Appendices for Online Publication

Appendix B: Supplementary Results for Inference

Proof of Theorem 1.

We prove that

$$\sqrt{T}\frac{\left(f\left(\hat{h}\right)-f\left(h_0\right)\right)}{v_T\left(f\right)}\xrightarrow{d}N\left(0,1\right) \tag{19}$$

The result then follows from Lemma 7 below. By Assumption 7(ii),

$$\sqrt{T}\frac{\left(f\left(\hat{h}\right)-f\left(h_0\right)\right)}{v_T\left(f\right)} \quad = \quad \sqrt{T}\frac{Df\left(h_0\right)[\hat{h}-h_0]}{v_T\left(f\right)}+o_p\left(\underbrace{\frac{\sqrt{T}}{v_T\left(f\right)}||\partial^{\alpha_1}\hat{h}-\partial^{\alpha_1}h_0||_\infty||\partial^{\alpha_2}\hat{h}-\partial^{\alpha_2}h_0||_\infty}_{c_T}\right)$$

By Assumption 7(iii), $c_T=o_p\left(1\right)$ and therefore,

$$\sqrt{T}\frac{\left(f\left(\hat{h}\right)-f\left(h_0\right)\right)}{v_T\left(f\right)} \quad = \quad \sqrt{T}\frac{Df\left(h_0\right)[\hat{h}-h_0]}{v_T\left(f\right)}+o_p\left(1\right) \tag{20}$$

Further, by Assumption 7(i)

$$Df\left(h_0\right)[\hat{h}-h_0]=Df\left(h_0\right)[\hat{h}-\tilde{h}]+Df\left(h_0\right)[\tilde{h}-h_0] \tag{21}$$

and

$$Df\left(h_0\right)[\tilde{h}-h_0]\lesssim||\partial^{\alpha}\tilde{h}-\partial^{\alpha}h_0||_\infty \tag{22}$$

By (22) and Assumption 7(iii),

$$\sqrt{T}\frac{Df\left(h_0\right)[\tilde{h}-h_0]}{v_T\left(f\right)}=o_p\left(1\right) \tag{23}$$

Combining (20), (21) and (23), we obtain

$$\sqrt{T}\frac{\left(f\left(\hat{h}\right)-f\left(h_0\right)\right)}{v_T\left(f\right)} \quad = \quad \sqrt{T}\frac{Df\left(h_0\right)[\hat{h}-\tilde{h}]}{v_T\left(f\right)}+o_p\left(1\right) \tag{24}$$

We define

$$R_T\left(w\right)=\frac{Df\left(h_0\right)[\psi_M]'\left(L'G_A^{-1}L\right)^{-1}L'G_A^{-1}a_K\left(w\right)}{v_T\left(f\right)}$$

and note that $\mathbb{E}\left[\left(R_T(W)\cdot[\xi_1,\cdots,\xi_J]'\right)^2\right]=1$. Then,

$$\begin{aligned}
\sqrt{T}\frac{Df(h_0)[\hat{h}-\tilde{h}]}{v_T(f)} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} R_T(w_t)\cdot[\xi_{1t},\cdots,\xi_{Jt}]' \\
&+ \frac{Df(h_0)[\psi_M]'\left(\left(\hat{L}'\hat{G}_A^{-}\hat{L}\right)^{-}\hat{L}'\hat{G}_A^{-}-\left(L'G_A^{-1}L\right)^{-1}L'G_A^{-1}\right)\left(A'\xi/\sqrt{T}\right)}{v_T(f)} \\
&\equiv T_1+T_2
\end{aligned}$$

where $\xi\equiv[\xi_{11},\cdots,\xi_{1T},\xi_{21},\cdots,\xi_{2T},\cdots,\xi_{J1},\cdots,\xi_{JT}]'$.
First, we show that $T_1\xrightarrow{d}N(0,1)$ by the Lindeberg-Feller theorem. The Lindeberg condition requires that, for every $\epsilon>0$,

$$C_{0,T}\equiv\mathbb{E}\left[\left(R_T(W)\cdot[\xi_1,\cdots,\xi_J]'\right)^2\underbrace{\mathbb{I}\left\{\left|R_T(W)\cdot[\xi_1\cdots\xi_J]'\right|>\epsilon\sqrt{T}\right\}}_{Q_T(W,\xi)}\right]=o(1) \tag{25}$$

To show that this condition holds, note that

$$\begin{aligned}
R_T(w_t)\cdot[\xi_{1t}\cdots\xi_{Jt}]' &= \sum_{i=1}^{J}\frac{Df(h_0)\left[\psi_{M_i}^{(i)}\right]'\left(L_i'G_{A,i}^{-1}L_i\right)^{-1}L_i'G_{A,i}^{-1}a_{K_i}^{(i)}(w_t)}{v_T(f)}\xi_{it} \\
&\equiv \sum_{i=1}^{J}R_T^{(i)}(w_t)\,\xi_{it}
\end{aligned}$$

Now, for $i\in\mathcal{J}$,

$$\begin{aligned}
\left|R_T^{(i)}(w_t)\right| &\leq \frac{\left|\left|Df(h_0)\left[\psi_{M_i}^{(i)}\right]'\left(L_i'G_{A,i}^{-1}L_i\right)^{-1}L_i'G_{A,i}^{-1/2}\right|\right|}{v_T(f)}\times\sup_{w\in\mathcal{W}}\left|\left|G_{A,i}^{-1/2}a_{K_i}^{(i)}(w)\right|\right| \\
&\equiv \lambda_i(T)\times\zeta_{A,i} \qquad (26)
\end{aligned}$$

by the Cauchy-Schwarz inequality and thus

$$\left|\sum_{j=1}^{J}R_T^{(j)}(w_t)\,\xi_{jt}\right|\leq\sum_{j=1}^{J}|\xi_{jt}|\;\;\times\max_i\left[\lambda_i(T)\times\zeta_{A,i}\right] \tag{27}$$

Equation (27) implies that, for all $w\in\mathcal{W}$ and all $\xi\in\Xi$,

$$Q_T(w,\xi)\leq\mathbb{I}\left\{\sum_{j=1}^{J}|\xi_j|>\frac{\epsilon\sqrt{T}}{\max_i\left[\lambda_i(T)\times\zeta_{A,i}\right]}\right\}\equiv\overline{Q}_T(\xi)$$

where $Q_T(w,\xi)$ was defined in (25). Therefore, using Cauchy-Schwarz and the law of iterated expectations,

$$
\begin{aligned}
C_{0,T} &\leq \mathbb{E}\left[\sum_{j=1}^{J}\left(R_T^{(j)}(W)\right)^2 \times \sum_{j=1}^{J}\xi_j^2 \times \overline{Q}_T(\xi)\right] \\
&\leq \sum_{j=1}^{J}\mathbb{E}\left[\left(R_T^{(j)}(W)\right)^2\right]\sum_{j=1}^{J}\sup_{w\in\mathcal{W}}\mathbb{E}\left[\xi_j^2 \times \overline{Q}_T(\xi)\,|w\right]
\end{aligned}
$$

Now, note that, for $i \in \mathcal{J}$,

$$
\limsup_{T\to\infty}\mathbb{E}\left[\left(R_T^{(i)}(W)\right)^2\right] = \limsup_{T\to\infty}\left(\lambda_i(T)\right)^2 < \infty
$$

where the inequality follows from Lemma 8 below. Further, $\sup_{w\in\mathcal{W}}\mathbb{E}\left[\xi_i^2\overline{Q}_T(\xi)\,|w\right] = o(1)$ by Assumption 3(iv) and the fact that, by Assumption 4(i) and Lemma 8, $\frac{\sqrt{T}}{\max_i[\lambda_i(T)\zeta_{A,i}]} \nearrow \infty$. Therefore, $C_{0,T} = o(1)$, the Lindeberg condition is verified, and $T_1 \xrightarrow{d} N(0,1)$.

Next, for T_2, we have

$$
\begin{aligned}
|T_2| &\leq v_T(f)^{-1}\left|\left|Df(h_0)[\psi_M]'\left(G_A^{-1/2}L\right)_l^-\right|\right|\;\left|\left|G_A^{-1/2}L\left\{\left(\hat{G}_A^{-1/2}\hat{L}\right)_l^-\hat{G}_A^{-1/2}G_A^{1/2} - \left(G_A^{-1/2}L\right)_l^-\right\}\right|\right|\;\left|\left|G_A^{-1/2}A'\xi/\sqrt{T}\right|\right| \\
&= \left[\sum_{j=1}^{J}\lambda_j(T)^2\right]^{1/2}\left|\left|G_A^{-1/2}L\left\{\left(\hat{G}_A^{-1/2}\hat{L}\right)_l^-\hat{G}_A^{-1/2}G_A^{1/2} - \left(G_A^{-1/2}L\right)_l^-\right\}\right|\right|\;\left|\left|G_A^{-1/2}A'\xi/\sqrt{T}\right|\right| \\
&\leq \left[\sum_{j=1}^{J}\lambda_j(T)^2\right]^{1/2}\max_{i\in\mathcal{J}}\left|\left|G_{A,i}^{-1/2}L_i\left\{\left(\hat{G}_{A,i}^{-1/2}\hat{L}_i\right)_l^-\hat{G}_{A,i}^{-1/2}G_{A,i}^{1/2} - \left(G_{A,i}^{-1/2}L_i\right)_l^-\right\}\right|\right|\;\left|\left|G_A^{-1/2}A'\xi/\sqrt{T}\right|\right| \\
&= O_p\left(\max_{i\in\mathcal{J}}\left[\tau_{M_i}^{(i)}\zeta_i\sqrt{M_i\log M_i/T}\right]\right)
\end{aligned}
$$

The first inequality follows from some algebra and the Cauchy-Schwarz inequality, the first equality is by the definition in (26), the second inequality holds by the definition of matrix norm, and the second equality is by Lemmas A.1, F.8 and F.10(c) in CC, Lemma 8 below and Assumption 4(iii). Therefore, by Assumption 4(i), we obtain $|T_2| = o_p(1)$. This completes the proof of (19).

□

Remark 1. *Note that I do not impose Assumption 4(i) in CC. This is because the assumption is automatically satisfied if the basis functions used for the endogenous variables and those used for the instruments form a Riesz basis for the conditional expectation operator. I follow CC in assuming that this is the case.*

Lemma 7. *Let* $||\hat{h} - h_0||_\infty = o_p(1)$ *and let Assumptions 3(i), 3(ii), 3(iii), 3(v), 3(vi), 4, 5, 7(iv) hold. Then*

$$
\left|\frac{\hat{v}_T(f)}{v_T(f)} - 1\right| = o_p(1)\,. \tag{28}
$$

Proof. Following the proof of Lemma G.2 in CC, we write

$$\frac{\hat{v}_T^2(f)}{v_T^2(f)} - 1 = \frac{(\hat{\gamma}_T - \gamma_T)' \Omega^o (\hat{\gamma}_T + \gamma_T)}{v_T^2(f)} + \frac{\hat{\gamma}_T' \left(\hat{\Omega}^o - \Omega^o\right) \hat{\gamma}_T}{v_T^2(f)} \equiv T_1 + T_2 \tag{29}$$

where

$$\hat{\Omega}^o = G_A^{-1/2} \hat{\Omega} G_A^{-1/2} \qquad \hat{\gamma}_T = G_A^{1/2} \hat{G}_A^{-1} \hat{L} \left(\hat{L}' \hat{G}_A^{-1} \hat{L}\right)^{-1} Df\left(\hat{h}\right)[\psi_M]$$

$$\Omega^o = G_A^{-1/2} \Omega G_A^{-1/2} \qquad \gamma_T = G_A^{-1/2} L \left(L' G_A^{-1} L\right)^{-1} Df(h_0)[\psi_M]$$

and note that $\frac{||\gamma_T||^2}{v_T^2(f)} = \sum_{j=1}^J \lambda_j(T)^2$ by the definition in (26).
We consider T_1 and T_2 in equation (29) in turn. Note that

$$\begin{aligned}
\frac{||\hat{\gamma}_T - \gamma_T||}{v_T(f)} &= \frac{1}{v_T(f)} \left|\left| Df\left(\hat{h}\right)[\psi_M]' \left(\hat{G}_A^{-1/2} \hat{L}\right)_l^- \hat{G}_A^{-1/2} G_A^{1/2} - Df(h_0)[\psi_M]' \left(G_A^{-1/2} L\right)_l^- \right|\right| \\
&\leq \frac{1}{v_T(f)} \left|\left| Df\left(\hat{h}\right)[\psi_M]' \left(G_A^{-1/2} L\right)_l^- \right|\right| \times \left|\left| G_A^{-1/2} L \left\{ \left(\hat{G}_A^{-1/2} \hat{L}\right)_l^- \hat{G}_A^{-1/2} G_A^{1/2} - \left(G_A^{-1/2} L\right)_l^- \right\} \right|\right| \\
&\quad + \frac{1}{v_T(f)} \left|\left| \left(Df\left(\hat{h}\right)[\psi_M]' - Df(h_0)[\psi_M]'\right) \left(G_A^{-1/2} L\right)_l^- \right|\right| \equiv T_1^{(1)} \times T_1^{(2)} + T_1^{(3)}
\end{aligned}$$

