Chen, Le-Yu; Lee, Sokbae

Working Paper

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

cemmap working paper, No. CWP26/15

Provided in Cooperation with:
Institute for Fiscal Studies (IFS), London

Suggested Citation: Chen, Le-Yu; Lee, Sokbae (2015) : Breaking the curse of dimensionality in conditional moment inequalities for discrete choice models, cemmap working paper, No. CWP26/15, Centre for Microdata Methods and Practice (cemmap), London, https://doi.org/10.1920/wp.cem.2015.2615

This Version is available at:
https://hdl.handle.net/10419/130038

cemmap
centre for microdata methods and practice

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 CWP26/15

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

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, Seoul National University

Centre for Microdata Methods and Practice, Institute for Fiscal Studies

June 2015

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 also extend this dimension reducing characterization result 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, preference parameters, curse of dimensionality*

*We thank Hidehiko Ichimura, Kengo Kato, Toru Kitagawa, Dennis Kristensen, Adam Rosen, Kyungchul Song and participants at SNU Workshop on Advances in Microeconometrics for helpful comments on this work. Financial support from the European Research Council (ERC) grant ERC-2009-StG-240910-ROMETA is gratefully acknowledged.

†E-mail: lychen@econ.sinica.edu.tw

‡E-mail: sokbae@snu.ac.kr

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 et al. (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 random 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 the issue of dimension reduction in the context of conditional moment inequalities.

When X contains several continuous covariates but their support is limited, in order to carry out inference based on the conditional moment inequalities in (1.2), we need to deal with the nonparametric conditional expectation $E(Y|X) = P(Y = 1|X)$. For example, Chernozhukov et al. (2013, henceforth CLR) plug in nonparametric (kernel or series based) estimators to form the sup-norm or one-sided Kolmogorov-Smirnov type statistic. Since 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, this can lead to misspecification that 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. For instance, when the method of CLR is utilized with (1.2), recall that 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 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. Section 6 concludes and Section A contains the proofs and some further results.

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 $Y = 1\{X'\beta \geq \varepsilon\}$. Let Γ_X denotes the support of X, and assume that at least one element of X is continuously distributed conditional on all the other elements. Write $X = (X_1, \widetilde{X})$ where the distribution of X_1 conditional on $\widetilde{X} = \widetilde{x}$ is absolutely continuous with respect to the Lebesgue measure for almost every realization $\widetilde{x}$.

Let b denote a generic element of Γ and write $b = (b_1, \widetilde{b})$. Assume that $b_1 = 1$ for all $b \in \Gamma$. Note that these assumptions ensure that $X'b$ is a continuous random variable for any $b \in \Gamma$. Let $Q_\tau(U|V)$ denote the τ quantile of the distribution of a random variable U conditional on a random vector V. We impose the following modeling restrictions.

Condition 1. *For all $x \in \Gamma_X$, the distribution of ε conditional on $X = x$ admits a density that is everywhere positive on $\mathbb{R}$ and satisfies that $Q_\tau(\varepsilon|X = x) = 0$ for some $\tau \in (0, 1)$.*

Condition 1, due to Manski (1985, 1988), allows for nonparametric specification of the preference shock with a general form of heteroskedasticity. It is known that point identification of β requires that X_1 should have sufficient variation conditional on the other covariates (see e.g., Manski (1985) and Horowitz (1998)). Nevertheless, the model induces restrictions on possible values of data generating preference parameters, which results in set identification of β. To see this, note that Condition 1 implies that for all $x \in \Gamma_X$,

$$P(Y = 1|X = x) > \tau \Leftrightarrow x'\beta > 0, \tag{2.1}$$

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

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

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

$$\Theta = \{b \in \Gamma : X'b\,[P(Y = 1|X) - \tau] \geq 0 \text{ almost surely}\}. \tag{2.4}$$

Given (2.1), (2.2) and (2.3), we also have that

$$\Theta = \{b \in \Gamma : b'XX'\beta \geq 0 \text{ almost surely}\}. \tag{2.5}$$

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.

