Supplemental Appendix for "Estimating Semi-parametric Panel Multinomial Choice Models Using Cyclic Monotonicity"

Xiaoxia Shi
University of Wisconsin-Madison

Matthew Shum
Caltech

Wei Song
Xiamen University

September 26, 2017

Abstract

This supplemental appendix contains three sections. Section B presents the proof for Theorem 6.1 in the main text, the convergence rate theorem for the aggregate data case. Section C presents a necessary condition for point identification of the binary choice model. Section D reports Monte Carlo simulation results for the estimator using the Andrews and Shi (2013)-type instrumental functions on the Cauchy designs described in the main text.

B Proof of Theorem 6.1: Convergence Rate for the Aggregate Data Case

Proof of Theorem 6.1. First we define the limiting version of $Q_n(b)$ as $n \to \infty$:

$$Q(b) = E\left\{[(b'\Delta\mathbf{X}_c)E(\Delta\mathbf{Y}_{ic}|\eta_c)]_-\right\}. \tag{B.1}$$

Let B_δ stand for the set $\{b \in R^{d_x} : b_1 = 1, \|b - \beta\| \leq \delta\}$. Below we show the folowing results: given a positive number δ, for all $b \in B_\delta$,

(i) for all $\eta > 0$ (regardless of how small), $|Q_n(b) - Q(b) - (Q_n(\beta) - Q(\beta))| \leq \eta\|b-\beta\|^2 + O_p(n^{-1})$, and

(ii) $Q(b) - Q(\beta) \geq c_2\|b - \beta\|^2/2$, where c_2 is the constant in Assumption 6.2(f).

Choose $\eta < c_2/2$. Results (i) and (ii) together with Assumption 6.2(c) imply that, with probability approaching one,

$$\begin{aligned} c_2\|\widehat{\beta}-\beta\|^2/2 &\leq \eta\|\widehat{\beta}-\beta\|^2 + O_p(n^{-1}) + Q_n(\widehat{\beta}) - Q_n(\beta) \\ &\leq \eta\|\widehat{\beta}-\beta\|^2 + O_p(n^{-1}). \end{aligned} \tag{B.2}$$

where the second inequality holds because $\widehat{\beta}$ minimizes $Q_n(\cdot)$ and hence $Q_n(\widehat{\beta}) - Q_n(\beta) \leq 0$. Thus,

$$(c_2/2-\eta)\|\widehat{\beta}-\beta\|^2 \leq O_p(n^{-1}). \tag{B.3}$$

This proves the theorem because $(c_2/2-\eta) > 0$ by design.

Now we show result (i). Note that

$$\begin{aligned} Q_n(b) - Q(b) = n^{-1}\sum_{c=1}^{n}[b'(\Delta\mathbf{X}_c)(\Delta\mathbf{S}_c)]_- - E\left\{[b'(\Delta\mathbf{X}_c)(\Delta\mathbf{S}_c)]_-\right\} \\ + E\left\{[b'(\Delta\mathbf{X}_c)(\Delta\mathbf{S}_c)]_-\right\} - E\left\{[b'(\Delta\mathbf{X}_c)E(\Delta\mathbf{Y}_{ic}|\eta_c)]_-\right\}. \end{aligned} \tag{B.4}$$

The first two summands on the right hand-side together form an empirical process which we now call $\nu_n(b)$ and will analyze later. The absolute value of the rest of the right hand-side is bounded by

$$\begin{aligned} &E\left\{\|b'(\Delta\mathbf{X}_c)\|\|(\Delta\mathbf{S}_c) - E(\Delta\mathbf{Y}_{ic}|\eta_c)\|\right\} \\ &\leq 2\max_{t=1,2}(E\|\mathbf{S}_{ct} - E(\mathbf{Y}_{ict}|\eta_c)\|^2)^{1/2}(E\|b'(\Delta\mathbf{X}_c)\|^2)^{1/2} \\ &\leq O(n^{-1})\|b\|(E\|\mathrm{vec}(\mathbf{X}_{c2}-\mathbf{X}_{c1})\|^2)^{1/2}, \end{aligned} \tag{B.5}$$

where the first inequality holds by the Cauchy-Schwarz inequality, the second inequality holds by Assumption 6.2(b) and the Cauchy-Schwarz inequality. The last line is $o_p(n^{-1})$ uniformly over a $o_p(1)$ neighborhood of β by Assumption 6.2(a). Therefore, we have, uniformly over a $o_p(1)$ neighborhood of β,

$$Q_n(b) - Q(b) - (Q_n(\beta) - Q(\beta)) = \nu_n(b) - \nu_n(\beta) + O(n^{-1}). \tag{B.6}$$

