ONLINE APPENDIX TO ACCOMPANY

Stress-Testing Structural Models of Unobserved Heterogeneity: Robust Inference on Optimal Nonlinear Pricing,

by Aaron Bodoh-Creed, Brent Hickman, John List, Ian Muir, and Greg Sun

Appendix A. Proofs of Propositions 1 and 2 (Under Standard Assumptions)

In Appendix A.1, we complete the proof of Proposition 1 by modifying our construction of $\overline{\mathcal{Q}}_{d_0}(q)$ to allow for the presence of mass points and show that our modified DGP attains the bound corresponding to $\overline{\mathcal{B}}_{d_0}(q;S,d)$. In Appendix A.2, we complete the proof of Proposition 2 by constructing $\underline{\mathcal{Q}}_{d_0}(q)$ and showing that it attains the bound corresponding to $\underline{\mathcal{B}}_{d_0}(q;S,d)$ in the case where the CDF difference $G_{d_0}(q)-G_c(q)$ is quasi-concave, and both distributions are absolutely continuous. Proofs of the lower bound for the cases where the CDF difference is not quasi-concave and where mass points exist build on the basic ideas here, but involve tedious technical complications, so we defer them to Online Appendix B.

As in the main text, let the CDF of potential outcomes in control, Q_c be G_c and the CDF of potential outcomes under some discount d, Q_d, be denoted G_d. As in the body of the paper, we denote random variables by upper-case letters, while realizations of random variables (or fixed numbers) are denoted by lower-case. Additionally, define the quantile functions $G_c^{-1}(r)=\inf\{q : G_c(q)\geq r\}$ and $G_d^{-1}(r)=\inf\{q : G_d(q)\geq r\}$, and note that these may represent either the inverses of the CDFs, if they exist, or the quasi-inverses otherwise. Throughout this appendix, we maintain the assumption that the underlying data-generating process satisfies the Law of Demand (LoD), and that the econometrician has access to a dataset (G_c, G_{d_0}), where $G_c(q)$ is observed demand under default price p_0, and $G_{d_0}(q)$ is observed demand under a particular discounted price $p_0(1-d_0)$. The econometric challenge here essentially stems from the fact that the copula of the joint distribution of (Q_c, Q_{d_0}) is unknown. This is because for each consumer we only observe *either* Q_c *or* Q_{d_0} (but never both), and (G_c, G_{d_0}) were generated from two separate samples of consumers having similar distributions of unobserved taste characteristics.

A.1. **Proof of Proposition 1 in the Presence of Mass Points in G_c.** Recall from equation (5) that $\overline{\mathcal{Q}}_{d_0}(q,v;d)\equiv G_d^{-1}\left(a(q)+b(q)v\right)$, where V is an independent uniform random variable, $a(q)\equiv\lim_{q'\to q^-}G_c(q)$ is the mass of consumers with baseline demand strictly below q, and $b(q)\equiv G_c(q)–\lim_{q'\to q^-}G_c(q)$ is the size of the mass point at $Q_c=q$. Intuitively, V is a device for "breaking ties" in rank that arise when a positive mass of consumers have the same baseline demand q. When G_c is left-continuous at a particular q then $b(q)=0$, and the upper-bound DGP reduces to the simpler form $\overline{\mathcal{Q}}_{d_0}(q,v;d)=G_d^{-1}(G_c(q))$. We break up our proof into two steps as follows.

Lemma 1. *$\overline{\mathcal{Q}}_{d_0}(Q_c, V; d_0)$ as defined in Equation (5) is an admissible DGP that cannot be ruled out by the dataset (G_c, G_{d_0}); that is* $\Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) \leq q\right] = G_{d_0}(q)$.

Proof. At any q where the quantile function G_c^{-1} is strictly increasing at $G_{d_0}(q)$, we have that $\Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) \leq q\right] = \Pr[G_d^{-1}(G_c(Q_c)) \leq q] = \Pr[G_c(Q_c) \leq G_d(q)] = G_d(q)$ where the last equality follows because $G_c(Q_c)$ is a $Uniform(0,1)$ random variable. Otherwise, the random variable Q_c has a mass point at $q_q^* \equiv G_c^{-1}(G_{d_0}(q)))$. By definition of a, $a(q_q^*) = \Pr[Q_c < q_q^*]$, so

$$\begin{aligned}
\Pr[\overline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) \leq q] &= \Pr\left[Q_c < q_q^* \text{ or } \left(Q_c = q_q^* \text{ and } V \leq \frac{G_{d_0}(q) - a(q_q^*)}{b(q_q^*)}\right)\right] \\
&= \Pr[Q_c < q_q^*] + \Pr\left[Q_c = q_q^* \text{ and } V \leq \frac{G_{d_0}(q) - a(q_q^*)}{b(q_q^*)}\right] \\
&= a(q_q^*) + \Pr[Q_c = q_q^*] \times \Pr\left[V \leq \frac{G_{d_0}(q) - a(q^*)}{b(q_q^*)}\right] \\
&= a(q_q^*) + b(q_q^*) \frac{G_{d_0}(q) - a(q_q^*)}{b(q_q^*)} = G_{d_0}(q). \qquad \square
\end{aligned} \tag{15}$$

Lemma 2. $\Pr[\overline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) \leq q | Q_c \geq q']$ *constitutes an upper bound (in the first-order dominance sense) on Strong Uptaker Distributions; that is,* $\Pr[\overline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) \leq q | Q_c \geq q'] \leq \Pr[Q_{d_0} \leq q | Q_c \geq q']$.

Proof. Consider the joint distribution of (Q_c, Q_d) with marginal distributions G_c and G_d. The upper bound property is equivalent to $\Pr\left[Q_d \leq q | Q_c \geq \frac{S}{p_0 \times d}\right] < \Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c; d, v) \leq q | Q_c \geq \frac{S}{p_0 \times d}\right]$ being impossible. Suppose for a contradiction that there exists a baseline consumption level $q' = \frac{S}{p_0 \times d}$ (under price p_0), and a counterfactual consumption level q (under price $p_0(1-d)$) satisfying this inequality. In that case,

$$\begin{aligned}
&\Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c; d, V) \leq q | Q_c \geq q'\right] \Pr[Q_c \geq q'] + \Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c; d, V) \leq q | Q_c < q'\right] \Pr[Q_c < q'] \\
&= \Pr[\overline{\mathcal{Q}}_{d_0}(Q_c; d, V) \leq q] = \Pr[Q_d \leq q] \\
&= \Pr\left[Q_d \leq q | Q_c \geq q'\right] \Pr[Q_c \geq q'] + \Pr\left[Q_d \leq q | Q_c < q'\right] \Pr[Q_c < q'], \\
\Rightarrow &\Pr[Q_c < q'] \left(\Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c; d, V) \leq q | Q_c < q'\right] - \Pr\left[Q_d \leq q | Q_c < q'\right]\right) \\
&\qquad = \Pr[Q_c \geq q'] \left(\Pr\left[Q_d \leq q | Q_c \geq q'\right] - \Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c; d, V) \leq q | Q_c \geq q'\right]\right),
\end{aligned} \tag{16}$$

where the first and third equalities follow from the law of total probability and the second follows from (15). Our supposition $\Pr\left[Q_d \leq q | Q_c \geq q'\right] < \Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c; d, V) \leq q | Q_c \geq q'\right]$ is thus equivalent to

$$\Pr\left[Q_d \leq q | Q_c < q'\right] > \Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c; d, V) \leq q | Q_c < q'\right], \tag{17}$$

since the last two lines of (16) have the same sign. By definition, $\overline{\mathcal{Q}}_{d_0}$ is non-decreasing; as a result, inequality (17) implies $\overline{\mathcal{Q}}_{d_0}(q'; d, v) = \overline{\mathcal{Q}}_{d_0}\left(\frac{S}{p_0 \times d}; d, v\right) > q$. To see why, note that if $\overline{\mathcal{Q}}_{d_0}(q'; d, v) \leq q$, then the conditioning event $Q_c \leq q'$ implies $\overline{\mathcal{Q}}_{d_0}(Q_c; d, v) \leq q$ as well, by monotonicity of $\overline{\mathcal{Q}}_{d_0}$. This in turn implies $\Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c; d, v) \leq q | Q_c \leq \frac{S}{p_0 \times d}\right] = 1$. But this would violate

(17) since $\Pr\left[Q_d \leq q | Q_c < \frac{S}{p_0 \times d}\right]$ cannot exceed 1. In other words, (17) requires that counterfactual consumption implied by the baseline consumption level $q' = \frac{S}{p_0 \times d}$ must weakly exceed the benchmark q. Furthermore,

$$\begin{aligned} G_d(q) = \Pr\left[Q_d \leq q\right] &= \Pr\left[Q_d \leq q | Q_c < q'\right] \Pr\left[Q_c < q'\right] + \Pr\left[Q_d \leq q, Q_c \geq q'\right] \\ &> \Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c; d, V) \leq q | Q_c < q'\right] \Pr\left[Q_c < q'\right] + \Pr\left[Q_d \leq q, Q_c \geq q'\right], \end{aligned} \tag{18}$$

where the equality follows from the law of total probability and the inequality follows from (17). This last expression can be re-written as

$$\begin{aligned} &\Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c; d, V) \leq q | Q_c < q'\right] \Pr\left[Q_c < q'\right] + \Pr\left[Q_d \leq q, Q_c \geq q'\right] \\ &= \Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c; d, V) \leq q, Q_c < q'\right] + \Pr\left[Q_d \leq q, Q_c \geq q'\right] \\ &= \Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c; d, V) \leq q\right] + \Pr\left[Q_d \leq q, Q_c \geq q'\right] \\ &= G_d(q) + \Pr\left[Q_d \leq q, Q_c \geq q'\right] \geq G_d(q). \end{aligned} \tag{19}$$

The second equality follows because if $\overline{\mathcal{Q}}_{d_0}(Q_c; d, v) \leq q$ then the event $Q_c \geq q' = \frac{S}{p_0 \times d}$ is impossible since otherwise $\overline{\mathcal{Q}}_{d_0}(Q_c; d, v) \geq \overline{\mathcal{Q}}_{d_0}(q'; d, v) > q$. Note, however, that (18) and (19) imply that $G_d(q)$ is strictly greater than itself, a contradiction. □

Taken together, Lemmas 1 and 2 imply that $\overline{\mathcal{Q}}_{d_0}$ is a sharp upper bound on the range of DGPs consistent with the dataset (G_c, G_{d_0}). Therefore $\overline{\mathcal{B}}_{d_0}(q; S, d_0) \equiv \Pr\left[\overline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) \leq q | Q_c \geq q'\right]$ in turn constitutes a sharp upper bound (in the first-order dominance sense) on strong uptaker distributions, or in other words, $\overline{\mathcal{B}}_{d_0}(q; S, d_0) \leq \Pr[Q_{d_0} \leq q | Q_c \geq q']$.

More formally, the two lemmas show that the rank-stable DGP is in fact the (sharp) upper bound on the set of DGPs that cannot be ruled out by the dataset (G_c, G_{d_0}). In light of Assumption 4, we can furthermore use the in sample distributions (G_c, G_{d_0}) to construct out-of-sample distribution $\overline{G}_d^{oos}(q; G_c, G_{d_0})$ for each value of d under consideration. Then, the logic of Lemmas 1 and 2 goes through exactly as before, but with $\overline{G}_d^{oos}$ replacing G_{d_0} in the definitions of $\overline{\mathcal{Q}}_{d_0}(q, v; d)$ and $\overline{\mathcal{B}}_{d_0}(q; S, d)$ and in equations (18) and (19). ■

A.2. **Proof of Proposition 2.** We split our proof into four cases. In *Case 1*, we complete the proof in the special case considered in the main text where G_c–G_{d_0} is quasi-concave and G_c, G_{d_0} are absolutely continuous. In all empirical applications considered in the body of this paper, this case appears to be the most empirically relevant of the three. For generality, in Online Appendix B we also consider *Case 2* and *Case 3* as well, where we relax the quasi-concavity and absolute continuity requirements, respectively. Throughout this section, we will again be making extensive use of a tie-breaking random variable, V, which is independent of Q_c and distributed $\text{Unif}(0, 1)$. We also refer the reader to Figure 1 for intuition on our proof construction; Panel A plots a hypothetical dataset from a pricing RCT, including control CDF G_c and treatment CDF G_{d_0}.

A.2.1. *Case 1:* Since the CDF difference $G_c(q) - G_{d_0}(q)$ is unimodal, it is weakly increasing below its smallest maximizer, q^*_{min}, and weakly decreasing above its largest maximizer, q^*_{max}. Let q_{max} denote the largest value at which G_c and G_{d_0} disagree, and define $\bar{q}_{d_0}(q') = \inf\left\{q \in [q^*_{max}, q_{max}] : G_c(q') - G_{d_0}(q') = G_c(q) - G_{d_0}(q)\right\}$. We also define a quasi-inverse of $\bar{q}_{d_0}(q')$, which we denote by $\underline{q}_{d_0}(q) = \sup\left\{q' \in [0, q^*_{min}] : G_c(q') - G_{d_0}(q') = G_c(q) - G_{d_0}(q)\right\}$. When G_c and G_{d_0} are absolutely continuous, the infimum and supremum defined above are attained, so $G_c(q) - G_{d_0}(q) = G_c(\bar{q}_{d_0}(q)) - G_{d_0}(\bar{q}_{d_0}(q))$ and $G_c(q) - G_{d_0}(q) = G_c(\underline{q}_{d_0}(q)) - G_{d_0}(\underline{q}_{d_0}(q))$. We also note that because G_c and G_{d_0} have no mass points, $\bar{q}_{d_0}$ and $\underline{q}_{d_0}$ are strictly decreasing on their respective domains. Given these preliminaries, recall from equation (10) that the lower-bound DGP $\underline{\mathcal{Q}}_{d_0}$ is defined by

$$\underline{\mathcal{Q}}_{d_0}(q; d, v) = \begin{cases} \bar{q}_{d_0}(q) \text{ if } q \leq q^*_{min},\ v \leq \frac{g_c(q) - g_{d_0}(q)}{g_c(q)}, \text{ and} \\ \underline{\mathcal{Q}}_{d_0}(q; d, v) = q \text{ otherwise.} \end{cases} \tag{20}$$

This definition implies that $\underline{\mathcal{Q}}_{d_0}(Q_c)$ respects the LoD, meaning $\underline{\mathcal{Q}}_{d_0}(Q_c) \geq Q_c$. In what follows, we will often refer to the individuals for whom $Q_c \leq q^*_{min}$ and $V \leq \frac{g_c(Q_c) - g_{d_0}(Q_c)}{g_c(Q_c)}$ as "jumpers". The maximal proportion of jumpers that could be consistent with the data (G_c, G_{d_0}), conditional on Q_c, is given by the quantity $\frac{g_c(Q_c) - g_{d_0}(Q_c)}{g_c(Q_c)}$ and is visualized in Panel B of Figure 1.

Lemma 3. $\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0)$ *as defined in* (20) *is an admissible DGP that cannot be ruled out by the dataset,* (G_c, G_{d_0})*; that is* $\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) \leq q] = G_{d_0}(q)$.