On the other hand, the restrictions (2.1), (2.2) and (2.3) imply that

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

In other words, the model induces the semiparametric restriction that with probability 1,

$$\text{sgn}[P(Y = 1|X) - \tau] = \text{sgn}[P(Y = 1|X'\beta) - \tau] = \text{sgn}[X'\beta], \tag{2.6}$$

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 (2.6) 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

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

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. *Assume $X'b$ is a continuous random variable for any $b \in \Gamma$. Then under Condition 1, we have that $\Theta = \widetilde{\Theta}$.*

To explain the characterization result of Lemma 1, note that the model (1.1) under

Condition 1 implies that for any $\gamma \in \Gamma$,

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

Thus, intuitively speaking, for any b that is observationally equivalent to β, equation (2.7) should also hold for b in place of β in the statement. Define

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

By similar arguments used in the proof of Lemma 1, it is straightforward to show that

$$\underline{\Theta} \subset \Theta = \widetilde{\Theta} \subset \overline{\Theta}. \tag{2.8}$$

It is interesting to note that the set inclusion in (2.8) can be strict as demonstrated in the examples of Appendix A.2. Namely, the set $\underline{\Theta}$ is too restrictive in the sense that it may exclude the true data generating parameter value β, whereas the set $\overline{\Theta}$ is not sharp and thus inference based on $\overline{\Theta}$ could admit some b values that are incompatible with the model restrictions given by (2.4).

The identifying relationship in (2.8) can be viewed as a conditional moment inequality analog of well-known index restrictions in semiparametric binary response models (e.g., Cosslett (1983), Powell et al. (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 heteroskedasticty. 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 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,*

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

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

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

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. 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

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

The following theorem generalizes the result of Lemma 1.

Theorem 1. *Given Assumptions 1 and 2, we have that $\Theta_0 = \Theta = \widetilde{\Theta}$.*

By taking $G(X, c, b) = X'b$ and $H(Y, c) = Y - \tau$, both being independent of c, it is easily seen that the binary choice model of Section 2 fits within this setting. Other examples are set forth as follows.

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

$$\lambda_{c-1} < X'\theta + \varepsilon \leq \lambda_c \tag{3.5}$$

where $\lambda_0 = -\infty < \lambda_1 < .... < \lambda_K < \lambda_{K+1} = \infty$. Let $\lambda \equiv (\lambda_1, ..., \lambda_K)$ be the vector of threshold parameters. Let Y be the observed choice. 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 1. Using this restriction, we see that Assumption 1 holds with $C = \{1, ..., K\}$, $H(Y, c) = 1\{Y \leq c\} - \tau$ and $G(X, c, b) = \widetilde{X}_c'b$ 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

$$U_j = {X_j}'\beta + \varepsilon_j \tag{3.6}$$

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

$${X_s}'\beta > X_t'\beta \iff P(Y = s|X) > P(Y = t|X). \tag{3.7}$$

Manski (1975), Matzkin (1993) and Fox (2007) used Condition 2 as an identifying restriction in the multinomial choice model to allow for nonparametric unobservables with unknown form of heteroskedasticity. Goeree et. al. (2005, Proposition 5, p. 359) showed that it suffices for Condition 2 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 2, Assumption 1 holds for this example by taking $C \equiv \{(s,t) \in \{1,...,K\}^2 : s < t\}$, $G(X,s,t,b) = (X_s - X_t)'b$ 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

$$Y_t = 1\{X_t'\beta + v \geq \varepsilon_t\},\ t \in \{1,...,T\} \tag{3.8}$$

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

Under Condition 3 and by Lemma 1 of Manski (1987), Assumption 1 holds for this example by taking $C \equiv \{(s,t) \in \{1,...,T\}^2 : s < t\}$, $G(X,s,t,b) = (X_s - X_t)'b$ 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

$$Y_t = \sum_{j=1}^{K+1} j1\{\lambda_{j-1} < X_t'\beta + v + \varepsilon_t \leq \lambda_j\}, \tag{3.9}$$

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 3. Under this restriction and by applying the law of iterated expectations, we see that Assumption 1 holds for this example by taking $C = \{(k,s,t) : k \in \{1,...,K\},\ (s,t) \in \{1,...,T\}^2 \text{ such that } s < t\}$, $G(X,k,s,t,b) = (X_t - X_s)'b$ 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 à la CLR. 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

