Supplement to "Moment inequalities for multinomial choice with fixed effects"

(*Quantitative Economics*, Vol. 15, No. 1, January 2024, 1–25)

Ariel Pakes
Department of Economics, Harvard University and NBER

Jack Porter
Department of Economics, University of Wisconsin

Appendix: Proofs

Proof of Theorem 2. Take $F_{y_s,y_t,x_s,x_t} \in \mathcal{F}_{\text{ob}}$. It is straightforward to show that $\Theta_S \subset \Theta_0$. So, we will focus on showing $\Theta_0 \subset \Theta_S$. Take $\theta \in \Theta_0$. For all x_s, x_t in the support of the joint covariate space, we will exhibit a conditional distribution $(\varepsilon_s^*, \varepsilon_t^*)|x_s, x_t$ satisfying Assumption 1(b) with $\lambda^* = 0$ and $F_{y_s^*,y_t^*|x_s,x_t} = F_{y_s,y_t|x_s,x_t}$ where $y_j^* = \mathrm{Y}(x_j, \lambda^* = 0, \varepsilon_j^*, \theta)$ for $j = s, t$.

Suppose we order the covariate indices for the parameter θ and there is a strict ordering: $[g_{(\mathcal{D})}(x_{(\mathcal{D}),s}, \theta) - g_{(\mathcal{D})}(x_{(\mathcal{D}),t}, \theta)] > [g_{(\mathcal{D}-1)}(x_{(\mathcal{D}-1),s}, \theta) - g_{(\mathcal{D}-1)}(x_{(\mathcal{D}-1),t}, \theta)] > \cdots > [g_{(0)}(x_{(0),s}, \theta) - g_{(0)}(x_{(0),t}, \theta)]$ (with some abuse of the order statistic subscript notation). Since $\theta \in \Theta_0$, the conditional moment inequalities imply

$$\Pr\bigl(y_s \in \bigl\{(\mathcal{D}), \dots, (d)\bigr\}|x_s, x_t\bigr) \geq \Pr\bigl(y_t \in \bigl\{(\mathcal{D}), \dots, (d)\bigr\}|x_s, x_t\bigr) \quad \forall d = 1, 2, \dots, \mathcal{D}.$$

Let $p_{d,d'} = \Pr(y_s = d, y_t = d'|x_s, x_t)$ and $p^*_{d,d'} = \Pr(y_s^* = d, y_t^* = d'|x_s, x_t)$. We need to find $(\varepsilon_s^*, \varepsilon_t^*)|x_s, x_t$ such that $p^*_{d,d'} = p_{d,d'}$ $\forall d, d'$ and $\varepsilon_s^*|x_s, x_t \sim \varepsilon_t^*|x_s, x_t$.

Define $R_{d;s} = \{\varepsilon^* : \mathrm{Y}(x_s, \lambda^* = 0, \varepsilon_s^*, \theta) = d\}$. The set inclusion obtained in the proof of Proposition 1 shows that

$$R_{(\mathcal{D});t} \cup \cdots \cup R_{(d);t} \subset R_{(\mathcal{D});s} \cup \cdots \cup R_{(d);s}, \quad \forall d \in \{1, \dots, \mathcal{D}\}. \tag{S1}$$

Since the sets $R_{(d);s}$ form a partition for $d = 0, \dots, \mathcal{D}$, the set inclusion (S1) implies that

$$R_{(d);s} \cap R_{(d');t} = \emptyset \quad \text{for } d' > d. \tag{S2}$$

Let $R_{d,d'} = R_{d;s} \cap R_{d';t}$, which is a set in the ε_s^*-space (or the ε_t^*-space). Cartesian products of these sets will form sets in the $(\varepsilon_s^*, \varepsilon_t^*)$-space, $R_{d,d'} \times R_{d'',d'''} = \{(\varepsilon_s^*, \varepsilon_t^*) : \varepsilon_s^* \in R_{d,d'}, \varepsilon_t^* \in R_{d'',d'''}\}$. Finally, let $q^*_{d,d'\times d'',d'''} = \Pr((\varepsilon_s^*, \varepsilon_t^*) \in R_{d,d'} \times R_{d'',d'''}|x_s, x_t)$. These