Now,

$$\begin{aligned}
T_1^{(1)} &\leq \frac{1}{v_T(f)} \left|\left| \left(Df\left(\hat{h}\right)[\psi_M]' - Df(h_0)[\psi_M]'\right) \left(G_A^{-1/2} L\right)_l^- \right|\right| + \frac{J}{v_T(f)} \max_{i \in \mathcal{J}} \left|\left| Df(h_0)\left[\psi_{M_i}^{(i)}\right]' \left(G_{A,i}^{-1/2} L_i\right)_l^- \right|\right| \\
&= O_p(1)
\end{aligned}$$

where the last step follows from Assumption 7(iv) and Lemma 8. Further, $T_1^{(2)} = o_p(1)$ by Lemmas F.10(c) and A.1 in CC and Assumption 4(i),and $T_1^{(3)} = o_p(1)$ by Assumption 7(iv). This implies that

$$\frac{||\hat{\gamma}_T - \gamma_T||}{v_T(f)} = o_p(1). \tag{30}$$

Therefore, by Cauchy-Schwarz,

$$|T_1| \leq \frac{||\hat{\gamma}_T - \gamma_T||}{v_T(f)} \times ||\Omega^o|| \times \frac{||\hat{\gamma}_T + \gamma_T||}{v_T(f)} \leq \frac{||\hat{\gamma}_T - \gamma_T||}{v_T(f)} \times ||\Omega^o|| \times \left(\frac{||\hat{\gamma}_T - \gamma_T|| + 2||\gamma_T||}{v_T(f)}\right) = o_p(1)$$

where in the last step we also use Lemma 8 and the fact that $||\Omega^o|| < \infty$ by Assumptions 3(i), 3(ii), 3(iii).
Turning to $|T_2|$, note that

$$\begin{aligned}
|T_2| &\leq \frac{||\hat{\gamma}_T||}{v_T(f)} \times ||\hat{\Omega}^o - \Omega^o|| \times \frac{||\hat{\gamma}_T||}{v_T(f)} \\
&\leq \frac{||\hat{\gamma}_T - \gamma_T|| + ||\gamma_T||}{v_T(f)} \times ||\hat{\Omega}^o - \Omega^o|| \times \frac{||\hat{\gamma}_T - \gamma_T|| + ||\gamma_T||}{v_T(f)} \\
&= O_p(1) \times ||\hat{\Omega}^o - \Omega^o|| \times O_p(1)
\end{aligned}$$

where the last step follows again from Lemma 8 and (30). We complete the proof by showing that $||\hat{\Omega}^o - \Omega^o|| =$

$o_p(1)$. Note that

$$\Omega^o = \begin{bmatrix} \Omega^o_{11} & \Omega^o_{12} & \cdots & \Omega^o_{1J} \\ \Omega^o_{21} & \Omega^o_{22} & \cdots & \Omega^o_{2J} \\ \vdots & \vdots & \ddots & \cdots \\ \Omega^o_{J1} & \Omega^o_{J2} & \cdots & \Omega^o_{JJ} \end{bmatrix} \qquad \hat{\Omega}^o = \begin{bmatrix} \hat{\Omega}^o_{11} & \hat{\Omega}^o_{12} & \cdots & \hat{\Omega}^o_{1J} \\ \hat{\Omega}^o_{21} & \hat{\Omega}^o_{22} & \cdots & \hat{\Omega}^o_{2J} \\ \vdots & \vdots & \ddots & \cdots \\ \hat{\Omega}^o_{J1} & \hat{\Omega}^o_{J2} & \cdots & \hat{\Omega}^o_{JJ} \end{bmatrix}$$

where, for $j, k \in \mathcal{J}$,

$$\Omega^o_{jk} = G^{-1/2}_{A,j} \Omega_{jk} G^{-1/2}_{A,k} \qquad \hat{\Omega}^o_{jk} = G^{-1/2}_{A,j} \hat{\Omega}_{jk} G^{-1/2}_{A,k}$$

Using this notation, we have that, for any $v = [v_1' \cdots v_J']'$, with $v_j \in \mathbb{R}^{K_j}$, $j \in \mathcal{J}$, and $||v|| = 1$,

$$\left|\left| \left(\hat{\Omega}^o - \Omega^o \right) v \right|\right| \quad = \quad \sum_{j=1}^{J} v_j' \left(\hat{\Omega}^o_{jj} - \Omega^o_{jj} \right) v_j + 2 \sum_{j=1}^{J} \sum_{k=1}^{j-1} v_j' \left(\hat{\Omega}^o_{jk} - \Omega^o_{jk} \right) v_k$$

and thus, by definition of matrix norm and Cauchy-Schwarz,

$$||\hat{\Omega}^o - \Omega^o|| \quad \leq \quad J \max_{j \in \mathcal{J}} ||\hat{\Omega}^o_{jj} - \Omega^o_{jj}|| + 2J^2 \max_{j,k \in \mathcal{J}, j \neq k} ||\hat{\Omega}^o_{jk} - \Omega^o_{jk}|| \equiv \tilde{T}_1 + \tilde{T}_2$$

Now, $\tilde{T}_1 = o_p(1)$ by Lemma G.3 in CC.For $\tilde{T}_2$, note that, by the triangle inequality, for all $j, k \in \mathcal{J}, j \neq k$,

$$\begin{aligned} ||\hat{\Omega}^o_{jk} - \Omega^o_{jk}|| \leq & \left|\left| G^{-1/2}_{A,j} \left[\frac{1}{T} \sum_{t=1}^{T} \xi_{jt} \xi_{kt} a^{(j)}_{K_j}(w_t) a^{(j)}_{K_j}(w_t)' - \mathbb{E}\left(\xi_j \xi_k a^{(j)}_{K_j}(W) a^{(j)}_{K_j}(W)' \right) \right] G^{-1/2}_{A,j} \right|\right| \\ & + \left|\left| G^{-1/2}_{A,j} \frac{1}{T} \sum_{t=1}^{T} \left[\left(\hat{\xi}_{jt} - \xi_{jt} \right) \xi_{kt} a^{(j)}_{K_j}(w_t) a^{(j)}_{K_j}(w_t)' \right] G^{-1/2}_{A,j} \right|\right| \\ & + \left|\left| G^{-1/2}_{A,j} \frac{1}{T} \sum_{t=1}^{T} \left[\left(\hat{\xi}_{jt} - \xi_{jt} \right) \left(\hat{\xi}_{kt} - \xi_{kt} \right) a^{(j)}_{K_j}(w_t) a^{(j)}_{K_j}(w_t)' \right] G^{-1/2}_{A,j} \right|\right| \\ & + \left|\left| G^{-1/2}_{A,j} \frac{1}{T} \sum_{t=1}^{T} \left[\xi_{j,t} \left(\hat{\xi}_{k,t} - \xi_{k,t} \right) a^{(j)}_{K_j}(w_t) a^{(j)}_{K_j}(w_t)' \right] G^{-1/2}_{A,j} \right|\right| \\ & \equiv ||T_{\Omega,1}|| + ||T_{\Omega,2}|| + ||T_{\Omega,3}|| + ||T_{\Omega,4}|| \end{aligned}$$

where we use the fact that $G_{A,j} = G_{A,k}$ and $a^{(j)}_{K_j} = a^{(k)}_{K_k}$ for all $j, k \in \mathcal{J}$ by Assumption 5. Using Lemma 9 below, we obtain $||T_{\Omega,1}|| = o_p(1)$. Further, $||T_{\Omega,2}|| = o_p(1)$ by $\left(\hat{\xi}_{jt} - \xi_{jt} \right) \xi_{kt} \leq ||\hat{h}_j - h_{0,j}||_{1,\infty} \left(1 + \xi^2_{kt} \right)$ and Lemma F.7 in CC. Similarly, $||T_{\Omega,4}|| = o_p(1)$. Finally, $||T_{\Omega,3}|| = o_p(1)$ by $\left(\hat{\xi}_{jt} - \xi_{jt} \right) \left(\hat{\xi}_{kt} - \xi_{kt} \right) \leq ||\hat{h} - h_0||^2_{\infty}$ and Lemma F.7 in CC.

□

Lemma 8. *For $i \in \mathcal{J}$, let $\lambda_i(T) \equiv \frac{\left|\left| Df(h_0) \left[\psi^{(i)}_{M_i} \right]' \left(L_i' G^{-1}_{A,i} L_i \right)^{-1} L_i' G^{-1/2}_{A,i} \right|\right|}{v_T(f)}$ and let Assumption 3(ii) hold. Then, $\limsup_{T \to \infty} \lambda_i(T) < \infty$.*

Proof. Note that

$$\begin{aligned} v_T^2(f) &= \sum_{i=1}^{J} Df(h_0)[\psi_{M_i}^{(i)}]' \left(L_i' G_{A,i}^{-1} L_i\right)^{-1} L_i' G_{A,i}^{-1} \Omega_{ii} G_{A,i}^{-1} L_i \left(L_i' G_{A,i}^{-1} L_i\right)^{-1} Df(h_0)[\psi_{M_i}^{(i)}] \\ &+ 2\sum_{j=1}^{J}\sum_{k=1}^{j-1} Df(h_0)[\psi_{M_j}^{(j)}]' \left(L_j' G_{A,j}^{-1} L_j\right)^{-1} L_j' G_{A,j}^{-1} \Omega_{jk} G_{A,k}^{-1} L_k \left(L_k' G_{A,k}^{-1} L_k\right)^{-1} Df(h_0)[\psi_{M_k}^{(k)}] \\ &\equiv \sum_{i=1}^{J} \sigma_{T,i}^2 + 2\sum_{j=1}^{J}\sum_{k=1}^{j-1} \sigma_{T,j,k} \end{aligned}$$

Further, by Assumption 3(ii)

$$\left|\left| Df(h_0)\left[\psi_{M_i}^{(i)}\right]' \left(L_i' G_{A,i}^{-1} L_i\right)^{-1} L_i' G_{A,i}^{-1/2} \right|\right|^2 \le \underline{\sigma}^{-2} \sigma_{T,i}^2$$

for $i \in \mathcal{J}$. Therefore, we can write

$$[\lambda_i(T)]^2 \le \frac{\underline{\sigma}^{-2}\sigma_{T,i}^2}{\sum_{i=1}^{J}\sigma_{T,i}^2 + 2\sum_{j=1}^{J}\sum_{k=1}^{j-1}\sigma_{T,j,k}}$$

Since we focus on the case in which the functional f is slower than $\sqrt{T}-$estimable, the denominator in the display above goes to infinity. Since the numerator is at most of the same order as the denominator, the result follows.

□

Lemma 9. *Let Assumptions 3(iii), 3(vi), 4(ii) and 5 hold. Then* $||T_{\Omega,1}|| = O_p(1)$*, where* $T_{\Omega,1}$ *is defined in the proof of Lemma 7.*

Proof. The proof adapts that of Lemma 3.1 in Chen and Christensen (2015). Let $C_T \asymp \zeta^{\left(1+\gamma^{(2)}\right)/\gamma^{(2)}}$ be a sequence of positive numbers with $\gamma^{(2)}$ defined in Assumption 3(vi), and let

$$T_{\Omega,1}^{(1)} \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\Xi_{1,t} - \mathbb{E}[\Xi_{1,t}]\right) \qquad T_{\Omega,1}^{(2)} \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\Xi_{2,t} - \mathbb{E}[\Xi_{2,t}]\right)$$

where

$$\begin{aligned} \Xi_{1,t} &\equiv \xi_{jt}\xi_{kt} G_{A,j}^{-1/2} a_{K_j}^{(j)}(w_t) a_{K_j}^{(j)}(w_t)' G_{A,j}^{-1/2} \mathbb{I}\left\{||\xi_{jt}\xi_{kt} G_{A,j}^{-1/2} a_{K_j}^{(j)}(w_t) a_{K_j}^{(j)}(w_t)' G_{A,j}^{-1/2}|| \le C_T^2\right\} \\ \Xi_{2,t} &\equiv \xi_{jt}\xi_{kt} G_{A,j}^{-1/2} a_{K_j}^{(j)}(w_t) a_{K_j}^{(j)}(w_t)' G_{A,j}^{-1/2} \mathbb{I}\left\{||\xi_{jt}\xi_{kt} G_{A,j}^{-1/2} a_{K_j}^{(j)}(w_t) a_{K_j}^{(j)}(w_t)' G_{A,j}^{-1/2}|| > C_T^2\right\} \end{aligned}$$

Note that $T_{\Omega,1} = T_{\Omega,1}^{(1)} + T_{\Omega,1}^{(2)}$, so that $||T_{\Omega,1}^{(1)}|| = o_p(1)$ and $||T_{\Omega,1}^{(2)}|| = o_p(1)$ imply the statement of the lemma. By definition, $||\Xi_{1,t}|| \le C_T^2$ and thus, by the triangle inequality and Jensen's inequality ($||\cdot||$ is convex), we have $||\Xi_{1,t} - \mathbb{E}(\Xi_{1,t})|| \le 2C_T^2$. Further, dropping the t subscripts,

$$\begin{aligned} &\mathbb{E}[\Xi_1 - \mathbb{E}(\Xi_1)]^2 \le \\ &\mathbb{E}\left[\xi_j^2\xi_k^2 ||G_{A,j}^{-1/2} a_{K_j}^{(j)}(W)||^2 G_{A,j}^{-1/2} a_{K_j}^{(j)}(W) a_{K_j}^{(j)}(W)' G_{A,j}^{-1/2} \mathbb{I}\left\{||\xi_j\xi_k G_{A,j}^{-1/2} a_{K_j}^{(j)}(W) a_{K_j}^{(j)}(W)' G_{A,j}^{-1/2}|| \le C_T^2\right\}\right] \le \\ &C_T^2 \mathbb{E}\left[|\xi_j\xi_k| G_{A,j}^{-1/2} a_{K_j}^{(j)}(W) a_{K_j}^{(j)}(W)' G_{A,j}^{-1/2} \mathbb{I}\left\{||\xi_j\xi_k G_{A,j}^{-1/2} a_{K_j}^{(j)}(W) a_{K_j}^{(j)}(W)' G_{A,j}^{-1/2}|| \le C_T^2\right\}\right] \le \end{aligned}$$

$$C_T^2 \mathbb{E}\left[\mathbb{E}\left(|\xi_j\xi_k||W\right) G_{A,j}^{-1/2} a_{K_j}^{(j)}\left(W\right) a_{K_j}^{(j)}\left(W\right)' G_{A,j}^{-1/2}\right] \lesssim$$
$$C_T^2 \mathbb{E}\left[G_{A,j}^{-1/2} a_{K_j}^{(j)}\left(W\right) a_{K_j}^{(j)}\left(W\right)' G_{A,j}^{-1/2}\right] = C_T^2 I_{K_j}$$

where the inequalities are in the sense of positive semi-definite matrices. Then, Corollary 4.1 in Chen and Christensen (2015) yields $||T_{\Omega,1}^{(1)}|| = O_p\left(C_T\sqrt{(\log K)/T}\right)$ and thus $||T_{\Omega,1}^{(1)}|| = o_p\left(1\right)$ by Assumption 4(ii). Turning to $||T_{\Omega,1}^{(2)}||$, since $||\Xi_{2,t}|| \leq \zeta^2|\xi_{jt}\xi_{kt}|\mathbb{I}\left\{|\xi_{jt}\xi_{kt}| \geq C_T^2/\zeta^2\right\}$, by the triangle inequality and Jensen's inequality ($||\cdot||$ is convex), we have

$$\mathbb{E}\left[||T_{\Omega,1}^{(2)}||\right] \leq 2\zeta^2 \mathbb{E}\left[|\xi_j\xi_k|\mathbb{I}\left\{|\xi_j\xi_k| \geq C_T^2/\zeta^2\right\}\right] \leq \frac{2\zeta^{2(1+\gamma^{(2)})}}{C_T^{2\gamma^{(2)}}}\mathbb{E}\left[|\xi_j\xi_k|^{1+\gamma^{(2)}}\mathbb{I}\left\{|\xi_j\xi_k| \geq C_T^2/\zeta^2\right\}\right] = o\left(1\right)$$

where the last step follows from Assumption 3(vi), the fact that $C_T^2/\zeta^2 \asymp \zeta^{2/\gamma^{(2)}} \to \infty$ and that $\zeta^{(1+\gamma^{(2)})}/C_T^{\gamma^{(2)}} \asymp 1$. Thus, $||T_{\Omega,1}^{(2)}|| = o_p\left(1\right)$ by Markov's inequality.

