E-Companion for

The Power and Limits of Predictive Approaches to Observational-Data-Driven Optimization: The Case of Pricing

by Dimitris Bertsimas and Nathan Kallus

EC.1. Omitted Proofs

Proof of Theorem 2 Consider the first part, where we assume $y(z)$ is linear. That $y(z)$ is linear and decreasing implies that $y(z)=y_0-\lambda(z-c)$ with $\lambda>0$. Hence, $R(z)=y_0(z-c)-\lambda(z-c)^2$, which is unimodal and uniquely maximized at $z^\star=c+y_0/(2\lambda)$ with value $R(z^\star)=y_0^2/(4\lambda)$. Let $\delta(z)=\mathbb{E}\left[\epsilon\middle|Z=z\right]$, $\eta=y_0\gamma$. Then $|\delta(z)|\le\eta$. Note that

$$\mathbb{E}\left[Y^{\text{obs}}\middle|Z=z\right]=\mathbb{E}\left[Y(z)\middle|Z=z\right]=\mathbb{E}\left[y(z)+\epsilon(z)\middle|Z=z\right]=y(z)+\mathbb{E}\left[\epsilon\middle|Z=z\right]=y_0-\lambda(z-c)+\delta(z).$$

Hence, the theorem is trivial if $\eta=0$ so let us assume $\eta>0$.

Next we ask the question, what is the largest and smallest that the maximizer $\tilde z$ of $\tilde R(z)$ can be. By assumption, $|\delta(z)|\le\eta$ for all $z\in\mathcal{Z}$. So, defining $\tilde R_{\delta_0}(z):=(z-c)\left(y_0-\lambda(z-c)+\delta_0(z)\right)$, we are interested in

$$\tilde z_{\max}=\sup\left\{\sup\left(\operatorname*{arg\,max}_{z\in\mathcal{Z}}\tilde R_{\delta_0}(z)\right):|\delta_0(z)|\le\eta\right\},\tag{EC.1}$$

$$\tilde z_{\min}=\inf\left\{\inf\left(\operatorname*{arg\,max}_{z\in\mathcal{Z}}\tilde R_{\delta_0}(z)\right):|\delta_0(z)|\le\eta\right\},\tag{EC.2}$$

where we define $\sup(\varnothing)=-\infty$ and $\inf(\varnothing)=\infty$ without loss of generality because we assumed an optimizer $\tilde z$ exists for $\tilde R(z)$ so we are only interested in those functions $\delta(z)$ that induce a nonempty argmax. In what follows, define $\tilde R_+(z)=(z-c)\left(y_0-\lambda(z-c)+\eta\right)$ and $\tilde R_-(z)=(z-c)\left(y_0-\lambda(z-c)-\eta\right)$, which are both unimodal and uniquely maximized at $\tilde z_+=c+(y_0+\eta)/(2\lambda)$ and $\tilde z_-=c+(y_0-\eta)/(2\lambda)$ respectively ($\tilde z_-<\tilde z_+$ because $\eta>0$). Notice that $\tilde R_-(z)\le\tilde R_{\delta_0}(z)\le\tilde R_+(z)$ whenever $|\delta_0(z)|\le\eta$ with equality when $\delta_0(z)=\pm\eta$ is extremal.

First we argue that the bounds (EC.1)–(EC.2) are finite. For any $z\ge z'=c+\left(y_0+\eta+2\sqrt{\lambda+y_0\eta}\right)/(2\lambda)$ and $|\delta_0(z)|\le\eta$, since $\tilde R_+(z)$ is decreasing past $\tilde z_+$ and $z'\ge\tilde z_+$, we have that

$$\tilde R_{\delta_0}(z)\le\tilde R_+(z)\le\tilde R_+(z')=(y_0-\eta)^2/(4\lambda)-1<(y_0-\eta)^2/(4\lambda)=\tilde R_-(\tilde z_-)\le\tilde R_{\delta_0}(\tilde z_-).$$

Since $\tilde z_-\le z'$ we conclude that $\tilde z_{\max}\le z'<\infty$. Finally, since $\delta_0(z)=0$ is feasible in (EC.1)–(EC.2), we have $c\le\tilde z_{\min}\le z^\star\le\tilde z_{\max}\le z'$.

Next we argue that in (EC.1) it is sufficient to consider functions $\delta_0(z)$ taking values in $\{-\eta, +\eta\}$ that are monotonic increasing, i.e. constant or step functions. Let $\delta_0(z)$ be feasible in (EC.1) and let $\tilde{z}_0 = \sup\left\{\arg\max_{z\in\mathcal{Z}} \tilde{R}_{\delta_0}(z)\right\}$. If at any $z_1 \geq \tilde{z}_0$ we have $\delta_0(z_1) < \eta$, then increasing the value of $\delta_0(z_1)$ to η can only increase the value of $\tilde{R}_{\delta_0}(z_1)$, which in turn may only increase the largest maximizer since $z_1 \geq \tilde{z}_0$. Moreover, if at any $z_1 < \tilde{z}_0$ we have $\delta_0(z_1) > -\eta$, then decreasing the value of $\delta_0(z_1)$ to $-\eta$ can only decrease the value of $\tilde{R}_{\delta_0}(z_1)$, which must already be at or below the maximal value and hence must leave the largest maximizer unchanged. The argument is unchanged even if $\tilde{z}_0$ is $\pm\infty$. A symmetric argument shows that in (EC.2) it is sufficient to consider functions $\delta_0(z)$ taking values in $\pm\eta$ that are monotonic decreasing.

