c,\gamma)G^2(X_i,c,b)K_n^2(X_i,v,b), \qquad (4.6)\\
\widehat{u}_i\,(b,c,\gamma) \quad &\equiv \quad H(Y_i,c) - \left[\sum_{j=1}^n K_n(X_j,(X_i,\gamma,c),b)\right]^{-1}\sum_{j=1}^n H(Y_j,c)K_n(X_j,(X_i,\gamma,c),b).
\end{aligned}$$

For a given value of b, we compare the test statistic $T(b)$ to a critical value to conclude whether there is significant evidence that the inequalities in (4.1) are violated for some $v \in \mathcal{V}$. By applying the test procedure to each candidate value of b, the estimator $\widehat{\Theta}$ is then the set comprising those b values not rejected under this pointwise testing rule.

Based on the CLR method, we estimate the critical value using simulations. Let B be the number of simulation repetitions. For each repetition $s \in \{1,...,B\}$, we draw an n dimensional vector of mutually independently standard normally distributed random variables which are also independent of the data. Let $\eta(s)$ denote this vector. For any compact set $\mathsf{V} \subseteq \mathcal{V}$, define

$$T_s^*(b;\mathsf{V}) \equiv \inf_{v\in\mathsf{V}}\left[\{nh_n(c,b)h_n(c,\gamma)\widehat{\sigma}_b(v)\}^{-1}\sum_{i=1}^n \eta_i(s)\widehat{u}_i\,(b,c,\gamma)\,G(X_i,c,b)K_n(X_i,v,b)\right]. \tag{4.7}$$

We approximate the distribution of $\inf_{v\in\mathsf{V}}[(\widehat{\sigma}_b(v))^{-1}\widehat{m}_b(v)]$ over $\mathsf{V} \subseteq \mathcal{V}$ by that of the simulated quantity $T_s^*(b;\mathsf{V})$. Let $\widehat{q}_p(b,\mathsf{V})$ be the p level empirical quantile based on the vector $(T_s^*(b;\mathsf{V}))_{s\in\{1,...,B\}}$. One could use $\widehat{q}_p(b,\mathcal{V})$ as the test critical value. However, following CLR, we can make sharper inference by incorporating the data driven inequality selection mechanism in the critical value estimation. Let

$$\widehat{V}_n(b) \equiv \{v \in \mathcal{V} : \widehat{m}_b(v) \leq -2\widehat{q}_{\gamma_n}(b, \mathcal{V})\widehat{\sigma}_b(v)\}, \tag{4.8}$$

where $\gamma_n \equiv 0.1/\log n$. Compared to $\widehat{q}_p(b, \mathcal{V})$, use of $\widehat{q}_\alpha(b, \widehat{V}_n(b))$ as the critical value results in a test procedure concentrating the inference on those points of v that are more informative for detecting violation of the non-negativity hypothesis on the function $m_b(v)$. In fact, the CLR test based on the set $\widehat{V}_n(b)$ is closely related to the power improvement methods such as the contact set idea (e.g., Linton, Song, and Whang (2010) and Lee, Song, and Whang (2017)), the generalized moment selection approach (e.g., Andrews and Soares (2010), Andrews and Shi (2013), and Chetverikov (2017)), and the iterative step-down approach (e.g., Chetverikov (2012)) employed in the literature on testing moment inequalities.

Assume that $0 < \alpha \leq 1/2$. Then we construct the $(1-\alpha)$ confidence set $\widehat{\Theta}$ by setting

$$\widehat{\Theta} \equiv \left\{ b \in \Gamma : T(b) \geq \widehat{q}_\alpha(b, \widehat{V}_n(b)) \right\}. \tag{4.9}$$

We can establish regularity conditions under which (4.2) holds by utilizing the general results of CLR. Since the main focus of this paper is identification, we omit the technical details for brevity.

In summary, our proposed algorithm takes the following form:

1. Specify $K(\cdot, \cdot)$, $h_n(c, \gamma)$ and generate $\{\eta(s) : s = 1, \ldots, B\}$, that is, $n \times B$ matrix of independent $N(0, 1)$.
2. Approximate Γ by a grid. For each value b in this grid,
   (a) compute $T(b)$ defined in (4.5) and $\widehat{V}_n(b)$ defined in (4.8),
   (b) simulate $T_s^*(b; \widehat{V}_n(b))$ defined in (4.7) for all $s = 1, \ldots, B$ to obtain the α quantile $\widehat{q}_\alpha(b, \widehat{V}_n(b))$,
   (c) include b in the $(1-\alpha)$ confidence set $\widehat{\Theta}$ if and only if $T(b) \geq \widehat{q}_\alpha(b, \widehat{V}_n(b))$.

When the dimension of β is high, it is computationally demanding to obtain $\widehat{\Theta}$ since it is necessary to cary out the pointwise test in (4.9) for a grid of Γ. However, this is a common problem in the literature when a confidence set is based on inverting a pointwise test. It is worth mentioning that there is additional computational complexity that is unique in our proposal compared to the case of conditioning on full covariates. In the proposed algorithm above, it is necessary to obtain the infimum over $v \equiv (x, \gamma, c)$. If the algorithm were based on conditioning on full covariates directly, it would be necessary to take the infimum over (x, c) only. In other words, to facilitate dimension reduction in nonparametric estimation, we need to find the infimum over a larger set of arguments. One way to deal with this complexity problem is to use the same number of random grid points between the index and full approaches, as we will demonstrate in a simulation study in the next section.

In practice, it is important to specify $K(\cdot,\cdot)$ and $h_n(c,\gamma)$. For the former, it is conventional to use the product of a univariate second-order kernel function, for example $K(u_1,u_2) = \widetilde{K}(u_1)\widetilde{K}(u_2)$ with

$$\widetilde{K}(u) \equiv \frac{15}{16}\left(1-u^2\right)^2 1\left\{|u| \leq 1\right\}. \tag{4.10}$$