□

Lemma 10. *Let Assumptions 3 and 8(i)-8(iii) hold. Then, for* $f \in \{f_\epsilon, f_{p_1}\}$,

$$\left[v_T\left(f\right)\right]^2 \lesssim \tau_M^2 M^4$$

Proof. We prove this for $f = f_\epsilon$. The proof for $f = f_{p_1}$ is identical. As shown in CC,[52] the maintained assumptions imply

$$\left[v_T\left(f_\epsilon\right)\right]^2 \asymp \tau_M^2 \sum_{m=1}^{M}\left(Df_\epsilon\left(h_0\right)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right]\right)^2 \tag{31}$$

where $\left(G_\psi^{-1/2}\psi_M\right)_m$ denotes the $m-$th row of the $M-$by$-2-$valued function $G_\psi^{-1/2}\psi_M$. Next,

$$\left|Df_\epsilon\left(h_0\right)\left[\left(G_\psi^{-1/2}\psi_M\right)_m\right]\right| \lesssim \max_{\tilde{\alpha}:|\tilde{\alpha}|=1}||\partial^{\tilde{\alpha}}\left(G_\psi^{-1/2}\psi_M\right)_m||_\infty \asymp M^{3/2}$$

where the first step follows from (12) and the second step follows from well-known properties of splines (see, e.g., Newey (1997)). Combining this and (31) completes the proof. □

Appendix C: Constraints

In this appendix, I provide more details on how to impose some of the constraints discussed in Section 3.2, and I introduce some additional constraints.

C.1 Imposing Exchangeability

First, I discuss how to impose the exchangeability restriction defined in Section 3.2 (see equation (5)). As in the main text, I consider the case where $x^{(2)}$ is a vector of product-specific characteristics each

[52] See pp.22-23. See also Chen and Pouzo (2015).

with dimension $\tilde{n}_{x^{(2)}}$. With J goods, the overall degree of the approximation is then $(2J + \tilde{n}_{x^{(2)}})m$. Let $v^s \equiv (v_1^s, \cdots, v_J^s)$ be a $J-$vector of integers between 0 and m, and define $v^p \equiv (v_1^p, \cdots, v_J^p)$, and $v^x \equiv (v_1^x, \cdots, v_J^x)$ similarly. Next, let $\theta_j\left(v_1^s, \cdots, v_J^s, v_1^p, \cdots, v_J^p, v_1^x, \cdots, v_J^x; m\right)$ denote the coefficient on the term $\Pi_{k=1}^J b_{v_k^s,m}\left(s_k\right) b_{v_k^p,m}\left(p_k\right) b_{v_k^x,m}\left(x_k^{(2)}\right)$ in the Bernstein approximation for σ_j^{-1}. Let $\pi : \{1,...,J\} \mapsto \{1,...,J\}$ be any permutation, π^{-1} be its inverse, and $\tilde{\pi}$ denote the function that, for any J-vector y, returns the reshuffled version of y obtained by permuting its subscripts according to π, i.e.

$$\tilde{\pi}\left(y_1, ..., y_J\right) = \left[y_{\pi(1)}, ..., y_{\pi(J)}\right]$$

$\tilde{\pi}^{-1}$ is defined similarly for π^{-1}. Then, exchangeability of the Bernstein approximation takes the form

$$\theta_j\left(v^s, v^p, v^x; m\right) = \theta_{\pi(j)}\left(\tilde{\pi}^{-1}\left(v^s\right), \tilde{\pi}^{-1}\left(v^p\right), \tilde{\pi}^{-1}\left(v^x\right); m\right) \tag{32}$$

for all $v_k^s, v_k^p, v_k^x \in \{0, 1, \cdots, m\}$. This is a set of linear constraints on the Bernstein coefficients that can be easily be enforced. In fact, one can directly embed the constraint into the definition of the vector of Bernstein coefficient, thus reducing the dimension of the program to be solved in estimation (equation (3)).

Without exchangeability, the number of coefficients to estimate for each demand function is equal to $(m+1)^{J\left(2+\tilde{n}_{x^{(2)}}\right)}$. In contrast, when exchangeability is imposed, that number is $\left[\frac{(m+J-1)!}{(J-1)!(m)!}\left(m+1\right)\right]^{2+\tilde{n}_{x^{(2)}}}$. To see this, note that θ_j in (32) has $J\left(2+\tilde{n}_{x^{((2))}}\right)$ arguments, of which $2+\tilde{n}_{x^{(2)}}$ are "own" argument (i.e. j's share, price and $x^{(2)}$ attributes) and $(J-1)\left(2+\tilde{n}_{x^{(2)}}\right)$ are other goods' arguments. Exchangeability of σ_j^{-1} means that the function is invariant to rearranging the rival goods's arguments, for any given value of the own arguments. Now, the number of ways $(J-1)$ numbers can be drawn with replacement from a set of size $m+1$ is $\frac{(m+J-1)!}{(J-1)!(m)!}$.[53] Repeating this for $m+1$ possible values of each own argument and for $2+\tilde{n}_{x^{(2)}}$ arguments per good (share, price and $x^{(2)}$ attributes), one obtains the total number of coefficients under exchangeability.

Finally, I consider the case where $x^{(2)}$ is a vector of market-level variables that are not product-specific (e.g. income). The corresponding definition of exchangeability was given in footnote 28. In this case, the analog of equation (32) is

$$\theta_j\left(v^s, v^p, v^x; m\right) = \theta_{\pi(j)}\left(\tilde{\pi}^{-1}\left(v^s\right), \tilde{\pi}^{-1}\left(v^p\right), v^x; m\right) \tag{33}$$

for all $v_k^s, v_k^p, v_k^x \in \{0, 1, \cdots, m\}$. Again, this is a set of equalities between pairs of Bernstein coefficients, which reduces the number of parameters to estimate.

C.2 Symmetry of the Jacobians

Next, I turn to an additional type of constraints that one might want to impose, namely symmetry of the Jacobians. Because these constraints are defined conditional on any given value of $x^{(2)}$, I drop this for notational convenience.

[53]This is sometimes called a multicombination.

Let $\mathbb{J}^{\delta}_{\sigma}(\delta, p)$ denote the Jacobian matrix of σ with respect to δ:

$$\mathbb{J}^{\delta}_{\sigma}(\delta, p) = \begin{bmatrix} \frac{\partial}{\partial \delta_1}\sigma_1(\delta, p) & \cdots & \frac{\partial}{\partial \delta_J}\sigma_1(\delta, p) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial \delta_1}\sigma_J(\delta, p) & \cdots & \frac{\partial}{\partial \delta_J}\sigma_J(\delta, p) \end{bmatrix}$$

In a discrete choice model where δ_j is interpreted as a quality index for good j, if one assumes that, for all j, δ_j enters the utility of good j linearly (and does not enter the utility of the other goods), then $\mathbb{J}^{\delta}_{\sigma}(\delta, p)$ must be symmetric. Conveniently, symmetry of $\mathbb{J}^{\delta}_{\sigma}(\delta, p)$ implies linear constraints on the Bernstein coefficients. To see this, note that by the implicit function theorem, for every (δ, p) and for $s = \sigma(\delta, p)$,

$$\mathbb{J}^{s}_{\sigma^{-1}}(s, p) = \left[\mathbb{J}^{\delta}_{\sigma}(\delta, p)\right]^{-1} \tag{34}$$

Because the inverse of a symmetric matrix is symmetric, symmetry of $\mathbb{J}^{\delta}_{\sigma}(\delta, p)$ implies symmetry of $\mathbb{J}^{s}_{\sigma^{-1}}(s, p)$. This, in turn, by Result 1 yields *linear* constraints on the Bernstein coefficients.

Similarly, let $\mathbb{J}^{p}_{\sigma}(\delta, p)$ denote the Jacobian matrix of σ with respect to p:

$$\mathbb{J}^{p}_{\sigma}(\delta, p) = \begin{bmatrix} \frac{\partial}{\partial p_1}\sigma_1(\delta, p) & \cdots & \frac{\partial}{\partial p_J}\sigma_1(\delta, p) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial p_1}\sigma_J(\delta, p) & \cdots & \frac{\partial}{\partial p_J}\sigma_J(\delta, p) \end{bmatrix}$$

This matrix is the Jacobian of the Marshallian demand system. Under the assumption that there are no income effects, it coincides with the Jacobian of the Hicksian demand by Slutsky equation and therefore it must be symmetric. Imposing symmetry of $\mathbb{J}^{p}_{\sigma}$ requires nonlinear, nonconvex constraints. This is because, by the implicit function theorem, for every (δ, p) and for $s = \sigma(\delta, p)$,

$$\mathbb{J}^{p}_{\sigma}(\delta, p) = -\left[\mathbb{J}^{s}_{\sigma^{-1}}(s, p)\right]^{-1} \mathbb{J}^{p}_{\sigma^{-1}}(s, p) \tag{35}$$

which shows that $\mathbb{J}^{p}_{\sigma}$ is a nonlinear function of the derivatives of σ^{-1} and therefore of the Bernstein coefficients. In the implementation, it might be convenient to rewrite (35) as

$$\mathbb{J}^{s}_{\sigma^{-1}}(s, p)\, \mathbb{J}^{p}_{\sigma}(\delta, p) = -\mathbb{J}^{p}_{\sigma^{-1}}(s, p)$$

Expressing $\mathbb{J}^{s}_{\sigma^{-1}}$ and $\mathbb{J}^{p}_{\sigma^{-1}}$ as linear combinations of the Bernstein polynomials and introducing extra parameters (call them γ) for the entries of $\mathbb{J}^{p}_{\sigma}$, one then obtains a set of constraints that are linear in the Bernstein coefficients, given γ, and linear in γ, given the Bernstein coefficients.[54]

C.3 Proofs for Results on Constraints

Proof of Lemma 1

First, we show that $\mathbb{J}^{\delta}_{\sigma}(\delta, p)$ belongs to the class of M-matrices. This follows from the fact that (a) under assumption 2 in BH, Theorem 2 in Berry et al. (2013) implies that $\mathbb{J}^{\delta}_{\sigma}(\delta, p)$ is a P-matrix for every (δ, p), i.e. a square matrix such that all of its principal minors are strictly positive; (b) by weak substitutability

[54] This is helpful especially when it comes to writing the analytic gradient of the constraints to input in the optimization problem.

$\mathbb{J}_{\sigma}^{\delta}(\delta, p)$ is also a Z-matrix, i.e. a matrix with non-positive off-diagonal entries; and (c) a Z-matrix which is also a P-matrix is an M-matrix (see, e.g., result 8.148 in Seber (2007)). Next, by the implicit function theorem $\mathbb{J}_{\sigma^{-1}}^{s}(s, p) = \left[\mathbb{J}_{\sigma}^{\delta}(\delta, p)\right]^{-1}$ for all δ, p and $s = \sigma(\delta, p)$. Thus, since $\mathbb{J}_{\sigma}^{\delta}(\delta, p)$ is an M-matrix, it follows that $\mathbb{J}_{\sigma^{-1}}^{s}(s, p)$ is an inverse M-matrix. Part (i) of the lemma then follows directly from the definition of an M-matrix. For part (ii) and (iii), note that $\left|\frac{\partial \sigma_j}{\partial \delta_j}(\delta, p)\right| \geq \sum_{k \neq j} \left|\frac{\partial \sigma_k}{\partial \delta_j}(\delta, p)\right|$ means that $\mathbb{J}_{\sigma}^{\delta}(\delta, p)$ is diagonally dominant of its columns. Then, Theorem 3.2 of McDonald et al. (1995) implies the result in (ii), where again we use the fact that $\mathbb{J}_{\sigma^{-1}}^{s}(s, p) = \left[\mathbb{J}_{\sigma}^{\delta}(\delta, p)\right]^{-1}$. Finally, by the implicit function theorem $\mathbb{J}_{\sigma^{-1}}^{p}(s, p) = -\mathbb{J}_{\sigma^{-1}}^{s}(s, p)\, \mathbb{J}_{\sigma}^{p}(\delta, p)$. The result in (iii) then follows from those in (i) and (ii) and the assumption that the own-price effects $\frac{\partial \sigma_j}{\partial p_j}$ be negative.

□

Proof of Lemma 2. For part (i), let $\pi : \{1, ..., J\} \to \{1, ..., J\}$ be any permutation with inverse π^{-1}. Further, let $\tilde{\pi}$ denote the function that, for any J-vector y, returns the reshuffled version of y obtained by permuting its subscripts according to π, i.e.

$$\tilde{\pi}(y_1, ..., y_J) = \left[y_{\pi(1)}, ..., y_{\pi(J)}\right]$$

and define $\tilde{\pi}^{-1}$ similarly for π^{-1}. Then, we can rewrite the definition of exchangeability for a generic $J-$valued function $g(y_1, y_2, y_3)$ of $3J$ arguments as

$$\tilde{\pi}^{-1}(g(y_1, y_2, y_3)) = g\left(\tilde{\pi}^{-1}(y_1), \tilde{\pi}^{-1}(y_2), \tilde{\pi}^{-1}(y_3)\right).$$

Now take any $\left(\delta, p, x^{(2)}\right)$ and let $s = \sigma\left(\delta, p, x^{(2)}\right)$. We can invert the demand system to obtain

$$\delta = \sigma^{-1}\left(s, p, x^{(2)}\right) \tag{36}$$

By exchangeability of σ,

$$\tilde{\pi}^{-1}(s) = \sigma\left(\tilde{\pi}^{-1}(\delta), \tilde{\pi}^{-1}(p), \tilde{\pi}^{-1}\left(x^{(2)}\right)\right)$$

Inverting this demand system, we obtain

$$\tilde{\pi}^{-1}(\delta) = \sigma^{-1}\left(\tilde{\pi}^{-1}(s), \tilde{\pi}^{-1}(p), \tilde{\pi}^{-1}\left(x^{(2)}\right)\right) \tag{37}$$

Combining (36) and (37),

$$\tilde{\pi}^{-1}\left(\sigma^{-1}\left(s, p, x^{(2)}\right)\right) = \sigma^{-1}\left(\tilde{\pi}^{-1}(s), \tilde{\pi}^{-1}(p), \tilde{\pi}^{-1}\left(x^{(2)}\right)\right)$$

which shows that σ^{-1} is exchangeable. This proves part (i).