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

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

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

Thus we have that

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

Assume that we observe a random sample of individual outcomes and covariates $(Y_i, X_i)_{i=1,...,n}$. 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. Let $K(.,.)$ denote a bivariate kernel function and h_n be a bandwidth sequence. To estimate the function m_b, we consider the following kernel type

estimator:

$$\widehat{m}_b(v) \equiv \left(nh_n^2\right)^{-1} \sum_{i=1}^{n} G(X_i, c, b) H(Y_i, c) K_n(X_i, v, b), \tag{4.3}$$

where

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

Let $u_i(b, c, \gamma) \equiv H(Y_i, c) - E(H(Y_i, c)|G(X_i, c, b), G(X_i, c, \gamma))$. Define

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

where

$$\begin{aligned}
\widehat{\sigma}_b^2(v) &\equiv n^{-2} h_n^{-4} \sum_{i=1}^{n} \widehat{u}_i^2(b, c, \gamma) G^2(X_i, c, b) K_n^2(X_i, v, b), \\
\widehat{u}_i(b, c, \gamma) &\equiv H(Y_i, c) - \left[\sum_{j=1}^{n} K_n(X_j, (X_i, \gamma, c), b)\right]^{-1} \sum_{j=1}^{n} H(Y_j, c) K_n(X_j, (X_i, \gamma, c), b).
\end{aligned}$$

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

$$T_s^*(b; \mathsf{V}) \equiv \inf_{v \in \mathsf{V}} \left[\left(nh^2 \widehat{\sigma}_b(v)\right)^{-1} \sum_{i=1}^{n} \eta_i(s) \widehat{u}_i(b, c, \gamma) G(X_i, c, b) K_n(X_i, v, b)\right]. \tag{4.6}$$

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

$$\widehat{V}_n(b) \equiv \{v \in \mathcal{V} : \widehat{m}_b(v) \leq -2\widehat{q}_{\gamma_n}(b, \mathcal{V})\widehat{\sigma}_b(v)\}, \tag{4.7}$$

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 (2014)), the generalized moment selection approach (e.g., Andrews and Soares (2010), Andrews and Shi (2013), and Chetverikov (2011)), and the iterative step-down approach (e.g., Chetverikov (2013)) 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

$$\widehat{\Theta} \equiv \left\{ b \in \Gamma : T(b) \geq \widehat{q}_\alpha(b, \widehat{V}_n(b)) \right\}. \tag{4.8}$$

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.

5 Simulation results

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 $d \geq 2$, and

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

where 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

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

where U is a uniformly distributed random variable on the interval $[0, 1]$ and is independent of $(\widetilde{X}, \xi)$. We set $\beta_1 = 1$ and $\beta_k = 0$ for $k \in \{2, ..., d\}$. The preference parameter space is specified to be

$$\Gamma \equiv \{b \in \mathbb{R}^d : b_1 = 1, (b_2, ..., b_d) \in [-1, 1]^{d-1}\}. \tag{5.2}$$

Under this setup, the sign of the true index $X'\beta = X_1$ is determined by X_2. By inspecting the formulation (2.5), the identified set Θ is thus given by

$$\Theta = \{b \in \Gamma : b_2 \geq 0 \text{ and } b_k = 0 \text{ for } k \in \{3, ..., d\}\}. \tag{5.3}$$

Recall that this 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 by

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

and $\widehat{s}(W)$ denotes the estimated standard deviation for the random variable W. The bandwidth sequence h_n is specified by

$$h_n = c_{Index} n^{-1/5} \tag{5.4}$$

where c_{Index} is a bandwidth scale. The rate considered in (5.4) corresponds to the

undersmoothing specification under the assumption that the true conditional expectation function is twice continuously differentiable.

