Supplemental Appendix for *Treatment Effects in Market Equilibrium*
Authors: Evan Munro, Xu Kuang, and Stefan Wager

B Details from the Simulations

B.1 Estimation

The model described in the main text captures the impact of a conditional cash transfer on children's health outcomes and the market for a perishable protein source in a remote village in the Philippines. It is designed based on the results in Filmer et al. (2023).

There are 10 coefficients of the model $\boldsymbol{\theta} = \begin{bmatrix}\theta_{d00} & \theta_{d01} & \theta_{dw} & \theta_{dp} & \theta_{s0} & \theta_{sp} & \theta_{y00} & \theta_{y01} & \theta_{yd} & \theta_{yw}\end{bmatrix}^\top$. The model is estimated to match the following 10 moments from the subsample of remote villages in Filmer et al. (2023). These moments are computed using data files found in the replication package for the paper, see Filmer et al. (2021).

- The average demand for eggs per week among eligible and ineligible children in control villages $(\hat{\mu}^d_{elig}, \hat{\mu}^d_{inelig})$.
- The average height-for-age Z-score among eligible and ineligible children in control villages $(\hat{\mu}^y_{elig}, \hat{\mu}^y_{inelig})$.
- The average price of eggs in treated and control villages ($\hat{p}^0$ and $\hat{p}^1$). The elasticity of demand for individuals in control villages in the model matches the elasticity of egg demand cited, although not directly estimated in Filmer et al. (2023) (η_{egg}).
- The effect of the treatment on eligible children on height-for-age Z-score and eggs per week $(\hat{\tau}^y_{elig}, \hat{\tau}^d_{elig})$, where treatment effects are defined as the differences-in-means for ineligible children in treated versus control villages.
- The ratio of treatment effects on height-for-age Z-score and eggs-per-week demand for ineligible children ($\hat{\pi} = \hat{\tau}^y_{inelig} / \hat{\tau}^d_{inelig}$)[12]. Again, treatment effects are defined as the differences-in-means for ineligible children in treated versus control villages.

We use these moments to estimate the model. The randomization of treatment makes estimating coefficients on treatments straightforward, because the errors in the demand equation are independent of $W_{h(i)}$. Under the assumptions in Filmer et al. (2023), the randomization of treatment also provides enough variation to estimate coefficients on prices in a linear demand and supply model. The coefficient on price in the demand model is provided directly by the elasticity of demand taken from the literature in Filmer et al. (2023). Given the demand elasticity, and under the assumption that $W_{h(i)}$ does not enter into the supply equation, then the change in prices in treated in and control villages provides an estimate for the supply elasticity under the market-clearing condition. The authors argue that treatments do not affect supply meaningfully, because most eggs sold in small villages are produced outside each village. Formally, each parameter has an estimator, which is a simple function of the moments.

[12] It is possible to estimate the model by matching the treatment effects on outcomes and demand directly for ineligible children, rather than their ratio. This would avoid using the elasticity of egg demand, which is taken from the literature rather than estimated from the data. However, the sample size of ineligible children in remote villages is small, so estimating the model to match these individual moments, which are imprecisely estimated, rather than their ratio, leads to an implied elasticity that does not seem reasonable.

$$\hat{\theta}_{d00} = \hat{\mu}^d_{inelig} - \hat{\theta}_{dp} \cdot \hat{p}^0, \qquad \hat{\theta}_{d01} = \hat{\mu}^d_{elig} - \hat{\theta}_{dp} \cdot \hat{p}^0,$$

$$\hat{\theta}_{dp} = \frac{\eta_{egg}}{1.01 \cdot \hat{p}_0} \cdot \frac{1}{n} \sum_{i=1}^{n} (E_i \hat{\mu}^d_{elig} + (1 - E_i)\hat{\mu}^d_{inelig}), \qquad \hat{\theta}_{dw} = \hat{\tau}^d_{elig} - \hat{\theta}_{dp}(\hat{p}_1 - \hat{p}_0),$$

$$\hat{\theta}_{sp} = \left(\frac{1}{n} \sum_{i=1}^{n} E_i \hat{\tau}^d_{elig} + \frac{1}{n} \sum_{i=1}^{n} (1 - E_i)(\hat{p}^1 - \hat{p}^0)\hat{\theta}_{dp} \right) \Big/ (\hat{p}^1 - \hat{p}^0)$$

$$\hat{\theta}_{s0} = \frac{1}{n} \sum_{i=1}^{n} (E_i \hat{\mu}^d_{elig} + (1 - E_i)\hat{\mu}^d_{inelig}) - \hat{\theta}_{sp}\hat{p}_0,$$

$$\hat{\theta}_{yd} = \frac{\hat{\tau}^y_{inelig}}{\hat{\tau}^d_{inelig}}, \qquad \hat{\theta}_{yw} = \hat{\tau}^y_{elig} - \hat{\theta}_{yd}\hat{\tau}^d_{inelig},$$

$$\hat{\theta}_{y00} = \hat{\mu}^y_{inelig} - \hat{\theta}_{yd}\hat{\mu}^d_{inelig}, \qquad \hat{\theta}_{y01} = \hat{\mu}^y_{elig} - \hat{\theta}_{yd}\hat{\mu}^d_{inelig}.$$

The moments used for estimating the model and the corresponding parameter estimates are described in Table 3.

Estimated Moments from Filmer et al. (2023)									
$\hat{\mu}^d_{elig}$	$\hat{\mu}^d_{inelig}$	$\hat{\mu}^y_{elig}$	$\hat{\mu}^y_{inelig}$	$\hat{p}^0$	$\hat{p}^1$	η_{egg}	$\hat{\tau}^y_{elig}$	$\hat{\tau}^d_{elig}$	$\frac{\hat{\tau}^y_{inelig}}{\hat{\tau}^d_{inelig}}$
1.6	2.2	-2.47	-1.44	6.07	6.26	-1.3	0.32	0.12	1.85
Estimated Parameters									
$\hat{\theta}_{d00}$	$\hat{\theta}_{d01}$	$\hat{\theta}_{dw}$	$\hat{\theta}_{dp}$	$\hat{\theta}_{s0}$	$\hat{\theta}_{sp}$	$\hat{\theta}_{y00}$	$\hat{\theta}_{y01}$	$\hat{\theta}_{yd}$	$\hat{\theta}_{yw}$
4.49	3.89	0.19	-0.38	-0.21	0.33	-5.51	-5.43	1.85	0.1

Table 3: Estimation of the Model in Section 3

Last, for simulating data from the model, C_i and $h(i)$ are determined by random draws from the empirical distribution of households with children between 0 and 14 (who are in eligible for the transfer if they meet a proxy means test) from the baseline survey in Filmer et al. (2023). E_i is drawn randomly to match the average percentage of eligible households in remote villages. The household and individual-level error terms in the demand equation are independent and mean-zero normally distributed variables with a standard deviation of 1/3rd. Similarly, the household level error terms in the outcome equation are independent and mean-zero normally distributed variables with a standard deviation of 1/3rd and the individual-level error terms have standard deviation of 1.

B.2 Adding Heterogeneity

The model in Section 5.1 is augmented to add some additional heterogeneity. For each household, a set of covariates $X_h \sim \mathcal{N}(0, 1)^{10}$ is drawn. Households now have type $\zeta_h \in \{A, B, C, D\}$. Households of type $\zeta_h = A$ have data generated by the model in the previous section. In response to the treatment, they increase their purchase of a perishable protein source, and children's health outcomes are improved on average.

$$D_i^A(W_h,p) = \theta_{d01}\cdot E_{h(i)} + \theta_{d00}(1-E_{h(i)}) + \theta_{dw}W_{h(i)}E_{h(i)} + \theta_{dp}\cdot p + \varepsilon_{d,h(i)} + \nu_{d,i}(W_i)$$
$$Y_i^A(W_{h(i)},p) = \theta_{y01}E_{h(i)} + \theta_{y00}(1-E_{h(i)}) + \theta_{yd}D_i^A(W_{h(i)},p) + \theta_{yw}W_{h(i)}E_{h(i)} + \varepsilon_{y,h(i)} + \nu_{yi}(W_i)$$

For the other three types, treatment effects are distinct from Type A. Households of type $\zeta_h = B$ increase their consumption of a perishable protein source and purchase a more expensive non-perishable protein source (such as canned fish). We assume for this second type of good, because it is non-perishable, supply is much more elastic and equilibrium effects are negligible. So, for these households, the data generating process is as follows:

$$D_i^B(W_h,p) = D_i^A(W_h,p) + 0.75\theta_{d01}\cdot E_{h(i)}\cdot W_{h(i)}$$
$$Y_i^B(W_{h(i)},p) = \theta_{y01}E_{h(i)} + \theta_{y00}(1-E_{h(i)}) + \theta_{yd}D_i^A(W_{h(i)},p) + (\theta_{yw}+0.2)W_{h(i)}E_{h(i)} + \varepsilon_{y,h(i)} + \nu_{yi}(W_i)$$

For households of type $\zeta_h = C$, the cash transfer does not have its intended positive effect. In response to the cash transfer, an increased ability of parents to afford addictive goods (such as cigarettes, or alcohol) leads to a decrease in the purchase of perishable protein, which has a large negative effect on children's health. Let Child_i be an indicator if individual i is under 5 years of age.