We now bound $\nu_n(b) - \nu_n(\beta)$. Let V_c denote $(\Delta\mathbf{X}_c)(\Delta\mathbf{S}_c)$, and let the space of V_c be denoted by $\mathcal{V}$.

We first show that the class of functions $\mathcal{F} = \{f : \mathcal{V} \to R | f(v) = [b'v]_-, b \in R^{d_x}, b_1 = 1\}$ is a Vapnik-Cervonenkis (VC)-subgraph class of functions. To begin, observe that a similar but different class of functions with $\mathcal{F}$: $\mathcal{F}_0 = \{f : \mathcal{V} \to R | f(v) = -b'v\}$ is a VC-subgraph of VC-index at most $d_x + 1$ because the space is a vector space of dimension equal to the dimension of the set $\{b \in R^{d_x} : b_1 = 1\}$. That implies that the collection of subgraphs $\{(v, a) : -b'v > a\}$ of functions in $\mathcal{F}_0$ forms a VC-class of sets of dimension at most $d_x + 1$. We use this to show that the collection of subgraphs of functions in $\mathcal{F}$ is also a VC-class with finite dimension. The subgraph of a function in $\mathcal{F}$ is

$$S(b) = \{(v, a) : a < [b'v]_-\} = \{(v, a) : a < 0\} \cup \{(v, a) : a \geq 0, -b'v > a\}. \tag{B.7}$$

Consider m points $(v_1, a_1), \ldots, (v_m, a_m)$. In order for $\{S(b) : b \in R^{d_x}, b_1 = 1\}$ to shatter these m points, there can at most be one $j \in \{1, \ldots, m\}$ such that $a_j < 0$ because $(v, a) \in S(b)$ for all b as long as $a < 0$ and thus $S(b)$ with different b cannot pick out two different points with $a < 0$. Suppose without loss of generality that $a_1 < 0$. Then, the collection of sets $\{\{(v, a) : -b'v > a\} : b \in R^{d_x}, b_1 = 1\}$ must shatter the set $\{(v_2, a_2), \ldots, (v_m, a_m)\}$. But this collection of sets is the collection of subgraphs of functions in $\mathcal{F}_0$, which is of VC-dimension at most $d_x + 1$. Therefore, $m - 1$ can at most be $d_x + 1$. This implies that $m \leq d_x + 2$. Thus, $\mathcal{F}$ is a VC-subgraph with VC-index at most $d_x + 2$.

Next define

$$\mathcal{F}_\delta = \{f : \mathcal{V} \to R | f(v) = [v'b]_- - [v'\beta]_-, b \in R^{d_x}, b_1 = 1, \|b - \beta\| \leq \delta\}. \tag{B.8}$$

This collection of functions is a VC-class with the same VC-index as $\mathcal{F}$ due to Lemma 2.6.18 of van der Vaart and Wellner (1996). Consider the envelope function $F_\delta(v) = \|v\|\delta$. Then, Theorem 2.6.7 of van der Vaart of Wellner (1996) gives the polynomial bound on the covering number of $\mathcal{F}_\delta$:

$$N(\varepsilon E_Q \|V_c\|^2 \delta^2, \mathcal{F}_\delta, L_2(Q)) \leq C\varepsilon^{-2d_x - 2}, \tag{B.9}$$