Proof. By definition of $\underline{\mathcal{Q}}_{d_0}$, we have the following:

$$\Pr\left[\underline{\mathcal{Q}}_{d_0}(Q_c, V) \leq q\right] = \begin{cases} A(q) & q \leq q^*_{min}, \\ A(q^*_{min}) + B(q) & q^*_{min} < q < q^*_{max}, \text{ and} \\ A(q^*_{min}) + B(q^*_{max}) + C(q) & q \geq q^*_{max}, \end{cases} \tag{21}$$

where $A(q) = \int_0^q \left[1 - \frac{g_c(x) - g_{d_0}(x)}{g_c(x)}\right] g_c(x)\, \mathrm{d}x$ covers the case where q is below the smallest maximizer q^*_{min}, $B(q) = \int_{q^*_{min}}^q g_c(x)\, \mathrm{d}x$ covers the case where q is between the smallest and largest maximizers, and $C(q) = \int_{q^*_{max}}^q g_c(x)\, \mathrm{d}x + \int_{\underline{q}_{d_0}(q)}^{q^*_{min}} \frac{g_c(x) - g_{d_0}(x)}{g_c(x)} g_c(x)\, \mathrm{d}x$ covers the case where q is above the largest maximizer q^*_{max}. $A(q)$ corresponds to the probability that $\underline{\mathcal{Q}}_{d_0}(Q_c, V) \leq q$ resulted because $Q_c \leq q^*_{min}$ and the values of (Q_c, V) do not imply a jumper. $B(q)$ and the first term of $C(q)$ together correspond to the case where $Q_c \in [q^*_{min}, q]$ since conditioning on $Q_c \in [q^*_{min}, q]$ implies $\underline{\mathcal{Q}}_{d_0}(Q_c, V) = Q_c \leq q$ with probability 1. Finally, the second term of $C(q)$ corresponds to jumpers for whom $Q_c \in [\underline{q}_{d_0}(q), q^*_{min}]$, in which case, despite jumping, it is still true that $\underline{\mathcal{Q}}_{d_0}(Q_c, V) \leq q$.

The expression for A can be simplified as $A(q) = \int_0^q g_{d_0}(x)\, \mathrm{d}x = G_{d_0}(q)$. On the other hand, for $q^*_{min} < q < q^*_{max}$, $G_c(q) - G_{d_0}(q)$ is constant (by unimodality), which implies $g_c(q) - g_{d_0}(q) = 0$. Thus,

$B(q)$ can also be written $\int_{q^*_{min}}^{q} g_{d_0}(x)\,\mathrm{d}x$, so $B(q)=G_{d_0}(q)-G_{d_0}(q^*_{min})$. Finally, we also have

$$\int_{\underline{q}_{d_0}(q)}^{q^*_{min}} \frac{g_c(x)-g_{d_0}(x)}{g_c(x)} g_c(x)\,\mathrm{d}x = G_c(q^*_{min})-G_{d_0}(q^*_{min})-(G_c(\underline{q}_{d_0}(q))-G_{d_0}(\underline{q}_{d_0}(q))$$
$$=G_c(q^*_{min})-G_{d_0}(q^*_{min})-(G_c(q)-G_{d_0}(q)),$$

which implies $C(q)=G_{d_0}(q)-G_{d_0}(q^*_{max})$. Plugging these identities into (21) shows that regardless of the value of q, $\Pr\left[\underline{\mathcal{Q}}_{d_0}(Q_c)\le q\right]=G_{d_0}(q)$. $\square$

Lemma 4. $\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)\le q|Q_c\ge q']$ *constitutes a lower bound (in the first-order dominance sense) on Strong Uptaker Distributions; that is,* $\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)\le q|Q_c\ge q']\ge \Pr[Q_{d_0}\le q|Q_c\ge q']$.

Proof. We must show that the inequality $\Pr[Q_{d_0}\le q|q_c\ge q'] > \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)\le q|Q_c\ge q']$ is impossible for any (q,q') pair. Suppose then, for a contradiction, that it holds for some (q,q') pair. By similar logic as in equation (16), this is equivalent to $\Pr[Q_{d_0}\le q|q_c<q'] < \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)\le q|Q_c<q']$. Since $\Pr[Q_{d_0}\le q|q_c<q']+\Pr[Q_{d_0}>q|q_c<q'] = 1 = \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)\le q|Q_c< q']+\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)>q|Q_c<q']$, our supposition is further equivalent to

$$\Pr[Q_{d_0}>q|Q_c<q'] > \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)>q|Q_c<q'], \tag{22}$$

which we now show is impossible. To reduce notational clutter, we denote the RHS of (22) as $RHS(q,q')\equiv\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)>q|Q_c<q']$. We further split Lemma 4 into the following steps.

Step 1: *If pair* (q,q') *satisfies* (22) *then* $\underline{\mathcal{Q}}_{d_0}(q',0)<q$*:* We will construct a proof of Step 1 by contrapositive by splitting the analysis into two further sub-cases, but it will be useful to first note the following. When $q'<q^*_{min}$, $\underline{\mathcal{Q}}_{d_0}(q',0;d_0)=\bar{q}_{d_0}(q')$ can be interpreted as the minimum value of counterfactual consumption among jumpers for whom $Q_c<q'$, or $\underline{\mathcal{Q}}_{d_0}(q',0;d_0;d_0)=\inf\{\underline{\mathcal{Q}}_{d_0}(q',v) : \underline{\mathcal{Q}}_{d_0}(q',v;d_0)>q'\}$. On the other hand, if $q'\ge q^*_{min}$, then $\underline{\mathcal{Q}}_{d_0}(q',0;d_0)=q'$.

Case 1.1: $q\le q'$*:* The Law of Demand implies that if $Q_{d_0}<q$, then $Q_c<q'$ also. This implies that

$$G_{d_0}(q)=\Pr[Q_{d_0}\le q]=\Pr[Q_{d_0}\le q, Q_c<q']=\Pr[Q_{d_0}\le q|Q_c<q']\Pr[Q_c<q'].$$

Dividing the above equation by $\Pr[Q_c<q']=G_c(q')$ shows that $\Pr[Q_{d_0}\le q|Q_c<q']=\frac{G_{d_0}(q)}{G_c(q')}$. Thus, $\Pr[Q_{d_0}>q|Q_c<q'] = 1-\Pr[Q_{d_0}\le q|Q_c<q'] = \frac{G_c(q')-G_{d_0}(q)}{G_c(q')}$. This line of reasoning relied *only* on the LoD, and applies to the pair of random variables $(Q_c, \underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0))$ as well, since $\underline{\mathcal{Q}}_{d_0}$ was constructed to satisfy the LoD. Thus, $RHS(q,q')=\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)>q|Q_c<q']=\frac{G_c(q')-G_{d_0}(q)}{G_c(q')}=\Pr[Q_{d_0}>q|Q_c<q']$ which contradicts (22) for the case where $q\le q'$.

Case 1.2: $q'<q\le\underline{\mathcal{Q}}_{d_0}(q',0;d_0)$*:* In this case, $RHS(q,q')$ is constant in its first argument for all q on the closed interval $\left[q',\underline{\mathcal{Q}}_{d_0}(q',0;d_0)\right]$. To see why, note that conditional on the event $Q_c<q'$, we know by the definition of $\underline{\mathcal{Q}}_{d_0}$ that either the value of V is high, so $\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)=\underline{\mathcal{Q}}_{d_0}(Q_c,1;d_0)=Q_c<q'$; or the value of V is low, in which case, $\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)=\underline{\mathcal{Q}}_{d_0}(Q_c,0;d_0)=\bar{q}_{d_0}(Q_c)>\bar{q}_{d_0}(q')=$

$\underline{\mathcal{Q}}_{d_0}(q',0;d_0)$, where the inequality follows from monotonicity of $\bar{q}_{d_0}$. In either case, $\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)$ lies outside of $\left[q',\underline{\mathcal{Q}}_{d_0}(q',0;d_0)\right]$ with certainty, so for any q on that interval, we have

$$\begin{aligned}0\leq RHS(q',q')-RHS(q,q')&=\Pr\left[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)\in[q',q]|Q_c>q'\right]\\&\leq \Pr\left[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)\in[q',\underline{\mathcal{Q}}_{d_0}(q',0;d_0)]|Q_c>q'\right]=0,\end{aligned}$$

where the first inequality and equality are by definition of RHS, and the second inequality is implied by the supposition of Case 1.2. Moreover, the logic employed in Case 1.1 above establishes that if we replace q with q' in the first argument of RHS, then we have $RHS(q',q')=\Pr\left[Q_{d_0}>q'|Q_c<q'\right]$. Therefore, since $RHS(q,q')$ is constant in its first argument for all $q\in\left[q',\underline{\mathcal{Q}}_{d_0}(q',0,d_0)\right]$, we have $RHS(q,q')=RHS(q',q')=\Pr\left[Q_{d_0}>q'|Q_c<q'\right]\leq\Pr\left[Q_{d_0}>q|Q_c<q'\right]$, which contradicts (22).

The contradictions in Cases 1.1 and 1.2 demonstrate that inequality (22) can only be satisfied when $q'\leq\underline{\mathcal{Q}}_{d_0}(q',0)<q$. The next step shows that this case leads to a contradiction as well.

Step 2: *Inequality* (22) *leads to a contradiction when* $q'\leq\underline{\mathcal{Q}}_{d_0}(q',0)<q$:

Using similar logic as in equations (18) and (19), we have

$$\begin{aligned}1-G_{d_0}(q)=\Pr[Q_{d_0}>q]&=\Pr[Q_{d_0}>q|Q_c<q']\Pr[Q_c<q']+\Pr[Q_{d_0}>q|Q_c\geq q']\Pr[Q_c\geq q']\\&>\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)>q|Q_c<q']\Pr[Q_c<q']+\Pr[Q_{d_0}>q,Q_c\geq q']\\&=\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)>q,Q_c<q']+\Pr[Q_{d_0}\geq q,Q_c\geq q'],\end{aligned}\tag{23}$$

where the strict inequality follows directly from (22). We now analyze each of the two terms on the RHS of (23) (last line) in turn. For the first term, note that the events $\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)>q$ and $Q_c<q'$ can simultaneously occur if and only if (Q_c,V) is a jumper *and* $\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)=\underline{\mathcal{Q}}_{d_0}(Q_c,0;d_0)=\bar{q}_{d_0}(Q_c)>q$.[44] But by monotonicity of $\underline{q}_{d_0}$, this is equivalent to $Q_c<\underline{q}_{d_0}(q)$, since $Q_c\leq\underline{q}_{d_0}(\bar{q}_{d_0}(Q_c))$, by definition—recall that $\underline{q}_{d_0}$ and $\bar{q}_{d_0}$ are quasi-inverses of each other—and since $\underline{q}_{d_0}(\bar{q}_{d_0}(Q_c))<\underline{q}_{d_0}(q)$, which follows from $q_L(Q_c)$ being strictly greater than q. This further implies that the first term on the RHS of (23) satisfies

$$\begin{aligned}\Pr\left[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)>q,Q_c<q'\right]&=\Pr\left[(Q_c,V)\text{ is a jumper}\cap Q_c<\underline{q}_{d_0}(q)\right]\\&=\int_0^{\underline{q}_{d_0}(q)}\Pr\left[V\leq\frac{g_c(x)-g_{d_0}(x)}{g_c(x)}\right]g_c(x)\,\mathrm{d}x\\&=\int_0^{\underline{q}_{d_0}(q)}\frac{g_c(x)-g_{d_0}(x)}{g_c(x)}g_c(x)\,\mathrm{d}x\\&=G_c(\underline{q}_{d_0}(q))-G_{d_0}(\underline{q}_{d_0}(q))=G_c(q)-G_{d_0}(q).\end{aligned}$$

[44]Note that $Q_c>q$ is inconsistent with the second event of the joint probability for the case considered in Step 2, where $Q_c<q'<q$.

Next, turning to the second joint probability on the RHS of (23), we have

$$\begin{aligned}\Pr[Q_{d_0}\geq q, Q_c\geq q'] &= \Pr[Q_{d_0}\geq q, Q_c\geq q] + \Pr[Q_{d_0}\geq q, q>Q_c\geq q'] \\ &= \Pr[Q_c\geq q] + \Pr[Q_{d_0}\geq q, q>Q_c>q'] \geq 1-G_c(q)\end{aligned}$$

where the first equality follows from the law of total probability and the supposition of Step 2, and the second equality is true because $Q_c\geq q \Rightarrow Q_{d_0}\geq q$ by the LoD. As a result, the last line of (23) is greater than or equal to $G_c(q)-G_{d_0}(q)+(1-G_c(q)) = 1-G_{d_0}(q)$, which combined with the rest of inequality (23), leads to the contradiction that $1-G_{d_0}(q)>1-G_{d_0}(q)$. □

Together, Lemmas 3 and 4 imply that $\underline{\mathcal{Q}}_{d_0}$ is a sharp upper bound on the range of DGPs that cannot be ruled out by the dataset (G_c, G_{d_0}). Therefore, $\underline{\mathcal{B}}_{d_0}(q;S,d_0)\equiv \Pr\left[\underline{\mathcal{Q}}_{d_0}(Q_c,V;d_0)\leq q|Q_c\geq q'\right]$ constitutes a sharp upper bound (in the first-order dominance sense) on strong uptaker distributions, or in other words, $\underline{\mathcal{B}}_{d_0}(q;S,d_0)\geq \Pr[Q_{d_0}\leq q|Q_c\geq q']$. ■

A.3. **Out-of-Sample Inference Using Only the Law of Demand.** In this Appendix, we further discuss the extent to which out-of-sample inference is possible without Assumption 4 but still main. If one assumes only the LoD, but is uncomfortable adding additional structure such as that discussed in Remark 1, then for out-of-sample discounts, counterfactual CDFs can be sharply bounded only as follows:

Proposition 4. *Under Assumption 1 and for arbitrary (S,d) pairs, if $G_c(q)$ and $G_{d_0}(q)$ are known and are discontinuous at countably many mass points, then the following constitute (identified) sharp bounds on SUDs:*

$$\overline{\mathcal{B}}^{LoD}_{d_0}(q;S,d)\equiv\begin{cases}\overline{\mathcal{B}}_{d_0}\left(q;\frac{Sd_0}{d},d_0\right), & \text{if } d<d_0,\\ \overline{\mathcal{B}}_{d_0}(q;S,d_0), & \text{if } d=d_0,\\ 0, & \text{if } d>d_0;\end{cases}\qquad \underline{\mathcal{B}}^{LoD}_{d_0}(q;S,d)\equiv\begin{cases}G_c\left(q|Q_c\geq\frac{S}{p_0\times d}\right), & \text{if } d<d_0,\\ \underline{\mathcal{B}}_{d_0}(q;S,d_0), & \text{if } d=d_0,\\ \underline{\mathcal{B}}_{d_0}\left(q;\frac{Sd_0}{d},d_0\right), & \text{if } d>d_0.\end{cases} \tag{24}$$

For some brief intuition on the bounds in Equation (24), note that when an out-of-sample discount is less generous than d_0—i.e., $0<d<d_0$—the LoD implies that G_d lies somewhere between G_c and G_{d_0} but does not give any further information. The upper bound on DGPs that cannot be ruled out by existing data thus corresponds to the case where we make the most optimistic possible assumption about shifts in consumer demand, i.e., that $G_d=G_{d_0}$ for *any* $d\in(0,d_0)$. On the other hand, the lower bound corresponds to the case where we make the most pessimistic assumption that $G_d=G_c$ for any $d\in(0,d_0)$. Similarly, when an out-of-sample discount is more generous than d_0—i.e., $d_0<d$—the lower bound corresponds to making the most pessimistic possible assumption that $G_d=G_{d_0}$, while the upper bound is completely uninformative, since we know only that G_d is located to the right of G_{d_0}, and demand is otherwise unconstrained from above by the LoD.