For the latter, we recommend using

$$h_n(c,\gamma) = C_{\text{bandwidth}} \times \widehat{s}(G(X,c,\gamma)) \times n^{-1/5}, \tag{4.11}$$

where $C_{\text{bandwidth}}$ is a constant, and $\widehat{s}(W)$ denotes the sample standard deviation for the random variable W. If the mapping

$$(u_1,u_2) \mapsto E\left(G(X,c,b)H(Y,c)|G(X,c,b)=u_1, G(X,c,\gamma)=u_2\right) f_{b,c,\gamma}\left(u_1,u_2\right)$$

is twice continuously differentiable for each c and γ, the optimal rate for $h_n(c,\gamma)$ in minimizing the mean squared error is proportional to $n^{-1/6}$. The rate of $n^{-1/5}$ in $h_n(c,\gamma)$ is chosen to ensure that the bias is asymptotically negligible due to undersmoothing. Although our suggested rule-of-thumb for $h_n(c,\gamma)$ in (4.11) is not completely data-driven, it has the advantage that its scale changes automatically as the scale of $G(X,c,\gamma)$ changes. It is a difficult task to choose $C_{\text{bandwidth}}$ optimally for our setup since it involves possibly higher-order comparison between the size and power of the test in (4.9). Moreover, one technical issue arising specifically from our setup is that when γ and b are close from each other, the two dimensional kernel function is close to the one dimensional kernel function. It might be better to choose a variable bandwidth that depends on the distance between γ and b. We leave the task of choosing the bandwidth optimally for future research.

We conclude this section by briefly remarking on an alternative form of the test statistic which can also be used in the algorithm above. Noting that

$$\begin{aligned} & E\left(G(X,c,b)H(Y,c)|G(X,c,b)=G(x,c,b),G(X,c,\gamma)=G(x,c,\gamma)\right) \\ = \; & G(x,c,b)E\left(H(Y,c)|G(X,c,b)=G(x,c,b),G(X,c,\gamma)=G(x,c,\gamma)\right), \end{aligned}$$

we can thus replace each individual specific index $G(X_i,c,b)$ in the summation term of (4.3) by the non-stochastic term $G(x,c,b)$, thereby resulting in an alternative definition of the estimator $\widehat{m}_b(v)$, which remains to be a consistent estimator for $m_b(v)$. Making such replacements as well in (4.6) and (4.7) for the definitions of $\widehat{\sigma}_b^2(v)$ and $T_s^*(b;\mathsf{V})$, respectively, we can then apply the algorithm above to obtain an alternative confidence set which also satisfies (4.2). Note that the standardized estimator $[\widehat{\sigma}_b(v)]^{-1}\,\widehat{m}_b(v)$ for this alternative approach becomes

$$\operatorname{sgn}\left(G(x,c,b)\right)\left[\sum\nolimits_{i=1}^{n}\widehat{u}_i^2(b,c,\gamma)K_n^2(X_i,v,b)\right]^{-1/2}\sum\nolimits_{i=1}^{n}H(Y_i,c)K_n(X_i,v,b),$$

which is discontinuous in the argument x. Operationally, this indicates that gradient based minimization algorithms become inapplicable for computing the statistics $T(b)$ and $T_s^*(b;\mathsf{V})$ defined under this alternative inference approach.

5 Simulation study

The main purpose of this simulation study is to compare finite-sample performance of the approach of conditioning on indexing variables with that of conditioning on full covariates. We use the binary response model set forth in Section 2 for the simulation design. The data is generated according to the following setup:

$$Y = 1\{X'\beta \geq \varepsilon\}, \tag{5.1}$$

where $X=(X_1,...,X_d)$ is a d dimensional covariate vector with $d \geq 2$,

$$\varepsilon = \left[1+\sum\nolimits_{k=1}^{d}X_k^2\right]^{1/2}\xi,$$

and ξ is standard normally distributed and independent of X. Let $\widetilde{X} = (X_2, ..., X_d)$ be a $(d-1)$ dimensional vector of mutually independently and uniformly distributed random variables on the interval $[-1, 1]$. The covariate X_1 is specified by

$$X_1 = \text{sgn}(X_2)U, \tag{5.2}$$

where U is a uniformly distributed random variable on the interval $[0, 1]$ and is independent of $(\widetilde{X}, \xi)$. We set

$$\beta_1 = 1 \text{ and } \beta_k = 0 \text{ for } k \in \{2, ..., d\}.$$

The preference parameter space is specified to be

$$\Gamma \equiv \{b \in \mathbb{R}^d : b_1 = 1, (b_2, ..., b_d) \in [-1, 1]^{d-1}\}. \tag{5.3}$$

Note that, by (5.2), $X'\beta = X_1$ so that the sign of the true index $X'\beta$ is determined by that of X_2 but the magnitude of $X'\beta$ is independent of $\widetilde{X}$. Using this fact and the simulation configurations, it is straightforward to see that the event $X'\beta > 0$ and $X'b < 0$ can occur with positive probability for any $b \in \Gamma$ such that either $b_2 < 0$ or $b_k \neq 0$ for some $k \in \{3, ..., d\}$. On the other hand, by (5.2), we also find that $X'\beta = X_1$ and $X'b = X_1 + X_2 b_2$ have the same sign almost surely for any $b \in \Gamma$ such that $b_2 \geq 0$ and $b_k = 0$ for $k \in \{3, ..., d\}$. Using these facts and by (2.6), the identified set Θ in this simulation setup is therefore given by

$$\Theta = \{b \in \Gamma : b_2 \geq 0 \text{ and } b_k = 0 \text{ for } k \in \{3, ..., d\}\}. \tag{5.4}$$

Recall that the present simulation design also satisfies the general framework of Section 3 by taking $G(X, c, b) = X'b$ and $H(Y, c) = Y - 0.5$. Let *Index* and *Full* be shorthand expressions for the index formulated and full covariate approaches, respectively. We implement the *Index* approach using the inference procedure of Section 4. We compute the term $K_n(X, v, b)$ using

$$K_n(X, v, b) = \widetilde{K}\left(\frac{x'b - X'b}{\widehat{s}(X'b)h_n}\right)\widetilde{K}\left(\frac{x'\gamma - X'\gamma}{\widehat{s}(X'\gamma)h_n}\right)$$

where $v = (x, \gamma)$, $\widetilde{K}(\cdot)$ is the univariate biweight kernel function defined in (4.10).

Recall that $\widehat{s}(W)$ denotes the estimated standard deviation for the random variable W. As suggested in the previous section, the bandwidth sequence h_n is specified by

$$h_n = c_{Index} n^{-1/5}, \tag{5.5}$$

where c_{Index} is a bandwidth scale.