$$D_i^C(W_h,p) = \theta_{d01}\cdot E_{h(i)}(1-0.75W_{h(i)}\text{Child}_i) + \theta_{d00}(1-E_{h(i)}) + \theta_{dp}\cdot p + \varepsilon_{d,h(i)} + \nu_{d,i}(W_i)$$
$$Y_i^C(W_{h(i)},p) = \theta_{y01}E_{h(i)} + \theta_{y00}(1-E_{h(i)}) + \theta_{yd}D_i(W_{h(i)},p) + \theta_{yw}W_{h(i)}E_{h(i)} + \varepsilon_{y,h(i)} + \nu_{yi}(W_i)$$

For households of type $\zeta_h = D$, there is an increase in food consumption of adults, but not of children.

$$D_i^D(W_h,p) = \theta_{d01}\cdot E_{h(i)} + \theta_{d00}(1-E_{h(i)}) + \theta_{dp}\cdot p + \varepsilon_{d,h(i)} + 1.5\theta_{dw}W_{h(i)}E_{h(i)}(1-\text{Child}_i) + \nu_{d,i}(W_i)$$
$$Y_i^D(W_{h(i)},p) = \theta_{y01}E_{h(i)} + \theta_{y00}(1-E_{h(i)}) + \theta_{yd}D_i(W_{h(i)},p) + 3\cdot\theta_{yw}W_{h(i)}E_{h(i)} + \varepsilon_{y,h(i)} + \nu_{yi}(W_i)$$

The household type is correlated with observed covariates as follows. The score for each household type is:

$$\begin{aligned}
U_{hA} &= 1 + X_{2h} - 0.5X_{3h} + \varepsilon_{hA},\\
U_{hB} &= X_{1h} + \varepsilon_{hB},\\
U_{hC} &= -1 + X_{3h} + \varepsilon_{hC},\\
U_{hD} &= X_{4h} + \varepsilon_{hD},
\end{aligned}$$

where each of the noise terms are drawn from a Type I extreme value distribution. $\zeta_h = \arg\max_{T\in\{A,B,C,D\}} U_{hT}$. The coefficients of the model are the same as in Table 3. In the simulation, 44% of the households are Type A, 22% are type B, 22% are Type D, and 12% are Type C. Demand and outcomes are aggregated by household as in Equation (39).

C Concentration Bounds

Consider a triangular array of random functions $F_{in}(p)$ with $i = 1, \ldots, n$, $n = 1, 2, \ldots$ and $p \in \mathbb{R}^J$. Suppose that $F_{in}(\cdot)$ with $i = 1, \ldots, n$ are independent and identically distributed, and let $f_n(p) = \mathbb{E}[F_{in}(p)]$. We say that the sample average of the $F_{in}(\cdot)$ is asymptotically equicontinuous at p^* if, for any sequence $\delta_n \to 0$, we have

$$\sup_{\|p-p^*\| \leq \delta_n} \left| \frac{1}{n} \sum_{i=1}^{n} \left(F_{in}(p) - F_{in}(p^*) - (f_n(p) - f_n(p^*))\right) \right| = o_p\left(\frac{1}{\sqrt{n}}\right). \tag{79}$$

The motivation behind establishing asymptotic equicontinuity expansions is that we may often be able to assume that the expected function $f_n(p)$ is differentiable even when the $F_{in}(p)$ themselves are not (e.g., if the $F_{in}(p)$ capture choices regarding supply and demand that may vary discontinuously in prices). Asymptotic equicontinuity then allows us to approximate sample averages of $F_{in}(p)$ by Taylor expanding $f_n(p)$.

The goal of this section is to establish asymptotic equicontinuity for a number of function classes used throughout the proofs. Our argument relies on technical tools from empirical process theory that go beyond what's used elsewhere in the paper, and here we will only give brief references to the results needed to establish our desired claims. Section 2 of van der Vaart & Wellner (1996) provides an excellent introduction and reference to the tools underlying the arguments made here.

Our approach to establishing asymptotic equicontinuity of relevant quantities in our marketplace model starts by bounding the bracketing number of approximately monotone functions. Let $F_{in} : \mathbb{R}^J \to \mathbb{R}$ be random functions drawn i.i.d. from a distribution Q_n, and let $\mathcal{S} \subset \mathbb{R}^J$ be the set of possible market equilibrium prices. We are interested in a bracket for the function class $\mathcal{F}_n = \{F_{in}(\cdot) \to F_{in}(p) : p \in \mathcal{S}\}$. For any pair p_-, p_+ in $\mathbb{R}^J$, we define a bracket as

$$[p_-, p_+] = \{p \in \mathcal{S} : F_{in}(p_-) \geq F_{in}(p) \geq F_{in}(p_+) \text{ for all } F_{in}(\cdot)\}. \tag{80}$$

We define the square of the $L_2(Q_n)$-length of the bracket as $d^2_{Q_n}(p_-, p_+) = \mathbb{E}_{Q_n}\left[(F_{in}(p_-) - F_{in}(p_+))^2\right]$, and the ε-bracketing number of $\mathcal{F}_n$ under Q_n as

$$N_{[]}(\varepsilon, \mathcal{F}_n, L_2(Q_n)) = \inf\left\{K : \mathcal{S} \subseteq \bigcup_{k=1}^{K} [p_-(k), p_+(k)] \text{ with } d_{Q_n}(p_-(k), p_+(k)) \leq \varepsilon \text{ for all } k\right\}. \tag{81}$$

The following result bounds the bracketing number of functions that are approximately monotone and weakly continuous in the sense of Assumption 4.

Lemma 5. *Consider a class of random functions $F_{in} : \mathbb{R}^J \to \mathbb{R}$ with $F_{in} \in \mathcal{G}_n$ almost surely, and $\mathcal{S}$ is a compact subset of $\mathbb{R}^J$. Let $\mathcal{F}_n = \{F_{in}(\cdot) \to F_{in}(p) : p \in \mathcal{S}\}$. Suppose that, for all $p \in \mathcal{S}$, $\varepsilon \leq 1$ and $\|\delta\|_2 \leq C\varepsilon$, we have*

$$F_{in}(p - \varepsilon e_j) \geq F_{in}(p + \delta) \geq F_{in}(p + \varepsilon e_j). \tag{82}$$

Suppose also that the F_{in} have a distribution Q_n under which

$$\mathbb{E}_{Q_n}\left[\left(F_{in}(p) - F_{in}(p')\right)^2\right] \leq L \left\|p - p'\right\|_2 \text{ for all } p, p' \in \mathbb{R}^J. \tag{83}$$

Then, there exists $\theta(\mathcal{S}, C, L)$ depending only on the geometry of $\mathcal{S}$ and the constants given above *such that, for all $\varepsilon \le 1$,*

$$N_{[]}(\varepsilon,\, \mathcal{F}_n,\, L_2(Q_n)) \le \theta(\mathcal{S},\, C,\, L)\, \varepsilon^{-2J}. \tag{84}$$

Proof. Let $B_\delta(p) = \{p' \in \mathcal{S} : \|p' - p\|_2 \le \delta\}$. By (82), we have

$$B_{C\zeta}(p) \subseteq [p - \zeta e_j,\, p + \zeta e_j].$$

Meanwhile, by (83) we have $d^2_{Q_n}(p - \zeta e_j,\, p + \zeta e_j) \le 2L\zeta$, and so $d_{Q_n}(p - \zeta e_j,\, p + \zeta e_j) \le \varepsilon$ if $\zeta = \varepsilon^2/(2L)$. These two facts together imply that

$$N_{[]}(\varepsilon,\, \mathcal{F}_n,\, L_2(Q_n)) \le \inf\left\{K : \mathcal{S} \subseteq \bigcup_{k=1}^{K} B_{C(\varepsilon^2/(2L))}(p_k) \text{ with } p_k \in \mathbb{R}^J\right\},$$

where the right-hand side quantity is the $C(\varepsilon^2/(2L))$-covering number of $\mathcal{S}$ under the usual norm $\|.\|_2$. For $\varepsilon \le 1$, this quantity can be further be bounded as (Polyanskiy & Wu 2024, Theorem 27.3)

$$\begin{aligned} &\inf\left\{K : \mathcal{S} \subseteq \bigcup_{k=1}^{K} B_{C(\varepsilon^2/(2L))}(p_k) \text{ with } p_k \in \mathbb{R}^J\right\} \\ &\qquad \le 3^J \left(C\left(\frac{\varepsilon^2}{2L}\right)\right)^{-J} \frac{\operatorname{vol}\left(\operatorname{conv}\left(\mathcal{S} * B_{C/(2L)}\right)\right)}{\operatorname{vol}\left(B_1\right)}, \end{aligned} \tag{85}$$

where B_r denotes the radius-r ball in $\mathbb{R}^J$ centered at 0, $\mathcal{A} * \mathcal{B} = \{a + b : a \in \mathcal{A},\, b \in \mathcal{B}\}$, conv(.) denotes the convex hull of a set and vol(.) denotes the volume of a set. The scaling in (85) establishes the desired claim. □