Next we evaluate $\tilde{z}_{\max}$. Fix $\tilde{z}' = c + (\sqrt{y_0} + \sqrt{\eta})^2/(2\lambda)$ and let us consider the step function $\delta_{\max}(z) = \eta\mathbb{I}[z \geq \tilde{z}'] - \eta\mathbb{I}[z < \tilde{z}']$. Since $\tilde{z}' > z_-$, $\tilde{R}_{\delta_{\max}}(z)$ is uniquely maximized on $(c, \tilde{z}')$ at z_-, with value $\tilde{R}_{\delta_{\max}}(z_-) = \tilde{R}_-(z_-) = (y_0 - \eta)^2/(4\lambda)$. Since $\tilde{z}' > z_+$, $R_{\delta_{\max}}(z)$ is uniquely maximized on $[\tilde{z}', \infty)$ at $\tilde{z}'$, with value $\tilde{R}_{\delta_{\max}}(\tilde{z}') = \tilde{R}_+(\tilde{z}') = (y_0 - \eta)^2/(4\lambda)$. Hence, $\arg\max_{z\in\mathcal{Z}} \tilde{R}_{\delta_{\max}}(z) = \{z_-, \tilde{z}'\}$ and $\sup\{z_-, \tilde{z}'\} = \tilde{z}'$. Now we show that it is impossible to achieve a higher maximizer with $|\delta(z)| \leq \eta$, which would lead to $\tilde{z}_{\max} = \tilde{z}'$. By our previous argument we need only consider functions $\delta(z)$ taking values in $\pm\eta$ that are monotonic increasing. The constant functions taking values in $\pm\eta$ induce the maxima $\tilde{z}_-$ and $\tilde{z}_+$, both of which are smaller than $\tilde{z}'$. Next, consider any step function $\delta_0(z) = \eta\mathbb{I}[z \geq \tilde{z}_0] - \eta\mathbb{I}[z < \tilde{z}_0]$ with $\tilde{z}_0 \neq z'$. If $\tilde{z}_0 \leq \tilde{z}_+$ then, for any $z \neq z_+$, we have that $\tilde{R}_{\delta_0}(\tilde{z}_+) = \tilde{R}_+(\tilde{z}_+) > \tilde{R}_+(z) \geq \tilde{R}_{\delta_0}(z)$ since $\tilde{z}_+$ is the unique maximizer of $\tilde{R}_+(z)$; hence $\tilde{z}_+ < z'$ is the unique maximum of $\tilde{R}_{\delta_0}(z)$. Consider $\tilde{z}_0 > \tilde{z}_+$. Then, since $\tilde{z}_0 > z_+ > z_-$, $\tilde{R}_{\delta_{\max}}(z)$ is uniquely maximized on $(c, \tilde{z}_0)$ at z_-, with value $\tilde{R}_{\delta_0}(z_-) = \tilde{R}_-(z_-) = (y_0 - \eta)^2/(4\lambda)$. Since $\tilde{z}_0 > z_+$, $R_{\delta_{\max}}(z)$ is uniquely maximized on $[\tilde{z}_0, \infty)$ at $\tilde{z}_0$, with value $\tilde{R}_{\delta_0}(\tilde{z}_0) = \tilde{R}_+(\tilde{z}_0)$. If $\tilde{z}_0 < \tilde{z}'$, then either of these potential maximizers are smaller than z'. If $\tilde{z}_0 > \tilde{z}'$ then, since $\tilde{R}_+(z)$ is strictly decreasing past z_+ and $\tilde{z}' \geq z_+$, we have $\tilde{R}_{\delta_0}(\tilde{z}_0) = \tilde{R}_+(\tilde{z}_0) < \tilde{R}_+(\tilde{z}') = (y_0 - \eta)^2/(4\lambda) = \tilde{R}_-(z_-) = \tilde{R}_{\delta_{\max}}(z_-)$. Hence $\tilde{z}_- < \tilde{z}'$ is the unique maximum of $\tilde{R}_{\delta_0}(z)$. When $\eta < y_0$, a symmetric argument applied to (EC.2) shows that $\tilde{z}_{\min} = c + \left(\sqrt{y_0} - \sqrt{\eta}\right)^2/(2\lambda)$. If $\eta \geq y_0$, the lower bound $\tilde{z}_{\min} = c$ is achieved by $\delta_-(z)$. Hence, $\tilde{z}_{\min} = c + \max\left\{0, \sqrt{y_0} - \sqrt{\eta}\right\}^2/(2\lambda)$.

To summarize, we conclude that since $\left|\mathbb{E}\left[\epsilon \middle| Z\right]\right| \leq \eta$, we must have

$$\tilde{z} \in [\tilde{z}_{\min}, \tilde{z}_{\max}] \quad \text{where} \quad \tilde{z}_{\min} = c + \frac{\max\left\{0, \sqrt{y_0} - \sqrt{\eta}\right\}^2}{2\lambda}, \; \tilde{z}_{\max} = c + \frac{(\sqrt{y_0} + \sqrt{\eta})^2}{2\lambda}.$$

Plugging these bounds into $R(z)$ we have

$$R(\tilde{z}_{\max}) = \frac{y_0^2 - 4y_0\eta - \eta^2 - 4\eta\sqrt{y_0\eta}}{4\lambda}, \qquad R(\tilde{z}_{\min}) = \begin{cases} \frac{y_0^2 - 4y_0\eta - \eta^2 + 4\eta\sqrt{y_0\eta}}{4\lambda} & \eta < y_0 \\ 0 & \eta \geq y_0 \end{cases}$$

