SALIENCE AND (NON-)BUYER'S REMORSE: OPTIMAL NONLINEAR PRICING WITH COGNITIVELY CONSTRAINED CONSUMERS

AARON BODOH-CREED, BRENT HICKMAN, JOHN LIST, IAN MUIR, AND GREGORY SUN

ABSTRACT. Nonlinear pricing theory predicts that firms can extract surplus by inducing heterogeneous consumers to self-sort across price contract offers that are ex-post optimal for them. We study subscription pricing when the frictionless sorting assumption fails. Using large-scale subscription experiments conducted by Lyft, we document systematic deviations from optimal self-selection: many high-demand consumers decline subscriptions that would have saved them money, while some subscribers fail to break even. We develop a structural model of intensive-margin demand in which consumers may exhibit salience failures, forecast errors about future demand, or impulsivity. We show that subscription uptake can be recast as one-sided noncompliance in a binary-instrument framework, allowing us to leverage LATE methods to identify counterfactual outcome distributions and a novel "uptake function" linking baseline outcomes to compliance behavior. Combining experimental price variation with this identification strategy, we recover utility primitives, demand heterogeneity, and behavioral parameters. Salience failures and forecast errors play quantitatively important roles. Counterfactual analyses show that optimal subscription pricing generates substantial gains relative to linear pricing, but these gains are highly sensitive to consumer deviations from ex-post optimal choice. Implementing nonlinear pricing therefore requires not only optimal contract design for consumer screening, but also coordinated efforts to mitigate behavioral frictions.

Keywords: nonlinear pricing, adverse selection, robust, experiment design

JEL Codes: C14, C51, C93, D4, L1

Contact Information:

Aaron Bodoh-Creed (Amazon Inc.): `abodoh.creed@gmail.com`;

Brent Hickman (*corresponding author*, WashU St. Louis-Olin): `hickmanbr@gmail.com`;

John List (University of Chicago): `jlist@uchicago.edu`;

Ian Muir (Walmart Inc.): `muir.ian.m@gmail.com`; and

Gregory Sun (WashU St. Louis-Olin) `14sungre@gmail.com`.

Original version: April 2025

Date: March, 2026.

Acknowledgements: We thank Michael Grubb, Bradley Larsen, John Horn, Michael Cuna, Ali Hortaçsu, and Xavier Gabaix for useful conversations and feedback, as well as seminar participants at WashU-Olin, the North American Summer Meetings of the Econometric Society, and the IIOC meetings.

1. Introduction

Economic theory of adverse selection and nonlinear pricing provides firms with a powerful toolkit for maximizing profits in the presence of consumer unobserved heterogeneity. Seminal papers by Mussa and Rosen (1978) and Maskin and Riley (1984, henceforth MR84) established how a monopolist facing privately known, idiosyncratic demand can design optimal schedules of consumer total expenditures, as a function of their observable demand choices. By carefully applying progressively larger volume discounts that respect incentive compatibility, the MR84 framework achieves a fully self-sorting equilibrium that facilitates substantial profit gains relative to linear pricing. In the canonical model, consumers are able to perfectly evaluate the costs and benefits of all options and select the contract offer that is ex-post optimal for their realized demand.

While nonlinear pricing takes many forms in the modern economy (e.g., Grubb (2009); Luo et al. (2011); Grubb and Osborne (2015); Luo et al. (2018)), one common implementation is a simplified version of the MR84 approach: subscription programs. Consider a firm seeking to implement nonlinear pricing for a volume-based good or service; in many settings the firm allows any consumer to purchase at a default linear price p, but also offers the option to pay a subscription fee $\$S$ in exchange for a discount d on all consumption units purchased within the ensuing billing cycle. Examples of firms using subscription pricing span a wide range of industries, including automotive services (e.g., Club Car Wash, Tesla, ChargePoint), ridesharing (e.g., Uber, Lyft), meal delivery (e.g., HelloFresh, Home Chef, Blue Apron), digital media (e.g., Audible), grocery and retail (e.g., Instacart, Costco, Sam's Club), Artificial Intelligence services (e.g., OpenAI and Anthropic), and even B2B platforms (e.g., Twilio), among others. Some firms offer multi-tiered subscription plans—with progressively higher fees paired with deeper discounts—to better approximate the continuous MR84 framework and enable more granular consumer self-selection.[1]

While this is a common pricing mechanism in many industries, it may be cognitively demanding for consumers to evaluate which price offering is best for them (e.g., Miravete (2003); Grubb (2009); Grubb and Osborne (2015)). To make an optimal subscription decision, a consumer must correctly forecast her near-term demand, compute whether expected savings exceed the fee, and compare this to the outside option. Real-world consumers may struggle with the task of cost-benefit analysis for various reasons: all price options may not