Lemma 6. *We say that a random function F_{in} is ζ-subgaussian over $\mathcal{S}$ if there exists a constant C such that, for all $n \ge 1$ and $t > 0$, that*

$$\mathbb{P}\left[\sup_{p \in \mathcal{S}} \left|\frac{1}{n}\sum_{i=1}^{n} F_{in}(p) - f(p)\right| > \frac{t}{\sqrt{n}}\right] \le C t^{\zeta} e^{-2t^2}. \tag{86}$$

Under Assumptions 4, 5 and 6, let U_i be random price perturbations with $\mathbb{E}[U_i] = 0$ and $|U_i| \le h_n$ almost surely for some sequence $h_n \to 0$. Suppose furthermore that W_i is Bernoulli-randomized as in Design 1. Then, the following random functions are ζ-subgaussian:

- *$Z_i(w,\, p + U_i)$ for $w = 0,\, 1$.*
- *$Z_i(W_i,\, p + U_i)$ and $(W_i/\pi_i - (1 - W_i)/(1 - \pi_i))Z_i(W_i,\, p + U_i)$.*
- *$Y_i(w,\, p + U_i)$ for $w = 0,\, 1$.*
- *$Y_i(W_i,\, p + U_i)$ and $(W_i/\pi_i - (1 - W_i)/(1 - \pi_i))Y_i(W_i,\, p + U_i)$.*

Proof. For $Z_i(w,\, p+U_i)$ and $Z_i(W_i,\, p+U_i)$, we can use the bracketing bounds in Lemma 18, which imply $N_{[]}(\varepsilon, \mathcal{F}_n, L_2(Q_n)) \le (K/\varepsilon)^{\zeta}$, for $\zeta = 2J$, and $\mathcal{F}_n$ is constructed by varying $Z_i(w,\, p + U_i)$ or $Z_i(W_i,\, p + U_i)$ over $p \in \mathcal{S}$. The tail bound follows from Theorem 2.14.9 of van der Vaart & Wellner (1996), given boundedness of the net demand functions and the polynomial bound on the ε-bracketing number. The result for $W_i/\pi_i\, Z_i(W_i,\, p + U_i)$, $(1 - W_i)/(1 - \pi_i)\, Z_i(W_i,\, p + U_i)$ and their differences follows by the same argument because W_i/π_i and $(1 - W_i)/(1 - \pi_i)$ are positive and bounded by η^{-1}, and thus these functions satisfy the conditions of Lemma 18 with a constant L/η in (83).

For outcomes, we have the decomposition $Y_i(w,p) = H_i(w,p) + \psi(\Gamma_i(w), Z_i(w,p), p)$. We first work with $H_i(w,p)$. Note that bracketing bounds for Lipschitz function classes are standard; see, e.g., Section 2.7.4 of van der Vaart & Wellner (1996). Adapting these results to our setting, we let $F_i(w,p) = H_i(w,p) + c$ and $\mathcal{F} = \{F_i(w,\cdot) \to F_i(w,p) : p \in \mathcal{S}\}$.

By the Lipschitz property, for any $p \in \mathbb{R}^J$,

$$\left\{(p',\, 0) : \left\|p - p'\right\|_2 \leq \varepsilon\right\} \leq [(p,\, -M\varepsilon),\, (p,\, +M\varepsilon)],$$

and $d_P((p,\, -M\varepsilon),\, (p,\, M\varepsilon)) \leq 2M\varepsilon$ for any distribution Q over the functions $H_i(w,p)$.

Thus, using notation from Lemma 18,

$$N_{[]}(\varepsilon,\, \mathcal{F},\, L_2(Q)) \leq \inf\left\{K : \mathcal{S} \subseteq \bigcup_{k=1}^{K} B_{\varepsilon/(2M)}(p_k) \text{ with } p_k \in \mathbb{R}^J\right\}, \tag{87}$$

which is in turn $O(\varepsilon^{-J})$ by the same argument as used in (85). This bracketing bound can be applied to $H_i(w,\, p + U_i)$, $H_i(W_i,\, p + U_i)$, etc., thus enabling us to again apply Theorem 2.14.9 of van der Vaart & Wellner (1996).

For the second component of the outcome function, it is useful to use a covering number bound rather than a bracketing number bound. For a function class $\mathcal{F}_n$, we can define the ε-covering number of $\mathcal{F}_n$ under a distribution Q_n as $N(\varepsilon, \mathcal{F}_n, L_2(Q_n))$. This is defined as the minimum number of balls of radius ε in $L_2(Q_n)$ norm required to cover $\mathcal{F}_n$. ε-covering numbers are bounded by 2ε-bracketing numbers, as shown in Section 2.1.1 of van der Vaart & Wellner (1996). Theorem 2.14.9 of van der Vaart & Wellner (1996) provides a tail bound as in the Lemma statement for function classes with

$$\sup_{Q_n} N(\varepsilon, \mathcal{F}_n, L_2(Q_n)) \leq \left(\frac{K}{\varepsilon}\right)^{\zeta}$$

for constants $K, \zeta > 0$. It remains to provide such a bound for the class $\{Y_i(\cdot) \to Y_i(w,p) : p \in \mathcal{S}\}$.

Let $G_{in} = \phi(F_{i1n}, \ldots, F_{iMn})$ be a composition of M random functions, drawn from distribution Q_n. Let $\mathcal{G}_n = \{\phi(F_{i1n}(\cdot), \ldots, F_{iMn}(\cdot)) \to \phi(F_{i1n}(p), \ldots, F_{iMn}(p)) : p \in \mathcal{S}\}$. and $\mathcal{F}_{mn} = \{F_{imn}(\cdot) \to F_{imn}(p) : p \in \mathcal{S}\}$. Lemma A.6 of Chernozhukov et al. (2014) indicates that as long as $\phi(\cdot)$ is Lipschitz in each of its M arguments, then if there is a polynomial bound on the ε-covering number of $\mathcal{F}_{mn}$ for $m \in \{1, \ldots, M\}$, there is also a polynomial bound on the ε-covering number of $\mathcal{G}$. Recall that a ε-covering number is bounded by a 2ε-bracketing number and our bracketing numbers do not depend on Q_n. We can now apply the composition result from Chernozhukov et al. (2014) to finish the proof for $Y_i(w,p)$, since we already showed that $H_i(w,p)$ has the right kind of bound on the bracketing number of its function class, $\psi(\cdot)$ is Lipschitz in each of its arguments, and each of its arguments are functions of p that have a polynomial bound on their ε-covering numbers. □

Lemma 7. *For $n = 1,\, 2,\, \ldots$, let $F_{in}(p)$ be IID random functions that are ζ-subgaussian and weakly continuous, and whose covariances converge pointwise to a finite limit Σ,*

$$\lim_{n\to\infty} \operatorname{Cov}\left[F_{in}(p),\; F_{in}(p')\right] = \Sigma(p,\, p') \text{ for all } p,\, p'. \tag{88}$$

Then, averages of that function are asymptotically equicontinuous, i.e., they satisfy (79) at all

$p \in \mathcal{S}$. Furthermore, the residuals in the asymptotic equicontinuty expansion,

$$R_n = \sup \left\{ \left| \frac{1}{n} \sum_{i=1}^{n} (F_{in}(p) - F_{in}(p^*) - (f_n(p) - f_n(p^*))) \right| : \|p - p^*\| \leq \delta_n, \; p \in \mathcal{S} \right\} \tag{89}$$

satisfy $\limsup_{n \to \infty} n \, \mathbb{E}\left[R_n^2\right] = 0$. *In particular, under the assumptions of Lemma 19, averages of the following random functions are asymptotically equicontinuous and have residuals in the asymptotic equicontinuity expansion meeting the above condition:*

- $Z_i(w, \, p + U_i)$ *for* $w = 0, \, 1$.
- $Z_i(W_i, \, p + U_i)$ *and* $(W_i/\pi_i - (1 - W_i)/(1 - \pi_i))Z_i(W_i, \, p + U_i)$.
- $Y_i(w, \, p + U_i)$ *for* $w = 0, \, 1$.
- $Y_i(W_i, \, p + U_i)$ *and* $(W_i/\pi_i - (1 - W_i)/(1 - \pi_i))Y_i(W_i, \, p + U_i)$.

Proof. Let Q_n define a distribution over random functions $F_{in}(p)$, for any of the random functions listed in the Lemma, and define the function class $\mathcal{F}_n = \{F_{in}(\cdot) \to F_{in}(p) : p \in \mathcal{S}\}$. Theorems 2.11.1 and Theorem 2.11.9 van der Vaart & Wellner (1996) imply weak convergence of the empirical process

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(F_{in}(p) - f_n(p) - F_{in}(p') - f_n(p')\right)$$

to a Gaussian process indexed by p when for each n, the ε-bracketing number or ε-covering number of $\mathcal{F}_n$ under Q_n is bounded by $(K/\varepsilon)^V$ for all $0 < \varepsilon < K$, for finite constants $K, V > 0$. Then, asymptotic equicontinuity follows from weak continuity, i.e. that for each $p \in \mathcal{S}$ and $p' \in \mathcal{S}$, $\mathbb{E}[(F_{in}(p) - F_{in}(p'))^2]$ is continuous in p. Weak continuity implies that the limiting Gaussian process has continuous sample paths.