Appendix B. Additional Proofs

B.1. **Case 2 (Proof of Propostition 2 with Violations of Quasi-Concavity):** In this section, we deal with violations of quasi-concavity of the CDF difference $G_c(q)-G_{d_0}(q)$ while maintaining the assumption that G_c and G_{d_0} are absolutely continuous with well-defined densities g_c and g_{d_0}. As in Appendix A.1, for expressions involving hypothetical values for both an individual's baseline demand (under price p_0) and her discounted demand (under price $p_0\times(1-d_0)$), we denote the former as q' and the latter as q. The proof logic here largely mirrors that of *Case 1* (Appendix A), but for graphical intuition on the difference between that and the current *Case 2*, we direct the reader's attention to Figure 10. By similar intuition as before, $\underline{\mathcal{Q}}_{d_0}$ represents a maximally adversarial DGP from the perspective of a naive market designer who optimized a subscription offer (S^*, d^*) using the rank-stable structural model of Section 2. The complication to the analogous procedure now is that the CDF difference $G_c(q)-G_{d_0}(q)$ may have a multiplicity of local maxima.

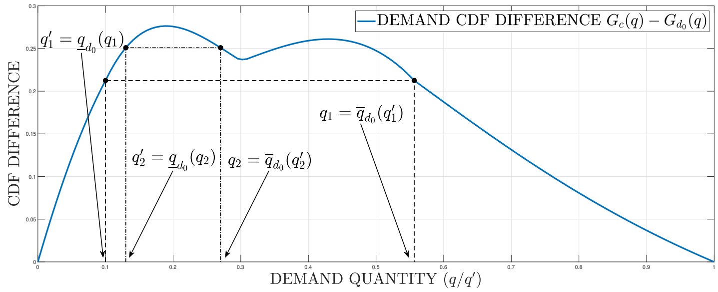
Thus, as depicted in the figure, we proceed as follows. For low values of baseline demand q', we re-allocate mass $g_c(q')-g_{d_0}(q')$ of consumers up to the *left-most value of counterfactual demand* q where the CDF difference matches $G_c(q')-G_{d_0}(q')$. We begin by continuously applying this procedure for values of q' beginning at 0 and working rightward until we reach the left-most local argmax of the CDF difference, as depicted in Panel A of the figure. Then, beginning with the left-most local argmin of the CDF difference, we continue the adversarial re-allocation of relatively low-q' consumers (with their masses still determined by the PDF difference at q') to the left-most values of counterfactual demand q where *both* the CDF difference matches $G_c(q')-G_{d_0}(q')$ *and* q respects the LoD (i.e., $q'\leq q$). This second intuitive phase of the procedure is depicted in Panel B of the figure. Finally, if there are any remaining local maxima further to the right (there can be at most countably many of them), we inductively apply this procedure from left to right until all local maxima of the CDF difference have been reached.[45]

More formally, we begin our proof by generalizing various objects used in *Case 1*. First, we re-define the functions $\bar{q}_{d_0}(q')\equiv\inf\{q>q' : G_c(q')-G_{d_0}(q')=G_c(q)-G_{d_0}(q)\}$ and $\underline{q}_{d_0}(q)\equiv\sup\{q'<q' : G_c(q')-G_{d_0}(q')=G_c(q)-G_{d_0}(q)\}$. Intuitively, $\bar{q}_{d_0}$ is a mapping from regions of baseline consumption q'-space where the CDF difference is increasing, into regions of counterfactual consumption q-space

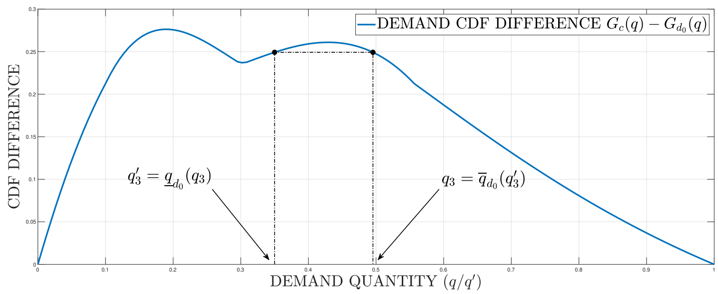
[45]One might wonder why treating local maxima in this left-to-right order achieves the *maximally adversarial* property, subject to the LoD and data. E.g., one might ask why re-allocating q_2' individuals (Panel A) up to q_3 (Panel B) might not imply even more extreme aggregate rank-stability violations. This technical hypothetical is ruled out by the logic of our proof, but for some rough intuition first recall that the LoD prevents downward counterfactual re-allocations of individuals with baseline demand q' to some counterfactual demand $q<q'$. Thus, jumping individuals at (or near) baseline demand q_2' up to (or near) counterfactual demand q_3 would do 2 things: (1) it would displace some or all of individuals having baseline demand at (or near) q_3' from rank jumping at all, and (2) it would mean that no low-q' consumers are rank jumping up to (or near) counterfactual demand q_2.

FIGURE 10. Proof Intuition For Case 2

(A)



(B)



Notes: This figure contains plots a hypothetical demand CDF difference violating the quasi-concavity assumption in Case 1 of Proposition 2. Panel A gives two examples of pairs of points q', q such that $q = \bar{q}_{d_0}(q')$ and $q' = \underline{q}_{d_0}(q)$ when q' lies before the first local mode of the CDF difference. Panel B provides a further example of a q', q pair such that $q = \bar{q}_{d_0}(q')$ and $q' = \underline{q}_{d_0}(q)$, but where q' lies after the first local mode of the CDF difference.

where the CDF difference is decreasing; $\underline{q}_{d_0}$ is a similar mapping, but in the opposite direction.[46] Moreover, $\underline{q}_{d_0}$ and $\bar{q}_{d_0}$ characterize maximal price responsiveness by consumers with low baseline demand, and by extension, minimal price responsiveness by consumers with high baseline demand, given the known aggregate shift from G_c to G_{d_0} under discount d_0. Now, we re-define $\underline{\mathcal{Q}}_{d_0}$ to be

$$\underline{\mathcal{Q}}_{d_0}(q, v; d_0) = \begin{cases} \text{if } \bar{q}_{d_0}(q) & g_c(q) - g_{d_0}(q) > 0, v \leq \frac{g_c(q) - g_{d_0}(q)}{g_c(q)} \\ q & \text{otherwise,} \end{cases} \tag{25}$$

where v is a realization of an independent uniform random variable as in *Case 1*. Following a similar structure as the proof of *Case 1*, we break our desired result up into a series of lemmas:

Lemma 5. $\underline{\mathcal{Q}}_{d_0}$ *satisfies the LoD.*

Lemma 6. $\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) \leq q] = G_{d_0}(q)$.

Lemma 7. $\Pr[Q_{d_0} > q | Q_c < q'] = \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) > q | Q_c < q']$ *for* $q < q'$.

[46]The definitions of $\bar{q}_{d_0}$ and $\underline{q}_{d_0}$ nest their respective definitions from the proof of *Case 1*.

Lemma 8. $\Pr[Q_{d_0} > q | Q_c < q'] > \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) > q | Q_c < q']$ *cannot occur if* $q \geq q'$ *and* $G_c(q) - G_{d_0}(q) > \inf\{G_c(x) - G_{d_0}(x) : q' \leq x \leq q\}$.

Lemma 9. $\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) > q, Q_c < q'] = G_c(q) - G_{d_0}(q)$ *if* $q \geq q'$ *and* $G_c(q) - G_{d_0}(q) \leq \inf\{G_c(x) - G_{d_0}(x) : q' \leq x \leq q\}$.

Lemma 10. $\Pr[Q_{d_0} \geq q, Q_c \geq q'] \geq 1 - G_c(q)$ *if* $q > q'$.

We begin by proving that the result of Proposition 2 must be true if Lemmas 5–10 are true. Since Lemma 5 is still true by construction as in *Case 1*, we will conclude our argument by showing that Lemmas 6–10 must be satisfied within our model, given the assumptions of Proposition 2.

Lemmas 5 and 6 ensure that $\underline{\mathcal{Q}}_{d_0}(Q_c)$ is a DGP consistent with the constraints imposed by the LoD and the observable distributions, (G_c, G_{d_0}). It thus suffices, as with the proof of *Case 1*, to show that Inequality (22), reproduced below, leads to a contradiction:

$$\Pr[Q_{d_0} > q | Q_c < q'] > \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) > q | Q_c < q'].$$

Supposing once again that this inequality is true, note that it must be violated if either (i) $q < q'$ (i.e., baseline demand is greater than discounted demand), by Lemma 7, or (ii) $q \geq q'$, and $G_c(q) - G_{d_0}(q) > \inf\{G_c(x) - G_{d_0}(x) : q' \leq x \leq q\}$, by Lemma 8. We thus need only consider the case (iii) $q \geq q'$ and $G_c(q) - G_{d_0}(q) \leq \inf\{G_c(x) - G_{d_0}(x) : q' \leq x \leq q\}$. Recall that the equalities and inequalities in (23) (Section A.2) follow only from general probability statements and the assumed inequality (22), both of which are still satisfied here. Therefore,

$$1 - G_{d_0}(q) > \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) > q, Q_c < q'] + \Pr[Q_{d_0} \geq q, Q_c \geq q'].$$

Lemma 9 ensures that in the current *Case 2* the first term on the right-hand side (RHS) satisfies $\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) > q, Q_c < q'] = G_c(q) - G_{d_0}(q)$ while Lemma 10 ensures that the second term on the RHS satisfies $\Pr[Q_{d_0} \geq q, Q_c \geq q'] = 1 - G_c(q)$. Combining these statements together implies $1 - G_{d_0}(q) > 1 - G_{d_0}(q)$, a contradiction. Therefore, since Inequality (22) must be false for *Case 2* (given Lemmas 6–10), then we have $\Pr[Q_{d_0} > q | Q_c < q'] \leq \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) > q | Q_c < q']$, which in turn implies the result of Proposition 2 that

$$\Pr\left[Q_{d_0} \leq q \middle| Q_c \geq \frac{S}{p_0 \times d_0}\right] \leq \Pr\left[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) \leq q \middle| Q_c \geq \frac{S}{p_0 \times d_0}\right].$$

To summarize, following similar logic as in *Case 1*, the result of Proposition 2 is true. What remains is to show that the basis of this logic, Lemmas 6–10, are true. The proof logic of *Step 1* of *Case 1* in Section A.2 above is still sufficient to deliver Lemma 7, since the argument there did not rely on quasi-concavity. Lemma 10 similarly is a consequence of the LoD and follows from the proof of *Step 2* of *Case 1* in A.2. The remainder of the lemmas to be proven—Lemmas 6, 8, and 9—make heavy use of Lemma 12 below, which in turn makes use of Lemma 11.

Lemma 11. *The following statements are true:*

(*i*) *If* $q_1' \in (q', \bar{q}_{d_0}(q'))$ *and* $g_c(q_1') - g_{d_0}(q_1') > 0$*, then* $\bar{q}_{d_0}(q_1') < \bar{q}_{d_0}(q')$*.*

(*ii*) *If* $q_1 \in (q', \bar{q}_{d_0}(q'))$ *and* $g_c(q_1) - g_{d_0}(q_1) < 0$*, then* $\underline{q}_{d_0}(q_1) > q'$*.*

(*iii*) *If* $q_1' \in (\underline{q}_{d_0}(q), q)$ *and* $g_c(q_1') - g_{d_0}(q_1') > 0$*, then* $\bar{q}_{d_0}(q_1') < q$*.*

(*iv*) *If* $q_1 \in (\underline{q}_{d_0}(q), q)$ *and* $g_c(q_1) - g_{d_0}(q_1) < 0$*, then* $q > \underline{q}_{d_0}(q_1)$*.*

Proof of Lemma 11. We will prove statement (i) of the Lemma here; the others are symmetric and proved in an analogous but tedious manner, so their explicit proofs are omitted.

By definition, $\bar{q}_{d_0}(q')$ is the infimal value of $x > q'$ such that $G_c(x) - G_{d_0}(x) = G_c(q') - G_{d_0}(q')$. We first show that $G_c(q_1') - G_{d_0}(q_1') > G_c(q') - G_{d_0}(q')$ must be true under the conditions required by statement (i). Suppose to the contrary that $G_c(q_1') - G_{d_0}(q_1') \leq G_c(q') - G_{d_0}(q')$. Because $g_c(q') - g_{d_0}(q') > 0$ (by definition of $\bar{q}_{d_0}$), there exists some $\tilde{q} \in (q', q_1']$ such that $G_c(\tilde{q}) - G_{d_0}(\tilde{q}) > G_c(q') - G_{d_0}(q')$. But then by intermediate value theorem, there exists $\tilde{q}^* \in (\tilde{q}, q_1'] \subseteq (\tilde{q}, \bar{q}_{d_0}(q'))$ such that $G_c(\tilde{q}^*) - G_{d_0}(\tilde{q}^*) = G_c(q') - G_{d_0}(q')$. Since $\tilde{q}^* < \bar{q}_{d_0}(q')$ this contradicts the definition of the latter, since $\bar{q}_{d_0}(q')$ could not be the infimal value of $x > q'$ where $G_c(x) - G_{d_0}(x) = G_c(q') - G_{d_0}(q')$. Thus, $G_c(q_1') - G_{d_0}(q_1') > G_c(q') - G_{d_0}(q')$ must be true under the conditions of statement (i).

Moreover, because $g_c(q_1') - g_{d_0}(q_1') > 0$, there must exist some $\tilde{q} \in (q_1', \bar{q}_{d_0}(q'))$ such that $G_c(\tilde{q}) - G_{d_0}(\tilde{q}) > G_c(q_1') - G_{d_0}(q_1')$. Thus, since the CDF difference at q' (and hence also at $\bar{q}_{d_0}(q')$) is strictly below the CDF difference at q_1', and since the CDF difference at $\tilde{q}$ is strictly above the CDF difference at q_1', then by the intermediate value theorem, there exists $\tilde{q}^* \in (\tilde{q}, \bar{q}_{d_0}(q'))$ such that the CDF difference there is the same, or $G_c(\tilde{q}^*) - G_{d_0}(\tilde{q}^*) = G_c(q_1') - G_{d_0}(q_1')$. Since $\bar{q}_{d_0}(q_1')$ is the infimal $x > q_1'$ for which $G_c(x) - G_{d_0}(x) = G_c(q_1') - G_{d_0}(q_1')$, we have $\bar{q}_{d_0}(q_1') \leq \tilde{q}^* < \bar{q}_{d_0}(q')$.