Part (ii) of the lemma follows directly from the definition of exchangeability and Result 1.

□

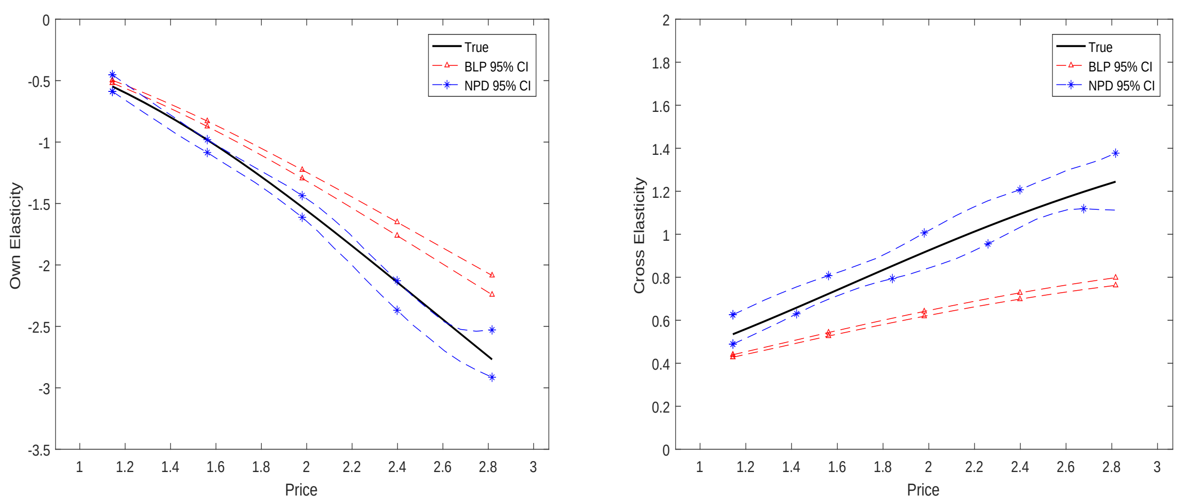
Appendix D: Additional Monte Carlo Simulations

D.1 Reference prices

Another type of behavior allowed by the NPD model is one where consumers like (dislike) a product more if its price is lower (higher) than its competitor's, all else equal. The idea is that consumers might enjoy the feeling of getting a bargain and, conversely, might be turned off if they perceive a good is over-priced. I formalize this by letting the utility for good j be a function not only of the price of j but also a (decreasing) function of the difference between the price of j and that of its competitor. I set the coefficient on the price difference to -0.15; the simulation design is otherwise the same as that in Section 4.1. As in the previous simulations, I compare the performance of the nonparametric approach with that of a mixed logit model. In this case, the latter is misspecified in that it only allows p_1, but not $p_1 - p_2$ to enter the utility of good 1, and similarly for good 2. In the nonparametric estimation, I impose the following constraints: monotonicity of σ^{-1}, diagonal dominance of $\mathbb{J}^{\delta}_{\sigma}$ and exchangeability.[55]

Figure 5 shows the own- and cross-price elasticity functions, respectively. While the nonparametric approach is on target, BLP tends to underestimate the magnitude of both due to the fact that it does not capture the reference pricing patterns in the data.

Figure 5: Reference Prices: Own-price (left) and cross-price (right) elasticity functions

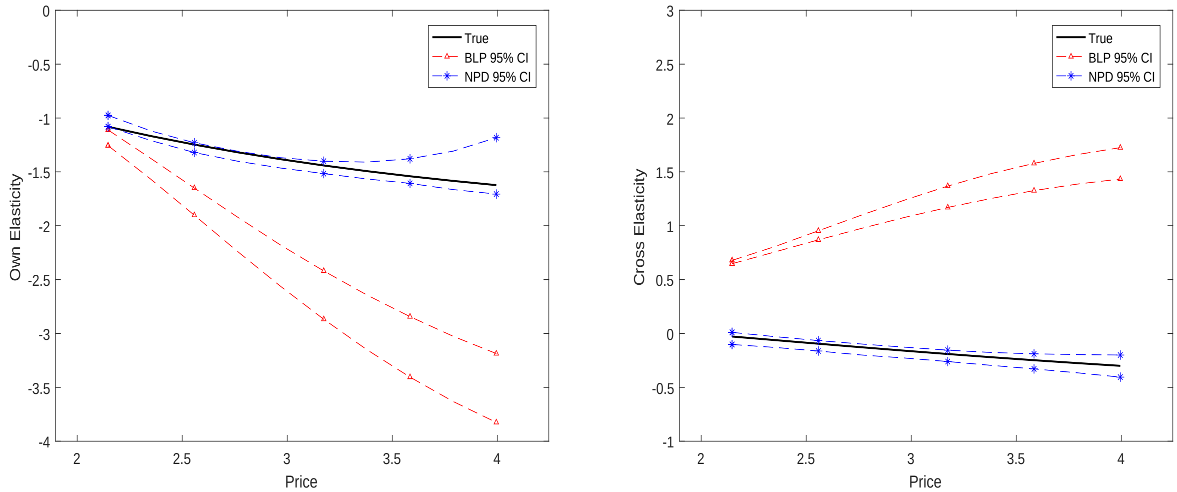


D.2 Smaller Sample Size

The simulations in Section 4 were based on sample sizes equal to 3,000. I now investigate how well the NPD estimator performs in a smaller sample size. Specifically, I focus on the complements example from Section 4.3 and repeat the simulation now using a sample of 500 observations.

[55]See Section 3.2 and Appendix C for a discussion of these constraints.

Figure 6: Complements, $T = 500$: Own-price (left) and cross-price (right) elasticity functions



D.3 Violation of the Index Restriction

The NPD estimator is based on the index restriction embedded in Equation (1). Here, I explore how robust the estimator is to violations of this assumption. Specifically, I generate the data from the mixed logit dgp as in Section 4.1, except that I let the coefficient on the covariate x be random and distributed $N(1, \sigma_{viol})$. Because the coefficient on the unobservable ξ is not random, this induces a violation of the index restriction which becomes more severe as σ_{viol} increases. Figures 7 to 9 show that, except for the own-price elasticity function at large values of own-price, the NPD estimator is quite robust to violations of the index assumption for this dgp. These results complement those on the median elasticities (Table 3 in the main text) by showing robustness of the entire own- and cross-elasticity functions.

Figure 7: Violation of Index Restriction, $\sigma_{viol} = 0.10$: Own-price (left) and cross-price (right) elasticity functions

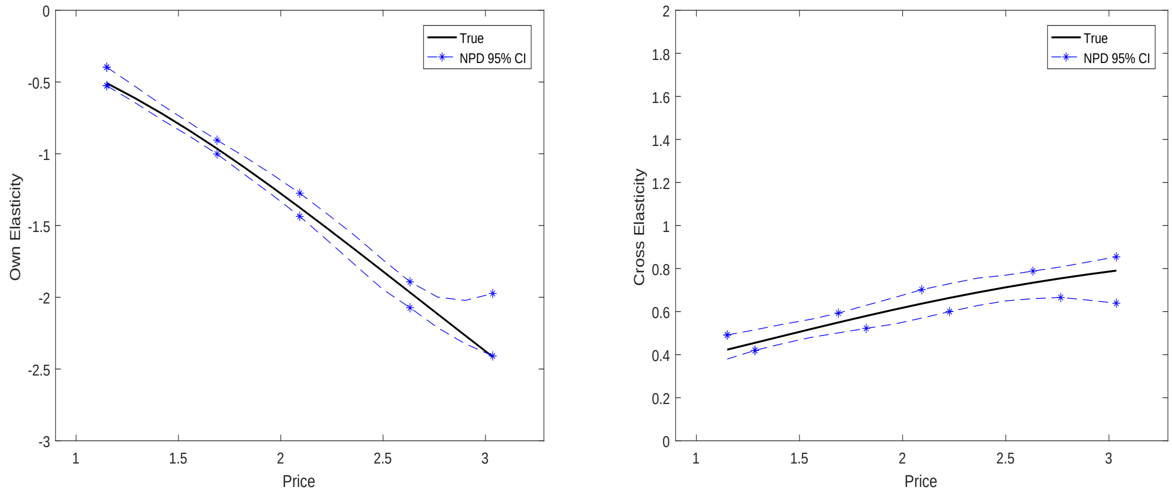


Figure 8: Violation of Index Restriction, $\sigma_{viol} = 0.50$: Own-price (left) and cross-price (right) elasticity functions

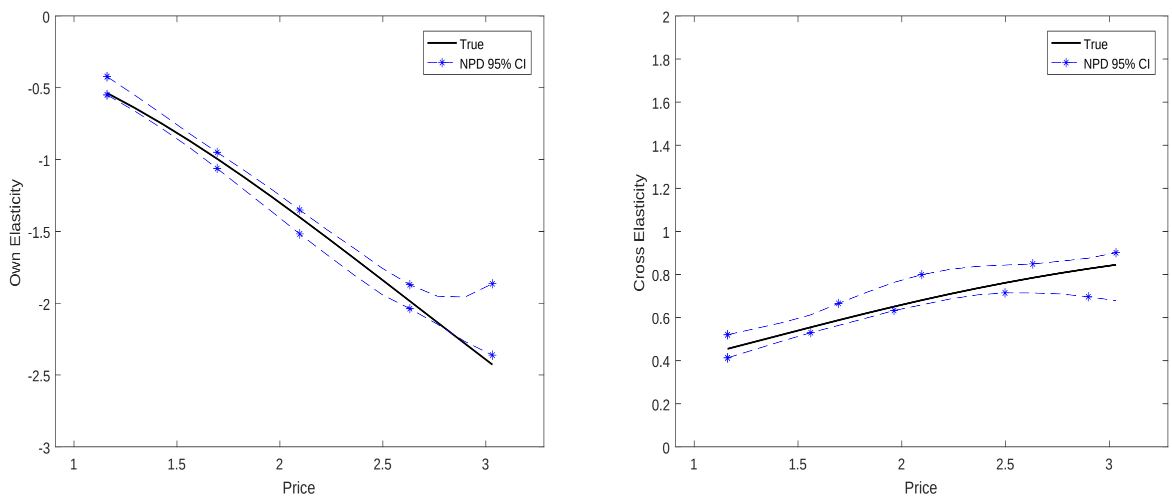
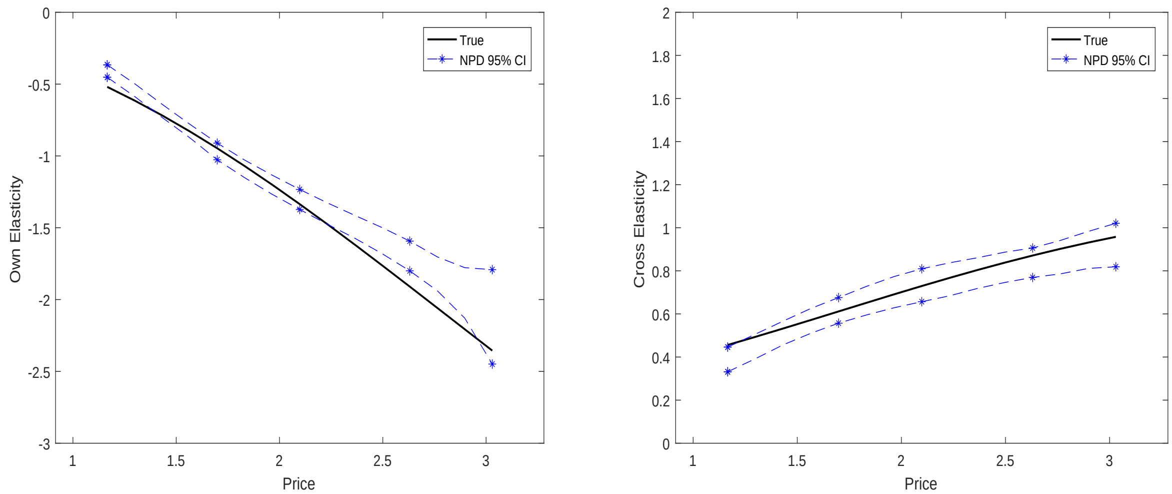


Figure 9: Violation of Index Restriction, $\sigma_{viol} = 1.50$: Own-price (left) and cross-price (right) elasticity functions



D.4 Sensitivity to the Choice of Polynomial Degree

To complement the results in Table 2 in the main text, I consider how the entire own- and cross-elasticity functions estimates vary as the degree for the polynomial approximation changes. I focus on the mixed logit dgp from Section 4.1 and the complements dgp from Section 4.3.