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

$$m_{b,Full}(x) \equiv E\left(X'b\left(Y-0.5\right)|X=x\right)f_X\left(x\right) \tag{5.5}$$

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

$$\widehat{m}_{b,Full}(x) \equiv \left(nh_n^d\right)^{-1}\sum_{i=1}^{n} X_i'b\left(Y_i-0.5\right)K_{n,Full}(X_i,x), \tag{5.6}$$

where

$$K_{n,Full}(X_i,x) \equiv \prod_{k=1}^{d}\widetilde{K}_{Full}\left(\frac{x_k-X_{i,k}}{\widehat{s}(X_{i,k})h_{n,Full}}\right), \tag{5.7}$$

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

$$h_{n,Full} = c_{Full}n^{-r} \tag{5.8}$$

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

$$T_{Full}(b) \equiv \inf_{x\in\Gamma_X}\frac{\widehat{m}_{b,Full}(x)}{\widehat{\sigma}_{b,Full}(x)}, \tag{5.9}$$

where

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

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 $d \in \{3, 4, 5, 10\}$. 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 CLR test power. 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 CLR test 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.

6 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 extend this dimension reducing characterization result 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.

We note that moment inequalities (3.4) for the general framework of Section 3 can also be applied to monotone transformation models (e.g., see Abrevaya (1999, 2000), Chen (2010) and Pakes and Porter (2014, Section 2)). Hence, our dimension reducing approach would also be useful in that context. 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 and Chan (2013), Armstrong (2014, 2015), Chetverikov (2011), Lee, Song, and Whang (2013, 2014) and Menzel (2014) among others. It will be an interesting further research topic to incorporate these alternative methods with the dimension reducing characterization result for set inference of the class of semiparametric models studied in Section 3 of this paper.

A Appendix

A.1 Proofs

Proof of Lemma 1. Lemma 1 can be proved by applying Theorem 1 and noting that Assumptions 1 and 2 of Theorem 1 are satisfied under Condition 1 and the assumption that $X'b$ is a continuous random variable for each $b \in \Gamma$. ∎

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 $\Theta_0 = \Theta$. Therefore, if $b \in \Theta$, 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, $b \in \widetilde{\Theta}$.

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 have that $b \in \Theta$. ∎

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 whose 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

$$F_\xi(t) \equiv \begin{cases} G_1(t) \ \text{ if } t \in (-\infty,-1] \\ \tau + ct \text{ if } t \in (-1,1] \\ G_2(t) \text{ if } t \in (1,\infty) \end{cases} \tag{A.3}$$

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

$$G_1(-1) = \tau - c, \lim_{t \longrightarrow -\infty} G_1(t) = 0, G_2(1) = \tau + c, \text{ and } \lim_{t \longrightarrow \infty} G_2(t) = 1. \tag{A.4}$$

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 the definition (2.5). Moreover for $s \geq 0$,

$$\begin{aligned}
& P(Y = 1 | X'\widetilde{b} = s) && \text{(A.5)} \\
= \; & E\left[F_\xi\left((1 + X_2^2)^{-1/2}(s + X_2)\right) | X_1 = s\right] && \text{(A.6)} \\
= \; & \int_{-1}^{1} F_\xi\left((1 + u^2)^{-1/2}(s + u)\right) du/2 && \text{(A.7)} \\
\geq \; & \int_{-1}^{1} F_\xi\left(u(1 + u^2)^{-1/2}\right) du/2. && \text{(A.8)}
\end{aligned}$$

Note that for each $u \in (-1, 1)$, $u(1 + u^2)^{-1/2}$ also falls within the interval $(-1, 1)$. Therefore by (A.3), the term on the right hand side of (A.8) equals

$$\int_{-1}^{1} \left[\tau + cu(1 + u^2)^{-1/2}\right] du/2 = \tau. \tag{A.9}$$

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

$$X_3 \equiv \begin{cases} \widetilde{X}_{3,1} \text{ if } X_1 + X_2 \geq 0 \\ \widetilde{X}_{3,2} \text{ if } X_1 + X_2 < 0 \end{cases} \tag{A.10}$$

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.3). 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.5) - (A.8) and the arguments yielding the bound (A.9) in Example

1, it also follows that

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

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

References

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

[2] Abrevaya, J. (2000), "Rank estimation of a generalized fixed-effects regression model," *Journal of Econometrics*, 95, 1-23.

[3] Andrews, D., and Shi X. (2013), "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609-666.

[4] Andrews, D. and Soares, G. (2010), "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119-157.

[5] Armstrong, T. B. and Chan H. P. (2013), "Multiscale adaptive inference on conditional moment inequalities," Working paper.