[1]E.g., Meal delivery firms Hello Fresh and Blue Apron offer a base service (2 meals/week), with decreasing per-meal costs as the consumer orders more weekly meals. Sam's Club offers default price consumption with no fee (through co-located Walmart stores), and two subscription options ("Club" and "Plus" membership) with increasing fee-discount pairs. Twilio, a B2B communication services provider, has a pricing schedule for SMS hosting equivalent to a baseline volumetric price plus a 5-tier subscription menu.

be fully salient to them, or they may have an imperfect capacity to forecast their future demand. Moreover, in order to optimize a nonlinear price menu, firms also require preliminary evidence about consumer price sensitivity. For this purpose, pricing randomized controlled trials (RCTs) have become a primary analytic tool among modern firms.

However, deviations from fully rational decisions by consumers creates two related challenges. First, they generate econometric difficulties: observed subscription choices may not reflect ex-post optimal sorting, complicating inference about underlying demand. Second, behavioral frictions create managerial problems: the profitability of nonlinear pricing may depend not only on optimal fee and discount design, but also on how effectively the firm can mitigate consumer mistakes. This paper derives methodology to (*i*) help interpret data coming out of these sorts of efforts, (*ii*) estimate both demand and the magnitudes of behavioral phenomena present in consumer decisions, and (*iii*) evaluate nonlinear (subscription) pricing efficacy when consumer decisions deviate from full ex-post optimality.

We study these issues in the context of an RCT conducted by the rideshare firm Lyft, which randomized approximately one million consumers into either a control group (with baseline pricing) or one of two treatment groups that were offered the opportunity to purchase a subscription for a fixed upfront fee, and yielding either a discount of 15% or 25% over the subsequent month. At the time of the experiment, Lyft had not yet fully committed to launching a subscription program; as a result, treated consumers received only light-touch email notifications and no large-scale marketing campaign. The experiment therefore varied price, but not the broader marketing environment that would accompany a full rollout. This makes it an ideal dataset, not just for studying price sensitivity, but also for estimating the baseline magnitudes of various behavioral frictions to be addressed by the firm.

The raw data strongly reject the frictionless benchmark. For example, about 15% of the experimental population had high enough baseline demand that a subscription would have unambiguously saved them money, yet only about 4.7% of them subscribed. Conversely, some consumers who paid the upfront subscription fee failed to break even, though sub-optimal non-subscribership was the larger problem. These patterns violate sharp predictions of a standard nonlinear pricing model under full rationality. They suggest that consumer subscription decisions partially reflect cognitive frictions rather than pure type-based selection.

To interpret data from such experiments, we develop an augmented structural model that embeds behavioral mistakes within an otherwise standard neoclassical intensive-margin demand framework (e.g., Maskin and Riley (1984); Luo et al. (2018)). We allow consumers making an ex-ante commitment decision to exhibit three types of behavioral frictions: *salience* failures, in which they do not notice or fully process the offer; *forecast errors*, in which they imperfectly predict their near-term demand; and *impulsivity*, whereby some subscribe without carefully weighing costs and benefits. Conditional on ex-ante subscription choice, we assume consumption decisions in real time are utility-maximizing.

We show that failures to purchase subscriptions within this framework can be reinterpreted as representing failures to "comply" with underlying treatment in the experiment. This reinterpretation allows us to make use of fairly standard results from the local average treatment effects (LATE) literature (Imbens and Angrist (1994); Imbens and Rubin (1997)) to identify the distribution of counterfactual outcomes for compliers. Building on this insight, we introduce and identify a nonparametric "uptake function" that links baseline consumption to subscription choice. In our setting, this object is central for recovering structural behavioral primitives, but we argue that it provides potentially useful inference in broader non-structural LATE environments as well. The uptake function represents individuals' compliance probability as a function of their baseline potential outcome. It allows the econometrician to characterize an intuitive link between selection patterns and outcomes, without layering additional structure onto the baseline LATE framework.

Combining experimentally induced price variation with these LATE-based tools, we nonparametrically identify the utility function and the distribution of consumer demand types, as well as behavioral parameters governing salience, impulsivity, and forecast error. We also propose a GMM estimator which mirrors the identification logic. With it we find strong evidence that salience failures and forecast errors play a quantitatively important role in the data, while impulsive purchasing is at most negligible.

We then use the estimated model to conduct counterfactual analyses of optimal nonlinear pricing. We show, in a baseline scenario with fully rational consumers, an optimal subscription program can generate significant excess profits (5.62%), relative to optimal linear pricing. A fully separating MR84-style continuum menu achieves maximal surplus extraction, but simpler single-offer and two-offer menus—closer to what executives often prefer for operational simplicity—capture 82% and 94% of maximal excess profits, respectively.