The $Full$ approach is based on inversion of the kernel-type CLR test for the inequalities that $m_{b,Full}(x) \geq 0$ for all $x \in \Gamma_X$, where

$$m_{b,Full}(x) \equiv E\left(X'b\left(Y - 0.5\right)|X = x\right) f_X\left(x\right) \tag{5.6}$$

and f_X denotes the joint density of X. As in the $Index$ approach, we consider the kernel type estimator

$$\widehat{m}_{b,Full}(x) \equiv \left(nh_n^d\right)^{-1} \sum_{i=1}^{n} X_i'b\left(Y_i - 0.5\right) K_{n,Full}(X_i, x), \tag{5.7}$$

where

$$K_{n,Full}(X_i, x) \equiv \prod\nolimits_{k=1}^{d} \widetilde{K}_{Full}\left(\frac{x_k - X_{i,k}}{\widehat{s}(X_{i,k}) h_{n,Full}}\right), \tag{5.8}$$

$\widetilde{K}_{Full}(\cdot)$ is the univariate pth order biweight kernel function (see Hansen (2005)), and $h_{n,Full}$ is a bandwidth sequence specifying by

$$h_{n,Full} = c_{Full} n^{-r}, \tag{5.9}$$

where c_{Full} and r denote the bandwidth scale and rate, respectively. The test statistic for the $Full$ approach is given by

$$T_{Full}(b) \equiv \inf\nolimits_{x \in \Gamma_X} \frac{\widehat{m}_{b,Full}(x)}{\widehat{\sigma}_{b,Full}(x)}, \tag{5.10}$$

where

$$\begin{aligned}
\widehat{\sigma}_{b,Full}^2(x) &\equiv n^{-2} h_{n,Full}^{-2d} \sum_{i=1}^{n} \widehat{u}_{i,Full}^2 \left(X_i'b\right)^2 K_{n,Full}^2(X_i, x), \\
\widehat{u}_{i,Full} &\equiv Y_i - \left[\sum_{j=1}^{n} K_{n,Full}(X_j, X_i)\right]^{-1} \sum_{j=1}^{n} Y_j K_{n,Full}(X_j, X_i).
\end{aligned}$$

We computed the simulated CLR test critical value that also embedded the inequality selection mechanism. By comparing $T_{Full}(b)$ to the test critical value, we constructed under the $Full$ approach the confidence set that also satisfies (4.2).

The nominal significance level α was set to be 0.05. Let $\widehat{\Theta}_{Index}$ and $\widehat{\Theta}_{Full}$ denote the $(1-\alpha)$ level confidence sets constructed under the $Index$ and $Full$ approaches, respectively. For $s \in \{Index, Full\}$ and for a fixed value of b, we calculated $\widehat{P}_s(b)$, which is the simulated finite-sample probability of the event $b \notin \widehat{\Theta}_s$ based on 1000 simulation repetitions. For each repetition, we generated $n \in \{250, 500, 1000\}$ observations according to the data generating design described above. We used 4000 simulation draws to calculate $\widehat{q}_\alpha(b, \widehat{V}_n(b))$ for the $Index$ approach and to estimate the CLR test critical value for the $Full$ approach. We implemented for the $Full$ approach the minimization operation based on grid search over 1000 grid points of x randomly drawn from the joint distribution of X. For the $Index$ approach, the minimization was implemented by grid search over 1000 grid points of (x, γ) for which x was also randomly drawn from the distribution of X, and γ was drawn from uniform distribution on the space Γ and independently of the search direction in x.

We conducted simulations for $d \in \{3, 4, 5, 10\}$. All simulation experiments were programmed in Gauss 9.0 and performed on a desktop PC (Windows 7) equipped with 32 GB RAM and a CPU processor (Intel i7-5930K) of 3.5 GHz. For the $Full$ approach, both the bandwidth rate r and the order p of $\widetilde{K}_{Full}$ depend on the covariate dimension. These were specified to fulfill the regularity conditions for the CLR kernel type conditional moment inequality tests (see discussions on Appendix F of CLR (pp. 7-9, Supplementary Material)). Note that for $b \in \Theta$, $\widehat{P}_{Index}(b)$ ($\widehat{P}_{Full}(b)$) is simulated null rejection probability of the corresponding CLR test under the $Index$ ($Full$) approach, whereas for $b \notin \Theta$, it is the power of the test. For simplicity, we computed $\widehat{P}_{Index}(b)$ and $\widehat{P}_{Full}(b)$ for b values specified as $b = (b_1, b_2, ..., b_d)$ where $b_1 = 1$, $b_2 \in \{0, 0.5, -1\}$, $b_k = 0$ for $k \in \{3, .., d\}$. For these candidate values of b, we experimented over various bandwidth scales to determine the value of c_{Index} (c_{Full}) with which the $Index$ ($Full$) approach exhibits the best overall performance in terms of its corresponding size and power. Table 1 presents the settings of r and p and the chosen bandwidth scales c_{Index} and c_{Full} in the simulation.

Table 1: Settings of r, p, c_{Index} and c_{Full}

d	3	4	5	10
r	11/70	1/9	21/220	1/21
p	2	4	4	6
	sample size 250			
c_{Index}	3.05	3.45	3.7	4.1
c_{Full}	2.65	4.8	5.6	8.35
	sample size 500			
c_{Index}	2.55	2.95	3.05	3.75
c_{Full}	2.35	4.3	4.9	8
	sample size 1000			
c_{Index}	2	2.5	2.75	3.5
c_{Full}	2.15	3.95	4.45	7.7

Tables 2 and 3 present the simulation results that compare performance of the *Index* and *Full* approaches.

Table 2: Simulated null rejection probabilities

d	3	4	5	10	3	4	5	10
	$b_2 = 0$				$b_2 = 0.5$			
	sample size 250							
$\widehat{P}_{Index}$	.034	.029	.034	.050	.051	.054	.052	.052
$\widehat{P}_{Full}$	.031	.043	.046	.050	.050	.053	.052	.055
	sample size 500							
$\widehat{P}_{Index}$	.030	.036	.039	.042	.051	.054	.052	.050
$\widehat{P}_{Full}$	.032	.034	.043	.044	.049	.048	.054	.053
	sample size 1000							
$\widehat{P}_{Index}$	.047	.045	.041	.048	.054	.053	.051	.054
$\widehat{P}_{Full}$	.029	.044	.041	.042	.046	.051	.047	.051