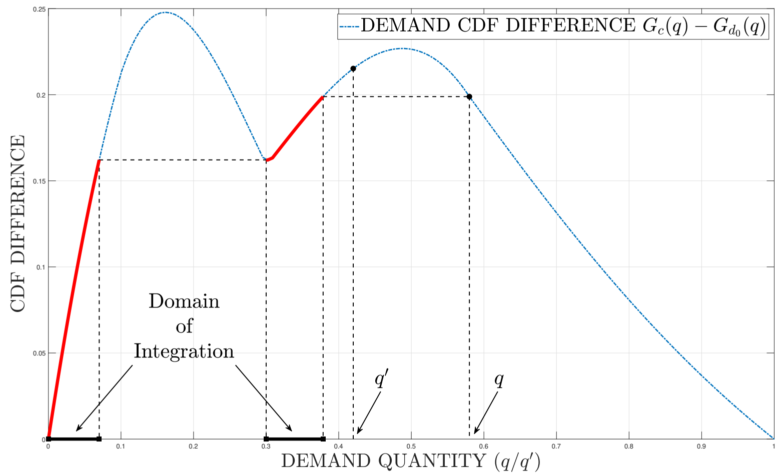
The proofs of statements (ii)–(iv) follow symmetric logic, illustrated by Figure 10. □

Lemma 11 can be thought of as establishing a local monotonicity property of $\bar{q}_{d_0}$ and its "inverse" function, $\underline{q}_{d_0}$. Specifically, recall that $\bar{q}_{d_0}(q')$ defines where jumpers consuming $Q_c = q'$ would jump under $\underline{\mathcal{Q}}_{d_0}$. Because our lower-bound DGP is meant to maximize jumping at the bottom, we should expect that jumps are larger for smaller values of q'. The first two statements of Lemma 11 establishes that $\bar{q}_{d_0}$ indeed satisfies this intuition. To understand the third and fourth statements, note that $\underline{q}_{d_0}(q)$ answers the following question: "if I see a consumer for whom $Q_{d_0} = q$, and if I know that this consumer is an adversarial jumper, what must her value of Q_c have been?" Again, because $\underline{\mathcal{Q}}_{d_0}$ is meant to maximize the aggregate degree of adversarial jumping (i.e, from low-q' values), the answer to this question should be smaller the larger q is.

Lemma 12. *If* (q', q) *are such that* $q' < q$ *and* $G_c(q) - G_{d_0}(q) < \inf\{G_c(x) - G_{d_0}(x) : q' < x < q\}$*, then*

$$\int_{\{x: g_c(x) - g_{d_0}(x) \geq 0, x \leq q', \bar{q}_{d_0}(x) \geq q\}} g_c(x) - g_{d_0}(x)\, \mathrm{d}x = G_c(q) - G_{d_0}(q). \tag{26}$$

FIGURE 11. Proof Intuition For Lemma 12



Notes: This Figure gives intuition for Lemma 12, for a hypothetical (q', q) pair in cases when the CDF difference is multi-modal. The domain of integration, $\{x : g_c(x) - g_{d_0}(x) \geq 0, x \leq q', \bar{q}_{d_0}(x) \geq q\}$, is depicted by bold lines.

Remark 8. The integration set in the statement of Lemma 12 is the set of possible realizations of baseline demand $q_c \leq q'$ such that (i) conditional on $Q_c = q_c$, there is non-zero probability of being a jumper, and (ii) conditional on being a jumper, the consumer jumps weakly *above* q, i.e., $\bar{q}_{d_0}(q_c) \geq q$. As we will see, the integral, depicted in Figure 11, represents the probability statement in Lemma 9, $\Pr[\underline{Q}_{d_0}(Q_c, V; d_0) > q, Q_c < q']$.

Proof of Lemma 12. Denote the integration domain in Lemma 12 by $R \equiv \{x : g_c(x) - g_{d_0}(x) > 0 \text{ and } x \leq q' \text{ and } \bar{q}_{d_0}(x) \geq q\}$. Consider the set $H = \{x : x < q\} \setminus \{x : g_c(x) - g_{d_0}(x) > 0 \text{ and } \bar{q}_{d_0}(x) \geq q\}$. This is the set of possible realizations of baseline demand $q_c \leq q'$ for whom there is either no chance of being a jumper or else they will not jump above q, $\underline{Q}_{d_0}(q_c, v; d_0) \geq q$, if they do rank jump.

Then R and H are disjoint and $R \cup H = \{x : x < q\}$. To see why, consider first the case where $y \in R \cup H$. Then either $y \in H$, which implies $y \in \{x : x < q\}$; or else $y \in R$, in which case $y \leq q' < q$ and $y \in \{x : x < q\}$ as well. On the other hand, suppose $y \in \{x : x < q\}$, but $y \notin R$. Then either (i) $g_c(y) - g_{d_0}(y) \leq 0$ or (ii) $q' < y < q$ or (iii) $\bar{q}_{d_0}(y) < q$. In case (i), $y \in H$ is clearly true. In case (ii), first note that $G_c(q) - G_{d_0}(q) < \inf\{G_c(x) - G_{d_0}(x) : q' < x < q\}$ (from the statement of Lemma 12), it must be true that $\underline{q}_{d_0}(q) \leq q' < y$. Lemma 11 then implies that $\bar{q}_{d_0}(y) < q$, hence, $y \in H$ must also be true. Finally, in case (iii), it is immediate from the definition of H that $y \in H$. Thus,

$$\int_R g_c(x) - g_{d_0}(x)\,\mathrm{d}x = \int_0^q g_c(x) - g_{d_0}(x)\,\mathrm{d}x - \int_H g_c(x) - g_{d_0}(x)\,\mathrm{d}x.$$

Since the first integral on the RHS above is exactly equal to $G_c(q) - G_{d_0}(q)$, the lemma follows if we can show that the second intergral on the RHS evaluates to 0. Graphically, it is readily apparent why this must be so in Figure 11. Note that integration of the PDF difference gives the

CDF difference depicted in the figure. Moreover, since the two "humps" of the CDF difference (with endpoints defined by the thin, dashed, horizontal lines) begin and end at the same functional value, it must be the case that integration of the PDF differences over the "hump" regions (which comprise set H) results in a value of zero. Although the general proof is somewhat involved, this is what it means within the depicted example for the integral $\int_H g_c(x)-g_{d_0}(x)\,\mathrm{d}x$ to equal zero.

More formally, we begin by partitioning H into disjoint subsets $H = H_1 \cup H_2$, where $x \in H_1$ implies that $g_c(x)-g_{d_0}(x) = 0$ while $H_2 = \bigcup_k I_k$ where the $I_k = [a_k, b_k]$ are a disjoint collection of non-singleton "hump" intervals such that $G_c(a_k)-G_{d_0}(a_k) = G_c(b_k)-G_{d_0}(b_k)$. We refer to this latter condition as the *hump property.* Intuitively, the subset H_1, which may be empty, is intended to account for flat regions of the CDF difference that do not occur at a local maximum. Note also that the definition of H_2 allows for it to contain "humps" with flat regions at local maxima. Our proof consists of a number of steps, which we state and prove in turn.

Step 1, Technical Preliminaries:

We show here that for each $x \in H$, if $g_c(x)-g_{d_0}(x) < 0$ (i.e., the CDF difference is *decreasing* at x) then the interval $[\underline{q}_{d_0}(x), x] \subseteq H$ while if $g_c(x)-g_c(x) > 0$ (i.e. *increasing* CDF difference at x), then $[x, \bar{q}_{d_0}(x)] \subseteq H$, and in either case, the hump property is satisfied by the resulting interval.

Step 1.1, $\boldsymbol{x \in H}$ and $\boldsymbol{g_c(x)-g_{d_0}(x) < 0}$: In this case, label the interval $I = [\underline{q}_{d_0}(x), x]$. By the definition of $\underline{q}_{d_0}$, I satisfies the hump property, so it suffices to show that $I \subseteq H$. For any $y \in I$, either $g_c(y)-g_{d_0}(y) \leq 0$, in which case $y \in H$ as well; or else $g_c(y)-g_{d_0}(y) > 0$. In this latter case, it must be true that either $y = \underline{q}_{d_0}(x)$, so $\bar{q}_{d_0}(y) = x < q$ and $y \in H$ or Lemma 11 implies that $\bar{q}_{d_0}(y) < x \leq q$, so $y \in H$ in this case as well.

Step 1.2, $\boldsymbol{x \in H}$ and $\boldsymbol{g_c(x)-g_{d_0}(x) > 0}$: The definition of H implies that $\bar{q}_{d_0}(x) < q$ must be true if $x \in H$ and $g_c(x)-g_{d_0}(x) > 0$. Label the interval $I = [x, \bar{q}_{d_0}(x)]$, and by definition of $\bar{q}_{d_0}$, note that I satisfies the hump property. If $y \in I$ then either $g_c(y)-g_{d_0}(y) \leq 0$ in which case $y \in H$ or $g_c(y)-g_{d_0}(y) > 0$, in which case, either $y = x$ so $y \in H$ or $y > x$, in which case Lemma 11 implies $\bar{q}_{d_0}(y) < \bar{q}_{d_0}(x) \leq q$, so $y \in H$ here as well.

Step 2, Construction of set $\boldsymbol{H_2}$: For each $x \in H$ where $g_c(x)-g_{d_0}(x) \neq 0$, define interval I_x as

$$I_x \equiv \begin{cases} \bigcup_{\{y: x \in [y, \bar{q}_{d_0}(y)] \subseteq H\}} [y, \bar{q}_{d_0}(y)], \text{ if } g_c(x) - g_{d_0}(x) > 0 \\ \bigcup_{\{y: x \in [\underline{q}_{d_0}(y), y] \subseteq H\}} [\underline{q}_{d_0}(y), y], \text{ if } g_c(x) - g_{d_0}(x) < 0, \end{cases} \tag{27}$$

and define $H_2 = \bigcup_{\{x \in D: g_c(x)-g_{d_0}(x) \neq 0\}} I_x$. For some intuition we refer the reader once again to the example depicted in Figure 11: for any $x_1 \in [\underline{q}_{d_0}(0.3), 0.3]$ the interval I_{x_1} will correspond to the left hump region $[\underline{q}_{d_0}(0.3), 0.3]$, and for any $x_2 \in [\underline{q}_{d_0}(q), q]$ the interval I_{x_2} will correspond to the right hump region $[\underline{q}_{d_0}(q), q]$. In this one example it is immediately obvious that $H = H_2 = I_{x_1} \cup I_{x_2}$ ($H_1 = \emptyset$ in the figure) and that I_{x_1} and I_{x_2} satisfy certain useful properties such as the hump property. Moreover, one can also see from the graph of the CDF difference (Figure 11) that therefore $\int_{H_2} g_c(x) - g_{d_0}(x)dx = 0$. However, a more general argument is needed to encompass all possible examples. Specifically, we proceed by showing that H_2 can generally be partitioned into a countable collection of intervals, where each interval in the partition satisfies the hump property. Since the integral of the PDF difference over an interval satisfying the hump property must be 0, this will in turn imply that the integral of the PDF difference over H_2 also must be 0.

More formally, we show that each member of the collection of I_x's defining H_2 satisfies the hump property, and that an equivalence relation on H_2 exists of the following form: $x \sim y$ if and only if there exists some $z \in H_2$ such that $x, y \in I_z$ is an equivalence relation on H_2.[47] Finally, we show that the equivalence classes generated by $\sim$ take the form I_x for $x \in H_2$.

Step 2.1, the $\boldsymbol{I_x}$ intervals satisfy the hump property: Assume that $g_c(x) - g_{d_0}(x) > 0$, and let $y_x^* \equiv \inf\{y\colon\ x \in [y, \bar{q}_{d_0}(y)]\}$. Let $\{\tilde{y}_n\}_{n=1}^{\infty}$ be a sequence such that such that $x \in [\tilde{y}_n, \bar{q}_{d_0}(\tilde{y}_n)] \subseteq H$ and $\tilde{y}_n \to y_x^*$. Such a sequence must exist by the definition of y_x^*. WLOG we can take y_n to be a decreasing subsequence, and show that $\bar{q}_{d_0}(y_n)$ is a (weakly) increasing sequence. Suppose for a contradiction that there is a pair $m > n$ such that $\bar{q}_{d_0}(y_m) < \bar{q}_{d_0}(y_n)$. If it is further the case that $\bar{q}_{d_0}(y_m) < y_n$, then by transitivity $\bar{q}_{d_0}(y_m) < x$, which would contradict $x \in [y_m, \bar{q}_{d_0}(y_m)]$. Thus, we must have $y_n \leq \bar{q}_{d_0}(y_m) < \bar{q}_{d_0}(y_n)$. But then the contra-positive of Lemma 11 part (i) implies that $y_m \geq y_n$, another contradiction. Therefore, $\{\bar{q}_{d_0}(y_n)\}_{n=1}^{\infty}$ must be a weakly increasing sequence.

Moreover, since it is also bounded from above by q (by construction) we also know this sequence converges to some limit, which we call z_x^*. Note also that by definition of the $\bar{q}_{d_0}$ operator, we know that $G_c(y_n) - g_{d_0}(y_n) = G_c(\bar{q}_{d_0}(y_n)) - g_{d_0}(\bar{q}_{d_0}(y_n))$, so each interval of the form $[y_n, \bar{q}_{d_0}(y_n)]$ satisfies the hump property, and therefore $[y_x^*, z_x^*]$ must also satisfy the hump property. Finally, note that y_x^* is the lower bound of the closure of the "hump" interval I_x containing x, and $\{y_n\}_{n=1}^{\infty}$ is a sequence of points chosen from I_x which are weakly below x and become arbitrarily close to y_x^*. Therefore, the closure of I_x satisfies $\mathrm{Cl}(I_x) = [y_x^*, z_x^*]$ and it must satisfy the hump property. Similar logic suffices to prove that I_x satisfies the hump property in the case where $g_c(x) - g_{d_0}(x) < 0$.

Step 2.2, properties of the $\sim$ operator defined above: First, note that $\sim$ defines an equivalence relation on the set H_2, with various important properties including reflexivity, transitivity, and symmetry. First, for reflexivity, note that for any $x \in H_2$, there exists $z \in H$ such that $x \in I_z$. This implies that $x \sim x$ for $x \in H_2$, or in other words, $\sim$ is reflexive.