Notice that if $\eta < y_0$ then $R(\tilde z_{\max}) = R(\tilde z_{\min}) - 2\eta\sqrt{y_0\eta}/\lambda \le R(\tilde z_{\min})$ and if $\eta \ge y_0$ then $R(\tilde z_{\max}) \le 0 = R(\tilde z_{\min})$. Therefore, $\min\{R(\tilde z_{\max}),\, R(\tilde z_{\min})\} = R(\tilde z_{\max})$. Since $R(z)$ is unimodal and $\tilde z \in [\tilde z_{\min},\, \tilde z_{\max}]$, we have

$$R(\tilde z) \ge \min\{R(\tilde z_{\max}),\, R(\tilde z_{\min})\} = R(\tilde z_{\max}) = \frac{y_0^2 - 4y_0\eta - \eta^2 - 4\eta\sqrt{y_0\eta}}{4\lambda}.$$

Finally, using $R(z^\star) = y_0^2/(4\lambda)$,

$$\frac{R(\tilde z)}{R(z^\star)} \ge 1 - 4\left(\frac{\eta}{y_0}\right) - 4\left(\frac{\eta}{y_0}\right)^{3/2} - \left(\frac{\eta}{y_0}\right)^2,$$

which is a univariate polynomial in $\sqrt{\eta/y_0}$.

Consider the second part, where we assume $y(z)$ is almost linear. Let $\check y(z) = y_0 - \lambda(z-c)$, $\check R(z) = y_0(z-c) - \lambda(z-c)^2$, $\check z = c + y_0/(2\lambda)$, and $\vartheta(z) = y(z) - \check y(z)$. By assumption, $|\vartheta(z)| \le y_0\psi$ and $|\tilde y(z) - \check y(z)| \le y_0(\psi+\gamma)$. Note that $R(\tilde z) - \check R(\tilde z) = \vartheta(\tilde z)(\tilde z - c)$ and that $R(z^*) - \check R(\check z) \le R(z^*) - \check R(z^*) = \vartheta(z^*)(z^* - c)$. Therefore,

$$\frac{R(\tilde z)}{R(z^\star)} \ge \frac{\check R(\tilde z) + \vartheta(\tilde z)(\tilde z - c)}{\check R(\check z) + \vartheta(z^*)(z^* - c)}.$$

We have shown above that $\tilde z \le c + \frac{(\sqrt{y_0} + \sqrt{y_0(\psi+\gamma)})^2}{2\lambda}$ and $\check R(\tilde z) \ge y_0^2 \frac{1 - 4\gamma - \gamma^2 - 4\gamma^{3/2}}{4\lambda}$. A symmetric argument also shows $z^* \le c + \frac{(\sqrt{y_0} + \sqrt{y_0\psi})^2}{2\lambda}$. Also recall that $R(z^\star) = y_0^2/(4\lambda)$. We conclude that

$$\frac{R(\tilde z)}{R(z^\star)} \ge \frac{y_0^2 \frac{1 - 4\gamma - \gamma^2 - 4\gamma^{3/2}}{4\lambda} - \psi y_0 \frac{(\sqrt{y_0} + \sqrt{y_0(\psi+\gamma)})^2}{2\lambda}}{\frac{y_0^2}{4\lambda} + \psi y_0 \frac{(\sqrt{y_0} + \sqrt{y_0\psi})^2}{2\lambda}}.$$

Simplifying yields the claim. □

Proof of Theorem 3 Since both $y(z)$ and $\tilde y(z)$ are linear, then so is $\mathbb{E}\left[\epsilon \middle| Z\right] = E(z) = \tilde y(z) - y(z)$. By simple linear regression we can write the linear $\mathbb{E}\left[\epsilon \middle| Z\right]$ as

$$\mathbb{E}\left[\epsilon \middle| Z\right] = \zeta(Z - \mu), \quad \text{where } \mu = \mathbb{E}[Z],\; \zeta = \frac{\operatorname{Cov}(\epsilon,\, Z)}{\operatorname{Var}(Z)}.$$

By assumption of non-positive correlation, $\zeta \le 0$.

That $y(z)$ is linear and decreasing implies that $y(z) = y_0 - \lambda(z-c)$ with $\lambda > 0$. Hence, $R(z) = y_0(z-c) - \lambda(z-c)^2$, which is unimodal and uniquely maximized at $z^\star = c + y_0/(2\lambda)$ with value $R(z^\star) = y_0^2/(4\lambda)$. Also, recall from the proof of Theorem 2 that $\mathbb{E}\left[Y^{\text{obs}} \middle| Z = z\right] = \mathbb{E}\left[Y(z) \middle| Z = z\right] = \mathbb{E}\left[y(z) + \epsilon(z) \middle| Z = z\right] = y(z) + \mathbb{E}\left[\epsilon \middle| Z = z\right]$ and hence $\tilde R(z) = y_0(z-c) - \lambda(z-c)^2 + \zeta(z-c)(z-\mu)$, which is unimodal and uniquely maximized at its critical point $\tilde z = (2\lambda c + y_0 - \zeta(c+\mu))/(2\lambda - 2\zeta)$ because it is feasible since $\lambda > 0 \ge \zeta$ and $(2\lambda - 2\zeta)(\tilde z - c) = y_0 - \zeta(\mu - c) \ge y_0 \ge 0$.