To verify the 2nd-moment bounds on the residuals in the expansions, we first note that

$$\begin{aligned} R_n = \sup_{\|p - p^*\| \leq \delta_n} & \left| \frac{1}{n} \sum_{i=1}^{n} (F_{in}(p) - F_{in}(p^*) - (f_n(p) - f_n(p^*))) \right| \\ & \leq 2 \sup_{p \in \mathcal{S}} \left| \frac{1}{n} \sum_{i=1}^{n} (F_{in}(p) - f_n(p)) \right|, \end{aligned}$$

which has a rapidly decaying tail by Lemma 19. Thus, in particular,

$$\limsup_{n \to \infty} n^2 \, \mathbb{E}\left[R_n^4\right] < \infty.$$

Now, pick any $c > 0$. By asymptotic equicontinuity, $\limsup_{n \to \infty} \mathbb{P}\left[|R_n| > c/\sqrt{n}\right] = 0$. Furthermore

$$\begin{aligned} \limsup_{n \to \infty} n \, \mathbb{E}\left[R_n^2\right] &= \limsup_{n \to \infty} n \, \mathbb{E}\left[1\left(\{|R_n| \leq c/\sqrt{n}\}\right) R_n^2\right] + n \, \mathbb{E}\left[1\left(\{|R_n| > c/\sqrt{n}\}\right) R_n^2\right] \\ &\leq c^2 + \limsup_{n \to \infty} \sqrt{\mathbb{P}\left[|R_n| > c/\sqrt{n}\right]} \sqrt{n^2 \, \mathbb{E}\left[R_n^4\right]} \\ &\leq c^2 + \limsup_{n \to \infty} \sqrt{\mathbb{P}\left[|R_n| > c/\sqrt{n}\right]} \limsup_{n \to \infty} \sqrt{n^2 \, \mathbb{E}\left[R_n^4\right]} \\ &= c^2, \end{aligned}$$

where for the second inequality we used Cauchy-Schwarz. Since we have shown that $0 \leq \limsup_{n\to\infty} n\,\mathbb{E}\left[R_n^2\right] \leq \epsilon$ for any $\epsilon > 0$, we see that in fact $\limsup_{n\to\infty} n\,\mathbb{E}\left[R_n^2\right] = 0$.

For all of the random functions listed, weak continuity holds by Assumption 5 and 6, and Lemma 19 provides the required subgaussian condition. A final check is htat

$$\lim_{n\to\infty} \mathrm{Cov}\left[Z_{ij}(w,\, p+U_i),\ Z_{ij}(w,\, p'+U_i)\right] = \mathrm{Cov}\left[Z_{ij}(w,\, p),\ Z_{ij}(w,\, p')\right], \tag{90}$$

which is finite, and the same holds for $Y_i(w, p+U_i)$ and $Y_i(W_i, p+U_i)$.

□

D Additional Results and Extensions

D.1 Consistency of Estimated Targeting Rule

When the optimal rule is deterministic and unique, it has the form $\nu^*(X_i) = \mathbb{1}(\tau^*_{\text{CADE}}(X_i) \geq c^*\tau^{*,z}_{\text{CADE}}(X_i))$. By Theorem 13, we can write the optimal rule as the unique solution to the following J-dimensional moment condition, where τ^* represents the vector of functions that concatenates $\tau^*_{\text{CADE}}(X_i)$ and $\tau^{*,z}_{\text{CADE}}(X_i)$ and τ is some bounded function with the same domain and codomain.

$$G(c^*; \tau^*) = 0, \text{ where } G(c;\tau) = \mathbb{E}\left[\Big(\mathbb{1}(\tau(X_i) \geq c^\top \tau^z(X_i)) - \pi(X_i)\Big)\tau^z(X_i)\right]$$

c^* is finite, so we can specify a ball in $\mathbb{R}^J$ such that the distance from c^* to the surface of this ball is at least M. Specify such a ball as $\mathcal{B}_M$. We can estimate the optimal rule by solving the following J-dimensional score condition. The conditional average direct effects are nuisance functions, since they need to be estimated, where $\hat{\tau}$ is some estimate of the population conditional average treatment effects on outcomes and net demand:

$$\hat{c} \in \{c \in \mathbb{R}^J : \hat{G}_n(c;\hat{\tau}) = o_p(1)\},$$
$$\hat{G}_n(c;\hat{\tau}) = \frac{1}{n}\sum_{i=1}^n (\mathbb{1}(\hat{\tau}(X_i) \geq c^\top \hat{\tau}^z(X_i) - \pi(X_i))\hat{\tau}^z(X_i).$$

The proof of Proposition 21 shows that the optimal-equilibrium neutral rule is the unique (and well-separated) solution to the J population constraints defined by $G(\cdot)$. Then, the consistency of the estimated rule follows from uniform convergence of the empirical constraint functions to the population constraint functions.

Proposition 4. *For every $v \in S^{J-1}$ and $b \in \mathbb{R}^+$, where S^{J-1} is the unit sphere in $\mathbb{R}^J$, assume that the random variable $\tau^*_{CADE}(X_i) - b \cdot v^\top \tau^{z,*}_{CADE}(X_i)$ has uniformly bounded density, and for every $v \in S^{J-1}$ that $v^\top \tau^{z,*}_{CADE}(X_i)$ has uniformly bounded density. In addition, assume that the covariance matrix of $\tau^{*,z}_{CADE}(X_i)$ is positive definite, and that $\hat{G}_n(\hat{c};\hat{\tau}) = o_p(1)$. Finally, assume that $\hat{\tau} : \mathcal{X} \to \mathbb{R}$ and $\hat{\tau}^z : \mathcal{X} \to \mathbb{R}$ are functions that are uniformly bounded and are consistent estimators of the population conditional average treatment effects in the following sense:*

$$\mathbb{E}_T[(\hat{\tau}(X_i) - \tau^*_{CADE}(X_i))^2] = o_p(1),$$
$$\mathbb{E}_T[||\hat{\tau}^z_j(X_i) - \tau^{*,z}_{j,CADE}(X_i))||_2^2] = o_p(1),$$

where $\mathbb{E}_T[\cdot]$ is the expectation over a random sample of test data, conditional on the training data used for estimation.

Then, the estimated targeting rule is consistent for the optimal equilibrium-stable targeting rule in the population,

$$\hat{c} = c^* + o_p(1).$$

First, we use a change of variables to make the parameter space compact. For every $c \in \mathbb{R}^J$, we can write $c = f(a) \cdot v$, where $v = c/||c||_2 \in S^{J-1}$ and $a = f^{-1}(||c||_2) \in [0, 1)$ for some continuous and strictly monotonic function $f : [0, 1] \to [0, +\infty]$ with $f(0) = 0$ and $f(1) = \infty$. For this proof, it will be useful to characterize properties of the population constraint function over the extended (and compact) space of $a \in [0, 1]$, even though the optimal rule, and all of the estimated rules are found within $a \in [0, 1)$. For a more compact notation define θ as the vector combining v and a, where $\theta \in \Theta$, and $\Theta = S^{J-1} \times [0, 1]$. Let

$$\nu(X_i; \theta, \tau) = \begin{cases} \mathbb{1}(\tau(X_i) > f(a)v^\top \tau^z(X_i)), & 0 \leq a < 1, \\ \mathbb{1}(0 > v^\top \tau^z(X_i)), & a = 1. \end{cases}$$

Now we can define the population and empirical constraint at a given treatment rule and for a given set of conditional mean functions:

$$G(\theta; \tau) = \mathbb{E}_T\left[\Big(\nu(X_i; \theta, \tau) - \pi(X_i)\Big)\tau^z(X_i)\right], \qquad G_n(\theta^*; \tau) = \frac{1}{n}\sum_{i=1}^{n}\Big(\nu(X_i; \theta, \tau) - \pi(X_i)\Big)\tau^z(X_i).$$

Any estimated treatment rule, characterized by $\hat{\theta}$, approximately satisfies the empirical constraint with conditional average treatment effects estimated by data-splitting, so that $G_n(\hat{\theta}; \hat{\tau}) = o_p(1)$.

Under our assumptions, Theorem 13 indicates that there is a unique and deterministic rule in the population characterized by $G(\theta^*; \tau^*) = 0$, where τ^* collects $\tau^*_{\text{CADE}}(X_i)$ and $\tau^{*,z}_{\text{CADE}}(X_i)$ and θ^* collects a^* and v^*. It is deterministic because the set $\{x \in \mathcal{X} : \tau^*_{\text{CADE}}(x) = f(a^*) \cdot v^{*\top} \tau^{z,*}_{\text{CADE}}(x)\}$ has measure 0. By Theorem 13, any optimal equilibrum-neutral rule that is deterministic has the structure $\mathbb{1}(\tau^*_{\text{CADE}}(x) > f(a^*)v^{*,\top}\tau^{z,*}_{\text{CADE}}(x))$. Since the covariance matrix of $\tau^{*,z}_{\text{CADE}}(X_i)$ is positive definite, there can be only one such rule that meets the population constraints (this rules out label swapping between two goods in the market, for example).