[6] Armstrong, T. B. (2014), "Weighted KS statistics for inference on conditional moment inequalities," *Journal of Econometrics*, 181, 92-116.

[7] Armstrong, T. B. (2015), "Asymptotically exact inference in conditional moment inequality models," *Journal of Econometrics*, 186, 51-65.

[8] Blevins, J. R. (2015), "Non-standard rates of convergence of criterion-function-based set estimators for binary response models," *Econometrics Journal*, forthcoming.

[9] Chen, S. (2010), "An integrated maximum score estimator for a generalized censored quantile regression model," *Journal of Econometrics*, 155, 90-98.

[10] Chernozhukov, V., Lee, S. and Rosen, A. M. (2013), "Intersection Bounds: Estimation and Inference," *Econometrica*, 81, 667-737.

[11] Chetverikov, D. (2011), "Adaptive Rate Optimal Test of Conditional Moment Inequalities," Working paper.

[12] Chetverikov, D. (2013), "Testing Regression Monotonicity in Econometric Models," Working paper.

[13] Coppejans, M. (2001), "Estimation of the Binary Response Model Using a Mixture of Distributions Estimator (MOD)," *Journal of Econometrics*, 102, 231-261.

[14] Cosslett, S. R. (1983), "Distribution-free Maximum Likelihood Estimator of the Binary Choice Model, " *Econometrica*, 51, 765-782.

[15] Fox, J. T. (2007), "Semiparametric estimation of multinomial discrete-choice models using a subset of choices," *RAND Journal of Economics*, 38, 1002-1019.

[16] Goeree, J. K., Holt, C. A., and Palfrey, T. R. (2005), "Regular Quantal Response Equilibrium," *Experimental Economics*, 8, 347-367.

[17] Han, A. (1987), "Nonparametric Analysis of a Generalized Regression Model," *Journal of Econometrics*, 35, 303-316.

[18] Hansen, B. E. (2005), "Exact Mean Integrated Squared Error of Higher Order Kernel Estimators," *Econometric Theory*, 21, 1031-1057.

[19] Horowitz, J. L. (1998), *Semiparametric Methods in Econometrics*. Vol. 131 of *Lecture Notes in Statistics*. New York: Springer-Verlag.

[20] Ichimura, H. (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models, " *Journal of Econometrics*, 58, 71-120.

[21] Klein, R. and Spady, R. (1993), "An Efficient Semiparametric Estimator for Binary Response Models, " *Econometrica*, 61, 387-421.

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

[23] Lee, M. J. (1992), "Median regression for ordered discrete response," *Journal of Econometrics*, 51, 59-77.

[24] Lee, S. Song, K. and Whang, Y. J. (2013), "Testing Functional Inequalities," *Journal of Econometrics*, 172, 14-32.

[25] Lee, S. Song, K. and Whang, Y. J. (2014), "Testing for a General Class of Functional Inequalities," Working paper.

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

[27] Manski, C. F. (1985), "Semiparametric analysis of discrete response. Asymptotic properties of the maximum score estimator," *Journal of Econometrics*, 27, 313-333.

[28] Manski, C. F. (1987), "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica*, 55, 357–362.

[29] Manski, C. F. (1988), "Identification of Binary Response Models," *Journal of the American Statistical Association*, 83, 729-738.

[30] Manski, C. F. (2003), *Partial Identification of Probability Distributions*. New York: Springer-Verlag.

[31] Manski, C. F. (2007), *Identification for Prediction and Decision*. Harvard University Press.

[32] Matzkin, R. (1993), "Nonparametric identification and estimation of polychotomous choice models," *Journal of Econometrics*, 58, 137-168.

[33] Menzel, K. (2014), "Consistent estimation with many moment inequalities," *Journal of Econometrics*, 182, 329-350.

[34] Pakes, A. and Porter, J. (2014), "Moment Inequalities for Multinomial Choice with Fixed Effects," Working Paper.

[35] Tamer, E. (2010), "Partial Identification in Econometrics," *Annual Review of Economics*, 2, 167-195.

[36] Powell, J. L., Stock, J. H., and T. M. Stoker (1989), "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403-1430.