Plugging $\tilde z$ into $R(z)$ we get

$$R(\tilde z) = \frac{(y_0 - \zeta(\mu - c))(y_0(\lambda - 2\zeta) + \lambda\zeta(\mu - c))}{4(\lambda - \zeta)^2}.$$

Rearranging and using $R(z^\star)=y_0^2/(4\lambda)$, we have

$$\frac{R(\tilde z)}{R(z^\star)}=1-\left(\frac{\zeta}{\lambda-\zeta}\right)^2\left(\frac{y_0+\lambda(c-\mu)}{y_0}\right)^2=1-\left(\frac{\zeta}{\lambda-\zeta}\right)^2\left(\frac{\mathbb{E}\left[Y^{\mathrm{obs}}\right]}{y_0}\right)^2,$$

where we plugged in $\mathbb{E}\left[Y^{\mathrm{obs}}\right]=\mathbb{E}\left[Y(Z)\right]=\mathbb{E}\left[y_0-\lambda(Z-c)\right]=y_0-\lambda(\mu-c)$. Moreover,

$$\sup_{\zeta\le 0}\left(\frac{\zeta}{\lambda-\zeta}\right)^2=\lim_{\zeta\to-\infty}\left(\frac{\zeta}{\lambda-\zeta}\right)^2=1.$$

Hence, we have the result in the statement of the theorem. □

Proof of Theorem 4 Consider the first part, where we assume $y(z)$ is linear. We repeat the proof of Theorem 2 but note that if $E(z)$ is non-increasing then in (EC.1) it is sufficient to consider *constant* functions $\delta_0(z)$ taking values in $\{-\eta,+\eta\}$, which implies $\tilde z_{\max}=\tilde z_+$. Therefore,

$$R(\tilde z_{\max})=\frac{y_0^2-\eta^2}{4\lambda},$$

since $\eta\le y_0$ is assumed. Since $R(z)$ is unimodal and $\tilde z\in[\tilde z_{\min},\tilde z_{\max}]$, we have

$$\begin{aligned}R(\tilde z)\ge\min\left\{R(\tilde z_{\max}),\,R(\tilde z_{\min})\right\}&=\min\left\{\frac{y_0^2-\eta^2}{4\lambda},\frac{y_0^2-4y_0\eta-\eta^2+4\eta\sqrt{y_0\eta}}{4\lambda}\right\}\\&\ge\frac{y_0^2-4y_0\eta-\eta^2+4\eta\sqrt{y_0\eta}}{4\lambda},\end{aligned}$$

since $\eta\le y_0$ is assumed. Finally, using $R(z^\star)=y_0^2/(4\lambda)$,

$$\frac{R(\tilde z)}{R(z^\star)}\ge 1-4\left(\frac{\eta}{y_0}\right)+4\left(\frac{\eta}{y_0}\right)^{3/2}-\left(\frac{\eta}{y_0}\right)^2,$$

completing the proof.

Consider the second part, where we assume $y(z)$ is almost linear. Let $\check y(z)=y_0-\lambda(z-c)$, $\check R(z)=y_0(z-c)-\lambda(z-c)^2$, $\check z=c+y_0/(2\lambda)$, and $\vartheta(z)=y(z)-\check y(z)$. By assumption, $|\vartheta(z)|\le y_0\psi$ and $|\tilde y(z)-\check y(z)|\le y_0(\psi+\gamma)$. Note that $R(\tilde z)-\check R(\tilde z)=\vartheta(\tilde z)(\tilde z-c)$ and that $R(z^*)-\check R(\check z)\le R(z^*)-\check R(z^*)=\vartheta(z^*)(z^*-c)$. Therefore,

$$\frac{R(\tilde z)}{R(z^\star)}\ge\frac{\check R(\tilde z)+\vartheta(\tilde z)(\tilde z-c)}{\check R(\check z)+\vartheta(z^*)(z^*-c)}.$$

We have shown above that $\check R(\tilde z)\ge\frac{y_0^2-4y_0\eta-\eta^2+4\eta\sqrt{y_0\eta}}{4\lambda}$. Moreover, following the same arguments in the proof of Theorem 2, we also conclude that $\tilde z\le c+\frac{(\sqrt{y_0}+\sqrt{y_0(\psi+\gamma)})^2}{2\lambda}$ and that $z^*\le c+\frac{(\sqrt{y_0}+\sqrt{y_0\psi})^2}{2\lambda}$ (in particular, since ϑ need not be non-increasing, we consider an unrestricted $(\psi+\gamma)$-bounded and ψ-bounded perturbation, respectively, to derive these bounds). Also recall that $R(z^\star)=y_0^2/(4\lambda)$. We conclude that

$$\frac{R(\tilde z)}{R(z^\star)}\ge\frac{y_0^2\frac{1-4\gamma-\gamma^2+4\gamma^{3/2}}{4\lambda}-\psi y_0\frac{(\sqrt{y_0}+\sqrt{y_0(\psi+\gamma)})^2}{2\lambda}}{\frac{y_0^2}{4\lambda}+\psi y_0\frac{(\sqrt{y_0}+\sqrt{y_0\psi})^2}{2\lambda}}.$$

Simplifying yields the claim. □

Proof of Theorem 5 For $u>0$ and $\lambda>0$ to be specified later, set $Z\sim\text{Exponential}(\lambda)$, $\epsilon=(2\mathbb{I}\left[Z\le\log(2)/\lambda\right]-1)\eta$, $Y(z)=(z-c+u)^{-2}+\epsilon$. Note that $\log(2)/\lambda$ is the median of Exponential(λ) so that $\mathbb{E}[\epsilon]=0$. Then $y(z)=(z-c+u)^{-2}$, which is non-increasing, $R(z)=(z-c+u)^{-2}(z-c)$, $z^*=c+u$, and $R(z^*)=(4u)^{-1}$. We can, in particular, set $u>0$ small enough so that $R(z^*)\ge\max\{1,\,M+1/M\}$.