Ariel Pakes: apakes@fas.harvard.edu
Jack Porter: jrporter@ssc.wisc.edu

 https://doi.org/10.3982/QE1776

probabilities form the basic building blocks for our constructed $(\varepsilon_s^*, \varepsilon_t^*)|x_s, x_t$ distribution, as $R_{d,d'} \times R_{d'',d'''}$ partitions the $(\varepsilon_s^*, \varepsilon_t^*)$-space. By (S2), $q^*_{(d),(d')\times(d''),(d''')} = 0$ if $d < d'$ or $d'' < d'''$, so

$$p^*_{(d),(d')} = \sum_{\underset{\sim}{d}=0}^{\mathcal{D}} \sum_{\tilde{d}=0}^{\mathcal{D}} q^*_{(d),(\underset{\sim}{d})\times(\tilde{d}),(d')} = \sum_{\underset{\sim}{d}=0}^{d} \sum_{\tilde{d}=d'}^{\mathcal{D}} q^*_{(d),(\underset{\sim}{d})\times(\tilde{d}),(d')}.$$

To get the constructed distribution to match the observed distribution, we will need to show that there exists $q^*_{d,d'\times d'',d'''}$ satisfying

$$p_{(d),(d')} = \sum_{\underset{\sim}{d}=0}^{d} \sum_{\tilde{d}=d'}^{\mathcal{D}} q^*_{(d),(\underset{\sim}{d})\times(\tilde{d}),(d')}, \tag{S3}$$

as well as ensuring that Assumption 1(b) holds for the constructed distribution. For each $R_{d,d'} \neq \emptyset$, choose a point $r_{d,d'} \in R_{d,d'}$. Define $(\varepsilon_s^*, \varepsilon_t^*)|x_s, x_t$ to be the discrete distribution on the support points $(r_{d,d'}, r_{d'',d'''})$, $\Pr((\varepsilon_s^*, \varepsilon_t^*) = (r_{d,d'}, r_{d'',d'''})|x_s, x_t) = q^*_{d,d'\times d'',d'''}$.[1] So, the marginal distribution is

$$\Pr\bigl(\varepsilon_s^* = r_{(d),(d')}|x_s, x_t\bigr) = \sum_{\underset{\sim}{d}=0}^{\mathcal{D}} \sum_{\tilde{d}=\underset{\sim}{d}}^{\mathcal{D}} q^*_{(d),(d')\times(\tilde{d}),(\underset{\sim}{d})}.$$

The marginal for $\varepsilon_t^*|x_s, x_t$ is similar. To ensure that Assumption 1(b) is satisfied, we will need the marginals to match, for $d \geq d'$,

$$0 = \sum_{\underset{\sim}{d}\leq\tilde{d}} \bigl(q^*_{(\tilde{d}),(\underset{\sim}{d})\times(d),(d')} - q^*_{(d),(d')\times(\tilde{d}),(\underset{\sim}{d})}\bigr). \tag{S4}$$

In addition to equations (S3) and (S4), the nonnegativity inequalities $q^*_{d,d'\times d'',d'''} \geq 0$ must hold. Let p denote the vector of joint probabilities, $p = (p_{(\mathcal{D}),(\mathcal{D})}, p_{(\mathcal{D}),(\mathcal{D}-1)}, \ldots)'$. Let q^* be the vector of probabilities $q^*_{(d),(d')\times(d''),(d''')}$ (where $d \geq d'$ and $d'' \geq d'''$), $q^* = (q^*_{(\mathcal{D}),(\mathcal{D})\times(\mathcal{D}),(\mathcal{D})}, \ldots)'$. And let Q_s be the matrix with elements in $\{0, 1\}$ such that equation (S3) can be restated as $p = Q_s q^*$, and let Q_p be the matrix with elements in $\{-1, 0, 1\}$ such that equation (S4) can be restated as $0 = Q_p q^*$.