D.4.1 Mixed logit dgp

Figure 10: Mixed Logit Data, degree = 16: Own-price (left) and cross-price (right) elasticity functions

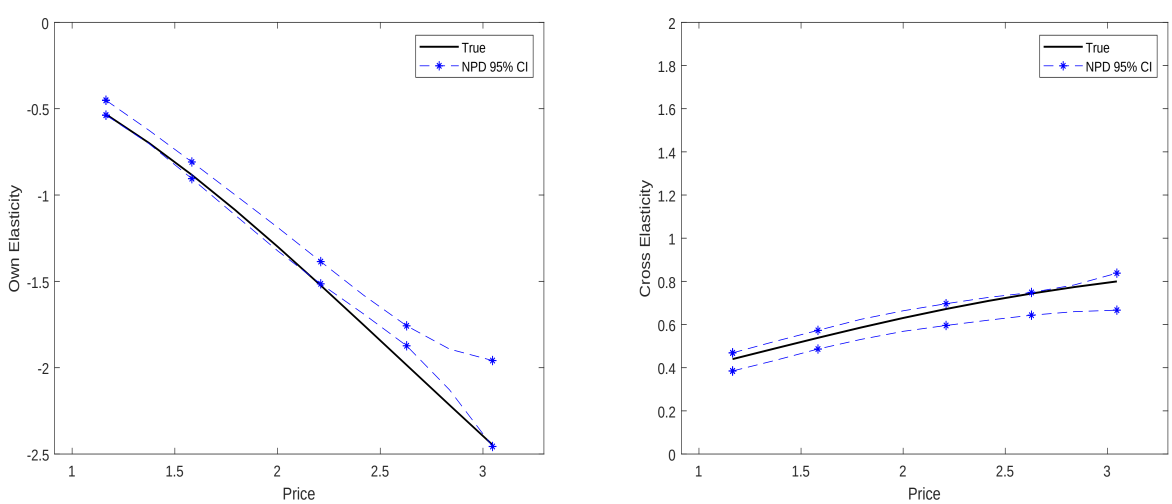


Figure 11: Mixed Logit Data, degree = 12: Own-price (left) and cross-price (right) elasticity functions

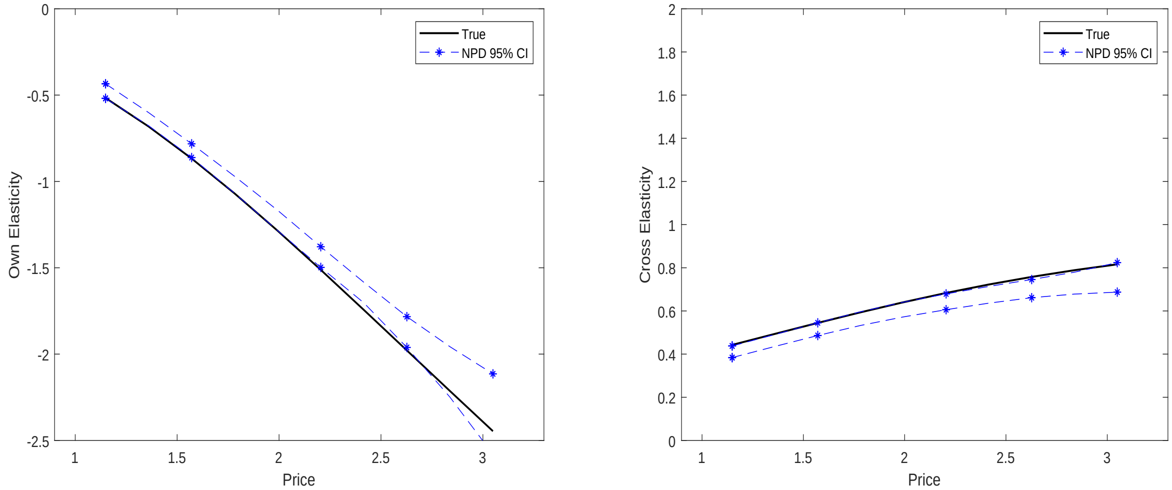


Figure 12: Mixed Logit Data, degree = 8: Own-price (left) and cross-price (right) elasticity functions

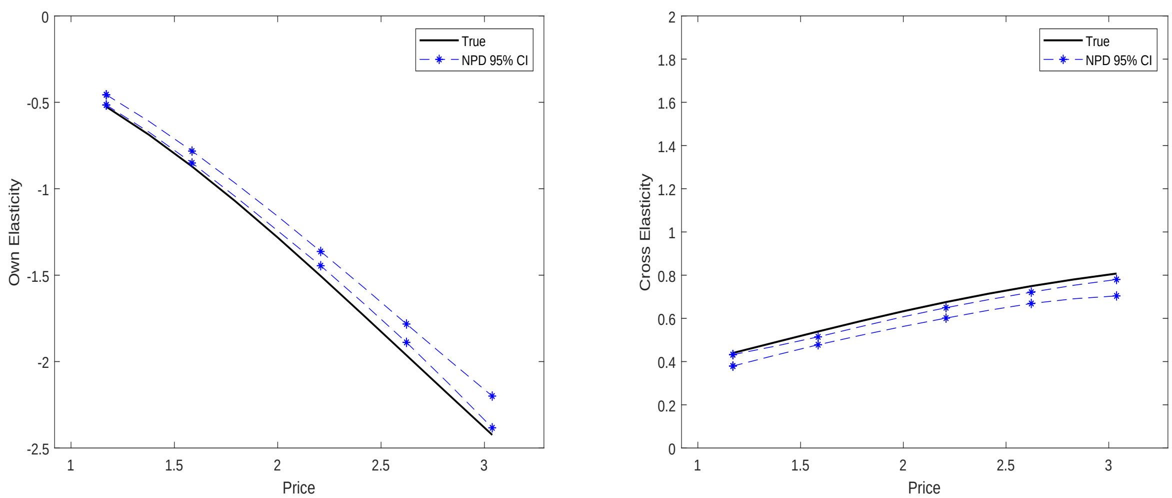


Figure 13: Mixed Logit Data, degree = 6: Own-price (left) and cross-price (right) elasticity functions

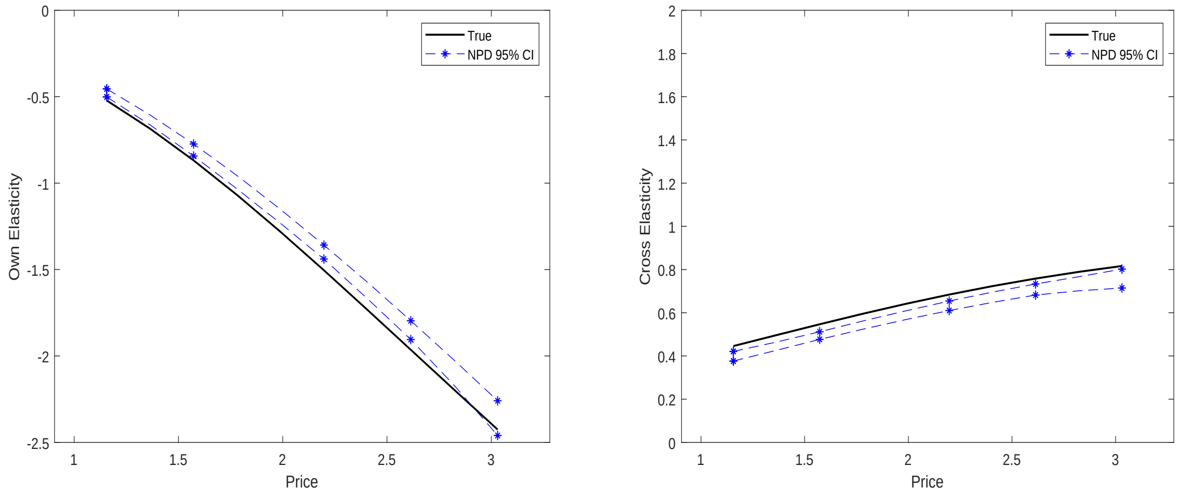
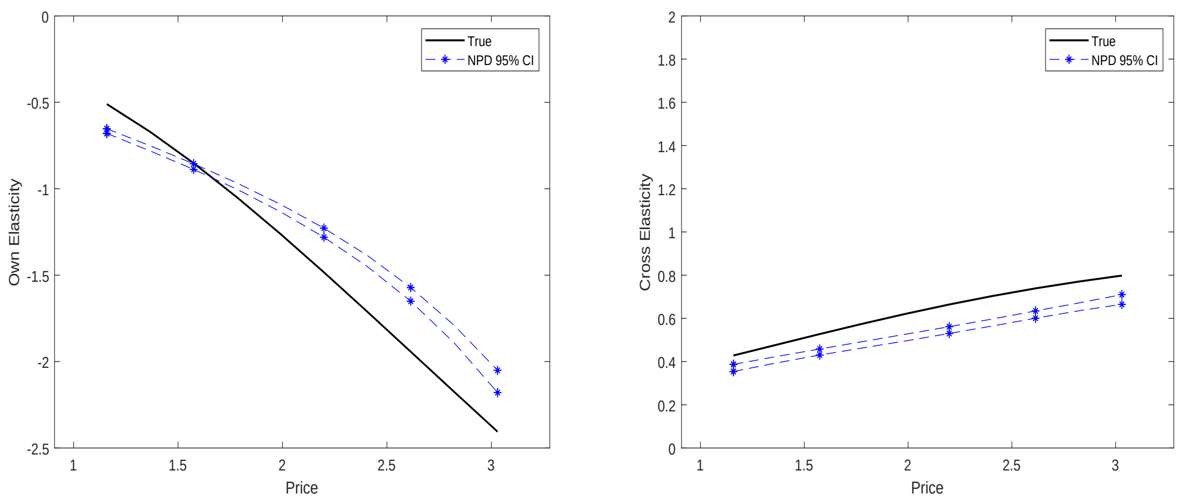


Figure 14: Mixed Logit Data, degree = 4: Own-price (left) and cross-price (right) elasticity functions



D.4.2 Complements dgp

Figure 15: Complements, degree = 16: Own-price (left) and cross-price (right) elasticity functions

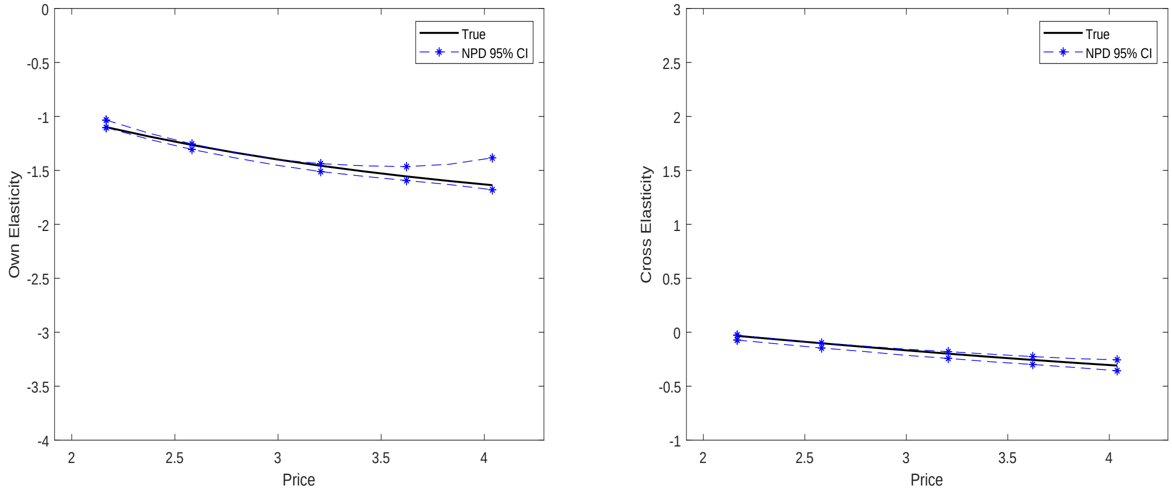


Figure 16: Complements, degree = 12: Own-price (left) and cross-price (right) elasticity functions

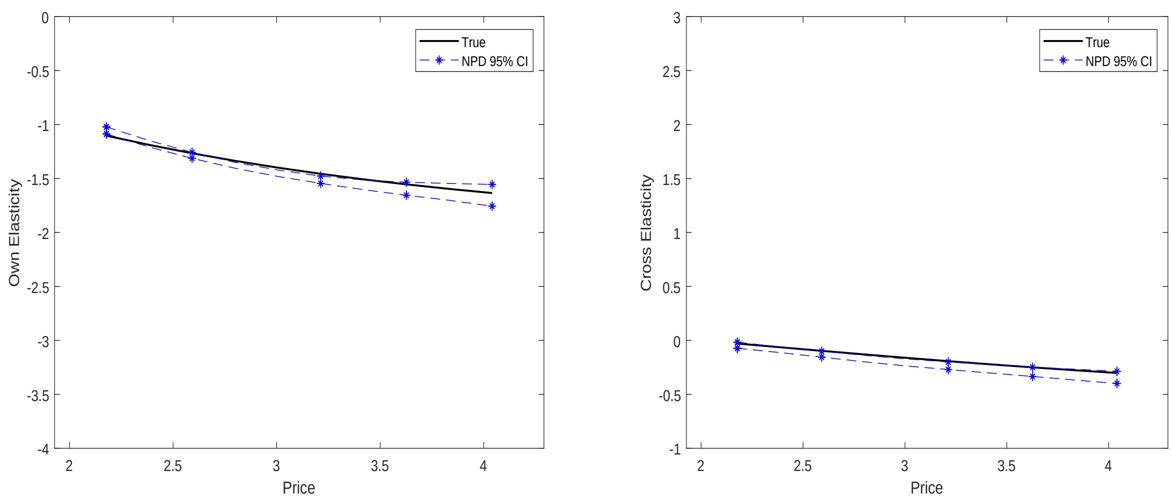


Figure 17: Complements, degree = 8: Own-price (left) and cross-price (right) elasticity functions

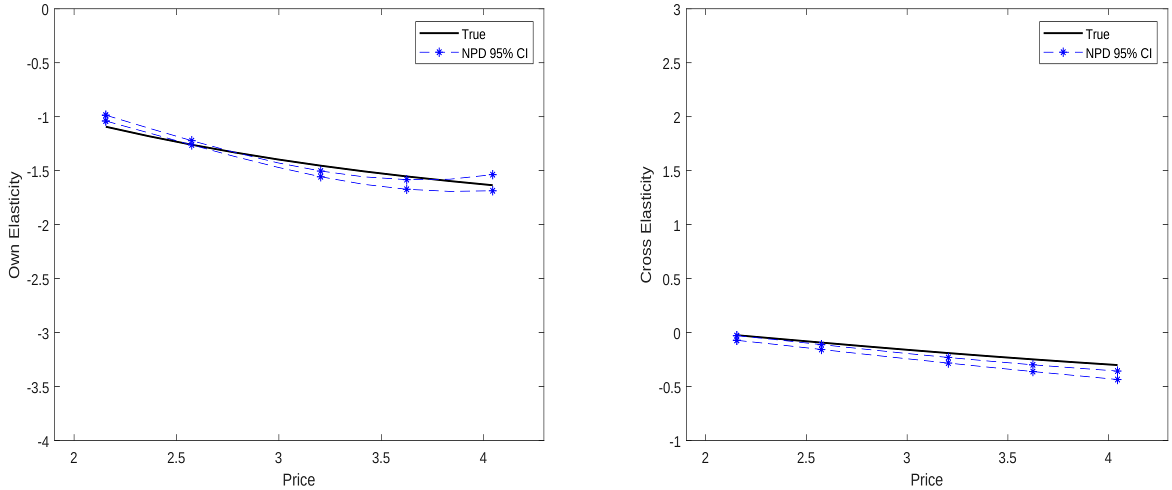


Figure 18: Complements, degree = 6: Own-price (left) and cross-price (right) elasticity functions