where $N(\varepsilon E_Q \|V_c\|^2 \delta^2, \mathcal{F}_\delta, L_2(Q))$ is the covering number of $\mathcal{F}_\delta$ by $L_2(Q)$ balls of radius $\varepsilon E_Q \|V_c\|^2 \delta^2$, and Q is an arbitrary probability measure on $\mathcal{V}$, and C is a universal constant that depends only on d_x. Next we apply Theorem 2.14.1 of van der Vaart and Wellner (1996) to bound $\nu_n(b) - \nu_n(\beta)$:

$$\begin{aligned} &E\left\{\sup_{b \in R^{d_x} : b_1 = 1, \|b - \beta\| \leq \delta} |\nu_n(b) - \nu_n(\beta)|^2\right\} \\ &\leq C \sup_Q \int_0^1 \sqrt{1 + \log N(\varepsilon E_Q \|V_c\|^2 \delta^2, \mathcal{F}_\delta, L_2(Q))} d\varepsilon \times E\|V_c\|^2 \delta^2 / n \end{aligned}$$

$$= C \int_0^1 \sqrt{1 + \log C - (2d_x + 2)\log \varepsilon} d\varepsilon \times E\|V_c\|^2 \delta^2/n, \tag{B.10}$$

where the $C's$ are universal constants which may not be the same each time it appears. A change of variable technique can be used to show that the integral is finite. That combined with $E(\|\text{vec}(\mathbf{X}_{ct})\|^2) < \infty$ (Assumption 6.2(a)) implies that

$$E\left\{ \sup_{b \in R^{d_x}: b_1 = 1, \|b-\beta\| \le \delta} |\nu_n(b) - \nu_n(\beta)|^2 \right\} \le C\delta^2/n. \tag{B.11}$$

Using this and the arguments used in the proof of Lemma 4.1 of Kim and Pollard (1990), we can show that for arbitrarily small η, we have for all b such that $b_1 = 1, \|b - \beta\| \le \delta$,

$$|\nu_n(b) - \nu_n(\beta)| \le \eta\|b - \beta\|^2 + O_p(n^{-1}). \tag{B.12}$$

This combined with (B.6) shows result (i).

Finally, we show result (ii). Consider any $h = (0, \tilde{h}')' \in R^{d_x}$ such that $\|h\| = 1$. Consider $Q(\beta + hz)$ as a function of z at a given β and h value. Below we show that for all $z \in [0, c_1)$,

$$\frac{\partial Q(\beta + zh)}{\partial z} = -E[\mathbf{W}_c' h 1\{\mathbf{W}_c'(\beta + zh) < 0\}], \tag{B.13}$$

and that this first derivative is continuous in z. Below we also show that for all $z \in (0, c_1)$,

$$\frac{\partial^2 Q(\beta + zh)}{\partial z^2} = E[(\tilde{\mathbf{W}}_c'\tilde{h})^2 f_{\mathbf{W}_c^1|\tilde{\mathbf{W}}_c}(-\tilde{\mathbf{W}}_c'\tilde{\beta} - z\tilde{\mathbf{W}}_c'\tilde{h}|\tilde{\mathbf{W}}_c)1(\tilde{\mathbf{W}}_c'\tilde{h} < 0)]. \tag{B.14}$$

Given those, for any $z \in (0, c_1)$, a Taylor expansion with Lagrangian remainder applies and gives[1]

$$\begin{aligned} Q(\beta + hz) &= Q(\beta) + \frac{\partial Q(\beta + 0h)}{\partial t} z + 2^{-1}\frac{\partial^2 Q(\beta + \tau h)}{\partial z^2} z^2 \\ &= 0 + 2^{-1} z^2 \tilde{h}' E[\tilde{\mathbf{W}}_c \tilde{\mathbf{W}}_c' f_{\mathbf{W}_c^1|\tilde{\mathbf{W}}_c}(-\tilde{\mathbf{W}}_c'\tilde{\beta} - \tau\tilde{\mathbf{W}}_c'\tilde{h}|\tilde{\mathbf{W}}_c)1(\tilde{\mathbf{W}}_c'\tilde{h} < 0)]\tilde{h} \\ &\ge c_2 z^2 \|\tilde{h}\|^2 = c_2 z^2 \end{aligned} \tag{B.15}$$

for a τ in between 0 and z, where the inequality holds by Assumption 6.2(f). For an arbitrary $b \in B_{c_1}$, let $z = \|b - \beta\|$, and let $h = (b - \beta)/\|b - \beta\|$. The above display shows Result (ii).