At the same time, $E(z)=(2\mathbb{I}\left[z\le\log(2)/\lambda\right]-1)\eta$ and $\tilde{y}(z)=(z-c+pu)^{-1+p}+(2\mathbb{I}\left[z\le\log(2)/\lambda\right]-1)\eta$, which are also both non-increasing. For $z\le\log(2)/\lambda$, we have $\tilde{R}(z)=(z-c+pu)^{-1+p}(z-c)+\eta(z-c)$. Ignoring for now the upper bound on z, the first term approaches zero and the second term grows linearly. Thus, choosing $\lambda>0$ sufficiently small so that $\log(2)/\lambda$ is sufficiently large, we have that $\tilde{z}=\log(2)/\lambda$. Since $R(z)$ is decreasing to zero, we can, in particular, set it so that $R(\tilde{z})\le 1/M$. This leads to the required suboptimality in both ratio and absolute difference. ☐

Proof of Theorem 7 Assumption 1 gives $R(z)=\mathbb{E}\left[\mathbb{E}\left[r(Z)Y^{\text{obs}}\middle|Z=z,X\right]\right]$, i.e. profit is given by taking a partial mean with $Z=z$ fixed of the regression of $r(Z)Y^{\text{obs}}$ on Z and X.

By Assumption 4 part i, there exists $\delta>0$ such that $[z^\star-\delta,\,z^\star+\delta]$ is contained inside the support of Z. By Assumption 4 part i, $f_{Z,X}(z,x)$ is bounded away from 0 on its support, and, by Assumption 4 part v, $\frac{\partial f_{Z,X}(z,x)}{\partial x}$ is bounded. Hence,

$$\left|\frac{\partial}{\partial x}\log(f_{Z,X}(z,x))\right|=\frac{\left|\frac{\partial f_{Z,X}(z,x)}{\partial x}\right|}{f_{Z,X}(z,x)}\le L<\infty$$

on the support of (Z,X). Therefore, we have that, for any x and $|u|\le\delta$,

$$\log(f_{Z,X}(z^\star+u,\,x))\le\log(f_{Z,X}(z^\star,\,x))+L\delta,$$

and consequently,

$$\int\sup_{|u|\le\delta}f_{Z,X}(z^\star+u,x)dx\le\int e^{L\delta}f_{Z,X}(z^\star,\,x)dx<\infty.$$

By Assumption 4 part iii, there exists M such that $\mathbb{E}\left[(Y^{\text{obs}})^4\middle|Z=z,X=x\right]\le M$ for all z,x. Combined, this yields

$$\int\sup_{|u|\le\delta}\left(1+\mathbb{E}\left[(r(Z)Y^{\text{obs}})^4\middle|Z=z^\star+u,X=x\right]\right)f_{Z,X}(z^\star+u,x)dx$$
$$\le(1+r(z^\star+\delta)^4M)\int\sup_{|u|\le\delta}f_{Z,X}(z^\star+u,x)dx<\infty.\qquad\text{(EC.3)}$$

To study our profit function estimator (11) for each fixed z, we employ Theorem 4.1 of Newey (1994) (henceforth, N in this proof), which provides convergence results for two-step kernel m-estimators, a specific case of which is our profit function estimator. We let the "trimming function" of N be $\tau(x)=\mathbb{I}\left[f_X(x)>0\right]$, an indicator for the compact support of X (where its density is assumed bounded away from zero). Since our estimator (11) has K both in the numerator and denominator, it is unchanged if we rescale K by a positive constant. Similarly, the conditions of Assumption 2 remain unchanged. Hence, by Assumption 2 part i, without loss of generality we may assume $\int_{\mathbb{R}^{1+k}} K = 1$. Then, Assumption K of N is satisfied with $\Delta = 2$ by Assumption 2 parts i-iv. Assumption H of N is satisfied with $d = s+1$ by Assumption 4 part v. These constitute condition (ii) of N's Theorem 4.1. By Assumption 4 part i, there exists $c < z_{\max} < \infty$ such that $Z \le z_{\max}$ almost surely. Let $r_{\max} = r(z_{\max}) < \infty$. Then, by Assumption 4 part iii, $\mathbb{E}\left[(r(Z)Y^{\text{obs}})^4\right] \le r_{\max}\mathbb{E}\left[(Y^{\text{obs}})^4\right] < \infty$ and $\mathbb{E}\left[(r(Z)Y^{\text{obs}})^4 \middle| Z = z, X = x\right] = r(z)\mathbb{E}\left[(Y^{\text{obs}})^4 \middle| Z = z, X = x\right]$ are bounded. Combined with Assumption 4 part ii, we satisfy condition (i) of N's Theorem 4.1. Condition (iii) of N's Theorem 4.1 is satisfied by our choice of $\tau(\cdot)$ and by Assumption 4 part i. The first clause of condition (iv) of N's Theorem 4.1 is satisfied by our choice of $\tau(\cdot)$ and by Assumption 4 part ii. The second clause is satisfied by Assumption 4 parts iv-v combined with the fact that for any $m > 0$, $\mathbb{E}\left[(r(Z)Y^{\text{obs}})^m \middle| Z = z, X = x\right] = r(z)^m\mathbb{E}\left[(Y^{\text{obs}})^m \middle| Z = z, X = x\right]$ and $r(z)^m$ is continuous. The third clause is satisfied by (EC.3). Since $X \in \mathbb{R}^k$ and $Z \in \mathbb{R}$, condition (v) of N's Theorem 4.1 is satisfied by Assumption 2 parts v-vi. Applying N's Theorem 4.1 for each fixed $z \in \mathcal{Z}$, we get

$$\sqrt{nh_n}(R(z) - \overline{R}_n(z)) \xrightarrow{d} \mathcal{N}\left(0,\, \eta_z\kappa\right) \ \forall z \in \mathcal{Z},$$

where $\eta_z\kappa$ is an algebraic simplification of the asymptotic variance in eq. (14) in N.