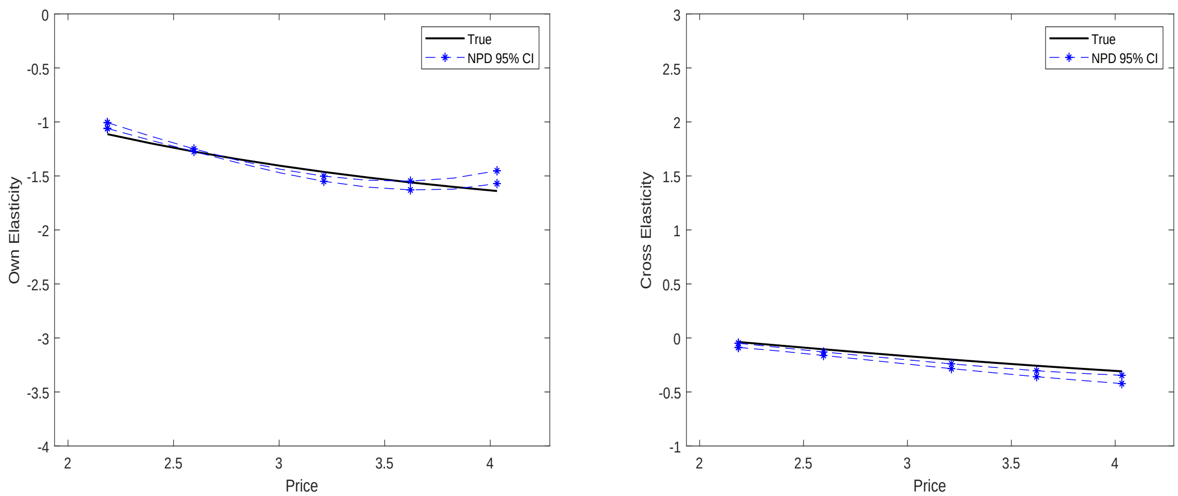
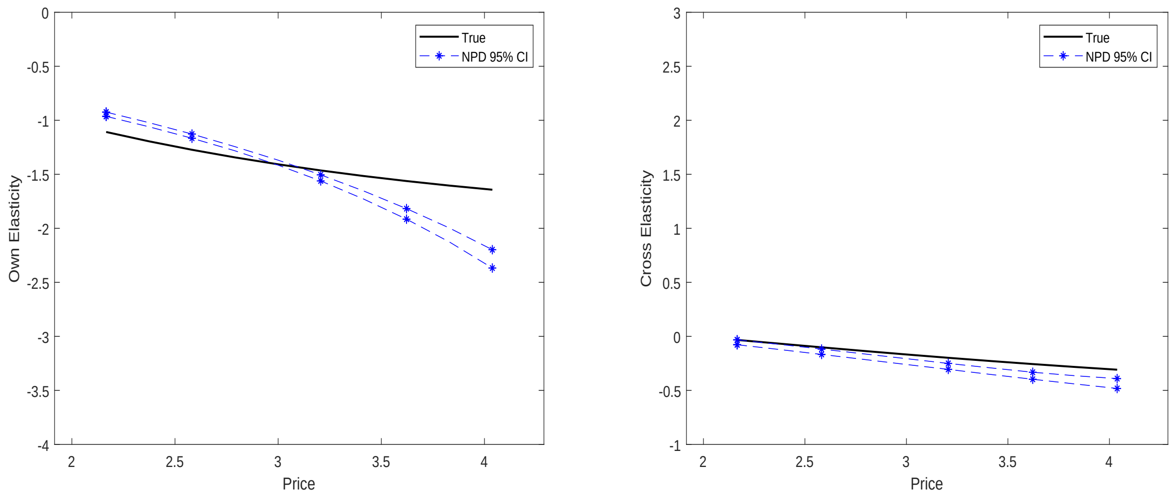


Figure 19: Complements, degree = 4: Own-price (left) and cross-price (right) elasticity functions



D.5 $J > 2$ goods

To complement the results in Table 3 in the main text, here I report estimates for the entire own- and cross-elasticity functions for the $J > 2$ goods case. I generate data from the logit model

$$u_{ij} = -p_j + x_j + \xi_j + \epsilon_{ij}$$

I choose this simple model as it means that I can put p_j into the linear index δ_j, which reduces the number of parameters to estimate. I report the own-price elasticity of good 1 and the elasticity of good 1 wrt the price of good 2 for $J = 3, J = 5$, and $J = 7$ below.[56]

[56]Since the dgp and the model are symmetric in the different goods, the remaining own- and cross-price elasticities are the same as those reported here.

Figure 20: Logit Data, $J = 3$: Own-price (left) and cross-price (right) elasticity functions

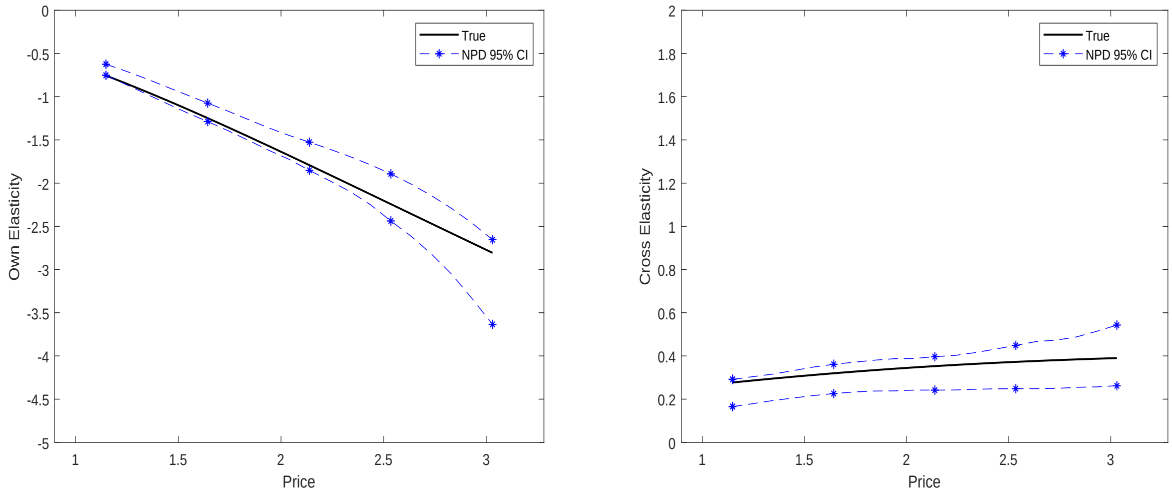


Figure 21: Logit Data, $J = 5$: Own-price (left) and cross-price (right) elasticity functions

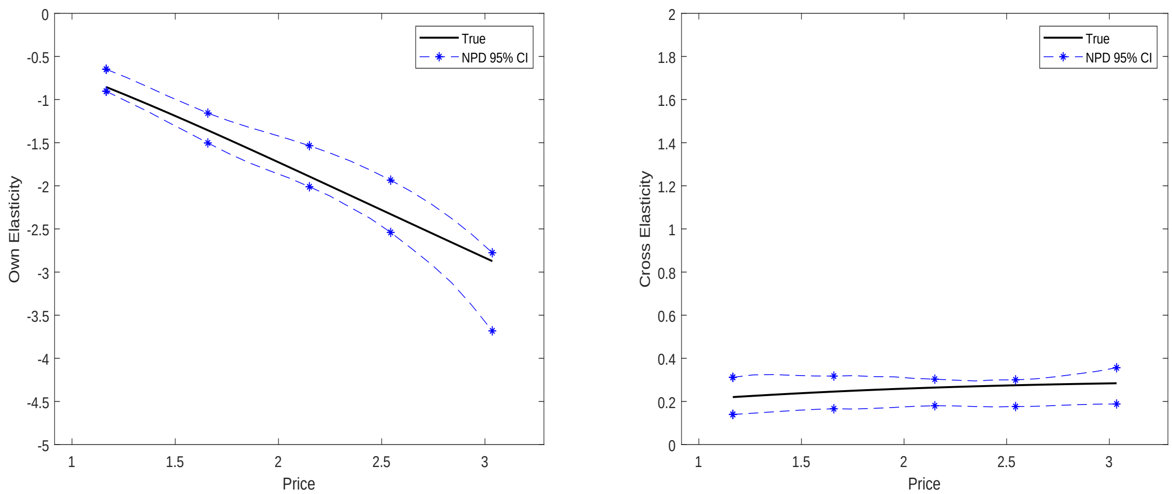
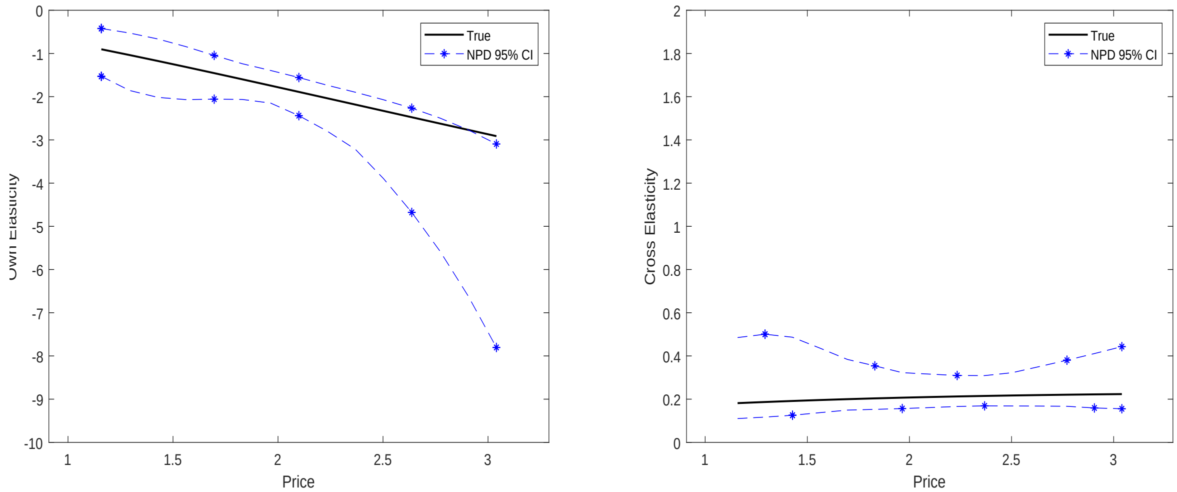


Figure 22: Logit Data, $J = 7$: Own-price (left) and cross-price (right) elasticity functions



D.6 Semiparametric Logit Model

Here, I consider a semiparametric version of the model that maintains the logit distributional assumption on the error terms, but is flexible on how prices and the $x^{(2)}$ covariates enter the demand functions. In particular, I consider the case where the $x^{(2)}$ covariates are product-specific characteristics and assume that demand functions take the form

$$\sigma_j\left(\delta_t, p_t, x_t^{(2)}\right) = \frac{\exp(\delta_{jt} + g\left(p_{jt}, x_{jt}^{(2)}\right))}{1 + \sum_{k=1}^{J} \exp(\delta_{kt} + g\left(p_{kt}, x_{kt}^{(2)}\right))}$$

for an unknown function g, which leads to

$$\log\left(\frac{s_{jt}}{s_{0t}}\right) = \delta_{jt} + g\left(p_{jt}, x_{jt}^{(2)}\right) \equiv \beta x_{jt}^{(1)} + \xi_{jt} + g\left(p_{jt}, x_{jt}^{(2)}\right)$$

Given instruments for price, one can estimate β and γ using the methods developed in the body of the paper.

Imposing the logit functional form substantially simplifies the problem. Specifically, the unknown function g now only depends on $1 + n_{x^{(2)}}$ arguments and so there is no curse of dimensionality in the number of goods.

To illustrate this, I generate data from the model with $\beta = 1$ and $g(p_{jt}, x_{jt}^{(2)}) = -p_{jt} + 0.5p_{jt}^2 - 0.25p_{jt}^3 + x_{jt}^{(2)} - 0.25\left(x_{jt}^{(2)}\right)^2 \exp(-x_{jt}^{(2)})$ and plot the own- and cross-price elasticity functions for $J = 20$ and $J = 50$ goods in Figures 23 and 24, respectively.