Our goal then can be summarized as showing that $\exists q^* \geq 0$ such that: (A) $p = Q_s q^*$; and (B) $0 = Q_p q^*$. Let z be a $(\mathcal{D}+1)^2$-dimensional vector conformable with p, $z = (z_{(\mathcal{D}),(\mathcal{D})}, z_{(\mathcal{D}),(\mathcal{D}-1)}, \ldots)'$. Let w be a $(\mathcal{D}+1)(\mathcal{D}+2)/2$-dimensional vector, $w = (\ldots, w_{(d),(d')}, \ldots)'$. Farkas' lemma states that if

$$\begin{pmatrix} z \\ w \end{pmatrix}' \begin{pmatrix} Q_s \\ Q_p \end{pmatrix} \geq 0 \quad \text{implies} \quad \begin{pmatrix} z \\ w \end{pmatrix}' \begin{pmatrix} p \\ 0 \end{pmatrix} = z'p \geq 0,$$

then $\exists q^* \geq 0$ satisfying (A) and (B) above.[2]

[1] A continuous distribution for $(\varepsilon_s^*, \varepsilon_t^*)|x_s, x_t$ could be obtained by defining the density on each $R_{d,d'} \neq \emptyset$ to be a constant chosen so that $\Pr((\varepsilon_s^*, \varepsilon_t^*) \in R_{d,d'}|x_s, x_t) = q^*_{d,d'\times d'',d'''}$.

[2] Farkas' lemma actually states that the condition provided is both necessary and sufficient; see Matoušek and Gärtner (2007, Proposition 6.4.1).

Each element $q^*_{(d),(d')\times(d''),(d''')}$ of q^* appears in exactly one equation from constraints (A) and either zero or two (with positive and negative signs) from constraints (B). In particular, elements of the form $q^*_{(d),(d')\times(d),(d')}$ with $d \geq d'$ appear in (A) but not (B). Hence, $\binom{z}{w}'\binom{Q_s}{Q_p} \geq 0$ implies $z_{(d),(d')} \geq 0$ for $d \geq d'$. Otherwise, $\binom{z}{w}'\binom{Q_s}{Q_p} \geq 0$ yields, for $d \neq d'$, $z_{(d),(d')} + w_{(\tilde{d}),(d')} - w_{(d),(\underset{\sim}{d})} \geq 0$ with $\tilde{d} \in \{d', \dots, \mathcal{D}\}$, $\underset{\sim}{d} \in \{0, \dots, d\}$. Define $\bar{w}_{(d),\cdot} = \max_{\underset{\sim}{d}\in\{0,\dots,d\}} w_{(d),(\underset{\sim}{d})}$ and $\underset{\sim}{w}_{\cdot,(d)} = \min_{\tilde{d}\in\{d,\dots,\mathcal{D}\}} w_{(\tilde{d}),(d)}$ for $d = 0, \dots, \mathcal{D}$. Then $z_{(d),(d')} \geq \bar{w}_{(d),\cdot} - \underset{\sim}{w}_{\cdot,(d')}$.

Also, the conditional moment inequalities yield $\Pr(y_s \in \{(\mathcal{D}), \dots, (d)\}|x_s, x_t) \geq \Pr(y_t \in \{(\mathcal{D}), \dots, (d)\}|x_s, x_t)$ for $d = 1, \dots, \mathcal{D}$, which implies

$$\sum_{\tilde{d}=d}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\tilde{d}),(\underset{\sim}{d})} \geq \sum_{\tilde{d}=d}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\underset{\sim}{d}),(\tilde{d})} \quad \text{for } d = 1, \dots, \mathcal{D}. \tag{S5}$$

For $d = 1, \dots, \mathcal{D}$, let $a_{(d)}$ denote the slackness in the inequalities in (S5),

$$a_{(d)} = \sum_{\tilde{d}=d}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\tilde{d}),(\underset{\sim}{d})} - \sum_{\tilde{d}=d}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\underset{\sim}{d}),(\tilde{d})},$$

so that $a_{(d)} \geq 0$ for $d = 1, \dots, \mathcal{D}$. We use these expressions to substitute for all terms of the form $p_{(d),(d-1)}$ in the equality (S6) below:[3]