To study the optimizer of our profit function estimator, we employ Flores (2005) (henceforth, F in this proof). Conditions (ii-vi) of F's Theorem 3 are satisfied in a similar way to the case of N's Theorem 4.1. Condition (i) of F's Theorem 3 is satisfied by Assumption 3 parts i and iii, condition (vii) by Assumption 4 part v, condition (viii) by Assumption 3 part iv, condition (ix) by Assumption 3 parts ii and iv, and finally condition (x) by Assumption 2 parts v-vi. Applying F's Theorem 3, we get

$$\sqrt{nh_n^3}(z^\star - \overline{z}_n) \xrightarrow{d} \mathcal{N}\left(0,\, \frac{\eta_{z^\star}\kappa'}{R''(z^\star)^2}\right), \tag{EC.4}$$

simplifying the asymptotic variance.

By Assumption 3 part iv and using Taylor's theorem to expand $R(z)$ around $z = z^\star$, there exists $z_n \in [\min(z^\star, \overline{z}_n), \max(z^\star, \overline{z}_n)]$ such that

$$R(\overline{z}_n) = R(z^\star) + R'(z^\star)(\overline{z}_n - z^\star) + \frac{1}{2}R''(z_n)(\overline{z}_n - z^\star)^2.$$

By first order optimality conditions, $R'(z^\star) = 0$. Hence, rearranging, we have

$$R(z^\star) - R(\overline{z}_n) = -\frac{1}{2}R''(z_n)(\overline{z}_n - z^\star)^2. \tag{EC.5}$$

By continuous transformation of eq. (EC.4), we have

$$\left(nh_n^3\right)\left(\overline{z}_n - z^\star\right)^2 \xrightarrow{d} \frac{\eta_{z^\star}\kappa'}{R''(z^\star)^2}\chi_1^2. \tag{EC.6}$$

Eq. (EC.4) also implies $\overline{z}_n \xrightarrow{\mathbb{P}} z^\star$, which also implies $z_n \xrightarrow{\mathbb{P}} z^\star$ since z_n is sandwiched between $\overline{z}_n$ and $z^\star$. Since $R''(z)$ is continuous, we also get by continuous transformation that

$$R''(z_n) \xrightarrow{\mathbb{P}} R''(z^\star). \tag{EC.7}$$

Combining eqs. (EC.5)–(EC.7), we get the desired result,

$$\left(nh_n^3\right)\left(R(z^\star) - R(\overline{z}_n)\right) \xrightarrow{d} \frac{-\eta_{z^\star}\kappa'}{2R''(z^\star)}\chi_1^2. \tag{EC.8}$$

If $nh_n^{2s+1} \to 0$, then we also satisfy the conditions of F's Theorem 4 with equal bandwidths. Applying F's Theorem 4, we get

$$\sqrt{nh_n}(R(z^\star) - \overline{R}_n(\overline{z}_n)) \xrightarrow{d} \mathcal{N}\left(0, \eta_{z^\star}\kappa\right),$$

simplifying the asymptotic variance. □

Proof of Theorem 8 Let $\overline{z}''_n \in \mathcal{Z}_n$ be a choice such that $|\overline{z}_n - \overline{z}''_n| \le \nu_n$. Then, by the optimality of $\overline{z}'_n$ over $\mathcal{Z}'_n$, Taylor's theorem, and first order conditions on $\overline{z}_n$ (each in order), we have that for some $\overline{z}'''_n \in [\min(\overline{z}_n, \overline{z}''_n), \max(\overline{z}_n, \overline{z}''_n)]$,

$$\begin{aligned}\overline{R}_n(\overline{z}_n) - \overline{R}_n(\overline{z}'_n) &\le \overline{R}_n(\overline{z}_n) - \overline{R}_n(\overline{z}''_n) \\ &= \overline{R}'_n(\overline{z}_n)(\overline{z}_n - \overline{z}''_n) + \frac{1}{2}\overline{R}''_n(\overline{z}'''_n)(\overline{z}_n - \overline{z}''_n)^2 \\ &\le \frac{1}{2}\left|\overline{R}''_n(\overline{z}'''_n)\right|\nu_n^2.\end{aligned}$$

As in equation (EC.11) in the proof of Theorem 9, we have $\overline{R}''_n(\overline{z}'''_n) \to_{\mathbb{P}} \overline{R}''_n(z^*)$. Hence, $\overline{R}_n(\overline{z}_n) - \overline{R}_n(\overline{z}'_n) = O_p(\nu_n^2)$.

Again, by Taylor's theorem and first order conditions on $\overline{z}_n$, we have that for some $\overline{z}''''_n \in [\min(\overline{z}_n, \overline{z}'_n), \max(\overline{z}_n, \overline{z}'_n)]$,

$$\overline{R}_n(\overline{z}_n) - \overline{R}_n(\overline{z}'_n) = \frac{1}{2}\overline{R}''_n(\overline{z}''''_n)(\overline{z}_n - \overline{z}'_n)^2.$$

Since $\overline{R}_n(\overline{z}_n) - \overline{R}_n(\overline{z}'_n) = O_p(\nu_n^2)$ and $\overline{R}''_n(\overline{z}''''_n) \to_{\mathbb{P}} \overline{R}''_n(z^*)$, we conclude that $\overline{z}_n - \overline{z}'_n = O_p(\nu_n)$. By Slutsky's theorem, combining with the second claim of Theorem 7, this ensures the first claim of this theorem holds.

Following precisely the same arguments in the proof of Theorem 7, starting in equation (EC.4) and ending in equation (EC.4), we obtain the second claim. □

Proof of Theorem 9 The proof borrows the outline of the proof of Theorem 2 of Besbes et al. (2010), but applied to our new testing case and causal estimators.