Second, for transitivity, suppose $x \sim y$ and $y \sim z$, so there exists $a, b \in H_2$ such that $x, y \in I_a$ and $y, z \in I_b$. We need to show that $x \in I_b$ as well. WLOG, we can choose a, b such that $g_c(a) - g_{d_0}(a) > 0$

[47] An equivalence relation $\sim$ on some set S is a binary relation satisfying three properties: reflexivity, transitivity, and symmetry. A binary equivalence relation $\sim$ over a set S, can equivalently be thought of as defining a partition, being a collection of equivalence classes $\mathcal{C}$ such that $S = \cup\{C\colon C \in \mathcal{C}\}$ and for any $x, y \in S$, $x \sim y$ if and only if there exists some $C \in \mathcal{C}$ such that $x, y \in C$. This will be useful later for proving Lemma 12.

and $g_c(b)–g_{d_0}(b)>0$.[48] Since $x\in I_a$, there is some w_x such that $x\in[w_x,\bar{q}_{d_0}(w_x)]$ and $a\in[w_x,\bar{q}_{d_0}(w_x)]$. Similarly, since $y\in I_a$ as well, there is some w_y such that $y\in[w_y,\bar{q}_{d_0}(w_y)]$ and $a\in[w_y,\bar{q}_{d_0}(w_y)]$. Take $w\equiv\min\{w_x,w_y\}$, and note that Lemma 11 implies that $x,y,a\in[w,\bar{q}_{d_0}(w)]\subseteq I_a$. Similarly, define w' such that $y,z,b\in[w',\bar{q}_{d_0}(w')]\subseteq I_b$, and note that by construction, since $y\in[w,\bar{q}_{d_0}(w)]$ and $y\in[w',\bar{q}_{d_0}(w')]$, these two intervals have a non-empty intersection.

We consider two cases: if $w\geq w'$, then since $[w,\bar{q}_{d_0}(w)]\cap[w',\bar{q}_{d_0}(w')]\neq\emptyset$, which implies that $w'\leq w\leq\bar{q}(w')$ must be true. Thus, by Lemma 11, $[w,\bar{q}_{d_0}(w)]\subseteq[w',\bar{q}_{d_0}(w')]$, so $x\in[w',\bar{q}_{d_0}(w')]$, and hence $x\in I_b$. Otherwise, if $w<w'$ then since $[w,\bar{q}_{d_0}(w)]\cap[w',\bar{q}_{d_0}(w')]\neq\emptyset$, we now have that $w<w'\leq\bar{q}_{d_0}(w)$. Again, by Lemma 11, this implies that $[w,\bar{q}_{d_0}(w)]\supseteq[w',\bar{q}_{d_0}(w')]\ni b$. This implies that $[w,\bar{q}_{d_0}(w)]\subseteq I_b$ since I_b is the set of all such intervals containing b, and again, we have $x\in I_b$.

Third, symmetry of $\sim$ is true by construction: $x\sim y$ is defined symmetrically for x and y. Thus, we have shown that $\sim$ constitutes an equivalence relation on H_2. As a result, H_2 can be partitioned into a collection $\mathcal{I}$ of disjoint, non-singleton intervals.

Step 2.3, each $\boldsymbol{I\in\mathcal{I}}$ is of the form $\boldsymbol{I_x}$ for some $\boldsymbol{x\in H_2}$. Fix $I\in\mathcal{I}$. Take $y\in I$ and let $x\in H_2$ be any x (possibly equal to y) such that $y\in I_x$. We wish to show that $I_x=I$. Clearly, $I_x\subseteq I$ because for any $z\in I_x$, we have that $y,z\in I_x$ and hence $z\sim y$, which in turn implies $z\in I$. Now suppose for the sake of contradiction that there exists some $z\in I$ such that $z\notin I_x$. Since $x,z\in I$, then $x\sim z$, and hence there exists w such that $x,z\in I_w$. In particular, the proof of the transitivity property of $\sim$ implies that $x,z\in[w',\bar{q}_{d_0}(w')]\subseteq I_w$ for some $w'\in H_2$. The definition of I_x then implies that $z\in I_x$, contradicting our initial supposition that $z\notin I_x$.

Step 3, completing the of proof of Lemma 12: A collection of non-singleton, disjoint intervals must be countable. We may enumerate the elements of $\mathcal{I}$ as $I_1,I_2,\ldots$, and express H_2 as $H_2=\bigcup_k I_k$. By *Step 2.3*, each interval k must be equal to $I_{x_k}\equiv[a_k,b_k]$ for some $x_k\in H_2$ and hence, by *Step 2.1* must satisfy the hump property: $G_c(a_k)-G_{d_0}(a_k)=G_c(b_k)-G_{d_0}(b_k)$. Thus,

$$\int_{H_2} g_c(x)-g_{d_0}(x)\,\mathrm{d}x=\sum_k\int_{I_k} g_c(x)-g_{d_0}(x)\,\mathrm{d}x=\sum_k G_c(b_k)-G_{d_0}(b_k)-(G_c(a_k)-G_{d_0}(a_k))=0$$

On the other hand, since $x\in H_2$ for all x such that $g_c(x)-g_{d_0}(x)\neq 0$ by construction, it must be the case that if $x\in H_1\equiv H\backslash H_2$, then $g_c(x)-g_{d_0}(x)=0$, which implies that $\int_{H_1} g_c(x)-g_{d_0}(x)\,\mathrm{d}x=0$. This shows that $\int_H g_c(x)-g_{d_0}(x)\,\mathrm{d}x=\int_{H_1} g_c(x)-g_{d_0}(x)\,\mathrm{d}x+\int_{H_2} g_c(x)-g_{d_0}(x)\,\mathrm{d}x=0$, and Lemma 12 is proved. □

Before completing the proof of *Case 2*, we provide some intuition on the role Lemma 12 plays. To understand the statement of Lemma 12, note that the condition that $q'<q$ and $G_c(q)–G_{d_0}(q)\leq\inf\{G_c(x)-G_{d_0}(x):q'<x<q\}$ is analogous to the conditions of Step 2 in the proof of *Case 1* that

[48] This is WLOG because for any a,b with $g_c(a)-g_{d_0}(a)<0$ or $g_c(b)-g_{d_0}(b)<0$, we could instead have picked $a'\equiv\underline{q}_{d_0}(a)$ and $b'\equiv\underline{q}_{d_0}(b)$ such that $x,y\in I_{a'}$ and $y,z\in I_{b'}$.

$q' < q$ and $q > \underline{\mathcal{Q}}_{d_0}(q', 0)$. The condition that $G_c(q) - G_{d_0}(q) \leq \inf\{G_c(x) - G_{d_0}(x) : q' < x < q\}$ in particular states that the difference in CDFs never dips below $G_c(q) - G_{d_0}(q)$ on the interval (q', q). In the quasi-concave case, this happens precisely when $G_c(q') - G_{d_0}(q') > G_c(q) - G_{d_0}(q)$, which occurs iff $q > \underline{\mathcal{Q}}_{d_0}(q', 0)$. The integral, on the other hand, counts the total mass of jumpers for whom $x \leq q'$ and who jump above q. It turns out that these jumpers are the *only* ones who jump above q, so Lemma 12 simply states that the mass of these jumpers is exactly sufficient to fill in the difference in mass between G_c and G_{d_0} above q, which is exactly the RHS of equation (26).

We now return to showing that lemmas 6, 8, and 9 are true.

Proof of Lemma 6. We simply show that an analog of Equation (21) from Appendix A continues to hold. Specifically, we have $\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c) \leq q'] = A(q') + B(q') + C(q')$ where

$$A(q') = \int_{\{x : g_c(x) - g_{d_0}(x) > 0, x \leq q'\}} \left[1 - \frac{g_c(x) - g_{d_0}(x)}{g_c(x)}\right] g_c(x)\,\mathrm{d}x = \int_{\{x : g_c(x) - g_{d_0}(x) > 0, x \leq q'\}} g_{d_0}(x)\,\mathrm{d}x,$$

$$B(q') = \int_{\{x : g_c(x) \leq g_{d_0}(x), x \leq q'\}} g_c(x)\,\mathrm{d}x, \quad \text{and}$$

$$C(q') = \int_{\{x : g_c(x) - g_{d_0}(x) > 0, \underline{q}_{d_0}(x) \leq q'\}} \frac{g_c(x) - g_{d_0}(x)}{g_c(x)} g_c(x)\,\mathrm{d}x.$$

Intuitively, A counts the fraction of individuals with baseline consumption $q_c \leq q'$ for whom (i) the proportion of jumpers conditional on $Q_c = q_c$ is non-zero but who (ii) do not themselves jump. B counts individuals for whom there is no probability of jumping based on q_c. Finally, C counts individuals who jump but for whom even conditional on jumping, their level of consumption after jumping still does not exceed q'. Using what we know from above we can now compute

$$\begin{aligned} A(q') + B(q') + C(q') &= \int_0^{q'} g_c(x)\,\mathrm{d}x + \int_{\{x : g_c(x) - g_{d_0}(x) \geq 0, \bar{q}_{d_0}(x) \leq q'\}} g_c(x) - g_{d_0}(x)\,\mathrm{d}x \\ &\quad + \int_{\{x : g_c(x) - g_{d_0}(x) \geq 0, x \leq q'\}} g_{d_0}(x) - g_c(x)\,\mathrm{d}x \\ &= \int_0^{q'} g_c(x)\,\mathrm{d}x + \int_{\{x : g_c(x) - g_{d_0}(x) \geq 0, x \leq q', \bar{q}_{d_0}(x) > q'\}} g_{d_0}(x) - g_c(x)\,\mathrm{d}x. \end{aligned}$$

The last integral simplifies to $G_{d_0}(q') - G_c(q')$ by Lemma 12 which establishes the desired result. $\square$

Proof of Lemma 8. Suppose $G_c(q) - G_{d_0}(q) > \inf\{G_c(x) - G_{d_0}(x) : q' \leq x \leq q\}$ and define

$$q^* = \min\{y \leq q' : G_c(y) - G_{d_0}(y) = \inf\{G_c(x) - G_{d_0}(x) : q' \leq x \leq q\}\}$$
$$q^{**} = \max\{y \leq q : G_c(y) - G_{d_0}(y) = \inf\{G_c(x) - G_{d_0}(x) : q' \leq x \leq q\}.$$

The above definitions ensure that $q^* \leq q' \leq q^{**} < q$. Now suppose for the sake of a contradiction that $\Pr[Q_{d_0} > q | Q_c < q'] > \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) > q | Q_c < q']$. Because the LHS and RHS condition on the same event, this is equivalent to assuming that $\Pr[Q_{d_0} > q, Q_c < q'] > \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) > q, Q_c < q']$. The definitions of q^* and q^{**} ensure that

$$\begin{aligned}\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) > q^{**}, Q_c < q^{**}] &= G_c(q^{**}) - G_{d_0}(q^{**}) \\ &= G_c(q^*) - G_{d_0}(q^*) \\ &= \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) > q^*, Q_c < q^*] = \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) > q, Q_c < q'],\end{aligned} \tag{28}$$

where the last equality is due to the fact that q^* was defined so that, conditional on $Q_c < q'$, the only jumpers for whom $\underline{\mathcal{Q}}_{d_0}(Q_c, V; d_0) > q$ are those for whom it is further the case that $Q_c < q^*$.

Therefore, the statement that $\Pr[Q_{d_0} > q, Q_c < q'] > \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c) > q, Q_c < q']$ continues to hold with q' and q both replaced by q^{**}, as the RHS does not change while the LHS can only increase (since q is replaced by a *smaller* quantity while q' is replaced by a *larger* quantity). But then

$$\begin{aligned}1 - G_{d_0}(q^{**}) &= \Pr[Q_c > q^{**}] + \Pr[Q_{d_0} > q^{**}, Q_c \leq q^{**}] \\ &> \Pr[Q_c > q^{**}] + \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c) > q^{**}, Q_c < q^{**}] \\ &= 1 - G_c(q^{**}) + G_c(q^{**}) - G_{d_0}(q^{**}) = 1 - G_{d_0}(q^{**}),\end{aligned}$$

where the first equality comes by splitting the event $Q_{d_0} > q^{**}$ into two disjoint events $Q_C > q^{**}$ and $(Q_{d_0} > q^{**}) \cap (Q_c \leq q^{**})$, and the second equality comes from substituting line one of equation (28). This yields a contradiction and establishes our desired result that $\Pr[Q_{d_0} > q | Q_c < q'] > \Pr[\underline{\mathcal{Q}}_{d_0}(Q_c) > q | Q_c < q']$ is impossible when $G_c(q) - G_{d_0}(q) > \inf\{G_c(x) - G_{d_0}(x) : q' \leq x \leq q\}$. □

Proof of Lemma 9.

$$\Pr[\underline{\mathcal{Q}}_{d_0}(Q_c) > q, Q_c < q'] = \int_{\{x : g_c(x) - g_{d_0}(x) \geq 0, x < q', \bar{q}_{d_0}(x) > q\}} \frac{g_c(x) - g_{d_0}(x)}{g_c(x)} g_c(x)\, \mathrm{d}x$$

Again, Lemma 12 implies that the RHS simplifies to $G_c(q) - G_{d_0}(q)$. □

B.2. **Case 3 (Proof of Proposition 2 with Mass Points):** We complete the proof of Proposition 2 by showing how absolute continuity may be relaxed. Under our assumptions, there are at most countably many mass points. Enumerate these mass points as $q_1 < q_2 < \ldots$. The idea of the proof is to reduce this case back to the absolutely continuous case by studying modified distributions G'_c, G'_{d_0} which are absolutely continuous, and showing that a lower bound on the SUDs of G'_c, G'_{d_0} can be mapped back into a lower bound on the SUDs of G_c, G_{d_0}. We do this essentially by "smoothing out" the distributions of Q_c and Q_{d_0} by adding in continuously distributed noise at the mass points of G_c and G_{d_0}.

Formally, let (V_c, V_{d_0}) be a pair of i.i.d. random variables $V_c, V_{d_0} \sim Unif(0,1)$, that are independent of (Q_c, Q_{d_0}), and define $Q'_c = Q_c + \sum_{k=1}^{\infty} \frac{1}{2^k} \left(\mathbb{1}\{Q_c < q_k\} + V_c \mathbb{1}\{Q_c = q_k\}\right)$, and $Q'_{d_0} =$

$Q_{d_0} + \sum_{k=1}^{\infty} \frac{1}{2^k} \left(\mathbb{1}\{Q_{d_0} < q_k\} + V_{d_0}\mathbb{1}\{Q_{d_0} = q_k\}\right)$.[49] The multiplier ensures that these random variables are finite. Additionally, knowledge of Q'_c and Q'_{d_0} identifies the underlying Q_c, Q_{d_0}. For Q_c, there are two cases to consider. First, if $Q'_c \in \left[q_k + \sum_{j:q_j<q_k} \frac{1}{2^j}, q_k + \sum_{j:q_j\le q_k} \frac{1}{2^j}\right]$ for some k, then $Q_c = q_k$. Otherwise, Q'_c is not a mass point in which case Q_c is the solution to the equation $Q'_c = Q_c + \sum_{k:k\le Q_c}^{\infty} \frac{1}{2^k}$. Since the RHS is (strictly) increasing in Q_c, the solution must be unique, provided that it exists. Since we are currently considering the case where Q'_c does not correspond to a mass point, the unique solution must exist and G'_c must be absolutely continuous. For any Q'_c not corresponding to a mass point, the previous argument shows that the density of $g'_c(Q'_c) = g_c(Q_c)$ for the unique Q_c corresponding to Q'_c. Otherwise, for some k, $Q'_c \in \left[q_k + \sum_{j:q_j<q_k} \frac{1}{2^j}, q_k + \sum_{j:q_j\le q_k} \frac{1}{2^j}\right]$, and the density is given by $2^k \Pr[Q_c = q_k]$. Similar logic applies to the Q_{d_0} case.