Now we derive the first derivative of $Q(\beta + hz)$ with respect to z. Its first derivative equals the limit of the following quantity as $\tau \to z$, if the limit exists:

$$\frac{E([\mathbf{W}_c'(\beta + \tau h)]_-) - E([\mathbf{W}_c'(\beta + zh)]_-)}{\tau - z} \tag{B.16}$$

[1] See, for example, Apostol (1967, Section 7.7).

By Assumption 6.2(d), we have, for small enough z ($z < c_1$) with probability one

$$\lim_{\tau \to z} \frac{[\mathbf{W}_c'(\beta + zh + (\tau - z)h)]_- - [\mathbf{W}_c'(\beta + zh)]_-}{\tau - z} = -\mathbf{W}_c' h 1\{\mathbf{W}_c'(\beta + zh) < 0\}. \tag{B.17}$$

Also observe that

$$\begin{aligned} \left| \frac{[\mathbf{W}_c'(\beta + \tau h)]_- - [\mathbf{W}_c'(\beta + zh)]_-}{\tau - z} \right| &\leq \left| \frac{\mathbf{W}_c'(\beta + \tau h) - \mathbf{W}_c'(\beta + zh)}{\tau - z} \right| \\ &= |\mathbf{W}_c' h|. \end{aligned} \tag{B.18}$$

Assumption 6.2(a) implies that the right hand-side has finite first moment. Equations (B.17) and (B.18) together combined with the dominated convergence theorem imply that,

$$\lim_{\tau \to z} E\left(\frac{[\mathbf{W}_c'(\beta + \tau h)]_- - [\mathbf{W}_c'(\beta + zh)]_-}{\tau - z} \right) = -E[\mathbf{W}_c' h 1\{\mathbf{W}_c'(\beta + zh) < 0\}]. \tag{B.19}$$

This shows (B.13). The derivative is continuous in z by a similar application of the dominated convergence theorem.

Next we derive the second derivative at $z \in (0, c_1)$. It is convenient to write $\partial Q(\beta + zh)/\partial z$ as

$$\frac{\partial Q(\beta + zh)}{\partial z} = -E\left[\tilde{\mathbf{W}}_c' \tilde{h} F_{\mathbf{W}_c^1 | \tilde{\mathbf{W}}_c}(-\tilde{\mathbf{W}}_c'(\tilde{\beta} + z\tilde{h}) | \tilde{\mathbf{W}}_c) \right] \tag{B.20}$$

The derivative of of this equals the limit of the following quantity as $\tau \to z$, if the limit exists:

$$E\left(\frac{\tilde{\mathbf{W}}_c' \tilde{h} F_{\mathbf{W}_c^1 | \tilde{\mathbf{W}}_c}(-\tilde{\mathbf{W}}_c'(\tilde{\beta} + \tau\tilde{h}) | \tilde{\mathbf{W}}_c) - \tilde{\mathbf{W}}_c' \tilde{h} F_{\mathbf{W}_c^1 | \tilde{\mathbf{W}}_c}(-\tilde{\mathbf{W}}_c'(\tilde{\beta} + z\tilde{h}) | \tilde{\mathbf{W}}_c)}{\tau - z} \right). \tag{B.21}$$

Under Assumption 6.2(e), the limit of the quantity inside the large brackets exists almost surely and equals

$$-(\tilde{\mathbf{W}}_c' \tilde{h})^2 f_{\mathbf{W}_c^1 | \tilde{\mathbf{W}}_c}(-\tilde{\mathbf{W}}_c' \tilde{\beta} - z\tilde{\mathbf{W}}_c' \tilde{h} | \tilde{\mathbf{W}}_c) 1(z\tilde{\mathbf{W}}_c' \tilde{h} < 0). \tag{B.22}$$