Decompose the test statistic ρ_n into three terms:

$$\rho_n = \overline{R}_n(\overline{z}_n) - \overline{R}_n(\hat{z}_n) = A_n + B_n + C_n,$$

where

$$A_n = \overline{R}_n(\overline{z}_n) - \overline{R}_n(z^\star),$$
$$B_n = \overline{R}_n(z^\star) - \overline{R}_n(\hat{z}),$$
$$C_n = \overline{R}_n(\hat{z}) - \overline{R}_n(\hat{z}_n).$$

We begin by showing that $(nh_n^3)\,A_n \xrightarrow{d} \Gamma\chi_1^2$. By Assumption 2 part iii, we have that $\overline{R}_n(z)$ is twice continuously differentiable. Thus, using Taylor's theorem to expand $\overline{R}_n(z)$ around $z = \overline{z}_n$, we get that there exists $z_n \in [\min(z^\star, \overline{z}_n), \max(z^\star, \overline{z}_n)]$ such that

$$\overline{R}_n(z^\star) = \overline{R}_n(\overline{z}_n) + \overline{R}_n'(\overline{z}_n)(z^\star - \overline{z}_n) + \frac{1}{2}\overline{R}_n''(z_n)(z^\star - \overline{z}_n)^2.$$

By first order optimality conditions, $\overline{R}_n'(\overline{z}_n) = 0$. Hence, rearranging, we have

$$A_n = -\frac{1}{2}\overline{R}_n''(z_n)(z^\star - \overline{z}_n)^2. \tag{EC.9}$$

Next we show that $\overline{R}_n''(z_n) \xrightarrow{\mathbb{P}} R''(z^\star)$. Note that

$$\left|\overline{R}_n''(z_n) - R''(z^\star)\right| \le \left|\overline{R}_n''(z_n) - R''(z_n)\right| + \left|R''(z_n) - R''(z^\star)\right|. \tag{EC.10}$$

As in the proof of Theorem 7, Assumptions 2, 3, and 4 imply the assumptions of Lemma 5.1 of Newey (1994) applied to $R''(z)$, which in turn yields the uniform convergence in probability of $\overline{R}_n''(z)$ over $\mathcal{Z}$ since $\mathcal{Z}$ is compact by Assumption 3 part i. Hence,

$$\left|\overline{R}_n''(z_n) - R''(z_n)\right| \le \sup_{z\in\mathcal{Z}}\left|\overline{R}_n''(z) - R''(z)\right| \xrightarrow{\mathbb{P}} 0. \tag{EC.11}$$

By Theorem 7, $\overline{z}_n \xrightarrow{\mathbb{P}} z^\star$. Because z_n is sandwiched between $\overline{z}_n$ and $z^\star$, we also get $z_n \xrightarrow{\mathbb{P}} z^\star$. Since $R''(z)$ is continuous by Assumption 3 part iv, we have

$$\left|R''(z_n) - R''(z^\star)\right| \xrightarrow{\mathbb{P}} 0 \tag{EC.12}$$

by continuous transformation of the former. Combining eqs. (EC.10)–(EC.12), we get

$$\overline{R}_n''(z_n) \xrightarrow{\mathbb{P}} R''(z^\star). \tag{EC.13}$$

By continuous transformation of the result of Theorem 7 (eq. (EC.4)), we have

$$\left(nh_n^3\right)(\overline{z}_n - z^\star)^2 \xrightarrow{d} \frac{\eta_{z^\star}\kappa'}{R''(z^\star)^2}\chi_1^2. \tag{EC.14}$$

Combining eqs. (EC.9)–(EC.14), we get

$$\left(nh_n^3\right)A_n \xrightarrow{d} \frac{-\eta_{z^\star}\kappa'}{2R''(z^\star)}\chi_1^2 = \Gamma\chi_1^2. \tag{EC.15}$$

Next, we show that $(nh_n^3)\, C_n \xrightarrow{\mathbb{P}} 0$. By Assumption 2 part iii, we have that $\overline{R}_n(z)$ is twice continuously differentiable. Thus, using Taylor's theorem to expand $\overline{R}_n(z)$ around $z = \hat{z}$, we get that there exists $z_n' \in [\min(\hat{z}, \hat{z}_n), \max(\hat{z}, \hat{z}_n)]$ such that

$$\overline{R}_n(\hat{z}_n) = \overline{R}_n(\hat{z}) + \overline{R}_n'(\hat{z})(\hat{z}_n - \hat{z}) + \frac{1}{2}\overline{R}_n''(z_n')(\hat{z}_n - \hat{z})^2.$$

Rearranging, we have

$$(nh_n^3)\, C_n = -\left(\sqrt{nh_n^3}\overline{R}_n'(\hat{z})\right)\left(\sqrt{nh_n^3}(\hat{z}_n - \hat{z})\right) - \frac{1}{2}\overline{R}_n''(z_n')\left(\sqrt{nh_n^3}(\hat{z}_n - \hat{z})\right)^2. \tag{EC.16}$$

By Assumption 5, we have that

$$\sqrt{nh_n^3}(\hat{z}_n - \hat{z}) = o_p(1), \text{ and hence also } \left(\sqrt{nh_n^3}(\hat{z}_n - \hat{z})\right)^2 = o_p(1). \tag{EC.17}$$

Applying Theorem 4 of Newey (1994) we get the convergence in distribution of $\sqrt{nh_n}\left(\overline{R}_n'(z) - R'(z)\right)$ for any fixed z, including $\hat{z}$ and hence, since $h_n \to 0$ we have

$$\sqrt{nh_n^3}\overline{R}_n'(\hat{z}) = o_p(1). \tag{EC.18}$$

Next we show that $\overline{R}_n''(z_n') = O_p(1)$. Note that

$$\left|\overline{R}_n''(z_n') - R''(\hat{z})\right| \le \left|\overline{R}_n''(z_n') - R''(z_n')\right| + \left|R''(z_n') - R''(\hat{z})\right|. \tag{EC.19}$$