We now just need to show that SUDs for G'_c, G'_{d_0} constitute SUDs for G_c, G_{d_0}. Clearly, this mapping holds when q, q' are chosen to be non-mass points of G_c, G_{d_0}. Otherwise, this fact follows simply from noting that the following events are equivalent (up to a measure zero set):

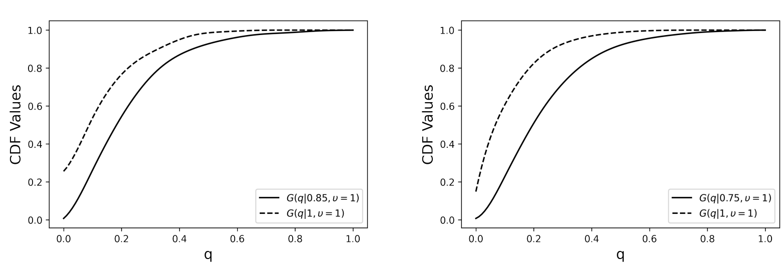
$$\{Q_c \le q_k\} \Leftrightarrow \left\{Q'_c \le q_k + \sum_{j:q_j\le q_k} \frac{1}{2^j}\right\}, \qquad \{Q_c \ge q_k\} \Leftrightarrow \left\{Q'_c \ge q_k + \sum_{j:q_j<q_k} \frac{1}{2^j}\right\},$$
$$\{Q_{d_0} \le q_k\} \Leftrightarrow \left\{Q'_{d_0} \le q_k + \sum_{j:q_j\le q_k} \frac{1}{2^j}\right\}, \text{ and } \{Q_{d_0} \ge q_k\} \Leftrightarrow \left\{Q'_{d_0} \ge q_k + \sum_{j:q_j<q_k} \frac{1}{2^j}\right\}.$$

■

Appendix C. Additional Tables and Figures

Figure 12. Stage-1 Estimation: Uptaker Demand CDFs

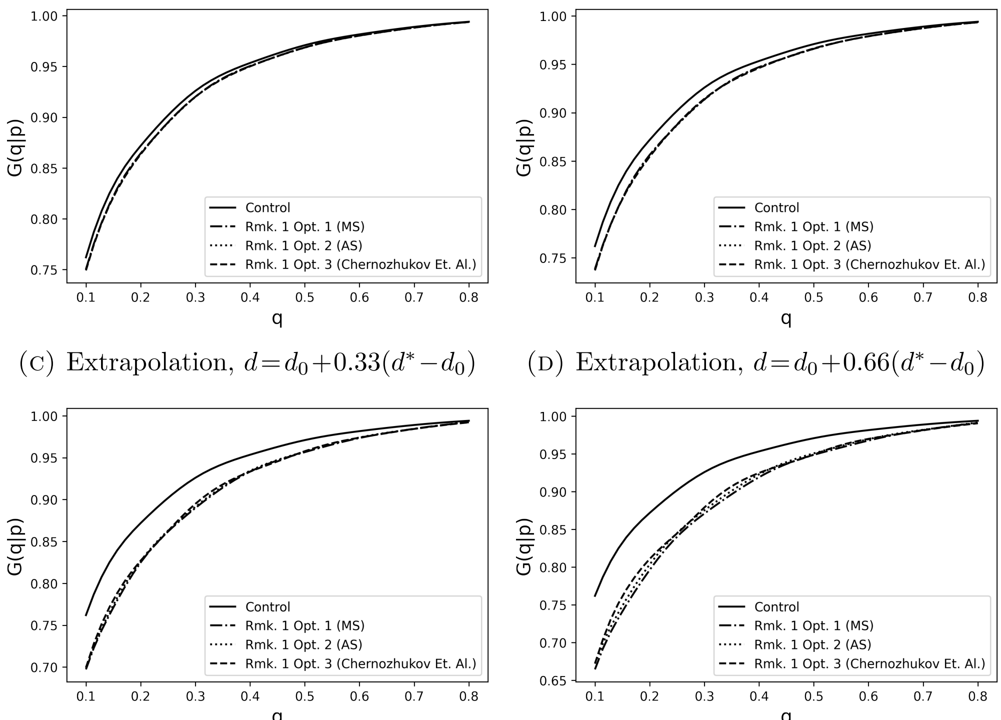
(a) 15% Off Experiment (b) 25% Off Experiment



Notes: This figure plots the CDFs of consumption with and without a discount in the subscription experiments. The left (right) Panel (C)orresponds to the experimental treatment arm offering a 15% (25%) discount.

[49] By way of clarification, these definitions are to be understood as saying that a given sample $\{q'_{cn}, q'_{d_0 n}\}_{n=1}^N$ drawn from the joint distribution of (Q'_c, Q'_{d_0}) depends on a corresponding sample $\{v_{cn}, v_{d_0 n}\}_{n=1}^N$. That is, for each n we have $q'_{cn} = q_{cn} + \sum_{k=1}^{\infty} \frac{1}{2^k} \left(\mathbb{1}\{Q_c < q_k\} + v_{cn}\mathbb{1}\{Q_c = q_k\}\right)$, and $q'_{d_0 n} = q_{d_0 n} + \sum_{k=1}^{\infty} \frac{1}{2^k} \left(\mathbb{1}\{Q_{d_0} < q_k\} + v_{d_0 n}\mathbb{1}\{Q_{d_0} = q_k\}\right)$, where $\{q_{cn}, q_{d_0 n}\}_{n=1}^N$ are drawn from the joint distribution of (Q_c, Q_{d_0}).

Figure 13. Robust Bounds



(A) Extrapolation, $d\!=\!0.33d_0$ (B) Extrapolation, $d\!=\!0.66d_0$

(C) Extrapolation, $d\!=\!d_0\!+\!0.33(d^*\!-\!d_0)$ (D) Extrapolation, $d\!=\!d_0\!+\!0.66(d^*\!-\!d_0)$

Notes: This figure compares reduced-form, aggregate demand extrapolations implied by options (1), (2), and (3) in Remark 1 for various out-of-sample discount levels d. Panels (A) and (B) plot extrapolated CDFs for prices at the middle thirds between p_0 and $p_0(1-d_0)$. Panels (C) and (D) plot extrapolated CDFs for prices at the middle thirds between $p_0(1-d_0)$ and $p_0(1-d^*)$.

Table 5. Fraction Subscriber Savings Retained: $(S_l^*(\lambda), d_l^*(\lambda))$ vs $(S^*(0), d^*(0))$

λ	Strong Uptaker Percentiles	0.1	0.25	0.5	0.75	0.9	Total
$\lambda = 0.35$	$SU(p_0, S^*(\lambda), d^*(\lambda))$	0.054	0.134	0.274	0.414	0.483	0.292
$\lambda = 0.35$	$SU(p_0, S^*(0), d^*(0))$	0	0	0	0	0.106	0.096
$\lambda = 0.16$	$SU(p_0, S^*(\lambda), d^*(\lambda))$	0.196	0.406	0.634	0.776	0.841	0.682
$\lambda = 0.16$	$SU(p_0, S^*(0), d^*(0))$	0	0	0.349	0.700	0.815	0.624

Notes: This table reports the retained savings ratio $\frac{q(r)d_l^*(\lambda)-S_l^*(\lambda)}{q(r)d^*(0)-S^*(0)}$ (*Extrapolation-Light* specification), where $q(r)$ is the r^{th} quantile of Q_c among strong uptakers, for $r \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$. The final column is aggregate retained savings, or $\frac{\int_0^1 q(r)d_l^*(\lambda)-S_l^*(\lambda)dr}{\int_0^1 q(r)d^*(0)-S^*(0)dr}$.

Table 6. Estimation of λ

Parameter	Estimate	95% Conf. Int.
$\mathcal{S}_c$ (Spearman Rank Corr.)	0.323	[0.320, 0.326]
$\mathcal{S}_t$ (Spearman Rank Corr.)	0.318	[0.315, 0.321]
$\mathcal{S}_a$ (Spearman Rank Corr.)	0.293	[0.291, 0.297]
λ (Adversarial Mass)	0.160	[0.026, 0.281]

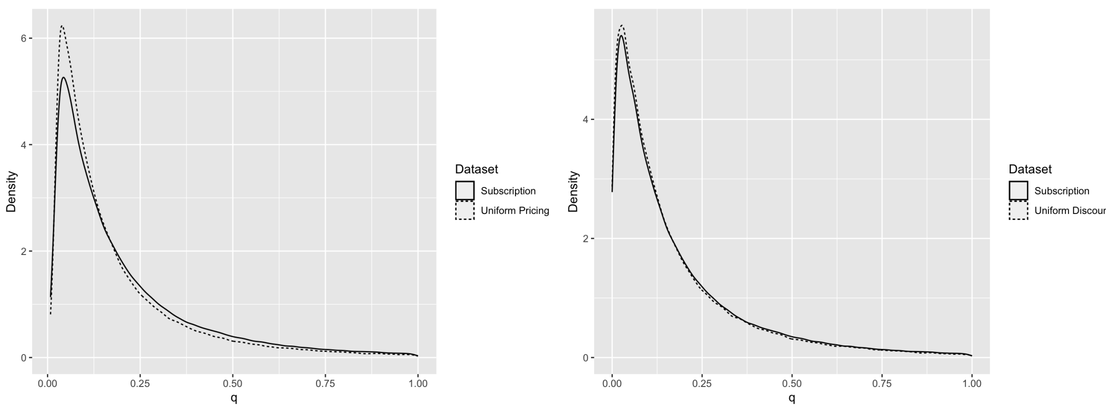
Appendix D. Comparability of Uniform Discount and Subscription Samples

In the main text, we showed that the elasticities measured in the subscription RCT were systematically higher than the elasticities measured in the uniform-discount RCT. In this section, we perform robustness checks to probe whether these differences are likely to be driven by the fact that the two RCTs draw on somewhat different sample populations. In Panel A of Figure 14, we plot boundary-corrected kernel density estimates of the density of q in the control groups of the respective RCTs. We see from this figure that the distribution of q in the uniform-discount RCT is shifted to the left relative to the subscription RCT. Looking at q in the period before the respective RCTs started, we find a similar trend. To correct for observable imbalances between the two sample populations, we re-run the estimator for the utility function in the uniform-discount RCT with sampling weights for each observation.

Figure 14. Distribution of q

(A) Subscriptions vs Unifrom Discount, Unweighted

(B) Subscriptions vs Uniform Discount, Re-weighted



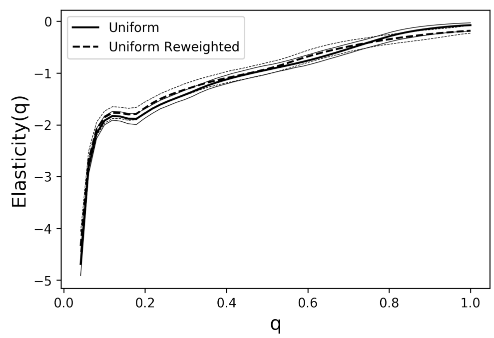
Notes: Kernel density estimates for the distribution of q (conditional on $q > 0$). The solid line corresponds to the Subscriptions experiment while the dashed line corresponds to the Uniform Discount experiment. In Panel A, we compare the raw distributions. In Panel B, we compare the raw distribution from the Subscriptions experiment to a re-weighted distribution from the Uniform Discount experiment.

We construct these weights to match the distribution of a number of pre-exposure observables the platform records daily for each individual. These observables capture information about the recent choices of passengers, such how many rides they took in the recent past, how much money was spent in the recent past, and which city a passenger took their last ride in. We will call this vector of passenger-level covariates $\boldsymbol{X}$ and balance on these covariates.

To do this, we begin by constructing a dataset where each row corresponds to a single consumer in a given RCT. We also include an indicator for whether that observation came from the uniform-discount RCT or from the subscriptions RCT, $Y = \mathbb{1}\{Subscriptions\}$. We next train a machine learning algorithm (gradient boosted decision trees) to predict $\Pr[Y = 1|\boldsymbol{X}]$. Let $\hat{P}$ be the fitted machine learning predictions. Then sampling weights as a function of pre-exposure covariates $\boldsymbol{X}$ can be constructed as $W(\boldsymbol{X}) \propto \frac{\hat{P}(Y=1|\boldsymbol{X})}{1-\hat{P}(Y=1|\boldsymbol{X})}$. These weights, when applied to the uniform-discount sample, will shift the distribution of covariates to look more like the distribution from the subscriptions RCT.

In Panel B of Figure 14, we plot the resulting re-weighted distribution and find that the two are now much more similar.[50] Having re-weighted our sample to make the subscriptions and uniform-discount RCT more comparable, we re-estimate the MS utility model, but with each observation weighted according to $W(\boldsymbol{X})$. In Figure 15, we compare the unweighted elasticity function to the weighted elasticity function. We find that the two are quantitatively similar, and despite the presence of tight confidence bounds, are not statistically distinguishable from one another. From this robustness check we conclude that the differences between structural estimates in the subscriptions experiment and uniform discount experiment depicted in Panel C of Figure 5 in the body cannot be attributed to sampling differences across the two data sets.

Figure 15. Elasticity from Weighted Sample



Notes: This figure compares the elasticity function estimated off of applying DF to an unweighted sample to DF estimated off of a sample weighted to make the uniform-discount RCT look more like the subscriptions experiment.

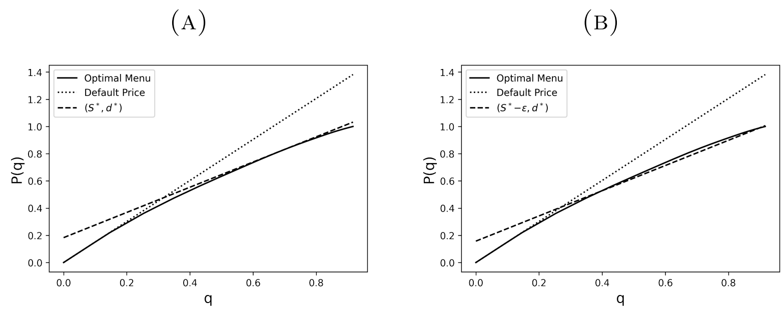
Appendix E. Contract Optimality: Additional Discussion

There is reason to believe that the high-profit, single-contract region in the heatmap of Figure 6 should generally be centered around the intermediate range of the optimal menu in other settings as well. In Online Appendix G.2 we show that the optimal menu $(S(\theta), d(\theta))$ is equivalent to a nonlinear pricing schedule $P(q)$, mapping consumption quantity q into total amount paid. Our constraint that default pricing is available implies that $P'(0) = p_0$ and no distortion at the top implies that $P'(q^{max}) = c$. Moreover, at any q^* with $P'(q^*) < p_0$, consider the tangent line to P at q^* defined by $S + qP'(q^*)$. Then the fee-discount pair, $(S, 1 - \frac{P'(q^*)}{p_0})$, is a contract in the menu of contracts implicitly defining $P(q)$. Consider offering only this single subscription instead of the entire menu. This mechanism implicitly defines a new price schedule $P_{q^*}(q)$ mapping q consumption units into the minimum purchase price: $P_{q^*}(q) = p_0 q$ if $q \leq \frac{p_0 - P'(q^*)}{S}$, and $P_{q^*}(q) = S + qP'(q^*)$ otherwise.