Next, we want to show that each element of $G(\theta; \tau^*)$ is a continuous function in θ for all $\theta \in \Theta$. Choose some a_0, v_0 and arbitrary $j \in \{1, \ldots, J\}$. For some sequence θ_n, where $\theta_n \to \theta_0 = (a_0, v_0)$, by the boundedness of $\tau^z(X_i)$, there is some finite M such that

$$\begin{aligned} |G_j(\theta_n; \tau^*) - G_j(\theta_0; \tau^*)| &\leq P(\nu(X_i; \theta_n, \tau^*) \neq \nu(X_i; \theta_0, \tau^*))M \\ &= O(||\theta_n - \theta_0||_2) \\ &= o(1) \end{aligned}$$

The first step is by Lemma 22. This proves that when $a = 1$, then $G(\theta; \tau^*)$ is continuous in θ. Since $G(\theta; \tau)$ is a continuous function in θ, $\hat{\theta}$ is also in a compact space, and θ^* is the unique solution to $G(\theta^*; \tau^*) = 0$, then we know that if $\epsilon > 0$, then $\sup_{\theta:||\theta-\theta^*||_2>\epsilon} ||G(\theta)||_2 > 0$. The final step is to show uniform consistency:

$$\sup_{\theta \in \Theta} |\hat{G}_n(\theta; \hat{\tau}) - G(\theta; \tau^*)| = o_p(1).$$

It will be useful to make the following decomposition:

$$|\hat{G}_n(\theta;\hat{\tau}) - g(\theta;\tau)| \leq |G_n(\theta;\hat{\tau}) - G(\theta;\hat{\tau})| + |G(\theta;\hat{\tau}) - G(\theta;\tau^*)|. \tag{91}$$

For the first term of (91), we write the empirical average using data-splitting, where I_k are the indexes of data in split k, and $\hat{\tau}^{k(i)}(X_i)$ are conditional average treatment effects estimated on data that is not in I_k. We drop the CADE subscripts here to keep the notation more manageable. Then, by treating $\hat{\tau}^k$ as fixed within each split, we use tail bounds for empirical processes indexed by VC-classes. The details are as follows:

$$\begin{aligned}
&\sup_{\theta\in\Theta} |G_n(\theta;\hat{\tau}) - G(\theta;\hat{\tau})| \\
&= \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}(\nu(X_i;\theta,\hat{\tau}) - \pi(X_i))\hat{\tau}^z(X_i) - \mathbb{E}_T[(\nu(X_i;\theta,\hat{\tau}) - \pi(X_i))\hat{\tau}^z(X_i)]\right| \\
&= \sup_{\theta\in\Theta}\left|\sum_{k=1}^{K} n_k/n \frac{1}{n_k}\sum_{i\in I_k}(\nu(X_i;\theta,\hat{\tau}^{k(i)}) - \pi(X_i))\hat{\tau}^{k(i),z}(X_i) - \mathbb{E}_T[(\nu(X_i;\theta,\hat{\tau}^{k(i)}) - \pi(X_i))\hat{\tau}^{k(i),z}(X_i)]\right| \\
&\leq \sum_{i=1}^{K}\frac{n_k}{n}\sup_{\theta\in\Theta}|A_{nk}(\theta)| \\
&\overset{(1)}{=} o_p(1)
\end{aligned}$$

Within each split, we can treat $\hat{\tau}^k$ as fixed. $\mathcal{F} = \{\mathbb{1}(\tau(X_i) \geq f(a)v^\top\tau^z(X_i))\tau^z(X_i) : a \in [0,1], v \in S^{J-1}\}$ for a fixed and uniformly bounded $\tau(X_i)$ is a VC-class multiplied by a bounded random variable. Then, the tail bound for empirical processes indexed by VC-classes (e.g. Theorem 2.6.7 of van der Vaart & Wellner (1996)) gives a tail bound for $\sup_{\theta\in\Theta}|A_{nk}(\theta)|$ that does not depend on a given realization of $\hat{\tau}$. This implies $\lim_{n\to\infty} Pr(\sup_{\theta\in\Theta}|A_{nk}| > t) = 0$ for any $t > 0$.

To handle the second term in (91),

$$\begin{aligned}
\sup_{\theta\in\Theta}|G(\theta;\hat{\tau}) - G(\theta;\tau^*)| &= \sup_{\theta\in\Theta}|\mathbb{E}_T[(\nu(X_i;\theta,\hat{\tau}) - \pi(X_i))\hat{\tau}^z(X_i)] - \mathbb{E}_T[(\nu(X_i;\theta,\tau^*) - \pi(X_i))\tau^{*,z}(X_i)]| \\
&\leq \sqrt{\mathbb{E}_T[(\hat{\tau}^z(X_i) - \tau^{*,z}(X_i))^2]} + \sup_{\theta\in\Theta} M\sqrt{\mathbb{E}_T[(\nu(X_i;\theta,\hat{\tau}) - \nu(X_i;\theta,\tau^*))^2]} \\
&\leq o_p(1) + \sup_{\theta\in\Theta} M\sqrt{P(\nu(X_i;\theta,\hat{\tau}) \neq \nu(X_i;\theta,\tau^*))} \\
&\overset{(1)}{=} o_p(1) + O(||\hat{\tau}^z(X_i) - \tau^{*,z}(X_i)||_2 + |\tau^*(X_i) - \hat{\tau}(X_i)|) \\
&= o_p(1)
\end{aligned}$$

where M is finite since net demand is bounded, (1) is from Lemma 22 and the final step is from the assumption on the mean-square convergence of the nuisance functions. We have now shown that $\sup_{\theta\in\Theta}|\hat{G}_n(\theta;\hat{\tau}) - G(\theta;\tau^*)| = o_p(1)$.

Now, that we have proved that the unique population treatment rule is well-separated and that the sample constraints converge uniformly to the population constraints, following Theorem 5.9 of van der Vaart (1998), we have consistency of $\hat{\theta}$ to θ^*. The details follow. Since $\hat{\theta}$ satisfies the empirical constraints, then $||G_n(\hat{\theta};\hat{\tau})||_2 \leq ||G_n(\theta^*;\tau^*)||_2 + o_p(1)$. By the convergence of the empirical constraints in the RHS, $||G_n(\hat{\theta};\hat{\tau})||_2 \leq ||G(\theta^*;\tau^*)||_2 + o_p(1)$. Adding and subtracting

$||G(\hat{\theta};\tau^*)||_2$, we have:

$$\begin{aligned}||G(\hat{\theta};\tau^*)||_2 - ||G_n(\hat{\theta};\hat{\tau})||_2 &\geq ||G(\hat{\theta};\tau^*)||_2 - ||G(\theta^*;\tau^*)||_2 - o_p(1)\\ ||G(\hat{\theta};\tau^*)||_2 - ||G(\theta^*;\tau^*)||_2 &\leq \sup_{\theta\in\Theta} ||G(\theta;\tau^*)||_2 - ||G_n(\theta;\hat{\tau})||_2\\ ||G(\hat{\theta};\tau^*)||_2 &\leq \sup_{\theta\in\Theta} ||G(\theta;\tau^*) - G_n(\theta;\hat{\tau})||_2\\ ||G(\hat{\theta};\tau^*)||_2 &\leq o_p(1)\end{aligned}$$

Since θ^* is well-separated, then for any $\epsilon > 0$, for any θ such that $||\theta - \theta^*|| > \epsilon$, then $||G(\theta)||_2 > \eta$. So, for any $\epsilon > 0$, $P(||\hat{\theta}_n - \theta^*|| > \epsilon) \leq P(||G(\hat{\theta};\tau^*)||_2 > \eta) \to_p 0$. Since $\hat{\theta}$ is just a re-parameterization of $\hat{c}$, then we have shown that $\hat{c} = c^* + o_p(1)$.