As before, $\overline{R}_n''(z)$ converges uniformly to $R''(z)$ in probability over $\mathcal{Z}$ and so

$$\left|\overline{R}_n''(z_n') - R''(z_n')\right| \le \sup_{z\in\mathcal{Z}}\left|\overline{R}_n''(z) - R''(z)\right| \xrightarrow{\mathbb{P}} 0. \tag{EC.20}$$

By Assumption 5, $\hat{z}_n \xrightarrow{\mathbb{P}} \hat{z}$. Because z_n' is sandwiched between $\hat{z}_n$ and $\hat{z}$, we also get $z_n' \xrightarrow{\mathbb{P}} \hat{z}$. Since $R''(z)$ is continuous by Assumption 3 part iv, we have

$$\left|R''(z_n') - R''(\hat{z})\right| \xrightarrow{\mathbb{P}} 0 \tag{EC.21}$$

by continuous transformation of the former. Combining eqs. (EC.19)–(EC.21), we get

$$\overline{R}_n''(z_n') \xrightarrow{\mathbb{P}} R''(\hat{z}). \tag{EC.22}$$

Combining eqs. (EC.16)–(EC.22) gives $(nh_n^3)\, C_n = -o_p(1)o_p(1) - O_p(1)o_p(1) = o_p(1)$.

Finally, we treat B_n. Under H_0, $B_n = 0$ because Assumption 3 part ii (unique optimizer) and H_0 ($R(z^\star) = R(\hat{z})$) imply that $z^\star = \hat{z}$. Next, we show that under H_1, $(nh_n^3)\, B_n \xrightarrow{\mathbb{P}} \infty$. By applying the first results of Theorem 7 to each term, we have that $B_n \xrightarrow{\mathbb{P}} R(z^\star) - R(\hat{z})$. Since $k \ge 0$, Assumption 2 part vi implies $nh_n^5/\log(n) \to \infty$, which, since we also assume $h_n \to 0$, implies $nh_n^3 \to \infty$. Hence, since $R(z^\star) - R(\hat{R}) > 0$ under H_1, we have that $(nh_n^3)\, B_n \xrightarrow{\mathbb{P}} \infty$. □

Proof of Theorem 10 Proven above. See eq. (EC.15). □

EC.2. More General Version of Theorem 9

ASSUMPTION EC.1 **(Convergent Decision-Making (relaxed)).** $\sqrt{nh_n^3}(\hat{z}_n - \hat{z}) \xrightarrow{d} \mathcal{N}(0, V)$ *for some* $V \geq 0$.

Note that Assumption 5 implies Assumption EC.1 with $V = 0$. In this sense. Assumption EC.1 is weaker and more general.

THEOREM EC.1. *Suppose Assumptions 1, 2, 3, 4, and EC.1 hold. Let* $\Gamma = \frac{-\eta_{z^\star}\kappa'}{2R''(z^\star)}$ *and* $\Gamma' = \frac{-V}{2}R''(z^\star)$*. Then,*

i. *under* H_0, $\limsup_{n\to\infty} \mathbb{P}\left((nh_n^3)\,\rho_n > t\right) \leq 1 - F_{\Gamma\chi_1^2+\Gamma'\chi_1^2}(t)$, *where* $F_{\Gamma\chi_1^2+\Gamma'\chi_1^2}$ *is the CDF of the weighted sum of two independent chi-squared random variables, and*

ii. *under* H_1, $(nh_n^3)\,\rho_n \xrightarrow{d} \infty$.

Proof. The only part of the proof of Theorem 9 that changes is the analysis of the term C_n. Following the arguments after eq. (EC.16) but using Assumption EC.1 we conclude that

$$(nh_n^3)C_n = \frac{-R''(\hat{z})}{2}(nh_n^3)(\hat{z} - \hat{z}_n)^2 + o_p(1).$$

As before, we had that

$$(nh_n^3)A_n = \frac{-R''(z^*)}{2}(nh_n^3)(z^* - \overline{z}_n)^2 + o_p(1).$$

Now, under H_0 and under Assumption 3, $R''(\hat{z}) = R''(z^*)$, which is indeed negative. Therefore, under H_1, $(nh_n^3)\,\rho_n \xrightarrow{d} \infty$, and, under H_0, $(nh_n^3)\,\rho_n \xrightarrow{d} H$, where $H = G_1^2 + G_2^2$ and (G_1, G_2) are jointly normal random variables with mean zero, variances Γ and Γ', and some covariance C. The distribution of H with some covariance C is stochastically dominated by the same with covariance 0, which yields the result. □

The implication is that if we use the p-value given by $1 - F_{\Gamma\chi_1^2+\Gamma'\chi_1^2}(nh_n^3\rho_n)$ then it would be a safe p-value in that it will still ensure at most α type-I error rate if we reject the null only when $p < \alpha$. Given Γ and Γ', we can simulate the distribution of $\Gamma\chi_1^2 + \Gamma'\chi_1^2$ to arbitrary precision or use formulae for the weighted sum of chi-squared random variables (Bausch 2013).

Finally, note that as in Theorem 10, we again have that $(nh_n^3)\mathbb{E}[C_n] \to \Gamma'$. Consequently, we can estimate Γ' using exactly the same bootstrap method but instead applied to $\hat{z}_n$.

References

Bausch J (2013) On the efficient calculation of a linear combination of chi-square random variables with an application in counting string vacua. *Journal of Physics A: Mathematical and Theoretical* 46(50):505202.

Besbes O, Phillips R, Zeevi A (2010) Testing the validity of a demand model: An operations perspective. *Manufacturing & Service Operations Management* 12(1):162–183.

Flores CA (2005) *Estimation of Dose-Response Functions and Optimal Treatment Doses with a Continuous Treatment*. Ph.D. thesis, University of California, Berkeley.

Newey WK (1994) Kernel estimation of partial means and a general variance estimator. *Econometric Theory* 10(02):1–21.