An important caveat emerges: counterfactual results suggest that fully optimizing a subscription program requires a high level of cooperation between at least three of the firm's internal teams. First, the *pricing team* is in charge of optimizing default (linear) price. Second, the *promotions team* is in charge of optimizing the terms of subscription offers. Third, the *marketing team* is tasked with the firm's communications strategy to make consumers fully aware of the benefits of the price offering that is best for them. Even in a frictionless environment, optimal nonlinear pricing requires coordination between the pricing and promotions teams to customize baseline price to non-subscriber demand and induce optimal selection into subscription offers by high-demand consumers. Our model simulations show that failure to integrate these functions eliminates one third of excess profits.

In our setting, where consumers exhibit salience and forecast errors when making ex-ante subscription decisions, full profitability further requires coordination among pricing, promotions, *and* marketing teams; failure of any one of these roles to coordinate with the others can significantly reduce incremental subscription profits. We quantify these complementarities by considering coordinated pricing-promotion optimality under various levels of marketing team efficacy, where salience failures and average forecast errors are reduced by a given percentage. Each efficacy level achieved by the marketing team implies a different joint objective function for the pricing and promotions teams to optimize.

We find that a relatively high degree of marketing efficacy is required to maintain 50% of maximal excess nonlinear pricing profits: 75% and 63%, respectively, for the optimal one-offer and two-offer subscription menus. We explore mechanisms behind these results, and show that multi-tiered subscription offerings are doubly advantageous for profits: not only do they provide a better approximation to the full MR84 continuum menu, but they mitigate the importance of consumer forecast errors by providing more locally-tailored options to potential subscribers. Our results are similar in spirit to recent findings by Hortaçsu et al. (2024), who studied intra-firm synergies in the airline industry, arising from cross-team information frictions. Our study shows how demand-side phenomena external to the firm—utility heterogeneity and behavioral frictions—can also generate substantial intra-firm synergies.

1.1. **Related Literature.** Our study aims to contribute to four literatures. First, our empirical approach builds on a recent literature which studies nonparametric identification of principal-agent models (D'Haultfoeuille and Février (2015); Torgovitsky (2015); Perrigne and Vuong (2011); Luo et al. (2018); D'Haultfœuille and Février (2020); Hedblom et al. (2022); Kang and Silveira (2021)). Identification in this literature derives from either exploiting the principal's first-order condition or from exogenous incentive variation. Our data come from an RCT, so we follow the "exogenous variation" approach for identification established by Guerre et al. (2009) and D'Haultfœuille and Février (2020). We extend this framework to account for behavioral mistakes that may arise when cognitive difficulties prevent consumers from making fully ex-post optimal choices.

Second, we show that our model of behavioral mistakes in subscription choice admits a reinterpretation as an example of the binary-instrument, binary-treatment instrumental variables (IV) framework pioneered by Imbens and Angrist (1994) and Imbens and Rubin (1997). Thus, standard results on the LATE interpretation of IV apply to our setting, enabling us to identify the distribution of counterfactual outcomes for compliers. Moreover, we contribute to this literature by developing nonparametric identification and estimation results for what we refer to as the "uptake function," which helps characterize the link between selection patterns and outcomes, especially in settings with one-sided non-compliance. Our methodology thus provides an example of the potential complementarities between so-called "reduced-form" and "structural" approaches to econometrics.

Third, our study contributes to a growing literature incorporating insights from behavioral economics into understanding nonlinear pricing Gabaix and Laibson (2006); DellaVigna and Malmendier (2004); Oster and Scott Morton (2005); Esteban et al. (2007); Grubb (2009). Our work is most similar to Grubb and Osborne (2015), who estimate a structural model of cellular service demand. Like us, they incorporate consumer forecast errors into a model of nonlinear pricing, though the identification strategy and focus of our paper is different. Their use of panel data allows them to focus attention on the dynamics of consumer learning. By contrast, our cross-sectional internal firm data with explicitly randomized pricing policy allows us to jointly identify a model that separates out forecast errors from other behavioral phenomena such as salience and impulsivity. In Section 5.5 we discuss how Grubb and Osborne (2015) and our paper produce complementary insights on consumers' tendency to misestimate their future demand when the structure of nonlinear contracts require varying levels of commitment at the ex-ante decision point.