Table 3: Simulated test power for $b_2 = -1$ ($ratio \equiv \widehat{P}_{Index}/\widehat{P}_{Full}$)

d	$\widehat{P}_{Index}$	$\widehat{P}_{Full}$	$ratio$	$\widehat{P}_{Index}$	$\widehat{P}_{Full}$	$ratio$	$\widehat{P}_{Index}$	$\widehat{P}_{Full}$	$ratio$
	$n = 250$			$n = 500$			$n = 1000$		
3	.583	.601	.970	.771	.731	1.05	.927	.828	1.11
4	.541	.530	1.02	.733	.653	1.12	.868	.758	1.14
5	.500	.393	1.27	.699	.624	1.12	.806	.738	1.09
10	.409	.216	1.89	.474	.212	2.23	.520	.225	2.31

From Table 2, we can see that all $\widehat{P}_{Index}$ and $\widehat{P}_{Full}$ values in all the simulation cases are either below or close to the nominal level 0.05 with the maximal value being 0.055 and occurring for the *Full* approach with sample size 250 under the setup of $d = 10$ and $b_2 = 0.5$. For both methods, there is slight over-rejection for the case of $b_2 = 0.5$. At the true data generating value ($b_2 = 0$), both $\widehat{P}_{Index}$ and $\widehat{P}_{Full}$ are well capped by 0.05 and the confidence sets $\widehat{\Theta}_{Index}$ and $\widehat{\Theta}_{Full}$ can hence cover the true parameter value with probability at least 0.95 in all simulations.

For the power of the test, we compare the *Index* and *Full* approaches under the same covariate configuration. Table 3 indicates that power of the *Index* approach dominates that of the *Full* approach in almost all simulation configurations. Moreover, at larger sample size ($n = 1000$), power of the *Index* approach exceeds 0.8 in almost all cases whereas that of the *Full* approach does so only for the case of $d = 3$. The power difference between these two approaches tends to increase as either the sample size or the covariate dimension increases. For the case of $d = 10$, it is noted that there is substantial power gain from using the *Index* approach. For this covariate specification, the curse of dimensionality for the *Full* approach is quite apparent because the corresponding $\widehat{P}_{Full}$ values vary only slightly across sample sizes. In short, the simulation results suggest that the *Index* approach may alleviate the problem associated with the curse of dimensionality and we could therefore make sharper inference by using the *Index* approach for a model with a high dimensional vector of covariates.

In practice, one will not carry out Monte Carlo experiments, but will compute a confidence set for the parameter. To do so, one must compute the statistic and its associated critical value for each value of the grid for the parameter space. However, this is a common problem in the literature that relies on inverting a pointwise test. To give a sense of the computation time for obtaining the confidence set, we now report the average computation time in the Monte Carlo experiments. It took about 10, 38, and 154 CPU seconds on average for a given value of b when the sample size is 250, 500 and 1000, respectively. These computation times were not sensitive to the covariate dimension d since we used the same number of random grid points of (x, γ). If we use 100 grid points for constructing the confidence set, then the resulting computation time will be 0.28, 1.06, and 4.27 CPU hours, respectively, for $n = 250, 500$, and 1000. In the implementation of our algorithm, there are two kinds of grid search: (i) the random grid for (x, γ) to evaluate $T(b)$ and its critical value for each b and (ii) the other grid for obtaining the confidence set for β. For the former, it might be desirable to use a larger random grid for (x, γ) or to adopt a more sophisticated optimization algorithm to compute the test statistic and its critical value as d gets large. For the latter, if the degree of precision is fixed, we will need more grid points as d gets large. Hence, in practice, it would be quite computationally demanding to construct the confidence set when $n = 1000$ and $d = 10$.

6 Application of the dimension reducing characterization approach to the monotone single index model

In this section, we discuss how to apply our dimension reducing approach to the single index model, which admits related yet different sign restrictions from those studied in Section 3. We consider the monotone single index model where the conditional mean of the outcome variable Y given a d dimensional covariate vector X satisfies

$$E(Y|X) = G(X'\beta) \tag{6.1}$$

for some unknown strictly increasing function G and a finite dimensional parameter vector $\beta \in \Gamma$, where $\Gamma \subset \mathbb{R}^d$ denotes the space of the index coefficients. Model (6.1) incorporates various semiparametric models such as the generalized regression model where $Y = G(X'\beta) + \varepsilon$ with $E(\varepsilon|X) = 0$, and the transformation model where $Y = H(X'\beta + \varepsilon)$ with the function H being an unknown strictly increasing transformation function and ε being a continuous unobservable that is independent of X. Other examples satisfying the restriction (6.1) also include the single-index binary choice model. See Han (1987) for further details.

For Model (6.1), it is known that, even with location and scale normalization, the true value β may remain non-identified provided that the index $X'\beta$ does not exhibit sufficient variation. This non-identification can arise even when the model admits a continuous covariate (see Example 2.4 of Horowitz (1998)).

In what follows, let (Y_1, X_1) and (Y_2, X_2) be two independent random vectors that are drawn from the joint distribution of (Y, X). By (6.1) and monotonicity of G, we have that, with probability 1,

$$X_1'\beta > X_2'\beta \Longleftrightarrow E(Y_1|X_1) > E(Y_2|X_2), \tag{6.2}$$

$$X_1'\beta = X_2'\beta \Longleftrightarrow E(Y_1|X_1) = E(Y_2|X_2), \tag{6.3}$$

$$X_1'\beta < X_2'\beta \Longleftrightarrow E(Y_1|X_1) < E(Y_2|X_2). \tag{6.4}$$

Given another parameter vector $b \in \Gamma$, we say that b is observationally equivalent to the true value β if and only if the sign equivalence restrictions (6.2), (6.3) and (6.4) hold with b in place of β. In other words, the set

$$\Theta_0 \equiv \{b \in \Gamma : (6.2), (6.3) \text{ and } (6.4) \text{ holds with } b \text{ in place of } \beta \text{ almost surely}\}$$

is the identified set of parameter values that are compatible with the restriction (6.1).

Define the set

$$\Theta \equiv \{b \in \Gamma : (X_1 - X_2)' \, b \, [E(Y_1|X_1) - E(Y_2|X_2)] \geq 0 \text{ almost surely}\}.$$

Condition 5. *For all $b \in \Gamma$, the event that $X_1'b = X_2'b$ occurs with zero probability.*

Condition 5 is a mild continuity assumption, which can hold if the covariate vector includes a continuously distributed component and, for all $b \in \Gamma$, the index coefficient associated with that component is non-zero. It is straightforward to see that Condition 5 implies $\Theta_0 = \Theta$ so that we can characterize the identified set using moment inequalities conditional on the covariates. Note that the sign restrictions (6.2), (6.3) and (6.4) are not nested in the general framework given by Assumption 1 of Section 3. Nonetheless, we can still apply the idea of conditioning on indexing variables to derive an equivalent yet dimension reducing characterization of the identified set.