Figure 23: Semiparametric Logit Model, $J = 20$: Own-price (left) and cross-price (right) elasticity functions

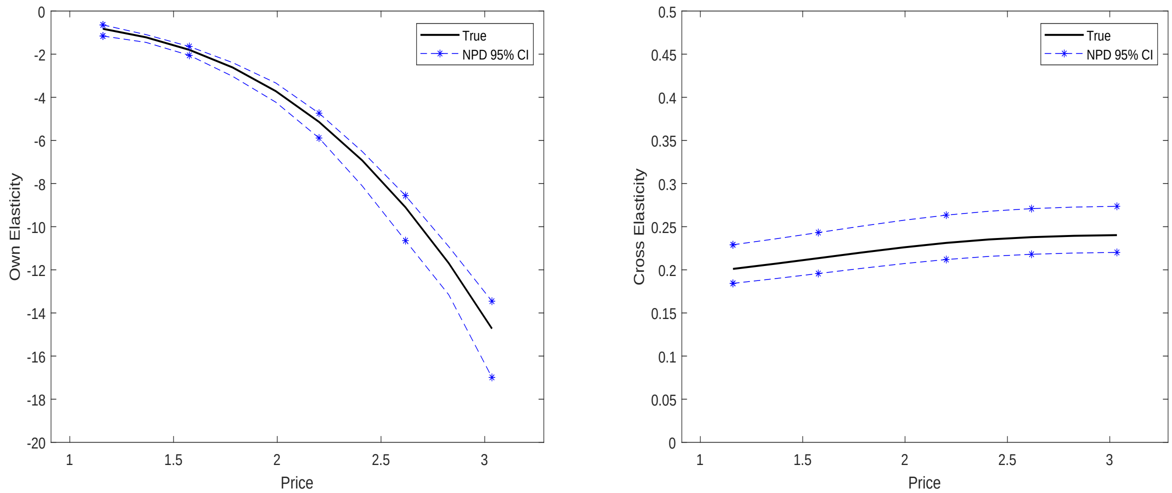
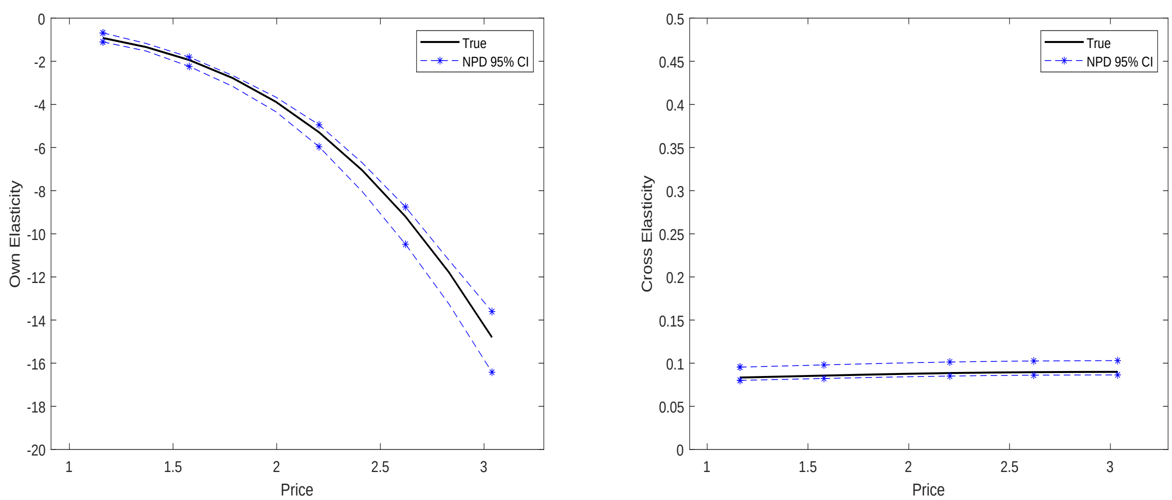


Figure 24: Semiparametric Logit Model, $J = 50$: Own-price (left) and cross-price (right) elasticity functions



Appendix E: Extension to Endogenous Demand Shifters

In this appendix, I consider violations of the exogeneity assumption that take the form $\mathbb{E}(\xi_j|x,z) = \gamma_j x_j$ for all j.[57] By Equation (2), for all j,

$$x_{jt} = \mathbb{E}\left[\frac{1}{\beta_j+\gamma_j}\sigma_j^{-1}\left(s_t, p_t, x_t^{(2)}\right)\Big|x,z\right] \equiv \mathbb{E}\left[\mu_j\left(s_t, p_t, x_t^{(2)}\right)\Big|x,z\right] \tag{38}$$

where I let $\mu \equiv [\mu_1, ..., \mu_J]' \equiv M_\mu \sigma^{-1}$ and M_μ is the diagonal matrix with (j,j) entry $\frac{1}{\beta_j+\gamma_j}$. Then, we can identify μ as in BH. Let $\mathbb{J}_\mu^s$ denote the Jacobian of μ wrt s, and similarly for $\mathbb{J}_\mu^p$, $\mathbb{J}_\mu^{x^{(1)}}$, and $\mathbb{J}_\mu^{x^{(2)}}$. Note that $\mathbb{J}_\sigma^p = -\left(\mathbb{J}_\mu^s\right)^{-1}\mathbb{J}_\mu^p$, so that $\mathbb{J}_\sigma^p$ is identified. An analogous argument applies to $\mathbb{J}_\sigma^{x^{(2)}}$. On the other hand, since $\mathbb{J}_\sigma^{x^{(1)}} = \left(\mathbb{J}_\mu^s\right)^{-1}\tilde{M}_\mu$, where $\tilde{M}_\mu$ is the diagonal matrix with (j,j) entry $\frac{\beta_j}{\beta_j+\gamma_j}$, identifying μ is not sufficient to recover $\mathbb{J}_\sigma^{x^{(1)}}$. In other words, the marginal effects of p and $x^{(2)}$ are identified in spite of the endogeneity of $x^{(1)}$, whereas—as one would expect—the marginal effects of $x^{(1)}$ are not. A corollary of this is that counterfactuals that only depend on derivatives wrt prices—such as those considered in Section 5.4—are robust to this type of endogeneity.

Appendix F: Data

I take a market to be a week/store combination.[58] Data on prices and quantities come from the 2014 Nielsen scanner data set For each market, the most granular unit of observation in the Nielsen data is a UPC (i.e. a specific bar code). I aggregate UPCs according to whether they bear or do not bear the USDA Organic Seal. When this information is missing, I assume the UPC is non-organic. The aggregate quantities are obtained by simply summing the quantities for the individual UPCs, whereas for prices I take a weighted average where the weights are determined by the yearly share of sales that a given UPC has in that store. Similarly, I aggregate across UPCs for selected non-strawberry fruits.[59] Specifically, I focus on the top four non-strawberry fruits according to Produce for Better Health Foundation (2015) in terms of per capita consumption nationwide, i.e. bananas, apples, other berries and oranges. For each of these fruits, I compute a price index (across UPCs) following the same procedure I used for strawberries. These fruit-level price indices are then aggregated even further into a single price index using weights that are proportional to the per capita eatings of each fruit and are normalized to sum to one.

Regarding Hausman instruments, I take the mean price of strawberries and the mean price index for the outside option, respectively, across the Californian supermarkets that are not in the same marketing area[60] as a given store. Excluding supermarkets in the same marketing area is meant to alleviate the usual concerns about Hausman instruments, i.e. that likely spatial correlation in the unobserved quality of the products might induce a violation of the exogeneity assumption.

[57]For simplicity, here I consider the case where $x_j^{(1)}$ is scalar, since that corresponds to the empirical settings in Section 5.

[58]I use the terms "store" and "retailer" interchangeably.

[59]In this case, however, I do not distinguish between organic and non-organic fruits.

[60]Here I follow the Nielsen partition of the United States into Designated Marketing Areas.

Table 8: Descriptive statistics

	Mean	Median	Min	Max
Quantity non-organic	735.33	581.00	6.00	5,729.00
Quantity organic	128.91	78.00	1.00	2,647.00
Price non-organic	2.97	2.89	0.93	4.99
Price organic	4.26	3.99	1.24	6.99
Price other fruit	3.95	3.80	1.30	13.88
Hausman non-organic	3.00	2.98	2.09	4.05
Hausman organic	4.28	4.07	2.95	5.50
Hausman other fruit	4.50	3.79	1.19	13.33
Spot non-organic	1.46	1.35	0.99	2.32
Spot organic	2.38	2.17	1.25	4.88
Quantity other fruit (per capita)	0.83	0.82	0.60	1.08
Share organic lettuce	0.08	0.06	0.00	0.41
Income	82.54	72.61	33.44	405.09
Sample size	38,800	38,800	38,800	38,800

Note: Prices in dollars per pound. Quantities in pounds. Income in thousands of dollars per household.

Spot prices for strawberries are obtained from the US Department of Agriculture website.[61] The data reports spot prices for the following shipping points: California, Texas, Florida, North Carolina, and Mexico. In absence of information on where supermarkets source their strawberries from, I take a simple average of the prices at the various shipping points in any given week.

I measure the availability of non-strawberry fresh fruit in any given week at the state level using the total sales of non-strawberry fruits at all stores included in the Nielsen data set in that week. To proxy for consumer tastes for organic produce at a given store, I compute the percentage of yearly organic lettuce sales over total yearly lettuce sales at the store.

Finally, data on income at the zip-code level is downloaded from the Internal Revenue Service website.[62]

The resulting data set has 38,800 markets. Table 8 reports descriptive statistics for each variable and Figure 25 shows the price pattern for a typical store over time. Both the retail price and the spot price exhibit strong seasonality. Moreover, the retail price sometimes displays a pattern in which it drops and then jumps back up to the initial level. This is typical of supermarket prices given the prevalence of periodic sales. However, in the case of strawberries, this pattern is much less marked than for other items, such as packaged goods. Therefore, the model does not explicitly account for sales.[63]

[61] http://cat.marketnews.usda.gov/cat/index.html.

[62] https://www.irs.gov/uac/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi.

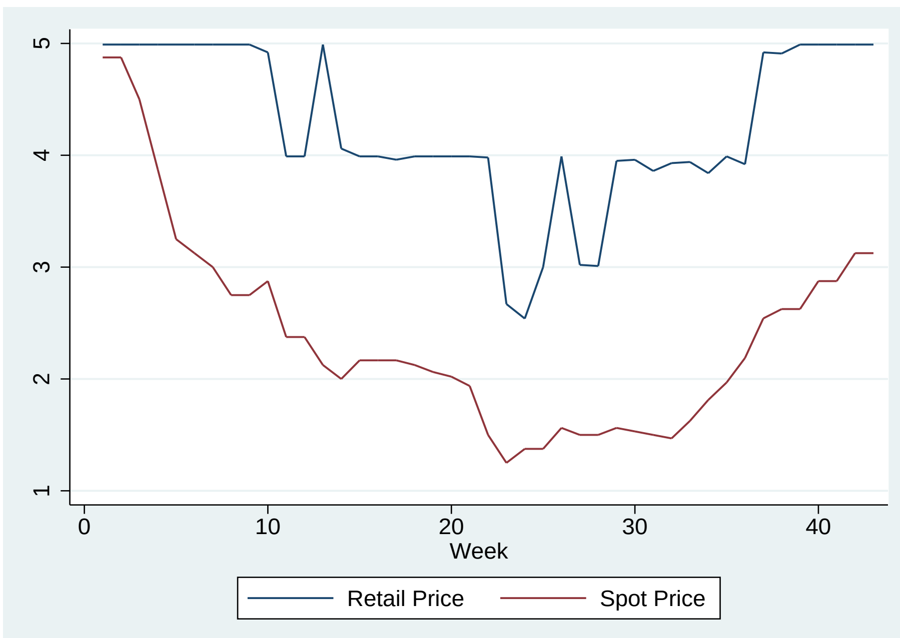
[63] Inventory is often invoked as a justification for sales in models of retail. However, because strawberries are so perishable, it is unlikely that inventory plays a first-order role in driving the retailer's pricing behavior.

Table 9: First-stage regressions

	Non-organic		Organic	
	Price	Share	Price	Share
Spot price (own)	0.12**	−0.68**	0.35**	−0.26**
Spot price (other)	0.04**	0.10**	−0.21**	0.22**
Hausman (own)	0.70**	−1.30**	0.46**	−0.19**
Hausman (other)	−0.01	0.25**	0.13**	0.22**
Hausman (out)	−0.01**	0.11**	−0.10**	0.04**
Availability other fruit	−0.01**	−0.07**	−0.02**	−0.01**
Share organic lettuce	0.08**	−0.20**	−0.01**	0.10**
Income	−0.02**	0.00**	0.01**	0.04**
R^2	0.46	0.27	0.52	0.16

Note: ** denotes significance at the 95% level. All variables are normalized to belong to the $[0, 1]$ interval.

Figure 25: Price patterns



Note: Prices in dollars per pound for organic strawberries sold at a representative store.

Next, I present the results of the first-stage regressions in Table 9. As expected, the retail prices significantly increase with the spot prices. Further, the share of organic strawberries increases with the taste for organic products, while the opposite is true of the non-organic share. Finally, the shares of both inside goods decrease with the availability of other fruit.

Appendix G: Microfoundation of the Empirical Model

This appendix shows how to map the model estimated on the Nielsen data in Section 5 into the general framework outlined in Section 2. Specifically, I outline two models of consumer behavior that yield the demand system in equation (6) and prove that the system is indeed invertible. It should be emphasized that these are only two out of many models that are compatible with (6) and invertibility, and that the estimation procedure does not rely on any of the parametric restrictions embedded in either model.[64]

G.1 Model 1

I first consider a standard discrete choice model. While the model is clearly at odds with the fact that consumers buying fresh fruit face an (at least partially) continuous choice, this serves as a building block for the more realistic model discussed in Section G.2. Moreover, given the prevalence of discrete choice models in the literature, it provides a connection between the demand system in (6) and a more familiar setup.

I assume that consumers face a discrete choice between one unit (say, one pound) of non-organic strawberries, one unit of organic strawberries and one unit of other fresh fruit. Consumer i's indirect utilities for each of these goods are, respectively

$$\begin{aligned} u_{i1} &= \theta_{str}\delta^*_{str} + \alpha_i p_1 + \epsilon_{i1} \\ u_{i2} &= \theta_{str}\delta^*_{str} + \theta_{org}\delta^*_{org} + \alpha_i p_2 + \epsilon_{i2} \\ u_{i0} &= \theta_{0,str}x^{(1)}_{str} + \theta_{0,org}\delta^*_{org} + \alpha_i p_0 + \epsilon_{i0} \end{aligned} \tag{39}$$

where

$$\begin{aligned} \delta^*_{str} &= \xi_{str} \\ \delta^*_{org} &= \theta_{1,org}x^{(1)}_{org} + \xi_{org} \end{aligned}$$

and p_1, p_2, p_0 denote the prices of non-organic strawberries, organic strawberries, and the price index for other fresh fruit, respectively. I interpret δ^*_{str} as the mean quality of all strawberries in the market and δ^*_{org} as the mean utility for organic products (including—but not limited to—organic strawberries). Because the outside option of buying other fresh fruit includes organic produce (e.g. organic apples), I let δ^*_{org} enter u_{i0}. In addition, u_{i0} also depends on the richness of the non-strawberry fruits offering, as captured by $x^{(1)}_{str}$. I use (ξ_{str}, ξ_{org}) to denote the unobserved quality levels for strawberries and organic produce, respectively, and $(\epsilon_{i2}, \epsilon_{i2})$ to denote taste shocks idiosyncratic to consumer i. Unlike BLP, I will not make any parametric assumptions on $(\epsilon_{i2}, \epsilon_{i2})$, nor on the distribution of the price coefficient α_i. In particular, note that the correlation structure of the vector $(\epsilon_{i2}, \epsilon_{i2}, \alpha_i)$ is unrestricted, which allows for patterns such as the fact that wealthier consumers may have a stronger preference for organic produce. Further, the distribution of α_i will be allowed to depend on other covariates such as mean income $x^{(2)}$ in the market.