Also, under Assumption 6.2(e), the absolute value of the quantity inside the large brackets in (B.21) is bounded above by $|C(\tilde{\mathbf{W}}_c' \tilde{h})^2|$, which has finite first moment by Assumption 6.2(a) and the fact that $\|\tilde{h}\| = 1$. Therefore, the bounded convergence theorem applies and shows that the limit of (B.21) exists and equals the expectation of (B.22). This concludes the proof of (B.14). □

C Appendix: Primitive necessary condition for point identification

In this section we characterize a primitive necessary condition for point identification, in the special case of a binary choice model.[2]

In the binary choice case, it is without loss to consider only cycles of length 2. Moreover, because $K = 1$, there is no need for the bold font on X_{it}, ϵ_{it}, A_i, v, and a. Similarly, there is also no need for the choice index superscript on these symbols. Thus, we omit them in this section.

Consider the G set defined in Section 3.2 and specialized to the binary case. Theorem C.1 below is the main result of this section. It shows that, if one regressor has finite support and all other regressors have bounded support, then point identification cannot be achieved at all values of β.

Assumption C.1. *For some $j = 1, \dots, d_x$,* (a) *G_j is a finite set, and*

(b) *G_{-j} is a bounded set.*

Theorem C.1 (**Necessary conditions for point identification).** *Under Assumptions* 3.1(a)-(b) *and* 3.2*, if Assumption* C.1 *holds, then it is not always true that $Q(b) > 0$ for all $b \in R^{d_x}$ such that $\|b\| = 1$ and $b \neq \beta$.*

Remark. According to the Theorem C.1, if one coordinate of $X_{is} - X_{it}$ has finite support for all s, t, then another coordinate of it must have unbounded support for some pair (s, t). The variable $X_{j,is} - X_{j,it}$ may have finite support, either when $X_{j,it}$ has finite support, or when the change of $X_{j,it}$ across time periods is restricted to a few grids. When that is the case, point identification requires that another regressor, say, $X_{j',it}$ to change unboundedly as t changes.

Theorem C.1 does not imply that β can never be point identified (up to scale normalization) under the conditions of the theorem. Point identification may still be achieved in part of the parameter space, but not on the whole space of β. In other words, there can be β values such that, when the population is generated from the model specified in (1.1) and (1.2) with β being one of those values, we have $Q(b) > 0$ for all $b \in \{b \in R^{d_x} : \|b\| = 1\}$ such that $b \neq \beta$. ■

[2] We were not able to obtain an analogous result in the more general multinomial choice case because (i) cycles longer than 2 would need to be considered, and (ii) the simultaneous variation of X_{it}^k for all k would also need to be taken into account.

Proof of Theorem C.1. It suffices to find at least one β value that generates a population for which point identification fails. Below we find such a value among β's that satisfy $\beta_j > 0$, $\beta_{j^*} > 0$ for some $j^* \neq j$, and $\beta_{j'} = 0$ for $j' \neq j, j^*$. It is useful to note that G is symmetric about the origin by definition. So are $G_{j'}$'s for all $j' = 1, \ldots, d_x$.

We discuss two cases. In the first case, $G_j \cap (-\infty, 0) = \emptyset$. Then $G_j = \{0\}$ because it is symmetric about the origin. Then $\mathcal{G}$ is contained in the subspace $\{g \in R^{d_x} : g_j = 0\}$. Let b^* be equal to β except that $b^*_j = 0$, and let $b = b^*/\|b^*\|$. Then $(b^*)'g = \beta' g$ for all $g \in \mathcal{G}$. This implies that $b'g \geq 0$ for all $g \in \mathcal{G}$, and thus $Q(b) = Q(\beta) = 0$.