Let

$$\widetilde{\Theta} \equiv \{b \in \Gamma : (X_1 - X_2)' b \, [E(Y_1 | X_1' b, X_1' \gamma) - E(Y_2 | X_2' b, X_2' \gamma)] \geq 0 \text{ almost surely for all } \gamma \in \Gamma\}.$$

Theorem 2. *Assume (6.1) and Condition 5. Then* $\Theta_0 = \Theta = \widetilde{\Theta}$.

Theorem 2 indicates that we can also derive the identified set using moment inequalities conditional on indexing variables. Using this result, we can construct a confidence set for the true value β based on the method of CLR. Implementation of such a confidence set is analogous to that described in Section 4 and its details are summarized in Appendix A.3 of the paper.

7 Conclusions

This paper studies inference of preference parameters in semiparametric discrete choice models when these parameters are not point identified and the identified set is characterized by a class of conditional moment inequalities. Exploring the semiparametric modeling restrictions, we show that the identified set can be equivalently formulated by moment inequalities conditional on only two continuous indexing variables. Such formulation holds regardless of the covariate dimension, thereby breaking the curse of dimensionality for nonparametric inference of the underlying conditional moment functions. We also apply this dimension reducing characterization approach to the monotone single index model and to a variety of semiparametric models under which the sign of conditional expectation of a certain transformation of the outcome is the same as that of the indexing variable.

There is a growing number of inference methods for conditional moment inequalities. The instrumental variable approach of Andrews and Shi (2013) does not rely on nonparametric estimation of conditional expectation. Nevertheless, the instruments required to convert the conditional moment inequalities to unconditional ones increase with the covariate dimension. In addition to the Andrews-Shi and CLR approaches, other existing inference procedures include Armstrong (2014, 2015), Armstrong and Chan (2016), Chetverikov (2017), Lee, Song, and Whang (2013, 2017) and Menzel (2014) among others. The performance of all of these methods are related to the dimension of conditioning variables. Armstrong (2016, see Tables 1 and 2) gives the local power properties of popular approaches in the literature and shows that the local power decreases as the dimension of conditional variables increases in each case that he considers. Thus, the curse of dimensionality problem is not limited to a particular test statistic. It will be an interesting further research topic to incorporate these alternative methods with the dimension reducing characterization result of this paper.

A Appendix

A.1 Proofs

Proof of Lemmas 1 and 2. Lemma 2 nests Lemma 1. So we focus on the proof of Lemma 2. To prove Lemma 2, we apply Theorem 1 with $G(X, c, b) = X'b$ and $H(Y, c) = Y - \tau$. Note that Assumptions 1 and 2 of Theorem 1 are both satisfied under Conditions 1 and 2. Hence, the result (2.11) follows from an application of Theorem 1. ∎

Proof of Theorem 1. By Assumptions 1 and 2, the event that $E\left(H(Y, c)|X\right) = 0$ also occurs with zero probability. It hence follows that $\Theta_0 = \Theta$.

We now show that $\Theta = \widetilde{\Theta}$. Suppose that $b \in \Theta$. Then with probability 1,

$$G(X, c, b) \geq 0 \Longleftrightarrow E\left(H(Y, c)|X\right) \geq 0. \tag{A.1}$$

Note that

$$E(H(Y, c)|G(X, c, b), G(X, c, \gamma)) = E(E(H(Y, c)|X)|G(X, c, b), G(X, c, \gamma)). \tag{A.2}$$

By (A.1), for any $\gamma \in \Gamma$, the right-hand side of (A.2) has the same sign as $G(X, c, b)$ does with probability 1. Hence, $b \in \widetilde{\Theta}$ and it follows that $\Theta \subset \widetilde{\Theta}$.

On the other hand, assume that $b \in \widetilde{\Theta}$. Since $\beta \in \Gamma$, we have that $G(X, c, b)$ and $E(H(Y, c)|G(X, c, b), G(X, c, \beta))$ have the same sign with probability 1. Using (A.2) and Assumption 1, we see that $E(H(Y, c)|G(X, c, b), G(X, c, \beta))$, $G(X, c, \beta)$ and $E\left(H(Y, c)|X\right)$ also have the same sign with probability 1. Therefore, we can deduce that $b \in \Theta$ and hence $\widetilde{\Theta} \subset \Theta$. Putting together all these results, we thus have that $\Theta_0 = \Theta = \widetilde{\Theta}$.

To complete the proof, it remains to show that $\underline{\Theta} \subset \Theta \subset \overline{\Theta}$. Note that, because

$\beta \in \Gamma$, we have that, for any $b \in \underline{\Theta}$, $G(X,c,b)$ and $E(H(Y,c)|G(X,c,\beta))$ have the same sign with probability 1. By Assumption 1, the law of iterated expectations, and using similar arguments in the proof above, it is straightforward to see that $E(H(Y,c)|G(X,c,\beta))$, $G(X,c,\beta)$ and $E\left(H(Y,c)|X\right)$ also have the same sign with probability 1. Hence, it follows that $b \in \Theta$ so that $\underline{\Theta} \subset \Theta$.

To verify that $\Theta \subset \overline{\Theta}$, note that, by the law of iterated expectations, the sign equivalence result (A.1) implies that $G(X,c,b)$ and $E(H(Y,c)|G(X,c,b))$ also have the same sign with probability 1. Thus, we can deduce that $\Theta \subset \overline{\Theta}$. Putting together all the proved results, we therefore have that $\underline{\Theta} \subset \Theta_0 = \Theta = \widetilde{\Theta} \subset \overline{\Theta}$. ∎

Proof of Theorem 2. By Condition 5 and (6.3), the event that $E(Y_1|X_1) = E(Y_2|X_2)$ also occurs with zero probability. It thus follows that $\Theta_0 = \Theta$.