Now I show that the demand system generated by the model above is identified under the following assumption (as well as the standard exogeneity and completeness assumptions discussed in Section 2).

[64]For instance, while Model 1 below assumes that prices enter linearly in utilities, this restriction is not needed for identification or estimation, given that I do not impose symmetry of the Jacobian of demand with respect to price.

Assumption 10. *The coefficients $\theta_{str}, \theta_{org}, \theta_{0,str}, \theta_{0,org}$ and $\theta_{1,org}$ are non-zero.*

Note that Assumption 10 is very mild. It is satisfied if (i) consumers care about the quality of strawberries ($\theta_{str} > 0$) and organic produce ($\theta_{org}, \theta_{0,org} > 0$), as well as the availability of non-strawberry fruit $\theta_{0,str} > 0$, when purchasing fresh fruit; and (ii) the variable $x^{(1)}_{org}$ is indeed a proxy for taste for organic produce ($\theta_{1,org} > 0$).

Lemma 11. *Under Assumption 10, the demand functions σ_1 and σ_2 generated by the model in (39) are point-identified under the same set of conditions used to obtain identification in BH.*

Proof. Since utility is ordinal, I can subtract $\theta_{0,str}x^{(1)}_{str} + \theta_{0,org}\delta^*_{org} + \alpha_i p_0$ from each equation in (39) and write

$$\begin{aligned} u_{i1} &= \tilde{\delta}_1 - \theta_{0,str}x^{(1)}_{str} + \alpha_i (p_1 - p_0) + \epsilon_{i1} \\ u_{i2} &= \tilde{\delta}_2 - \theta_{0,str}x^{(1)}_{str} + \alpha_i (p_2 - p_0) + \epsilon_{i2} \\ u_{i0} &= \epsilon_{i0}, \end{aligned} \tag{40}$$

where

$$\begin{aligned} \tilde{\delta}_1 &\equiv \theta_{str}\delta^*_{str} - \theta_{0,org}\delta^*_{org} \\ \tilde{\delta}_2 &\equiv \theta_{str}\delta^*_{str} + (\theta_{org} - \theta_{0,org})\,\delta^*_{org} \end{aligned}$$

Using (40) and the fact that the distribution of α_i is allowed to depend on $x^{(2)}$, we can write the demand system as

$$s = \tilde{\sigma}\left(\tilde{\delta}_1 - \theta_{0,str}x^{(1)}_{str}, \tilde{\delta}_2 - \theta_{0,str}x^{(1)}_{str}, p, x^{(2)}\right), \tag{41}$$

where $p \equiv (p_0, p_1, p_2)$, $s \equiv (s_1, s_2)'$ is the vector of market shares and $\tilde{\sigma}$ is a function from $\mathbb{R}^2 \times \mathbb{R}^4_+$ to the unit 2-simplex. Next, by Theorem 1 of Berry et al. (2013), we can invert the system in (41) for the mean utility levels as follows

$$\begin{aligned} \tilde{\delta}_1 &= \tilde{\sigma}^{-1}_1\left(s, p, x^{(2)}\right) + \theta_{0,str}x^{(1)}_{str} \\ \tilde{\delta}_2 &= \tilde{\sigma}^{-1}_2\left(s, p, x^{(2)}\right) + \theta_{0,str}x^{(1)}_{str}, \end{aligned} \tag{42}$$

where $\tilde{\sigma}^{-1}_k$ denotes the $k-$th element of the inverse, $\tilde{\sigma}^{-1}$, of $\tilde{\sigma}$. I now show that there is a one-to-one mapping between $\left(\delta^*_{str}, \delta^*_{org}\right)$ and $\left(\tilde{\delta}_1, \tilde{\delta}_2\right)$. Letting $\delta^* \equiv \left(\delta^*_{str}, \delta^*_{org}\right)'$ and $\tilde{\delta} \equiv \left(\tilde{\delta}_1, \tilde{\delta}_2\right)'$, we have

$$\tilde{\delta} = A\delta^*,$$

where

$$A \equiv \begin{bmatrix} \theta_{str} & -\theta_{0,org} \\ \theta_{str} & \theta_{org} - \theta_{0,org} \end{bmatrix}$$

Since $\det(A) = \theta_{str}\theta_{org} \neq 0$ under Assumption 10, we can rewrite (42) as

$$\delta^* = A^{-1}\tilde{\sigma}^{-1}\left(s, p, x^{(2)}\right) + A^{-1}\cdot[1 \quad 1]' \times \theta_{0,str}x^{(1)}_{str} \tag{43}$$

or, equivalently,

$$\begin{aligned}\delta^*_{str} &= \sigma_1^{-1}\left(s, p, x^{(2)}\right) + \theta_1 x^{(1)}_{str} \\ \delta^*_{org} &= \sigma_2^{-1}\left(s, p, x^{(2)}\right) + \theta_2 x^{(1)}_{str},\end{aligned} \tag{44}$$

for functions $\sigma_i^{-1} : \Delta^2 \times \mathbb{R}^4_+ \to \mathbb{R}^2$, $i = 1, 2$, where Δ^2 denotes the unit 2-simplex. Now I derive expressions for the coefficients θ_1 and θ_2 in terms of the model primitives. Note that

$$A^{-1} = \frac{1}{\theta_{org}} \begin{bmatrix} \frac{\theta_{org}-\theta_{0,org}}{\theta_{str}} & \frac{\theta_{0,org}}{\theta_{str}} \\ -1 & 1 \end{bmatrix}$$

and thus

$$A^{-1} \cdot [1 \quad 1]' = \begin{bmatrix} \frac{1}{\theta_{str}} & 0 \end{bmatrix}',$$

i.e.

$$\begin{aligned}\theta_1 &= \frac{\theta_{0,str}}{\theta_{str}} \\ \theta_2 &= 0\end{aligned}$$

Plugging this into (44) and using the definitions of δ^*_{str} and δ^*_{org}, we obtain

$$\begin{aligned}\xi_{str} &= \sigma_1^{-1}\left(s, p, x^{(2)}\right) + \frac{\theta_{0,str}}{\theta_{str}} x^{(1)}_{str} \\ \theta_{1,org} x^{(1)}_{org} + \xi_{org} &= \sigma_2^{-1}\left(s, p, x^{(2)}\right)\end{aligned} \tag{45}$$

The final step is to show that we can identify the system in (45), given the instruments available. Because we are free to normalize the scale of ξ_{str} and ξ_{org} in the display above, we can divide the first equation of (45) by $\frac{\theta_{0,str}}{\theta_{str}}$ and the second equation by $\theta_{1,org}$ without loss,[65] and rearrange terms as follows

$$-x^{(1)}_{str} = \sigma_1^{-1}\left(s, p, x^{(2)}\right) - \xi_{str} \tag{46}$$

$$x^{(1)}_{org} = \sigma_2^{-1}\left(s, p, x^{(2)}\right) - \xi_{org}, \tag{47}$$

Equations (46) and (47) are in the same form as Equation (6) in BH and thus we can follow their argument to show that σ_1 and σ_2 are identified. Further, note that inverting the system in (46) and (47) yields the demand system in equation (6) that was estimated on the Nielsen data (after normalizations).

□

[65] These divisions are well-defined operations as $\frac{\theta_{0,out}}{\theta_{str}}$ and $\theta_{1,org}$ are nonzero by Assumption 10.

G.2 Model 2

I now turn to a model of continuous choice that is likely a closer approximation to the behavior of consumers buying fresh fruit. Let consumer i face the following maximization problem

$$\begin{aligned} &\max_{q_0,q_1,q_2} U_i\left(q_0,q_1,q_2\right) \\ &\text{s.t.} \;\; p_0q_0+p_1q_1+p_2q_2 \leq y_i^{inc} \end{aligned} \tag{48}$$

where y_i^{inc} denotes the income consumer i allocates to fresh fruit, q_0 is the quantity of non-strawberry fresh fruit, q_1 is the quantity of non-organic strawberries and q_2 is the quantity of organic strawberries, and similarly for prices p_0, p_1, p_2. One could think of y_i^{inc} as being the outcome of a higher-level optimization problem in which the consumer chooses how to allocate total income across different product categories, including fresh fruit. Assume U_i takes the Cobb-Douglas form

$$U_i\left(q_0,q_1,q_2\right) = q_0^{d_0\epsilon_{i,0}} q_1^{d_1\epsilon_{i,1}} q_2^{d_2\epsilon_{i,2}},$$

for positive $d \equiv (d_0, d_1, d_2)$ and $\epsilon_i \equiv (\epsilon_{i,0}, \epsilon_{i,1}, \epsilon_{i,2})$. Then, the optimal quantities chosen by the consumer are

$$q_j^*\left(d,p,y_i^{inc},\epsilon_i\right) = \frac{y_i^{inc}}{p_j}\cdot\frac{d_j\epsilon_{i,j}}{\sum_{k=0}^{2} d_k\epsilon_{i,k}} \qquad j=0,1,2 \tag{49}$$

where $d \equiv (d_0, d_1, d_2)$ and $p \equiv (p_0, p_1, p_2)$. Now assume that

$$\begin{aligned} d_0 &= \gamma_{org}^{\theta_{0,org}}\tilde{x}_{str}^{\theta_{0,str}} \\ d_1 &= \gamma_{str}^{\theta_{str}} \\ d_2 &= \gamma_{str}^{\theta_{str}}\gamma_{org}^{\theta_{org}} \end{aligned}$$

where

$$\begin{aligned} \gamma_{str} &\equiv \exp\left\{\delta_{str}^*\right\} \\ \gamma_{org} &\equiv \exp\left\{\delta_{org}^*\right\} \\ \tilde{x}_{str} &\equiv \exp\left\{x_{str}^{(1)}\right\} \end{aligned}$$

and $\delta_{str}^*, \delta_{org}^*$ are defined as in Section G.1. I can then re-write (49) as

$$q_j^*\left(\tilde{d},p,y_i^{inc},\epsilon_i\right) = \frac{y_i^{inc}}{p_j}\cdot\frac{\tilde{d}_j\epsilon_{i,j}}{\sum_{k=0}^{2} \tilde{d}_k\epsilon_{i,k}} \qquad j=0,1,2 \tag{50}$$

where

$$\begin{aligned} \tilde{d}_0 &\equiv 1 \\ \tilde{d}_1 &\equiv \gamma_{str}^{\theta_{str}}\gamma_{org}^{-\theta_{0,org}}\tilde{x}_{str}^{-\theta_{0,str}} \\ \tilde{d}_2 &\equiv \gamma_{str}^{\theta_{str}}\gamma_{org}^{\theta_{org}-\theta_{0,org}}\tilde{x}_{str}^{-\theta_{0,str}} \end{aligned}$$

and $\tilde{d} \equiv \left(\tilde{d}_0, \tilde{d}_1, \tilde{d}_2\right)$.

Next, let $F_{Y,\epsilon}$ denote the joint distribution of y_i^{inc} and ϵ_i in the market, and define[66]

$$Q_j^* \left(\tilde{d}, p, x^{(2)}\right) = \int q_j^* \left(\tilde{d}, p, y, \epsilon\right) dF_{Y,\epsilon} \left(y, \epsilon; x^{(2)}\right) \qquad j = 0, 1, 2$$

$Q_j^* \left(\tilde{d}, p, x^{(2)}\right)$ is the model counterpart to the market-level quantity Q_j observed in the data.

The last step is to show that there exists a mapping of quantities into market shares such that the resulting demand system is invertible. For $j = 0, 1, 2$, define

$$\tilde{\tilde{\sigma}}_j \left(\tilde{d}, p, x^{(2)}\right) = \frac{Q_j^* \left(\tilde{d}, p, x^{(2)}\right)}{\sum_{k=0}^{2} Q_k^* \left(\tilde{d}, p, x^{(2)}\right)}$$

and

$$s_j = \frac{Q_j}{\sum_{k=0}^{2} Q_k}$$

Then, equating observed shares to their model counterparts, we obtain the system

$$s = \tilde{\tilde{\sigma}} \left(\tilde{d}, p, x^{(2)}\right) \tag{51}$$

where $s \equiv (s_0, s_1, s_2)'$ and $\tilde{\tilde{\sigma}} \left(\tilde{d}, p, x^{(2)}\right) \equiv \left(\tilde{\tilde{\sigma}}_0 \left(\tilde{d}, p, x^{(2)}\right), \tilde{\tilde{\sigma}}_1 \left(\tilde{d}, p, x^{(2)}\right), \tilde{\tilde{\sigma}}_2 \left(\tilde{d}, p, x^{(2)}\right)\right)'$.

Because $\tilde{\tilde{\sigma}}_j$ is strictly decreasing in $\tilde{d}_k$ for all j and all $k > 0, k \neq j$, by Theorem 1 in Berry et al. (2013), we can invert (51) as follows

$$\tilde{d} = \tilde{\tilde{\sigma}}^{-1} \left(s, p, x^{(2)}\right)$$

and, taking logs, we can write

$$\begin{aligned} \theta_{str} \delta_{str}^* - \theta_{0,org} \delta_{org}^* &= \tilde{\sigma}_1^{-1} \left(s, p, x^{(2)}\right) + \theta_{0,str} x_{str}^{(1)} \\ \theta_{str} \delta_{str}^* + \left(\theta_{org} - \theta_{0,org}\right) \delta_{org}^* &= \tilde{\sigma}_2^{-1} \left(s, p, x^{(2)}\right) + \theta_{0,str} x_{str}^{(1)} \end{aligned} \tag{52}$$

where $\tilde{\sigma}_j^{-1} \left(s, p, x^{(2)}\right) \equiv \log \left(\tilde{\tilde{\sigma}}_j^{-1} \left(s, p, x^{(2)}\right)\right)$ for $j = 1, 2$.

Note that (52) has the exact same form as (42). Therefore, we can use the argument in Section G.1 to show that the demand system is identified.

[66] Note that I let $F_{Y,\epsilon}$ be a function of mean income $x^{(2)}$, consistently with the information available in the data.