In the second case, $G_j \cap (-\infty, 0) \neq \emptyset$. Assumption C.1(a) implies that G_j is a finite set. Then $\eta \equiv \max(G_j \cap (-\infty, 0))$ is well defined and $\eta < 0$. Assumption C.1(b) implies that there is a positive constant C such that $G_{j^*} \subseteq [-C, C]$. Let β further satisfy $\beta_{j^*}/\beta_j < -\eta/C$. Consider an arbitrary $g \in G$ with $g_j < 0$. Then $g_j \leq \eta$ and $g_{j^*} \leq C$, which implies that

$$\beta' g = \beta_j g_j + \beta_{j^*} g_{j^*} \leq \beta_j \eta + \beta_{j^*} C < 0.$$

Next we use this to show that $\mathcal{G}$ does not contain any g such that $g_j < 0$. We show this by contradiction. Suppose that it does contain a g^* such that $g^*_j < 0$. Then $g^* = E[\Delta Y_i | X_{i1} = x_1, X_{i2} = x_2]\Delta X_i = x_2 - x_1$ for some values x_1, x_2 of X_{i1}, X_{i2}. It must not be that $E[\Delta Y_i | X_{i1} = x_1, X_{i2} = x_2] = 0$ because $g^* \neq 0$. If $\lambda := E[\Delta Y_i | X_{i1} = x_1, X_{i2} = x_2] > 0$, we have $g := \lambda^{-1} g^* \in \text{supp}(\Delta X_i) \subseteq G$, and $g_j < 0$. Then (C) implies that $\beta' g < 0$, which in turn implies that $\beta' g^* < 0$, which contradicts the fact that $\mathcal{G} \subseteq \{b \in R^{d_x} : \beta' g \geq 0\}$ (implied by equation (3.8)). Similar arguments (using the fact that $\text{supp} {-\Delta X_i} \subseteq G$ leads to the same contradiction if $\lambda < 0$.

Therefore, $\mathcal{G}$ does not contain any point whose jth element is negative. Let b^* be the same as β except that $b^*_j > \beta_j$. Let $b = b^*/\|b^*\|$. Then $(b^*)'g \geq \beta' g$ for all $g \in \mathcal{G}$. Because $\beta' g \geq 0$ for all $g \in \mathcal{G}$, we have $(b^*)'g \geq 0$ for all $g \in \mathcal{G}$, and thus $b'g \geq 0$ for all $g \in \mathcal{G}$. This implies that $Q(b) = 0$ and we have constructed a $b \neq \beta$ that is not identifiable from β. □

D Appendix: Monte Carlo Results for Instrumental Function-Based Estimator

In this section we report the Khan and Tamer (2009)-variant of the moment inequality estimator. In this approach, rather than estimating the conditional choice probabilities and plugging them into the CM inequalities, we transform the conditional moment inequalities into unconditional moment inequalities for estimation.

The instrumental functions are indicator functions of hypercubes in the space of $\mathbf{X}_i$, where $\mathbf{X}_i = (\text{vec}(\mathbf{X}_{i1})', \text{vec}(\mathbf{X}_{i2})')'$. There are many ways to choose and weight the hypercubes to use, among which Khan and Tamer (2009) suggests to use the hypercubes formed by pairs of observations in the data. That suggests the criterion function below:

$$Q_n^{IF}(b) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} [\bar{g}_n(\mathbf{X}_i, \mathbf{X}_j)'b]_-,$$

where

$$\bar{g}_n(\underline{\mathbf{x}}, \bar{\mathbf{x}}) = n^{-1} \sum_{i=1}^{n} g_i(\underline{\mathbf{x}}, \bar{\mathbf{x}})$$

$$g_i(\underline{\mathbf{x}}, \bar{\mathbf{x}}) = \Delta \mathbf{X}_i \Delta \mathbf{Y}_i 1\{\underline{\mathbf{x}} \leq \mathbf{X}_i < \bar{\mathbf{x}}\}.$$

When implementing this approach, we were faced with two problems: (1) there are too many pairs involved for our sample sizes (e.g. 499,500 pairs for $n = 1000$), which makes computation very difficult, and (2) most of the hypercubes end up being empty simply due to the the fact that our $\mathbf{X}_i$ is 12 dimensional (3 variables × 2 time periods × 2 inside alternatives), which means that the criterion function often does not give a meaningful estimate.