We now show that $\Theta = \widetilde{\Theta}$. Suppose that $b \in \Theta$. Then with probability 1,

$$X_1'b \geq X_2'b \Longleftrightarrow E(Y_1|X_1) \geq E(Y_2|X_2). \tag{A.3}$$

Note that, for all $\gamma \in \Gamma$,

$$\begin{aligned} & E(Y_1|X_1'b, X_1'\gamma, X_2'b, X_2'\gamma) \\ = \; & E\left[\left(E\left(Y_1|X_1,X_2\right) - E\left(Y_2|X_1,X_2\right)\right) + E\left(Y_2|X_1,X_2\right)|X_1'b, X_1'\gamma, X_2'b, X_2'\gamma\right] \\ = \; & E\left[\left(E\left(Y_1|X_1\right) - E\left(Y_2|X_2\right)\right) + E\left(Y_2|X_2\right)|X_1'b, X_1'\gamma, X_2'b, X_2'\gamma\right] \end{aligned} \tag{A.4}$$

where (A.4) follows from statistical independence between (Y_1, X_1) and (Y_2, X_2). Using (A.3) and (A.4), we then have that, with probability 1,

$$X_1'b \geq X_2'b \Longleftrightarrow E(Y_1|X_1'b, X_1'\gamma, X_2'b, X_2'\gamma) \geq E[E\left(Y_2|X_2\right)|X_1'b, X_1'\gamma, X_2'b, X_2'\gamma]. \tag{A.5}$$

Using again the independence between (Y_1, X_1) and (Y_2, X_2), it follows that, for all b and $\gamma \in \Gamma$,

$$E(Y_1|X_1'b, X_1'\gamma, X_2'b, X_2'\gamma) = E(Y_1|X_1'b, X_1'\gamma) \tag{A.6}$$

and

$$E[E\left(Y_2|X_2\right)|X_1'b, X_1'\gamma, X_2'b, X_2'\gamma] = E\left[E\left(Y_2|X_2\right)|X_2'b, X_2'\gamma\right] = E(Y_2|X_2'b, X_2'\gamma). \tag{A.7}$$

Putting together (A.5), (A.6) and (A.7), we can deduce that $b \in \widetilde{\Theta}$ and thus $\Theta \subset \widetilde{\Theta}$.

It remains to show that $\widetilde{\Theta} \subset \Theta$. Note that, because $\beta \in \Gamma$, it follows from (A.4) that, for all $b \in \Gamma$,

$$\begin{aligned} & E(Y_1|X_1'b, X_1'\beta, X_2'b, X_2'\beta) \\ = \; & E\left[\left(E\left(Y_1|X_1\right) - E\left(Y_2|X_2\right)\right) + E\left(Y_2|X_2\right)|X_1'b, X_1'\beta, X_2'b, X_2'\beta\right]. \end{aligned}$$

Thus, using (6.2), (6.3), (6.4), (A.6) and (A.7), we have that, with probability 1,

$$X_1'\beta > X_2'\beta \Longleftrightarrow E(Y_1|X_1'b, X_1'\beta) > E(Y_2|X_2'b, X_2'\beta), \tag{A.8}$$
$$X_1'\beta = X_2'\beta \Longleftrightarrow E(Y_1|X_1'b, X_1'\beta) = E(Y_2|X_2'b, X_2'\beta), \tag{A.9}$$
$$X_1'\beta < X_2'\beta \Longleftrightarrow E(Y_1|X_1'b, X_1'\beta) < E(Y_2|X_2'b, X_2'\beta). \tag{A.10}$$

Hence, for all $b \in \Gamma$, we have that, with probability 1,

$$X_1'\beta \geq X_2'\beta \Longleftrightarrow E(Y_1|X_1'b, X_1'\beta) \geq E(Y_2|X_2'b, X_2'\beta). \tag{A.11}$$

Note that, by Condition 5 and (A.9), the event that $E(Y_1|X_1'b, X_1'\beta) = E(Y_2|X_2'b, X_2'\beta)$ also occurs with zero probability. Using (A.11) and the presumption that $\beta \in \Gamma$, we have that, for any $b \in \widetilde{\Theta}$,

$$X_1'b \geq X_2'b \Longleftrightarrow E(Y_1|X_1'b, X_1'\beta) \geq E(Y_2|X_2'b, X_2'\beta) \Longleftrightarrow X_1'\beta \geq X_2'\beta. \tag{A.12}$$

Because (A.3) holds when $b = \beta$, it therefore follows from (A.12) that $\widetilde{\Theta} \subset \Theta$. ∎

A.2 Illustrating examples for non-equivalence of the sets $\underline{\Theta}$, Θ and $\overline{\Theta}$

Recall that Γ denotes the space of preference parameter vectors b of which the magnitude of the first element is equal to 1.

Example 1: Θ can be a proper subset of $\overline{\Theta}$

Let $X = (X_1, X_2)$ be a bivariate vector where $X_1 \sim U(0,1)$, $X_2 \sim U(-1,1)$ and X_1 is stochastically independent of X_2. Assume that $\beta = (1,1)$ and $\varepsilon = \sqrt{1 + X_2^2}\xi$ where ξ is a random variable independent of X and has distribution function $F_\xi(t)$ defined as

$$F_\xi(t) \equiv \begin{cases} G_1(t) \ \text{ if } t \in (-\infty, -1] \\ \tau + ct \text{ if } t \in (-1, 1] \\ G_2(t) \text{ if } t \in (1, \infty) \end{cases} \tag{A.13}$$

where $c \in (0, \min\{\tau, 1-\tau\})$ is a fixed real constant, G_1 and G_2 are continuous differentiable and strictly increasing functions defined on the domains that include the intervals $(-\infty, -1]$ and $(1, \infty)$, respectively, and satisfy that

$$G_1(-1) = \tau - c, \ \lim_{t \longrightarrow -\infty} G_1(t) = 0, \ G_2(1) = \tau + c, \text{ and} \lim_{t \longrightarrow \infty} G_2(t) = 1. \tag{A.14}$$

Consider the value $\widetilde{b} \equiv (1, 0)$. Note that $X'\beta = X_1 + X_2$ can take negative value with positive probability but $X' \widetilde{b} = X_1$ is almost surely positive. It hence follows that $\widetilde{b} \notin \Theta$ by (2.6). Moreover, for each s in the support of the distribution of $X'\widetilde{b}$,

$$\begin{aligned} & P(Y = 1 | X'\widetilde{b} = s) & \text{(A.15)} \\ = \ & E\left[F_\xi\left((1 + X_2^2)^{-1/2}(s + X_2)\right) | X_1 = s\right] & \text{(A.16)} \\ = \ & \int_{-1}^{1} F_\xi\left((1 + u^2)^{-1/2}(s + u)\right) du/2 & \text{(A.17)} \\ \geq \ & \int_{-1}^{1} F_\xi\left(u(1 + u^2)^{-1/2}\right) du/2, & \text{(A.18)} \end{aligned}$$

where (A.18) follows from the fact that $X' \widetilde{b} = X_1 \sim U(0, 1)$ so that $s \geq 0$. Note that for each $u \in (-1, 1)$, $u(1 + u^2)^{-1/2}$ also falls within the interval $(-1, 1)$. Therefore by (A.13), the term on the right hand side of (A.18) equals

$$\int_{-1}^{1} \left[\tau + cu(1 + u^2)^{-1/2}\right] du/2 = \tau. \tag{A.19}$$

Hence, $\text{sgn}[X'\widetilde{b}] = \text{sgn}[P(Y = 1|X'\widetilde{b}) - \tau]$ almost surely and we have that $\widetilde{b} \in \overline{\Theta}$.