$$\begin{aligned}
z'p &= \sum_{d=0}^{\mathcal{D}} z_{(d),(d)} p_{(d),(d)} + \sum_{d>d'} \left(z_{(d),(d')} p_{(d),(d')} + z_{(d'),(d)} p_{(d'),(d)} \right) \\
&\geq \sum_{d>d'} \left(z_{(d),(d')} p_{(d),(d')} + z_{(d'),(d)} p_{(d'),(d)} \right) \\
&\geq \sum_{d>d'} \left((\bar{w}_{(d),\cdot} - \underset{\sim}{w}_{\cdot,(d')}) p_{(d),(d')} + (\bar{w}_{(d'),\cdot} - \underset{\sim}{w}_{\cdot,(d)}) p_{(d'),(d)} \right) \\
&= \Bigg[a_{(\mathcal{D})} \bar{w}_{(\mathcal{D}),\cdot} - a_{(1)} \underset{\sim}{w}_{\cdot,(0)} + \sum_{d=1}^{\mathcal{D}-1} \left(a_{(d)} - (a_{(d+1)} \wedge a_{(d)}) \right) \bar{w}_{(d),\cdot} \qquad \text{(S6)} \\
&\quad - \sum_{d=1}^{\mathcal{D}-1} \left(a_{(d+1)} - (a_{(d+1)} \wedge a_{(d)}) \right) \underset{\sim}{w}_{\cdot,(d)} \Bigg] \\
&\quad + (\bar{w}_{(0),\cdot} - \underset{\sim}{w}_{\cdot,(0)}) \sum_{d=1}^{\mathcal{D}} p_{(0),(d)} + (\bar{w}_{(\mathcal{D}),\cdot} - \underset{\sim}{w}_{\cdot,(\mathcal{D})}) \sum_{d=0}^{\mathcal{D}-1} p_{(d),(\mathcal{D})} \\
&\quad + \sum_{d=1}^{\mathcal{D}-1} (\bar{w}_{(d),\cdot} - \underset{\sim}{w}_{\cdot,(d)}) \Bigg[(a_{(d+1)} \wedge a_{(d)}) + \sum_{\tilde{d}=d}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\underset{\sim}{d}),(\tilde{d})} + \sum_{\tilde{d}=d+1}^{\mathcal{D}} p_{(d),(\tilde{d})} \qquad \text{(S7)}
\end{aligned}$$

[3]The expressions below apply to the case $\mathcal{D}+1 = 2$ with the convention that summations are taken to be zero when the upper limit is strictly less than the lower limit.

$$- \sum_{\tilde{d}=d+1}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\tilde{d}),(\underset{\sim}{d})} \Bigg].$$

Now we can show directly that the expression in (S6) is nonnegative. First, note that for all $d \geq d'$, $\bar{w}_{(d),\cdot} - \underline{w}_{\cdot,(d')} \geq w_{(d),(d')} - w_{(d),(d')} = 0$.

Next, show that the term in square brackets in (S6) is nonnegative. For each $d = 1, \dots, \mathcal{D}$,

$$a_{(\mathcal{D})} + \sum_{d'=d}^{\mathcal{D}-1} \big(a_{(d')} - (a_{(d'+1)} \wedge a_{(d')}) \big) = a_{(d)} + \left[\sum_{d'=d}^{\mathcal{D}-1} \big(a_{(d'+1)} - (a_{(d'+1)} \wedge a_{(d')}) \big) \right]. \tag{S8}$$

The term on the left is the sum of coefficients on $\bar{w}_{(d),\cdot}, \dots, \bar{w}_{(\mathcal{D}),\cdot}$, and the term in the square brackets on the right is the sum of coefficients on $\underline{w}_{\cdot,(d)}, \dots, \underline{w}_{\cdot,(\mathcal{D}-1)}$ for $d = 1, \dots, \mathcal{D}$. Moreover, when $d = 1$, the expression on the right is the sum of coefficients on $\underline{w}_{\cdot,(0)}, \dots, \underline{w}_{\cdot,(\mathcal{D}-1)}$, which is exactly equal to the sum of coefficients on $\bar{w}_{(1),\cdot}, \dots, \bar{w}_{(\mathcal{D}),\cdot}$. These relationships are used to rearrange the initial expression. In particular, using (S8) we can find nonnegative values $\underline{b}_{(d'),(d)}$ such that (i) $a_{(\mathcal{D})} = \sum_{d'=0}^{\mathcal{D}-1} b_{(\mathcal{D}),(d')}$ and $a_{(d)} - (a_{(d+1)} \wedge a_{(d)}) = \sum_{d'=0}^{d} b_{(d),(d')}$ for $d = 1, \dots, \mathcal{D}-1$ and (ii) $a_{(1)} = \sum_{d'=1}^{\mathcal{D}} b_{(d'),(0)}$ and $a_{(d+1)} - (a_{(d+1)} \wedge a_{(d)}) = \sum_{d'=d}^{\mathcal{D}} b_{(d'),(d)}$ for $d = 1, \dots, \mathcal{D}-1$. Hence, we can write