[50]Note that this improved balance is not a mechanical consequence of matching. The q being plotted in these figures is consumption *during* the RCT whereas we matched only on *pre-experiment* covariates. The fact that matching only on pre-experiment covariates greatly improves balance in outcomes during the RCT gives us confidence that our balancing is accounting for most of the important differences between the two samples.

Because $P(q)$ is concave, $P_{q^*}(q)$ is an upper bound for $P(q)$ and can be thought of as the function resulting from approximating $P(q)$ by the more accurate of two Taylor approximations, which we now specify. First, the Taylor expansion of $P(q)$ at $q=0$ corresponds to $p_0 \times q$ and determines how much the consumer pays in total under default pricing when consuming q. Second, the Taylor expansion of $P(q)$ at $q=q^*$ determines total payments after buying a subscription, $(S, 1-\frac{P'(q^*)}{p_0})$, and consuming q. For $P(q)$ without too much curvature (which seems to be the case for our data), this approximation will be quite good and is depicted in Panel (A) of Figure 16.

FIGURE 16. Approximating the Optimal Menu



Notes: Panels (A) and (B) compare the $(S(\theta), d(\theta))$-equivalent nonlinear pricing schedule, $P(q)$, to a single contract.

But we may be able to do better. By concavity of $P(q)$, we have $P_{q^*}(q) \geq P(q)$. Thus, if we used the single fee-discount pair $\left(S-\varepsilon, 1-\frac{P'(q^*)}{p_0}\right)$ for sufficiently small $\varepsilon > 0$, the resulting minimum cost schedule would do an even better job of approximating $P(q)$ on average. This is depicted in Panel (B) of Figure 16, and explains why the very best single contracts lay just below the optimal menu in Figure 6.

APPENDIX F. EXTRAPOLATION IN THE φ-SEPARABLE FAMILY

In this Appendix, we provide some additional discussion on the φ-separable family of utility functions mentioned in Remark 1. This family nests the MS utility model as a special case and enables a menu of possible ways to perform out-of-sample counterfactual extrapolations. As a motivating example, consider the additively-separable (AS) utility form, $U(q;\theta) = u(q) + \theta q$. This functional form can be found in much of the early theoretical literature on the principal-agent problem (e.g., see Maskin and Riley (1984) and Laffont and Tirole (1986)). The main difference between multiplicative and additive separability is their implications for extrapolation. Multiplicative separability implies that the price *elasticity* of demand, $\frac{\partial q^*(p;\theta)}{\partial p}\frac{p}{q^*(p;\theta)}$ depends on p and θ only through their implied level of demand $q^*(p;\theta)$ while additive separability implies the same property for the price *derivative* of demand, $\frac{\partial q^*(p;\theta)}{\partial p}$. Between additive and multiplicative separability is a continuum of other potential forms of separability, which we refer to as φ-separability:

$$U(q;\theta,\varphi) = \begin{cases} \int_0^q \left((1-\varphi)u(t)+\theta\right)^{\frac{1}{1-\varphi}} \mathrm{d}t, & \varphi<1 \\ \int_0^q \exp(u(t)+\theta)\,\mathrm{d}t, & \varphi=1. \end{cases}$$

Denoting the demand function as $q^*(p;\theta,\varphi) = \underset{q}{\operatorname{argmax}}\, U(q;\theta,\varphi)-pq$, within the φ-separable class, the *generalized elasticity* measure, $\frac{\partial q^*(p;\theta)}{\partial p}p^{\varphi}$, depends on p and θ only through their implied level of demand.[51] This family nests both MS utility ($\varphi=1$) and AS utility ($\varphi=0$) as special cases.

In this appendix, we demonstrate three facts about the φ-separable utility model as it relates to our present work. First, given rank stability, a single exogenous price change, and an ex-ante known value of φ, the structural primitives, including the function $u(q)$, are identified. Second, given rank stability and two exogenous price changes, φ can itself be identified as well. Third, our main empirical results are robust to how we extrapolate in the φ-separable family in the sense that our results remain quantitatively similar whether one adopts the polar opposite extremes of a $\varphi=1$ (MS) utility model, as we do in the body of the paper or a $\varphi=0$ (AS) utility model. A reader who is interested in a broader discussion of separability and the resulting implications for out-of-sample inference should consult Sun (2023b).

F.1. **(Rank-Stable) Identification with known φ and one exogenous price change.** For known φ, it suffices to prove is that $u(q)$ is identified in the φ-separable model. This follows immediately from Proposition 3 in Sun (2023b), which implies D'Haultfoeuille and Février's (2011) argument that $u(q)$ is identified under MS utility can be adapted to any φ-separable utility model with known φ.

F.2. **(Rank-Stable) Identification of φ with two exogenous price changes.** We now assume the econometrician has data (G_c, G_{d_0}) on demand distributions for control and treatment groups as before, but now, WLOG there is also a second treatment arm available, G_{d_1}, where $0<d_1<d_0$. What this means is that between d_0 and d_1 the observed overall demand response under price level $p_0\times(1-d_1)$ gives a more accurate picture of the derivative of the demand curve within a neighborhood of baseline price level p_0. In particular, as $d_1\to 0$, data from (G_c, G_{d_1}), along with the RS assumption imply that the derivatives of demand at price level p_0 are identified. Specifically, if θ_r is the r^{th}-quantile θ type then regardless of the true value of φ, we know that $\frac{\partial q^*(p_0;\theta_r,\varphi)}{\partial p} = \lim_{d_1\to 0}\frac{G_c^{-1}(r)-G_{d_1}^{-1}(r)}{p_0 d_1}$. Throughout the remainder of our discussion, we therefore treat

[51] When $\varphi=1$, we have MS utility and the generalized elasticity corresponds to the usual price elasticity measure. When $\varphi=0$, we have AS utility, and the generalized elasticity is the derivative of demand. Within the φ-separable utility family, the generalized elasticity polynomially interpolates between these two extremes, in similar fashion as a Box-Cox transformation.

d_1 as being close to 0 and assume that $\frac{\partial q^*(p_0;\theta_r,\varphi)}{\partial p}$ is identified. We now show how this derivative information, along with the distributions (G_c, G_{d_0}) suffice to identify φ.

WLOG, normalize default price $p_0 = 1$. The FOC of the φ-separable utility model implies that $1 = [(1-\varphi)u(q^*(1,\theta_r,\varphi) + \theta_r]^{\frac{1}{1-\varphi}}$ if $\varphi < 1$ and $1 = \exp(u(q) + \theta_r)$ if $\varphi = 1$. In either case, applying the implicit function theorem and rearranging shows that $\frac{1}{u'(q^*(p;\theta,\varphi))} = \frac{\partial q^*(p,\theta,\varphi)}{\partial p}$. Since $\frac{\partial q^*(p,\theta,\varphi)}{\partial p}$ is assumed to be identified, $u'(q)$ is identified. Adopting the location normalization that $u(0) = 0$ implies that $u(q)$ is identified as well.

Therefore, the model is identified if there exists a unique value of φ consistent with the data. Fix a quantile rank r and let $\theta_r(\varphi)$ be the r^{th}-quantile θ type implied by assuming that utility is φ-separable with utility index $u(q)$. Let $(q_c^*, q_{d_0}^*)$ respectively denote this individual's consumption under default price 1 and discounted price $1-d_0$. These quantities are identified by $q_c^* = G_c^{-1}(r)$ and $q_{d_0}^* = G_{d_0}^{-1}(r)$. For $\varphi < 1$, such an individual must satisfy the two FOCs $1 = [(1-\varphi)u_\varphi(q_c^*) + \theta_r(\varphi)]^{1/(1-\varphi)}$ and $1 - d_0 = [(1-\varphi)u_\varphi(q_{d_0}^*) + \theta_r(\varphi)]^{1/(1-\varphi)}$. Taking both sides of both equations to the $(1-\varphi)^{th}$ power and subtracting implies that $A(\varphi) \equiv \frac{(1-d_0)^{1-\varphi}-1}{1-\varphi} = u(q_{d_0}^*) - u(q_c^*)$. Similarly, at $\varphi = 1$, such an individual must satisfy the FOC $1 = \exp(u_1(q_c^*) + \theta_r(\varphi))$ and $1 - d_0 = \exp(u_1(q_{d_0}^*) + \theta_r(\varphi))$, which implies that $A(1) \equiv \log(1-d_0) = u(q_{d_0}^*) - u(q_c^*)$. Note that $\lim_{\varphi\to 1} A(\varphi) = A(1)$ and that $A(\varphi)$ is decreasing in φ.[52] Thus, there is at most one value of φ such that $A(\varphi) = u(q_{d_0}^*) - u(q_c^*)$. $\square$

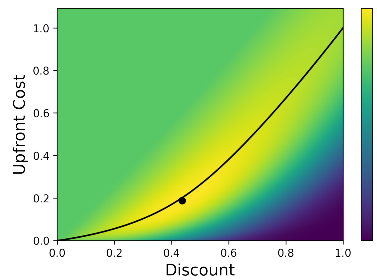
F.3. **Counterfactuals Under $\varphi = 0$ (Additive) Separability.** In Section 4 of the main body, we executed our empirical market design analysis under the MS ($\varphi = 1$) utility model. Here, we perform a robustness check by re-doing estimation and counterfactuals using the polar opposite case within the φ-separable family, additively separable utility ($\varphi = 0$). In doing so, we will demonstrate robustness to certain forms of mis-specification of the underlying demand system by showing that our results remain qualitatively unchanged and quantitatively similar as before.

In Figure 17, we plot the analog of Figure 6, but where we fit the data to an additively separable model. Compared to the MS case, the optimal single subscription in the AS case features a slightly lower discount. However, this comparison is not entirely informative, as it is impacted by the different imputed costs/markups under the AS structural model.[53] For a more direct comparison of optimal policy prescriptions, note that the ratios of the optimal discount to the firm's (imputed) markup under baseline pricing are quite close, being 0.5 under MS utility, and 0.45 under AS utility.

[52] Note that $A(\varphi)$ is exactly the Box-Cox transformation with parameter $\lambda^{bc} \equiv 1-\varphi$, so the above facts about $A(\varphi)$ follows from the well-known behavior of the Box-Cox transformation.

[53] Recall that in our empirical application we used cost imputation techniques common to the demand estimation literature (assuming constant marginal costs, see Section 4.0.1 and Appendix G.1), rather than actual costs, in order to protect Lyft's internal data confidentiality. In general, practitioners, consultants, and/or researchers working with internal firm data will have no need to impute costs, making this caveat less notable.

FIGURE 17. Profitability of Subscription Offers, Additive Model



Notes: Discount d is expressed as a fraction of the markup under default pricing. Upfront fee S is expressed as a fraction of the maximal upfront fee in the optimal menu. Lighter shades in the heatmap denote higher profitability. We also plot the optimal menu of subscription prices, and a point representing the optimal single subscription.

As for the robust market design exercise of Section 4.3, in Figure 19, we reproduce the analog of Figure 8, but using the AS utility model and the extrapolation-full lower bound. For sufficiently low values of λ, we again find broad agreement between the AS and MS extremes of the φ-separable utility family. Recall that MS structural estimates implied meaningful profits to be had from nonlinear pricing for values of $\lambda \leq 0.35$. Here, AS structural estimates imply a tipping point at roughly $\lambda = 0.42$, where nonlinear pricing profits begin to precipitously decline. In a slight change from the MS estimates, we see that AS structural estimates imply a somewhat more optimistic outlook, where optimal profitability of a nonlinear pricing plan remains (very low but) non-negligible as long as $\lambda \leq 0.61$.

Still, robust subscription recommendations are similar as before up to the initial tipping point. We still find that as λ grows larger, the optimal robust regime moves toward subscriptions with higher upfront costs and lower discounts, as can be seen in Figure 19. Since the value of $\hat{\lambda} = 0.16$ estimated from auxiliary data (Section 4.3.2) does not depend on utility specification, it remains unchanged, and in this case it still falls well within the viability range for meaningful nonlinear pricing profits (see black triangle depicted in Panel (A) of Figure 19). Although the AS and MS models are not identical in their counterfactual predictions, they are sufficiently similar within the range of plausible λ values to produce very similar counterfactual policy prescriptions and robust profit projections.

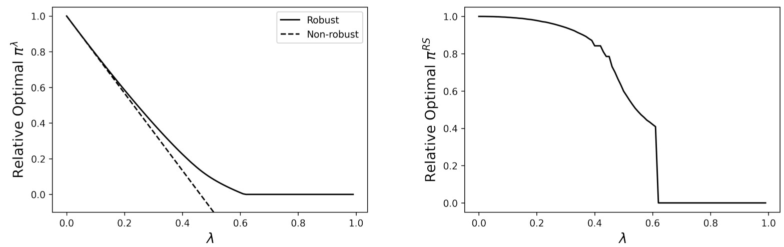
Appendix G. Details for Counterfactual Analysis

G.1. **Imputation of Marginal Costs.** Under constant marginal costs, the optimal uniform price is obtained by choosing p to solve for the firm's first order condition. Formally, letting $\theta_m(p)$ be the type who is on the margin between 0 consumption and positive consumption, we can write:

$$p = \frac{\varepsilon(p)}{1+\varepsilon(p)} c, \quad \varepsilon(p) = \frac{1}{\int_{\underline{\theta}}^{\bar{\theta}} q^*(p,\theta) \mathrm{d}F(\theta)} \int_{\theta_m(p)}^{\bar{\theta}} \frac{u'(q(p,\theta)) \mathrm{d}F(\theta)}{u''(q(p,\theta))} \tag{29}$$

FIGURE 18. Robustness Tests, Additive Model

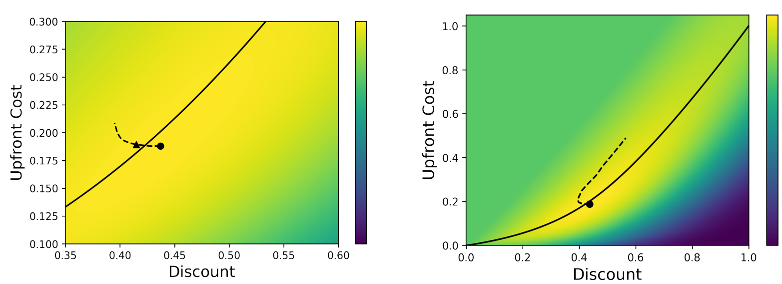
(A) Relative Adversarial Profits (B) Robust Profits Under RS



Notes: Panel (A) shows how optimal robust profits vary by λ, relative to RS profits—solid line, i.e., $\frac{\pi^{\lambda}(S^*(\lambda),d^*(\lambda))-\pi^{\lambda}(0,0)}{\pi^{rs}(S^*(0),d^*(0))-\pi^{rs}(0,0)}$—and how naive relative profits vary by λ—dashed line, i.e., $\frac{\pi^{\lambda}(S^*(0),d^*(0))-\pi^{\lambda}(0,0)}{\pi^{\lambda}(S^*(\lambda),d^*(\lambda))-\pi^{\lambda}(0,0)}$. Panel (B) depicts the cost of adopting a robust policy $(S^*(\lambda), d^*(\lambda))$ when the DGP is actually rank stable; i.e., $\frac{\pi^{rs}(S^*(\lambda),d^*(\lambda))-\pi^{rs}(0,0)}{\pi^{rs}(S^*(0),d^*(0))-\pi^{rs}(0,0)}$.