Lemma 8. *Under the Assumptions (and notation) of Proposition 21,*

$$\begin{aligned}&P(\nu(X_i;\theta_n,\tau^*) \neq \nu(X_i;\theta_0,\tau^*)) = O(||\theta_n - \theta_0||_2),\\ &\sup_{\theta\in\Theta} P(\nu(X_i;\theta,\hat{\tau}) \neq \nu(X_i;\theta,\tau^*)) = O(||\hat{\tau}^z(X_i) - \tau^{*,z}(X_i)||_2 + |\tau^*(X_i) - \hat{\tau}(X_i)|).\end{aligned}$$

Proof. For $P(\nu(X_i;\theta_n,\tau^*) \neq \nu(X_i;\theta_0,\tau^*))$ we divide into two cases. In the first case, $a_0 < 1$. We can choose large enough n (to ensure that $a_n \neq 1$) so that

$$\begin{aligned}&P(\nu(X_i;\theta_n,\tau^*) \neq \nu(X_i;\theta_0;\tau^*))\\ &= P(\min\{f(a_0)v_0^\top\tau^{*,z}(X_i), f(a_n)v_n^\top\tau^{*,z}(X_i)\} < \tau^*(X_i) < \max\{f(a_0)v_0^\top\tau^{*,z}(X_i), f(a_n)v_n^\top\tau^{*,z}(X_i)\})\\ &\leq P\Big(|\tau^*(X_i) - f(a_0)v_0^\top\tau^{*,z}(X_i)| < |(f(a_n)v_n - f(a_0)v_0)\tau^{*,z}(X_i)|\Big)\\ &\leq P\Big(|\tau^*(X_i) - f(a_0)v_0^\top\tau^{*,z}(X_i)| < |f(a_n)v_n - f(a_0)v_0|M\Big)\\ &\leq C||\theta_n - \theta_0||_2,\end{aligned}$$

for finite C, where M is finite since net demand is bounded. The last step is because $\tau^*(X_i) - av^\top\tau^{*,z}(X_i)$ has bounded density, so its distribution function is Lipschitz, and $f(\cdot)$ is continuous. Next, for $a_0 = 1$, we can prove something similar. We can choose large enough (to ensure that $a_n \neq 0$) so that

$$\begin{aligned}&P(\nu(X_i;\theta_n,\tau^*) \neq \nu(X_i;\theta_0;\tau^*))\\ &= P(\min\{0, (v_0^\top - v_n^\top)\tau^{*,z}(X_i) + \tau^*(X_i)/a_n\} < v_0^\top\tau^{*,z}(X_i) < \max\{0, (v_0^\top - v_n^\top)\tau^{*,z}(X_i) + \tau^*(X_i)/a_n\})\\ &\leq P\Big(|v_0^\top\tau^{*,z}(X_i)| < |(v_0^\top - v_n^\top)\tau^{*,z}(X_i) + \tau^*(X_i)/f(a_n)|)\\ &\leq P\Big(|v_0^\top\tau^{*,z}(X_i)| < |(v_0^\top - v_n^\top)M + M/f(a_n)|)\\ &\leq C||\theta_n - \theta_0||_2\end{aligned}$$

for some finite C, where M is finite by boundedness of outcomes and net demand. The last step is because $v^\top\tau^{*,z}(X_i)$ has bounded density, so its distribution function is Lipschitz, and $f(\cdot)$ is continuous.

For the second part of the Lemma, also split into two cases. First, when $0 \leq f(a) \leq 2$ (denote the product of this restricted space for a and S^{J-1} as Θ^+), let $b_n = av^\top(\hat{\tau}^z(X_i) - \tau^{*,z}(X_i)) +$

$\tau^*(X_i) - \hat{\tau}(X_i)$.

$$
\begin{aligned}
&\sup_{\theta\in\Theta^+} P(\nu(X_i;\theta,\hat{\tau}) \neq \nu(X_i;\theta,\tau^*)) \\
&= P(\min\{b_n, 0\} < \tau^*(X_i) - f(a)v^\top \tau^{*,z}(X_i) < \max\{b_n, 0\}) \\
&\leq \sup_{\theta\in\Theta^+} P(|\tau^*(X_i) - f(a)v^\top \tau^{*,z}(X_i)| \leq |f(a)v^\top(\hat{\tau}^z(X_i) - \tau^{*,z}(X_i)) + \tau^*(X_i) - \hat{\tau}(X_i)|) \\
&\leq \sup_{\theta\in\Theta^+} P(|\tau^*(X_i) - f(a)v^\top \tau^{*,z}(X_i)| \leq 2||\hat{\tau}^z(X_i) - \tau^{*,z}(X_i)||_2 + |\tau^*(X_i) - \hat{\tau}(X_i)|) \\
&= O(||\hat{\tau}^z(X_i) - \tau^{*,z}(X_i)||_2 + |\tau^*(X_i) - \hat{\tau}(X_i)|)
\end{aligned}
$$

(1) is because $\tau^*(X_i) - av^\top \tau^{*,z}(X_i)$ has a density uniformly bounded over a and v. This means that the distribution function of $\tau^*(X_i) - av^\top \tau^{*,z}(X_i)$, which we can call $F(t)$ is Lipschitz in t, so

$$
F(2||\hat{\tau}^z(X_i) - \tau^{*,z}(X_i)||_2 + |\tau^*(X_i) - \hat{\tau}(X_i)|) - F(0) \leq 2M||\hat{\tau}^z(X_i) - \tau^{*,z}(X_i)||_2 + M|\tau^*(X_i) - \hat{\tau}(X_i)|,
$$

where M does not depend on a or v. When $a \geq 2$, we can argue similarly, where we use Θ^- to denote the product of this restricted space and S^{J-1}. Let $c_n = v^\top(\tau^{*,z}(X_i) - \hat{\tau}^{*,z}(X_i)) + 1/f(a)(\hat{\tau}^*(X_i) - \tau^*(X_i))$.

$$
\begin{aligned}
&\sup_{\theta\in\Theta^-} P(\nu(X_i;\theta,\hat{\tau}) \neq \nu(X_i;\theta,\tau^*)) \\
&= P(\min\{c_n, 0\} < av^\top \tau^{*,z}(X_i) - 1/a\tau^*(X_i) < \max\{c_n, 0\}) \\
&\leq \sup_{\theta\in\Theta^-} P(v^\top \tau^{*,z}(X_i) - (1/f(a))\tau^*(X_i) \leq ||\hat{\tau}^z(X_i) - \tau^{*,z}(X_i)||_2 + |\tau^*(X_i) - \hat{\tau}(X_i)|) \\
&\leq O(||\hat{\tau}^z(X_i) - \tau^{*,z}(X_i)||_2 + |\tau^*(X_i) - \hat{\tau}(X_i)|)
\end{aligned}
$$

Where for the last step this is because under our assumptions, for any $f(a) \in [2, \infty]$ the density of $v^\top \tau^{*,z}(X_i) - 1/f(a)\tau^*(X_i)$ is uniformly bounded (the bound does not depend on v or a). □

D.2 Market-Clearing

Proposition 5. ***Single-Good Market.*** *In a market with $J = 1$ goods, assume that outcomes and net demand functions are bounded, and for all $p \in \mathcal{S}$ and $w \in \{0, 1\}$, $\mathbb{P}(Z_i(w, p)$ is continuous at $p) = 1$ and $\mathbb{P}(Y_i(w, p)$ is continuous at $p) = 1$. In addition, assume that*

- *$\mathbb{E}[Z_i(w, 0)] > 0$ and there exists a finite $b > 0$ such that $\mathbb{E}[Z_i(w, b)] < 0$ for $w \in \{0, 1\}$.*
- *With probability 1, $Z_i(w, p)$ is monotonically non-increasing in p for $w \in \{0, 1\}$.*

Let $Z_n(\boldsymbol{w}, p) = \frac{1}{n}\sum_{i=1}^{n} Z_i(w_i, p)$ and $P_n(\boldsymbol{w}) = \arg\min_p |Z_n(\boldsymbol{w}, p)|$ for $\boldsymbol{w} \in \{0, 1\}^n$. In cases where there are multiple minimizers, choose one by some deterministic rule.

Then, Assumption 2 holds, with $U_i = 0$. That is, there exist a sequence a_n with $\lim_{n\to\infty} a_n \sqrt{n} = 0$ and a constant $c > 0$ such that, given any treatment vector $\mathbf{w} \in \{0, 1\}^n$,

$$
\mathcal{S}_{\mathbf{w}} = \left\{ p \in \mathbb{R}^J : \left\| \frac{1}{n} \sum_{i=1}^{n} Z_i(w_i, p) \right\|_2 \leq a_n \right\}. \tag{92}
$$

is non-empty with probability $1 - e^{-c_1 n}$. *And* $P_n(\boldsymbol{w}) \in \mathcal{S}_{\mathbf{w}}$ *when it is non-empty.*

Proof. Denote by $\mathcal{E}_n$ the event that the realized $Z_n(\boldsymbol{w}, p)$ crosses zero during the interval $(0-s, b+s)$:

$$\mathcal{E}_n = \{Z_n(0) > 0, \text{ and } Z_n(b) < 0\}, \tag{93}$$

and denote by E_n the indicator variable for $\mathcal{E}_n$: $E_n = \mathbb{I}(\mathcal{E}_n)$.

Suppose that $E_n = 1$. Here, either $Z_n(P_n) = 0$, or $|Z_n(P_n)| \neq 0$, and we will focus on the latter case. Suppose $|Z_n(P_n)| \neq 0$. Because we condition on $E_n = 1$, we have that $Z_n(0) > 0$ and $Z_n(b) < 0$. Since P_n is chosen to be a minimizer of $|Z_n(p)|$, the only possibility is that the function $Z_n(\cdot)$ crosses zero at a point of discontinuity, P', that is, $Z_n(\tilde{p}) < 0$ for all $\tilde{p} > P'$, and $Z_n(\bar{p}) > 0$ for all $\bar{p} < P'$.