$$\begin{aligned}
& a_{(\mathcal{D})} \bar{w}_{(\mathcal{D}),\cdot} - a_{(1)} \underline{w}_{\cdot,(0)} + \sum_{d=1}^{\mathcal{D}-1} \big(a_{(d)} - (a_{(d+1)} \wedge a_{(d)}) \big) \bar{w}_{(d),\cdot} \\
&\quad - \sum_{d=1}^{\mathcal{D}-1} \big(a_{(d+1)} - (a_{(d+1)} \wedge a_{(d)}) \big) \underline{w}_{\cdot,(d)} \\
&= \left(\sum_{d'=0}^{\mathcal{D}-1} b_{(\mathcal{D}),(d')} \right) \bar{w}_{(\mathcal{D}),\cdot} + \sum_{d=1}^{\mathcal{D}-1} \left(\sum_{d'=0}^{d} b_{(d),(d')} \right) \bar{w}_{(d),\cdot} \\
&\quad - \left[\sum_{d=1}^{\mathcal{D}-1} \left(\sum_{d'=d}^{\mathcal{D}} b_{(d'),(d)} \right) \underline{w}_{\cdot,(d)} + \left(\sum_{d'=1}^{\mathcal{D}} b_{(d'),(0)} \right) \underline{w}_{\cdot,(0)} \right] \\
&= \sum_{d=1}^{\mathcal{D}-1} \left(\sum_{d'=d}^{\mathcal{D}} b_{(d'),(d)} (\bar{w}_{(d'),\cdot} - \underline{w}_{\cdot,(d)}) \right) + \sum_{d'=1}^{\mathcal{D}} b_{(d'),(0)} (\bar{w}_{(d'),\cdot} - \underline{w}_{\cdot,(0)}) \\
&\geq 0,
\end{aligned}$$

where the final inequality follows from the terms in each sum being nonnegative.

Finally, show that the term in square brackets in (S7) is nonnegative. This term is the minimum of the following expressions in (S9) and (S10):

$$a_{(d+1)} + \sum_{\tilde{d}=d}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\underset{\sim}{d}),(\tilde{d})} + \sum_{\tilde{d}=d+1}^{\mathcal{D}} p_{(d),(\tilde{d})} - \sum_{\tilde{d}=d+1}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\tilde{d}),(\underset{\sim}{d})} \tag{S9}$$

$$
\begin{aligned}
&= \sum_{\tilde{d}=d+1}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d} p_{(\tilde{d}),(\underset{\sim}{d})} - \sum_{\tilde{d}=d+1}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d} p_{(\underset{\sim}{d}),(\tilde{d})} \\
&\quad + \sum_{\tilde{d}=d+1}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\underset{\sim}{d}),(\tilde{d})} + \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\underset{\sim}{d}),(d)} + \sum_{\tilde{d}=d+1}^{\mathcal{D}} p_{(d),(\tilde{d})} - \sum_{\tilde{d}=d+1}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\tilde{d}),(\underset{\sim}{d})} \\
&= \sum_{\tilde{d}=d+1}^{\mathcal{D}} p_{(\tilde{d}),(d)} + \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\underset{\sim}{d}),(d)}
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
a_{(d)} &+ \sum_{\tilde{d}=d}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\underset{\sim}{d}),(\tilde{d})} + \sum_{\tilde{d}=d+1}^{\mathcal{D}} p_{(d),(\tilde{d})} - \sum_{\tilde{d}=d+1}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\tilde{d}),(\underset{\sim}{d})} \\
&= \sum_{\underset{\sim}{d}=0}^{d-1} p_{(d),(\underset{\sim}{d})} + \sum_{\tilde{d}=d+1}^{\mathcal{D}} p_{(d),(\tilde{d})}.
\end{aligned}
\tag{S10}
$$

Both terms (S9) and (S10) are seen to be sum of probabilities, and hence nonnegative, so the minimum of these two terms is also nonnegative:

$$
0 \leq (a_{(d+1)} \wedge a_{(d)}) + \sum_{\tilde{d}=d}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\underset{\sim}{d}),(\tilde{d})} + \sum_{\tilde{d}=d+1}^{\mathcal{D}} p_{(d),(\tilde{d})} - \sum_{\tilde{d}=d+1}^{\mathcal{D}} \sum_{\underset{\sim}{d}=0}^{d-1} p_{(\tilde{d}),(\underset{\sim}{d})}.
$$