Example 2: $\underline{\Theta}$ can be a proper subset of Θ

Let $X = (X_1, X_2, X_3)$ be a trivariate vector where $X_1 \sim U(-1, 1)$, $X_2 \sim U(-1, 1)$ and

$$X_3 \equiv \begin{cases} \widetilde{X}_{3,1} \text{ if } X_1 + X_2 \geq 0 \\ \widetilde{X}_{3,2} \text{ if } X_1 + X_2 < 0 \end{cases} \tag{A.20}$$

where $\widetilde{X}_{3,1} \sim U(1,2)$, $\widetilde{X}_{3,2} \sim U(-2,-1)$ and the random variables X_1, X_2, $\widetilde{X}_{3,1}$ and $\widetilde{X}_{3,2}$ are independent. Assume that $\beta = (1,1,0)$ and $\varepsilon = \sqrt{1+X_2^2}\xi$ where ξ is a random variable independent of X and has the same distribution function F_ξ as defined by (A.13). Consider the value $\widetilde{b} \equiv (1,0,1)$. By design, $X'\beta$ and $X'\widetilde{b}$ have the same sign almost surely and hence $\widetilde{b} \in \Theta$. Now consider the vector $\gamma \equiv (1,0,0)$. Since $X'\gamma = X_1$, by (A.15) - (A.18) and the arguments yielding the bound (A.19) in Example 1, it also follows that

$$P(Y=1|X'\gamma = s) \geq \tau \text{ for } s \geq 0.$$

Note that the event $\{X'\widetilde{b} < 0 \text{ and } X_1 > 0\}$ can occur with positive probability. Therefore we have that $\widetilde{b} \notin \underline{\Theta}$.

A.3 Construction of a confidence set for the true value β in the monotone single index model

In this section, we briefly discuss how to construct a confidence set for the true value β in the monotone single index model. Let Γ_X denote the support of the distribution of X. Let $v \equiv (s,t,\gamma)$ and $\mathcal{V} \equiv \{(s,t,\gamma) : (s,t) \in \Gamma_X \times \Gamma_X, \gamma \in \Gamma\}$. Assume the set $\mathcal{V}$ is nonempty and compact. Define

$$\begin{aligned} m_b(v) \;\;\equiv\;\; & (s-t)'\,b\,[E(Y|X'b = s'b, X'\gamma = s'\gamma) - E(Y|X'b = t'b, X'\gamma = t'\gamma)] \\ & \times f_{b,\gamma}(s'b, s'\gamma)\, f_{b,\gamma}(t'b, t'\gamma), \end{aligned}$$

where the function $f_{b,\gamma}$ denotes the joint density function of the indexing variables $(X'b, X'\gamma)$. Note that for almost every $v \in \mathcal{V}$,

$$\begin{aligned} & m_b(v) \geq 0 \\ & \Longleftrightarrow (s-t)'\,b\,[E(Y|X'b = s'b, X'\gamma = s'\gamma) - E(Y|X'b = t'b, X'\gamma = t'\gamma)] \geq 0. \end{aligned}$$

Thus the set $\widetilde{\Theta}$ defined in Section 6 can be equivalently formulated as the following set

$$\{b \in \Gamma : m_b(v) \geq 0 \text{ for almost every } v \in \mathcal{V}\}. \tag{A.21}$$

Assume that we observe a random sample of individual outcomes and covariates $(Y_i, X_i)_{i=1,...,n}$ that are generated from a monotone single index model defined by (6.1). We now construct a set estimator $\widehat{\Theta}$ at the $(1-\alpha)$ confidence level such that

$$\liminf_{n \longrightarrow \infty} P(\beta \in \widehat{\Theta}) \geq 1 - \alpha$$

by inverting a CLR based test of the conditional moment inequalities in (A.21). The confidence set construction principle here is analogous to that described in Section 4. To avoid repetition, we mainly present the formulae for the relevant components in the implementation of $\widehat{\Theta}$.

Let

$$\widehat{m}_b(v) \equiv \{nh_n(b)h_n(\gamma)\}^{-1} (s-t)' b \sum_{i=1}^{n} \left[Y_i K_n(X_i, s, b, \gamma) \widehat{f}_{b,\gamma}(t) - Y_i K_n(X_i, t, b, \gamma) \widehat{f}_{b,\gamma}(s) \right],$$

where

$$\begin{aligned} \widehat{f}_{b,\gamma}(x) &\equiv \{nh_n(b)h_n(\gamma)\}^{-1} \sum_{i=1}^{n} K_n(X_i, x, b, \gamma), \\ K_n(X_i, x, b, \gamma) &\equiv K\left(\frac{x'b - X_i'b}{h_n(b)}, \frac{x'\gamma - X_i'\gamma}{h_n(\gamma)} \right), \end{aligned}$$

$K(\cdot, \cdot)$ is a bivariate kernel function, and $h_n(\gamma)$ is a sequence of bandwidths for each γ. Define

$$T(b) \equiv \inf_{v \in \mathcal{V}} \frac{\widehat{m}_b(v)}{\widehat{\sigma}_b(v)},$$

where

$$\begin{aligned} \widehat{\sigma}_b^2(v) &\equiv n^{-2}[h_n(b)]^{-2}[h_n(\gamma)]^{-2} \left((s-t)' b\right)^2 \sum_{i=1}^{n} \widehat{\omega}_i^2(b, v), \\ \widehat{\omega}_i(b, v) &\equiv \widehat{u}_i(b, \gamma) \left[K_n(X_i, s, b, \gamma) \widehat{f}_{b,\gamma}(t) - K_n(X_i, t, b, \gamma) \widehat{f}_{b,\gamma}(s) \right], \\ \widehat{u}_i(b, \gamma) &\equiv Y_i - \left[\sum_{j=1}^{n} K_n(X_j, X_i, b, \gamma) \right]^{-1} \sum_{j=1}^{n} Y_i K_n(X_j, X_i, b, \gamma). \end{aligned}$$