We have assumed that $\mathbb{P}(Z_i(W_i, \cdot)$ is discontinuous at $p) = 0$ for any $p > 0$, the Z_i's are drawn independently, and by the boundedness assumption they are bounded in magnitude by M. These facts further imply that with probability one, any jump in $Z_n(\cdot)$ cannot exceed magnitude M/n, for otherwise it would have required at least two separate Z_i to have a point of discontinuity at the same exact location. Formally, conditional on $E_n = 1$, we have that with probability one:

$$|Z_n(P_n)| \leq |\lim_{p \uparrow P'} Z_n(p) - \lim_{p \downarrow P'} Z_n(p)| \leq M/n, \tag{94}$$

Choosing $a_n = M/n$, we note that $\lim_{n\to\infty} a_n \sqrt{n} = 0$, and that we have that the specified $P_n(\boldsymbol{w}) \in \mathcal{S}_{\mathbf{w}}$ whenever $\mathcal{S}_{\mathbf{w}}$ is non-empty. It is non-empty whenever $E_n = 1$, so to finish the proof we now check the probability that $E_n = 1$.

If $E_n = 0$, there are four possible cases. The first two cases are $Z_n(0) = 0$ or $Z_n(b) = 0$, in which case $Z_n(P_n) = 0$. The last two are $Z_n(p) > 0$ for all $p \in \mathcal{S}$, or $Z_n(p) < 0$ for all $p \in \mathcal{S}$, in which case $Z_n(P_n) \leq c_1$ for some constant $c_1 > 0$, since each net demand function is bounded. Note that c_1 is not less than the specified a_n, so to verify the proposition we need that $\mathbb{P}(E_n = 0)$ is exponentially small. To bound $\mathbb{P}(E_n = 0)$,

$$\begin{aligned}
\mathbb{P}(E_n = 0) &\leq \mathbb{P}(Z_n(0) \leq 0) + \mathbb{P}(Z_n(b) \geq 0) \\
&= \mathbb{P}(Z_n(0) - \mathbb{E}[Z_n(0)] \leq -\mathbb{E}[Z_n(0)]) + \mathbb{P}(Z_n(b) - \mathbb{E}[Z_n(b)] \geq -\mathbb{E}[Z_n(b)]) \\
&\leq \mathbb{P}(|Z_n(0) - \mathbb{E}[Z_n(0)]| \geq \mathbb{E}[Z_n(0)]) + \mathbb{P}(|Z_n(b) - \mathbb{E}[Z_n(b)]| \geq -\mathbb{E}[Z_n(b)]) \\
&\leq \mathbb{P}(|Z_n(0) - \mathbb{E}[Z_n(0)]| \geq \varepsilon) + \mathbb{P}(|Z_n(b) - \mathbb{E}[Z_n(b)]| \geq \varepsilon),
\end{aligned}$$

for $\varepsilon = \min\{\mathbb{E}[Z_n(0), -\mathbb{E}[Z_n(c)]\} > 0$. We can now use Hoeffding's inequality, and obtain that for any n and $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}(E_n = 0) &\leq 2 \cdot \exp\left(\frac{-2n\varepsilon^2}{4c_1^2}\right), \\
&= 2 \cdot \exp\left(\frac{-n\varepsilon^2}{2c_1^2}\right).
\end{aligned} \tag{95}$$

We have now verified that $\mathcal{S}_{\mathbf{w}}$ is non-empty with probability $1 - e^{-c_1 n}$, for some $c_1 > 0$, which completes the proof.

□

D.3 Tighter Conservative Bound for $\bar{\sigma}_D^2$

First, introduce the notation $a_i(1) = \epsilon_i(1) - \pi\Delta_i(1, p_\pi^*)$, and $a_i(0) = \epsilon_i(0) + (1-\pi)\Delta_i(0, p_\pi^*)$. Under the augmented randomized experiment, an estimator is available for the following variance.

$$\tilde{\sigma}_D^2 = \mathbb{E}\left[\frac{(1-\pi)}{\pi}a_i^2(1) + \frac{\pi}{1-\pi}a_i^2(0) + 2\mathbb{E}[a_i(1)^2]^{1/2}\mathbb{E}[a_i(0)^2]^{1/2}\right].$$

In an extension of the classical results in Neyman (1923), and described in detail in Section 2.1 of Aronow et al. (2014), we can show that this variance $\tilde{\sigma}_D^2$ is conservative for $\bar{\sigma}_D^2$ and is tighter than σ_D^2.

$$\begin{aligned}
\sigma_D^2 &= \mathbb{E}\left[\left(\frac{W_i\varepsilon_i(1)}{\pi} - \frac{(1-W_i)\varepsilon_i(0)}{1-\pi} - \Delta_i(W_i, p_\pi^*)\right)^2\right], \\
&= \mathbb{E}\left[\left(\frac{W_i a_i(1)}{\pi} - \frac{(1-W_i)a_i(0)}{1-\pi}\right)^2\right], \\
&= \mathbb{E}\left[\frac{a_i(1)^2}{\pi}\right] + \mathbb{E}\left[\frac{a_i(0)^2}{1-\pi}\right], \\
&= \mathbb{E}\left[\left(\frac{(1-\pi)a_i^2(1)}{\pi} + \frac{\pi a_i^2(0)}{1-\pi} + a_i^2(1) + a_i^2(0)\right)\right], \\
&\geq \mathbb{E}\left[\left(\frac{(1-\pi)a_i^2(1)}{\pi} + \frac{\pi a_i^2(0)}{1-\pi} + 2\mathbb{E}[a_i(1)^2]^{1/2}\mathbb{E}[a_i(0)^2]^{1/2}\right)\right], \\
&= \tilde{\sigma}_D^2,
\end{aligned}$$

where the inequality is from the AM-GM inequality. Additionally, to show this is conservative for $\bar{\sigma}_D^2$

$$\begin{aligned}
\bar{\sigma}_D^2 &= \pi(1-\pi)\,\mathbb{E}\left[\left(\frac{\varepsilon_i(1)}{\pi} + \frac{\varepsilon_i(0)}{1-\pi} - (\Delta_i(1, p_\pi^*) - \Delta_i(0, p_\pi^*))\right)^2\right], \\
&= \pi(1-\pi)\,\mathbb{E}\left[\left(\frac{a_i(1)}{\pi} + \frac{a_i(0)}{1-\pi}\right)^2\right], \\
&= \mathbb{E}\left[\frac{(1-\pi)a_i^2(1)}{\pi} + \frac{\pi a_i^2(0)}{1-\pi} + 2a_i(1)a_i(0)\right], \\
&\leq \mathbb{E}\left[\frac{(1-\pi)a_i^2(1)}{\pi} + \frac{\pi a_i^2(0)}{1-\pi} + 2\sqrt{\mathbb{E}[a_i^2(1)]\mathbb{E}[a_i^2(0)]}\right], \\
&= \tilde{\sigma}_D^2,
\end{aligned}$$

where the inequality is from the Cauchy-Schwarz inequality.

D.4 Convergence Rate of $\tau_{\textbf{AIE}}$ to $\tau_{\textbf{AIE}}^*$

In this section, we provide a simple example where $Z_i(w,p)$ and $Y_i(w,p)$ are differentiable in p. In this example, the asymptotic representation of $\tau_{\text{AIE}} - \tau_{\text{AIE}}^*$ depends on $\frac{1}{n}\sum_{i=1}^{n} \nabla_p Y_i(w,p) - y(w,p)$, and it is straightforward to extend this example so that it also depends on the concentration of $\frac{1}{n}\sum_{i=1}^{n} \nabla_p Z_i(w,p)$ around its expected derivative. This suggests that it is not possible to get $\sqrt{n}$ converge of τ_{AIE} to τ^*_{AIE} under Assumption 6, which allows for discontinuous individual-level demand and outcome functions.

Define bounded random variables $(Y_i(1), Y_i(0), Z_i(1), \beta_i)$ that are sampled IID from some distribution. $Z_i(0) = 0$ always. This distribution is such that $\mathbb{E}[Z_i(1)] = z(1)$, $\mathbb{E}[Y_i(1) - Y_i(0)] = y(1) - y(0)]$ and $\mathbb{E}[\beta_i] = \beta$. W_i is drawn IID from Bernoulli(π). The rest of the data-generating process is:

$$Z_i(w,p) = \theta_0 - p + Z_i(W_i), \qquad Y_i(w,p) = -\beta_i W_i p + Y_i(W_i).$$

In this model, $p^*_\pi = \theta_0 + \pi \cdot z(1)$. The population AIE is

$$\tau^*_{\text{AIE}} = -\pi \cdot \beta \cdot z(1).$$

This model implies that $P_n(W_i = 1; \boldsymbol{W}_{-i}) - P_n(W_i = 0; \boldsymbol{W}_{-i}) = \frac{1}{n} Z_i(1)$ and $Y_j(W_j, P_n(W_j = 1; \boldsymbol{W}_{-i}) - Y_j(W_j, P_n(W_j = 0; \boldsymbol{W}_{-i}) = -\beta_j W_j \cdot \frac{1}{n} Z_i(1)$.