It follows that $z'p \geq 0$, which completes the argument for the case where there is a strict covariate index ordering. When the covariate index ordering is weak (includes some ties), the analogous argument applies imposing additional restrictions as in (S2) on the partition $R_{d,d'}$ and using the information provided by the additional implied moment inequalities to verify Farkas' lemma as above. Finally, we can conclude that a constructed disturbance distribution exists that satisfies Assumption 1 and generates a constructed outcome and covariate distribution that matches the observed distribution, so that $\Theta_0 \subset \Theta_S$ and Θ_0 is sharp. □

Proof of Proposition 3. In binary choice, it will be useful to note that when $\Pr(y_s = 1 | x_s, x_t) = \Pr(y_t = 1 | x_s, x_t)$ then $\Pr(y_s = 0 | x_s, x_t) = \Pr(y_t = 0 | x_s, x_t)$. In this case, take an arbitrary θ. Then $\overline{\mathbb{D}}(x_s, x_t, \theta) = \{\{0\}\}, \{\{1\}\}$, or $\{\{0\}, \{1\}\}$. In any of these cases, $H(x_s, x_t, \theta) = 0$, so $H(x_s, x_t, \theta) = 0\ \forall \theta$ when $\Pr(y_s = 1 | x_s, x_t) = \Pr(y_t = 1 | x_s, x_t)$.

Another useful finding is that for any θ such that $\overline{\mathbb{D}}(x_s, x_t, \theta) = \{\{0\}, \{1\}\}$, $H(x_s, x_t, \theta) = \sum_{D \in \overline{\mathbb{D}}(x_s, x_t, \theta)} E[\mathbf{1}\{y_s \in D\} - \mathbf{1}\{y_t \in D\} | x_s, x_t] = [\Pr(y_s = 0 | x_s, x_t) - \Pr(y_t = 0 | x_s, x_t)] + [\Pr(y_s = 1 | x_s, x_t) - \Pr(y_t = 1 | x_s, x_t)] = 0$.

Take $\theta \in \Theta_0$ and show that $H(x_s, x_t, \theta) = H(x_s, x_t, \theta_0)$ a.s., so that $H(\theta) = H(\theta_0)$. Cases:

(a) $\Delta g(x_s, x_t, \theta) > 0$. Then $\overline{\mathbb{D}}(x_s, x_t, \theta) = \{\{1\}\}$, so $H(x_s, x_t, \theta) = \Pr(y_s = 1 | x_s, x_t) - \Pr(y_t = 1 | x_s, x_t)$. Since $\theta \in \Theta_0$, $\Pr(y_s = 1 | x_s, x_t) - \Pr(y_t = 1 | x_s, x_t) \geq 0$. If $\Pr(y_s =$

$1|x_s, x_t) - \Pr(y_t = 1|x_s, x_t) > 0$, then we must have by Proposition 1, $\overline{\mathbb{D}}(x_s, x_t, \theta_0) = \{\{1\}\}$, so $H(x_s, x_t, \theta) = H(x_s, x_t, \theta_0)$. If $\Pr(y_s = 1|x_s, x_t) - \Pr(y_t = 1|x_s, x_t) = 0$, then as noted above $H(x_s, x_t, \theta) = 0 = H(x_s, x_t, \theta_0)$.

(b) $\Delta g(x_s, x_t, \theta) = 0$. Then $\overline{\mathbb{D}}(x_s, x_t, \theta) = \{\{0\}, \{1\}\}$, so as noted above $H(x_s, x_t, \theta) = 0$. Since $\theta \in \Theta_0$, we must have $\Pr(y_s = 0|x_s, x_t) - \Pr(y_t = 0|x_s, x_t) \geq 0$ and $\Pr(y_s = 1|x_s, x_t) - \Pr(y_t = 1|x_s, x_t) \geq 0$. So, $\Pr(y_s = 1|x_s, x_t) = \Pr(y_t = 1|x_s, x_t)$, and $H(x_s, x_t, \theta_0) = 0$. Hence, $H(x_s, x_t, \theta) = H(x_s, x_t, \theta_0)$.