For those reasons, we use the high-dimensional version of the hypercubes suggested in Andrews and Shi (2013) instead. In our design our variables are supported in the unit interval. Thus, we first evenly divide $[0, 1]$ into q subintervals ($q = 3$ for $n = 250$, 4 for $n = 500$, 5 for $n = 1000$, and 6 for $n = 2000$). Then use all the hypercubes that are the Cartesian products of two such sub-intervals and ten copies of $[0, 1]$. Let the collection of all such hypercubes be denoted by $\mathcal{C}$. Specifically, we form

$$Q_n^{IF}(b) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{C \in \mathcal{C}} [\bar{g}_n(C)'b]_-,$$

where $\bar{g}_n(C) = n^{-1}\sum_{i=1}^{n} g_i(C)$ and $g_i(C) = \Delta X_i \Delta Y_i 1\{X_i \in C\}$. We do not divide up all dimensions of the space of $\mathbf{X}_i$ precisely to avoid the same difficulties that arise with the Khan and Tamer (2009) instrumental functions discussed above.

We take the Cauchy design in Section 4.1 and report statistics of the instrumental function based estimator of β_2 in Table VII below. Comparing to Table II, we can see that the instrumental function-based CM estimator has larger bias and standard deviation. In addition, the standard deviation decreases slower with the sample size. For this reason, we focus on the estimator based on the nonparametric estimator of $\mathbf{p}(\cdot,\cdot)$ in the main text.

Table VII: Monte Carlo Performance of Estimators of β_2 (Cauchy Design, $\beta_{0,2} = 0.5$)

n	BIAS	SD	rMSE	25% quantile	median	75% quantile
	Instrumental Function-Based CM Estimator					
500	-0.1132	0.1496	0.1876	0.2847	0.3831	0.4830
1000	-0.0808	0.1172	0.1424	0.3402	0.4162	0.4955
2000	-0.0471	0.0970	0.1078	0.3853	0.4501	0.5151

E A Lemma Used in the Proof of Theorem 3.1

Lemma E.1. *Suppose that $\{g \in R^J : \beta' g \geq 0\} \subseteq \{g \in R^{d_x} : b'g \geq 0\}$ and that $\|\beta\| = \|b\| = 1$. Then $b = \beta$.*

Proof. Let $(\beta, g^1, \ldots, g^{J-1})$ be an orthonormal basis of R^J. Then $ag^j \in \{g \in R^J : \beta' g \geq 0\}$ for all $a \in R$ and $j = 1, \ldots, J-1$ because

$$\beta'(ag^j) = a(\beta' g^j) = a \times 0 = 0.$$

The condition of the lemma implies that $ag^j \in \{g \in R^J : b'g \geq 0\}$ for all $a \in R$ and $j = 1, \ldots, J-1$. With a unrestricted, this means that

$$b'g^j = 0 \text{ for all } j = 1, \ldots, J-1.$$

Let $c, d_1, \ldots, d_{J-1}$ be the constants such that $b = c\beta + \sum_{j=1}^{J-1} d_j g^j$. Then,

$$0 = b'g^j = c\beta' g^j + \sum_{j'=1}^{J-1} d_j((g^j)' g^j)$$

$$= d_j \|g^j\|^2 = d_j \text{ for all } j = 1, \ldots, J-1. \tag{E.1}$$

Therefore

$$b = c\beta.$$

That and $\|b\| = \|\beta\| = 1$ implies that $|c| = 1$. Now we just need to rule out $c = -1$. Suppose without loss of generality that $\beta_1 > 0$, then $e_1 = (1, 0, \ldots, 0)' \in \{g \in R^J : \beta' g \geq 0\}$. Thus, $e_1 = (1, 0, \ldots, 0)' \in \{g \in R^J : b' g \geq 0\}$. Therefore, $b_1 > 0$, which rules out $c = -1$.

□

References

[1] D. Andrews and X. Shi. Inference based on conditional moment inequalities. *Econometrica*, 81: 609-666, 2013.

[2] T. M. Apostol. *Calculus Volume 1: One-Variable Calculus, with an Introduction to Linear Algebra.* John Wiley and Sons, Inc. Second Edition. 1967.

[3] S. Khan and E. Tamer Inference on Endogenously Censored Regression Models Using Conditional Moment Inequalities. *Journal of Econometrics*, 152:104-119, 2009.

[4] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer, 1996.