FIGURE 19. Path of Optimal Single Subscriptions

(A) Before Discontinuity (B) After Discontinuity



Notes: Discounts d are expressed as a fraction of the firm's markup under default pricing. Upfront fees S are expressed as a fraction of the maximal upfront fee in the RS-optimal continuum menu. This figure plots evolution of robust optimal subscriptions as λ varies. In Panels (A) and (B), the black dot is the RS-optimal subscription offer, and the black triangle is $(S^*(0.16), d^*(0.16))$ for the calibrated value of $\lambda = 0.16$.

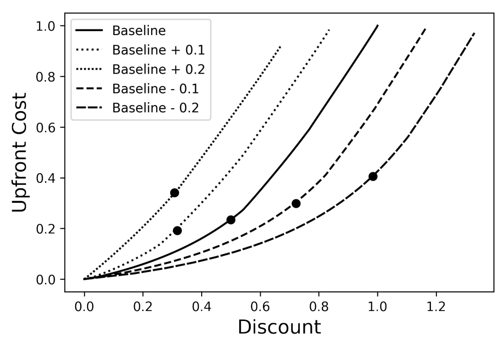
Here, $\varepsilon(p)$ is the average pointwise elasticity of the population at price p implied by the MS model. Inferences about the optimal linear price to charge are robust against RS violations. To see why, note that the monopolist's first order condition is more generally described by the condition that $\frac{\mathrm{d}\mathbb{E}[Q|p]}{\mathrm{d}p}(p-c)+\mathbb{E}[Q|p]=0$. Because under linear pricing everyone faces the same marginal incentive, we only need to know about the *average* effect of a price change on demand, which is identified given linear pricing variation.[54] The issue of measuring marginal costs can be somewhat tricky. Because we wish to focus on the general issues arising in the design of subscriptions pricing rather than the specific details of Lyft's subscription program, in the subsequent analysis, we proceed as follows. We treat $p=1$ as the optimal uniform price and calibrate c according to Equation (29). Doing this also has the benefit that it allows us to minimize the degree to which we need to extrapolate beyond

[54]This contrasts to the case of subscriptions design, because under a subscriptions program, not everyone gets a discount. Heterogeneity in elasticities is thus policy relevant in that case, hence the need to impose the RS for our subsequent nonlinear pricing counterfactuals.

the support of the available data. Our primary motivation for taking this imputation approach, based on the simplifying assumption of constant marginal costs, was to preserve the confidentiality of Lyft's internal profit information, which was a pre-requisite for us to gain access to the data for our empirical case study.

In order to give the reader a sense of how different marginal cost levels might impact the primary market-design objectives of this paper, Figure 20 plots the implied optimal menu of subscriptions and optimal single subscription for alternative assumptions about marginal cost. Most notably, the "no distortion at the top" property implies that the maximal discount offered should be equal to marginal cost. Nonetheless, the qualitative insights are fairly stable across specifications. We find in all cases that the optimal single subscription is very close to a subscriptions offering on the optimal menu, and that the discount is generally not far from half of marginal cost.

Figure 20. Sensitivity of Subscriptions to Marginal Cost Assumptions



Notes: This figure plots the optimal menu of subscriptions, as well as the single best subscriptions offer form a number of alternative assumptions on marginal cost. Specifically, we consider cases where marginal cost differs from our calibrated marginal cost by ± 0.1 or ± 0.2.

G.2. **Optimal Menu of Subscriptions Derivation.** Our goal in this Appendix is to derive the optimal menu of two-part tariffs as implemented by a menu of subscription offerings. In line with a large mechanism design literature (e.g., see Maskin and Riley (1984)) on second degree price discrimination, we find that our optimal mechanism can be represented as a nonlinear mapping from quantity consumed q to price paid $P(q)$. As we will see, the fact that we wish to implement $P(q)$ as a menu of subscriptions will impose the constraint that $P(q)$ must be increasing and concave. The highest consuming individuals will be charged at marginal cost, $P'(q^{max}) = c$, a result that is referred to in the mechanism-design literature as "no distortions at the top."

The optimal mechanism is a continuum menu of subscriptions, one for each type, θ. By the revelation principle, we can focus on solving for a subscriptions program implemented via an incentive-compatible mechanism. In this mechanism, agents reveal their types θ, and in return pay $S(\theta)$ upfront for discount $d(\theta)$. To make notation more standard, we will derive our results in terms of $p(\theta) = 1 - d(\theta)$. We add a participation constraint that consumers can always choose to not buy a two-part tariff and just consume at price $p = 1$.

To begin our derivation, first note that the firm's objective has the following form:

$$\int_{\underline{\theta}}^{\overline{\theta}} [S(\theta) + (p(\theta) - c)q(p, \theta)] f_\theta(\theta) d\theta \tag{30}$$

Let $U(\theta) = \theta u(q(p(\theta), \theta)) - p(\theta)q^*(p(\theta), \theta) - S(\theta)$ be the maximized utility of a type-θ individual under a revelation mechanism. The envelope theorem implies that, because truth-telling is incentive compatible, we have $U'(\theta) = u(q(p(\theta), \theta))$. We can then rewrite equation (30) as

$$\begin{aligned} &\int_{\underline{\theta}}^{\overline{\theta}} [\theta u(q(p(\theta), \theta)) - pq^*(p(\theta), \theta) - U(\theta) + (p(\theta) - c)q(p(\theta), \theta)] f_\theta(\theta) d\theta \\ &\qquad = \int_{\underline{\theta}}^{\overline{\theta}} [\theta u(q(p(\theta), \theta)) - cq(p(\theta), \theta)) - U(\theta)] f_\theta(\theta) d\theta \end{aligned} \tag{31}$$

On the RHS of Equation (31), $\theta u(q(p(\theta), \theta)) - cq(p(\theta), \theta))$ represents total surplus for the type θ segment of the market while $U(\theta)$ represents type θ consumer surplus. Thus, their difference represents producer surplus, i.e., firm profit, from the type θ segment of the market. Compared to the more familiar approaches to nonlinear pricing in Mussa and Rosen (1978) or Maskin and Riley (1984), we must impose a stronger incentive compatibility constraint: $p'(\theta) \leq 0$. To see why this is necessary, we first optimize out the choice of q given (p, θ) by defining $q(p, \theta) = u^{-1}(\theta/p)$. Then, given some schedule of contracts $(p(\theta), S(\theta))$, an agent in the revelation mechanism reports their type $\hat{\theta}$ to solve

$$\max_{\widehat{\theta}} \theta u(q(p(\widehat{\theta}), \theta))) - p(\widehat{\theta})q(p(\widehat{\theta}), \theta) - S(\widehat{\theta}). \tag{32}$$

The first order condition is

$$\theta u'(q(p(\theta), \theta)))q_p(p(\theta), \theta)p'(\theta) - p'(\theta)q(p(\theta), \theta) - p(\theta)q_p(p(\widehat{\theta}), \theta)p'(\theta) - S'(\theta) = 0. \tag{33}$$

Using the FOC of the implicit optimization problem defining $q(p, \theta)$, we can reduce this to

$$S'(\theta) = -p'(\theta)q(p(\theta), \theta). \tag{34}$$

We first remark that $p'(\theta) < 0$ at least somewhere. To see why, note that familiar results from mechanism design theory imply that $p(\bar{\theta}) = c$. On the other hand, $p(\theta) \leq c$ for all θ cannot be optimal, since, for example, it is dominated by the simple mechanism of no price discrimination at all and just choosing a single optimal monopoly price $p(\theta) = 1 > c$. Now, suppose that in addition to points where $p'(\theta) < 0$, there is some region where $p'(\theta) > 0$. This implies that there are some $\theta_1 < \theta_2 < \theta_3$ where $p'(\theta) > 0$ either from θ_1 to θ_2 or θ_2 to θ_3 and $p'(\theta) \leq 0$ in the other region, with strict inequality somewhere. Using the intermediate value theorem, we can find some $\theta_1 < \theta_4 < \theta_2 < \theta_5 < \theta_3$ such that $p(\theta_4) = p(\theta_5)$. Incentive compatibility implies that $S(\theta_4) = S(\theta_5)$. Additionally, we have $\frac{S'(\theta_4)}{p'(\theta_4)} = q^*(p(\theta_4), \theta_4) < q^*(p(\theta_5), \theta_5) = \frac{S'(\theta_5)}{p'(\theta_5)}$, where the inequality follows from the fact that $p(\theta_4)=p(\theta_5)$ and q is monotone increasing in θ for fixed p. The quantities S'/p' define the slope of the relationship between S and p around θ_4 and θ_5 respectively. Because they vary continuously in θ, we can find θ_6 and θ_7 in sufficiently small neighborhoods of θ_4 and θ_5, respectively, such that $p(\theta_6) = p(\theta_7)$, but either $S(\theta_6) < S(\theta_7)$ if $p(\theta_6) > p(\theta_4)$; or $S(\theta_6) > S(\theta_7)$ if $p(\theta_6) < p(\theta_4)$. As above, this yields a contradiction to incentive compatibility, since in the former case, type θ_7 agents could strictly improve their utility by reporting type θ_6, and in the latter case, type θ_6 agents could strictly improve their utility by reporting type θ_7.[55]

Given the above constraints, the maximization of profits becomes an optimal control problem with state variable $U(\theta)$, where the Hamiltonian has the following form:

$$\begin{aligned}\mathcal{H} = [\theta u(q(p(\theta),\theta)) - cq(p(\theta),\theta) - U(\theta)] f_\theta(\theta) + \nu_1(\theta) u(q(p(\theta),\theta)) \\ + \nu_2(\theta)\,[1-p(\theta)] + \nu_3(\theta)\,[U(\theta) - \theta u(q(1,\theta)) - q^*(1,\theta)] + \nu_4(\theta)p'(\theta)\end{aligned} \tag{35}$$

Recall from the discussion above that the first term of $\mathcal{H}$ represents producer surplus (i.e., profits) from type θ individuals. The term involving the multiplier $\nu_1(\theta)$ represents the law of motion of $U(\theta)$ implied by the incentive compatibility constraint. The $\nu_2(\theta)$ and $\nu_3(\theta)$ terms are participation constraints (PCs): the former enforces the constraint that marginal price must be smaller than 1 while the latter enforces the constraint that consumers may always pick default pricing. Finally, the term involving $\nu_4(\theta)$ enforces the incentive compatibility constraint that marginal prices must be (weakly) decreasing in θ. Standard results in optimal control theory imply that the solution satisfies the following five necessary conditions:

$$\mathcal{H}_U : -f_\theta(\theta) + \nu_3(\theta) = -\nu_1'(\theta) \tag{36}$$

$$\begin{aligned}\mathcal{H}_p : [\theta u'(q(p(\theta),\theta)) q_p(p(\theta),\theta) - cq_p(p(\theta),\theta)] f_\theta(\theta) + \nu_1(\theta) u'(q(p(\theta),\theta)) q_p(p(\theta),\theta) \\ - \nu_2(\theta) = 0.\end{aligned} \tag{37}$$

$$PC_1 : \nu_2(\theta)\,[1-p(\theta)] = 0 \tag{38}$$

$$PC_2 : \nu_3(\theta)\,[U(\theta) - \theta u(q(1,\theta)) - q^*(1,\theta)] = 0 \tag{39}$$

$$IC : \nu_4(\theta) p'(\theta) = 0 \tag{40}$$

Note that if any type θ buys a subscription, it must be the case that the highest type, $\overline{\theta}$, does so as well. In that case, neither of the participation constraints (PC_1 and PC_2) bind (i.e., $p(\overline{\theta}) < 1$ and type $\overline{\theta}$ gets an information rent). Therefore, $\nu_2(\overline{\theta}) = \nu_3(\overline{\theta}) = 0$. Since $\nu_3(\theta) = 0$ for any θ type

[55] Note that by the envelope theorem, the constraint $p'(\theta) \leq 0$ immediately implies the more typical incentive compatibility constraint $U'(\theta) \geq 0$. The strengthening of the usual incentive compatibility arises because under subscriptions, individuals have an extra degree of freedom to choose q after picking a subscription structure.

buying a contract, Equation (36) reduces to $-f_\theta(\theta) = -\nu_1'(\theta)$, and we must have $\nu_1(\theta) = F_\theta(\theta) - 1$ for any θ type that chooses to buy a contract.

Consider θ types sufficiently large so that buying a contract is optimal. Canceling the $q_p(p,\theta)$ terms in Equation (37) yields $\mathcal{H}_p : [\theta u'(q(p,\theta)) - c] f_\theta(\theta) - (1 - F_\theta(\theta)) u'(q(p(\theta),\theta)) = 0$ (for interior solutions). Solving this in turn yields

$$\left(\theta - \frac{1 - F_\theta(\theta)}{f_\theta(\theta)}\right) u'(q(p(\theta))) = c. \tag{41}$$

To invert this equation and solve for $p(\theta)$, we use the fact that by the consumer's FOC, $p(\theta) = \theta u'(q(p(\theta)))$, so we substitute $p(\theta)/\theta$ for $u'(q(p(\theta),\theta))$ in equation 41 to get

$$p^*(\theta) = \frac{c}{1 - \frac{1-F_\theta(\theta)}{\theta f_\theta(\theta)}}. \tag{42}$$

Thus, assuming an interior solution to the firm's optimal subscriptions problem, type θ consumers receive discount $d^*(\theta) = 1 - p^*(\theta) = 1 - \frac{c}{1 - \frac{1-F_\theta(\theta)}{\theta f_\theta(\theta)}}$.

The constraint that $(p^*)'(\theta) \leq 0$ is equivalent to the constraint that the inverse generalized failure rate, $\frac{1-F_\theta(\theta)}{\theta f_\theta(\theta)}$ is increasing.[56] Provided that this condition is met, (42) defines the optimal menu of subscriptions without modification. Otherwise, the usual ironing arguments must be used, and there will be pooling of contracts. The condition that $\frac{1-F_\theta(\theta)}{\theta f_\theta(\theta)}$ is increasing is a strictly stronger condition than the usual mechanism design assumption that the virtual valuation $\theta - \frac{1-F_\theta(\theta)}{f_\theta(\theta)}$ is decreasing, and as argued above, is due to our focus on nonlinear prices that can be implemented as a menu of subscriptions. In our empirical applications, we always find that $p(\theta)$ as defined by Equation (42) is indeed decreasing in θ, so even with our strengthened incentive compatibility constraint, no ironing is necessary and our solution is equivalent to the Maskin and Riley (1984) solution applied to our multiplicatively separable utility model. $\square$

[56]This terminology comes from Lariviere (2006).