Let B be the number of simulation repetitions. For each repetition $r \in \{1, ..., B\}$, we draw an n dimensional vector of mutually independently standard normally distributed random variables which are also independent of the data. Let $\eta(r)$ denote this vector. For any compact set $\mathsf{V} \subseteq \mathcal{V}$, define

$$T_r^*(b;\mathsf{V}) \equiv \inf_{v\in\mathsf{V}} \left[\{nh_n(b)h_n(\gamma)\widehat{\sigma}_b(v)\}^{-1} (s-t)' \, b \sum_{i=1}^{n} \eta_i(r)\widehat{\omega}_i(b,v) \right].$$

Let $\widehat{q}_p(b,\mathsf{V})$ be the p level empirical quantile based on the vector $(T_r^*(b;\mathsf{V}))_{r\in\{1,\ldots,B\}}$. Let

$$\widehat{V}_n(b) \equiv \{v \in \mathcal{V} : \widehat{m}_b(v) \le -2\widehat{q}_{\gamma_n}(b,\mathcal{V})\widehat{\sigma}_b(v)\},$$

where $\gamma_n \equiv 0.1/\log n$. Then as before, we construct the $(1-\alpha)$ confidence set $\widehat{\Theta}$ for the true value β in the monotone single index model by setting

$$\widehat{\Theta} \equiv \left\{ b \in \Gamma : T(b) \ge \widehat{q}_\alpha(b, \widehat{V}_n(b)) \right\}.$$

References

Abrevaya, J. (1999): "Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable," *Journal of Econometrics*, 93(2), 203–228.

——— (2000): "Rank estimation of a generalized fixed-effects regression model," *Journal of Econometrics*, 95(1), 1–23.

Andrews, D. W., and X. Shi (2013): "Inference based on conditional moment inequalities," *Econometrica*, 81(2), 609–666.

Andrews, D. W., and G. Soares (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78(1), 119–157.

Armstrong, T. B. (2014): "Weighted KS statistics for inference on conditional moment inequalities," *Journal of Econometrics*, 181(2), 92–116.

——— (2015): "Asymptotically exact inference in conditional moment inequality models," *Journal of Econometrics*, 186(1), 51–65.

——— (2016): "On the choice of test statistic for conditional moment inequalities," Cowles Foundation Discussion Paper No. 1960R.

Armstrong, T. B., and H. P. Chan (2016): "Multiscale adaptive inference on conditional moment inequalities," *Journal of Econometrics*, 194(1), 24–43.

Blevins, J. R. (2015): "Non-Standard Rates of Convergence of Criterion-Function-Based Set Estimators," *Econometrics Journal*, 18, 172–199.

Chen, S. (2010): "An integrated maximum score estimator for a generalized censored quantile regression model," *Journal of Econometrics*, 155(1), 90–98.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection bounds: estimation and inference," *Econometrica*, 81(2), 667–737.

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

——— (2017): "Adaptive Tests of Conditional Moment Inequalities," *Econometric Theory*, pp. 1–42.

Coppejans, M. (2001): "Estimation of the binary response model using a mixture of distributions estimator (mod)," *Journal of Econometrics*, 102(2), 231–269.

Cosslett, S. R. (1983): "Distribution-free maximum likelihood estimator of the binary choice model," *Econometrica*, pp. 765–782.

Fox, J. T. (2007): "Semiparametric estimation of multinomial discrete-choice models using a subset of choices," *RAND Journal of Economics*, pp. 1002–1019.

Goeree, J. K., C. A. Holt, and T. R. Palfrey (2005): "Regular quantal response equilibrium," *Experimental Economics*, 8(4), 347–367.

Han, A. K. (1987): "Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator," *Journal of Econometrics*, 35(2-3), 303–316.

Horowitz, J. L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60(3), 505–531.

——— (1998): *Semiparametric Methods in Econometrics*. Springer.

Ichimura, H. (1993): "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models," *Journal of Econometrics*, 58(1-2), 71–120.

Klein, R. W., and R. H. Spady (1993): "An efficient semiparametric estimator for binary response models," *Econometrica*, pp. 387–421.

Komarova, T. (2013): "Binary choice models with discrete regressors: Identification and misspecification," *Journal of Econometrics*, 177(1), 14–33.

Lee, M.-j. (1992): "Median regression for ordered discrete response," *Journal of Econometrics*, 51(1-2), 59–77.

Lee, S., K. Song, and Y.-J. Whang (2013): "Testing functional inequalities," *Journal of Econometrics*, 172(1), 14–32.

——— (2017): "Testing for a general class of functional inequalities," *Econometric Theory*, forthcoming.

Linton, O., K. Song, and Y.-J. Whang (2010): "An improved bootstrap test of stochastic dominance," *Journal of Econometrics*, 154(2), 186–202.

Manski, C. F. (1975): "Maximum score estimation of the stochastic utility model of choice," *Journal of Econometrics*, 3(3), 205–228.

——— (1985): "Semiparametric analysis of discrete response. Asymptotic properties of the maximum score estimator," *Journal of Econometrics*, 27(3), 313–333.

——— (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55(2), 357–362.

——— (1988): "Identification of binary response models," *Journal of the American Statistical Association*, 83(403), 729–738.

Manski, C. F. (2003): *Partial identification of probability distributions*. Springer Science & Business Media.

——— (2007): *Identification for prediction and decision*. Harvard University Press.

Matzkin, R. L. (1993): "Nonparametric identification and estimation of polychotomous choice models," *Journal of Econometrics*, 58(1-2), 137–168.

Menzel, K. (2014): "Consistent estimation with many moment inequalities," *Journal of Econometrics*, 182(2), 329–350.

PAKES, A., AND J. PORTER (2016): "Moment inequalities for multinomial choice with fixed effects," Discussion paper, National Bureau of Economic Research.

POWELL, J. L., J. H. STOCK, AND T. M. STOKER (1989): "Semiparametric estimation of index coefficients," *Econometrica: Journal of the Econometric Society*, pp. 1403–1430.

TAMER, E. (2010): "Partial identification in econometrics," *Annu. Rev. Econ.*, 2(1), 167–195.