This implies that τ_{AIE} in this model is:

$$\begin{aligned}
\tau_{\text{AIE}} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} \pi \cdot \beta_j \frac{1}{n}(Z_i(1) - Z_i(0)) \\
&= \frac{1}{n}\sum_{i=1}^{n} Z_i(1) \frac{1}{n}\sum_{j=1}^{n} \pi \cdot \beta_j \\
&= \tau^*_{\text{AIE}} + \frac{1}{n}\sum_{i=1}^{n} Z_i(1) \frac{1}{n}\sum_{j=1}^{n} \pi \cdot (\beta_j - \beta) + \frac{1}{n}\sum_{i=1}^{n}(Z_i(1) - Z_i(0)) - z(1) - z(0))\pi \cdot \beta \\
&\overset{(1)}{=} \tau^*_{\text{AIE}} + z(1)\frac{1}{n}\sum_{j=1}^{n} \pi \cdot (\beta_j - \beta) + \frac{1}{n}\sum_{i=1}^{n}(Z_i(1) - z(1))\pi \cdot \beta + o_p(n^{-1/2}).
\end{aligned}$$

The last line is by the application of the LLN and recognizing that the CLT applies to the term $\frac{1}{n}\sum_{j=1}^{n} \pi \cdot (\beta_j - \beta)$. In this case there is $\sqrt{n}$ convergence of τ_{AIE} to τ^*_{AIE}, but the distribution depends on the variance of $\nabla_p Y_j(1, p^*_\pi)$. In a more general derivation with non-differentiable Y_i, the analogous term would not converge at a $\sqrt{n}$ rate, so the scaling of τ_{AIE} to τ^*_{AIE} will be slower than $\sqrt{n}$ under the assumptions in the paper (just as the convergence of $\hat{\tau}_{\text{AIE}}$ to τ^*_{AIE} is slower than $\sqrt{n}$).

D.5 Observing a Sub-Sample of the Market

The estimators for τ^*_{ADE} and τ^*_{AIE} are unchanged when a sub-sample of the n agents in the finite-sample market are observed rather than all of them. However, the expansion for $\hat{\tau}_{\text{ADE}}$ described in Theorem 5 is affected slightly, which changes the asymptotic variance. This appendix provides some detail on these claims which were made in the main text, under the assumptions of Theorem 5.

Let n be the size of the finite-sample market. Let $\mathcal{M} \subset \{1, \ldots, n\}$ index the m total number of observations sampled without replacement from the finite-sample market. We assume that the market price clears the finite-sample market, but we only observe Y_j and Z_j for $j \in \mathcal{M}$. Let

$\hat{s} = \frac{1}{m}\sum_{j\in M} W_j$ be the fraction of individuals treated in the sub-sample, $\hat{t} = \frac{m}{n}$ be the fraction of individuals observed, and $\hat{\pi} = \frac{1}{n}\sum_{i=1}^{n} W_i$ be the fraction of individuals treated in the entire finite market. We assume that for all agents that are not in $\mathcal{M}$, then $W_i = 0$. The sub-sample is a fixed fraction of the finite market, so that as $n \to \infty$, $\hat{s} \to s > 0$, $\hat{\pi} \to \pi > 0$ and $\hat{t} \to t > 0$.

The estimators $\hat{\tau}_{\text{ADE}}$ and $\hat{\tau}_{\text{AIE}}$ are exactly the same as in the main text, except the observations used are in $\mathcal{M}$ rather than in $\{1, \dots n\}$. Let h_n remain the same as in the main text, where the shrinkage rate of the price perturbations depend on n.

$$\hat{\tau}_{\text{m, ADE}} = \frac{1}{m}\sum_{j\in\mathcal{M}}\left[\frac{W_jY_j}{\hat{s}} - \frac{(1-W_j)Y_j}{1-\hat{s}}\right]$$
$$\hat{\tau}_{\text{m, AIE}} = -\hat{\gamma}^\top\frac{1}{m}\sum_{j\in\mathcal{M}}\left[\frac{W_jZ_j}{\hat{s}} - \frac{(1-W_j)Z_j}{1-\hat{s}}\right].$$

Let $\boldsymbol{Y}$ be the m-length vector of observed outcomes, $\boldsymbol{U}$ is the $m \times J$ matrix of observed price perturbations, and $\boldsymbol{Z}$ is the $m \times J$ matrix of observed net demand. Then, $\hat{\gamma} = (\boldsymbol{U}^\top\boldsymbol{Z})(\boldsymbol{U}^\top\boldsymbol{Y})$.

The asymptotic expansion for $\hat{\tau}_{m,\text{AIE}}$ is unchanged, apart from depending on m rather than n, since the variance of $P_n - p^*_\pi$ does not impact the limiting distribution.

$$\hat{\tau}_{m,\text{AIE}} = \tau^*_{\text{AIE}} - Q_z^\top\frac{1}{\sqrt{m}h_n^2}\sum_{j\in\mathcal{M}} U_j\nu_j(W_j) + o_p(1),$$

where $\nu_j(W_j) = Y_j(W_j, p^*_\pi) - Z_j(W_j, p^*_\pi)^\top[\xi_z^{-1}]^\top\xi_y$ and $Q_z = \xi_z^{-1}\tau^{*,z}_{\text{ADE}}$. The asymptotic variance can still be estimated by $\hat{\sigma}_I$, where each component is estimated using the observed data only.

The asymptotic expansion for $\hat{\tau}_{\text{ADE}}$ is affected, since $P_n - p^*_\pi$ does impact the asymptotic expansion, and the finite sample market price depends on n rather than m.

$$\begin{aligned}
\hat{\tau}_{m,\text{ADE}} &\overset{(1)}{=} \tau^*_{\text{ADE}} + \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\mathbb{1}(i\in\mathcal{M})}{\hat{t}}\left(\frac{W_i\varepsilon_i(1)}{\hat{s}} - \frac{(1-W_i)\varepsilon_i(0)}{1-\hat{s}}\right) - \Delta_i(W_i, p^*_\pi)\right] + o_p(n^{-0.5}) \\
&= \tau^*_{\text{ADE}} + \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\mathbb{1}(i\in\mathcal{M})}{t}\left(\frac{W_i\varepsilon_i(1)}{s} - \frac{(1-W_i)\varepsilon_i(0)}{1-s}\right) - \Delta_i(W_i, p^*_\pi)\right] + o_p(n^{-0.5}) \\
&\overset{(2)}{=} \tau^*_{\text{ADE}} + \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}(i\in\mathcal{M})}{t}\left(\frac{W_i\varepsilon_i(1)}{s} - \frac{(1-W_i)\varepsilon_i(0)}{1-s} - t\Delta_i(W_i, p^*_\pi)\right) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(i\notin\mathcal{M})\Delta_i(W_i, p^*_\pi) + o_p(n^{-0.5})
\end{aligned}$$

(1) comes from following the same steps as in the proof of Theorem 5. (2) comes from splitting up the sum for observations in $\mathcal{M}$ and those not in $\mathcal{M}$. As before $\varepsilon_i(w) = Y_i(w, p^*_\pi) - y(w, p^*_\pi)$.

The CLT now applies to this expansion:

$$\begin{aligned}
&\sqrt{n}(\hat{\tau}_{m,\text{ADE}} - \tau^*_{\text{ADE}}) \Rightarrow \mathcal{N}(0, \sigma^2_{t,D}) \\
&\sigma^2_{t,D} = \frac{1}{t}\mathbb{E}\left[\left(\frac{W_i\varepsilon_i(1)}{s} - \frac{(1-W_i)\varepsilon_i(0)}{(1-s)} - t\Delta_i(W_i, p^*_\pi)\right)^2\right] + (1-t)\mathbb{E}\left[\Delta_i(W_i, p^*_\pi)^2\right]
\end{aligned}$$

For inference on τ^*_{ADE} using the sub-sample only, we can estimate each component of $\sigma^2_{t,D}$ using only observations in $\mathcal{M}$.

D.6 Alternative Estimator of AIE Variance

In analyzing the asymptotic properties of σ_I^2 in the proof of Theorem 7 in Appendix A.5, we dropped a term that was asymptotically negligible, which was $\sqrt{n}h_n(\xi_y\xi_z^{-1}(\hat{\tau}^z_{\text{ADE}} - \tau^{*,z}_{\text{ADE}}))$.

$\tilde{\sigma}_I^2$ adds a plug-in estimator for the second term, including an estimator for the variance of $\hat{\tau}^z_{\text{ADE}}$. $\tilde{\sigma}_I^2 = \hat{\sigma}_I^2 + h_n^2\hat{\gamma}^\top\hat{\sigma}^2_{z,D}\hat{\gamma}$, where $\hat{\sigma}^2_{z,D} = \frac{1}{n}\sum_{i=1}^n B_iB_i'$ and the $J \times 1$ vector $\hat{B}_i = \frac{W_i\hat{\varepsilon}^z_i(1)}{\hat{\pi}} - \frac{(1-W_i)\hat{\varepsilon}^z_i(0)}{1-\hat{\pi}} - [\hat{\xi}_{z1} - \hat{\xi}_{z0}]^\top\hat{\xi}_z^{-1}Z_i$. $\varepsilon^z_i(1) = Z_i - \frac{1}{n_w}\sum_{i:W_i=w} Z_i$. Last, the $J \times J$ matrix $\hat{\xi}_{zw}$ for $w \in \{0,1\}$ are estimated from regressions of Z_i on U_i using only observations such that $W_i = w$. We found in simulations that this second-order correction leads to better coverage properties at smaller sample sizes.