(c) $\Delta g(x_s, x_t, \theta) < 0$. Similar to case (a), $H(x_s, x_t, \theta) = H(x_s, x_t, \theta_0)$.

So, we have shown that $H(x_s, x_t, \theta) = H(x_s, x_t, \theta_0)$ a.s., and so $H(\theta) = H(\theta_0)$.

Now take $\theta \notin \Theta_0$ and show $H(\theta) < H(\theta_0)$. Let

$$
\begin{aligned}
\mathcal{A} &= \big\{(x_s, x_t) : E\big[m_D(y_s, y_t, x_s, x_t, \theta)\big|x_s, x_t\big] < 0 \text{ for some } D \in \mathbb{D}\big\} \\
&= \big\{(x_s, x_t) : E\big[\mathbf{1}\{y_s \in D\} - \mathbf{1}\{y_t \in D\}\big|x_s, x_t\big] < 0 \text{ for some } D \in \overline{\mathbb{D}}(x_s, x_t, \theta)\big\}.
\end{aligned}
$$

Since $\theta \notin \Theta_0$, $\Pr(\mathcal{A}) > 0$.[4] Consider $(x_s, x_t) \in \mathcal{A}$.

(a) $\Delta g(x_s, x_t, \theta) > 0$. But $\Pr(y_s = 1|x_s, x_t) < \Pr(y_t = 1|x_s, x_t)$. Then $\Pr(y_s = 0|x_s, x_t) > \Pr(y_t = 0|x_s, x_t)$ and $\Delta g(x_s, x_t, \theta_0) < 0$. Hence, $\overline{\mathbb{D}}(x_s, x_t, \theta) = \{\{1\}\}$ and $\overline{\mathbb{D}}(x_s, x_t, \theta_0) = \{\{0\}\}$, and $H(x_s, x_t, \theta) = \Pr(y_s = 1|x_s, x_t) - \Pr(y_t = 1|x_s, x_t) < 0 < \Pr(y_s = 0|x_s, x_t) - \Pr(y_t = 0|x_s, x_t) = H(x_s, x_t, \theta_0)$.

(b) $\Delta g(x_s, x_t, \theta) = 0$. Then $\overline{\mathbb{D}}(x_s, x_t, \theta) = \{\{0\}, \{1\}\}$, so $H(x_s, x_t, \theta) = 0$. But since $(x_s, x_t) \in \mathcal{A}$, $\Pr(y_s = 1|x_s, x_t) \neq \Pr(y_t = 1|x_s, x_t)$. Suppose $\Pr(y_s = 1|x_s, x_t) > \Pr(y_t = 1|x_s, x_t)$, then $\Delta g(x_s, x_t, \theta_0) > 0$, and $H(x_s, x_t, \theta_0) = \Pr(y_s = 1|x_s, x_t) - \Pr(y_t = 1|x_s, x_t) > 0 = H(x_s, x_t, \theta)$. The argument holds similarly when $\Pr(y_s = 1|x_s, x_t) < \Pr(y_t = 1|x_s, x_t)$. So $H(x_s, x_t, \theta_0) > H(x_s, x_t, \theta)$.

(c) $\Delta g(x_s, x_t, \theta) < 0$. Arguing similar to (a), $H(x_s, x_t, \theta_0) > H(x_s, x_t, \theta)$.

So, we have shown that $H(x_s, x_t, \theta_0) > H(x_s, x_t, \theta)$ for $(x_s, x_t) \in \mathcal{A}$. For $(x_s, x_t) \notin \mathcal{A}$, it is straightforward to show $H(x_s, x_t, \theta) = H(x_s, x_t, \theta_0)$ using previous arguments. Hence, $H(\theta_0) - H(\theta) = E[\mathbf{1}\{(x_s, x_t) \in \mathcal{A}\}(H(x_s, x_t, \theta_0) - H(x_s, x_t, \theta))] > 0$. The result follows. □

Reference

Matoušek, Jiří and Bernd Gärtner (2007), *Understanding and Using Linear Programming*. Springer-Verlag, Berlin. [2]

Co-editor Andres Santos handled this manuscript.

Manuscript received 23 November, 2020; final version accepted 12 May, 2022; available online 24 February, 2023.

[4]Suppose $\mathcal{A}$ is measurable or contains a measurable set of positive measure.