Fourth, our results speak to an organizational economics literature on the tradeoff between incentives and coordination within firms. The idea of the "M-form" firm—with internal segmentation into semi-autonomous divisions—was first proposed in economics by Williamson (1975). A number of more recent theoretical frameworks in this literature (Bolton and Farrell, 1990; Alonso et al., 2008; Dessein et al., 2010) are motivated by a fundamental tension between decentralized decision making—which leads to stronger incentives and more effective use of local knowledge—and increased coordination costs. Hortaçsu et al. (2024) presented empirical evidence of synergies across interdependent departments of an airline company, who best-respond to the actions of their intra-firm counterparts. They find that no individual department can improve firm profits by acting unilaterally. Our work contributes by analyzing an empirical model where cross-team coordination incentives exist as a consequence of external demand-side phenomena, rather than intra-firm information frictions. We quantify the degree of demand-driven synergies across teams, and find that they can be large: full intra-firm coordination allows for a significant boost in profits from nonlinear pricing, but its potential drops substantially if any one team's contribution is absent.

2. A Theoretical Framework For Nonlinear Pricing

To set the stage for our empirical analyses we define a baseline theoretical framework for intensive-margin demand. Following prior studies on nonlinear pricing (e.g., Luo et al. (2018)), we assume that consumers are heterogeneous along a single dimension, with a demand-intensity type $\theta \sim F_\theta$.[2] They choose the quantity q of a good/service to purchase from a single producer each period, trading off consumption utility against monetary costs.

Formally, if the per-unit price is p, consumer i chooses q to solve $\max_q U(q;\theta_i) - pq$, where $U(q;\theta_i)$ is monetized utility (in dollars) from q units of consumption if i's demand type is θ_i. We assume the following about preferences:

Assumption 1. Utility U and the distribution of types F_θ satisfy the following conditions:

(1) *Multiplicative Separability*: $U(q;\theta) = \theta u(q)$.
(2) *Utility Regularity*: u is twice continuously differentiable, increasing, and strictly concave. Additionally, $\lim_{q\to\infty} u'(q) = 0$
(3) *Utility Normalizations*: $u(0) = 0$ and $u'(0) = 1$.
(4) *Type Distribution Regularity*: F_θ is absolutely continuous, with strictly positive density $f_\theta(\theta)$ on a compact support $[\underline{\theta}, \bar{\theta}]$.

Multiplicatively separable utility is the most common specification in empirical studies of principal-agent problems and nonlinear pricing (see, e.g., Luo et al. (2018); D'Haultfœuille and Février (2020); Brugues (2020); Attanasio and Pastorino (2020); Kang and Silveira (2021); Hedblom et al. (2022); Chen et al. (2025); Cotton et al. (2026)). As explored in Sun (2025), this condition provides a means of forecasting the distribution of demand under prices not seen in the data.[3] The shape restrictions on u within the utility regularity condition ensure that each consumer's demand curve is well-defined and continuous. The normalizations are innocuous, with the first one implying that all utilities are measured relative to the outside option of zero consumption. The boundary derivative condition establishes a scale for utility types θ, which can be thought of as representing a consumer's *choke price*, or the

[2]In a related paper, Hickman et al. (2026) explore partial identification within a more general setting where consumers are k-dimensionally heterogeneous. For the purpose of this study, we abstract from these issues, with three notes. First, Hickman et al. (2026) show that the point-identified 1-dimensional model is a useful empirical benchmark, even if mis-specified, as it provides a sharp upper bound on the set of data-generating processes that cannot be ruled out by the data. Second, Hickman et al.'s bounds-based estimator of the optimal policy can in principle be combined straightforwardly with our estimator here to simultaneously account for behavioral frictions and multi-dimensional heterogeneity. Third, using a similar Lyft dataset, Hickman et al. (2026) find that the optimal subscription offer that is fully robust to residual selection driven by high-dimensional heterogeneity is close to the one implied by the 1-dimensional demand model used here.

[3]Hickman et al. (2026) show that the essential role of multiplicatively separable utility is to require smooth demand transitions between different prices, and to rule out pathological latent demand systems that are locally satiated precisely at observed prices.

minimum price above which they purchase $q=0$. Finally, the regularity conditions on F_θ (combined with smoothness of u) ensure that the distribution of q is smooth.

Individual demand curves are defined by the solution to a consumer's first-order condition, which under Assumption 1 implies $q^*(p;\theta)=(u')^{-1}(p/\theta)$. Proposition 1 formalizes some key implications of the demand model which will be central to our empirical exercise.

Proposition 1. *Under Assumption 1, demand curves $q^*(p;\theta)$ have the following properties:*

1. *Law of Demand: $q^*(p;\theta)$ is non-increasing in p and strictly decreasing for all prices p such that $q^*(p;\theta)>0$.*
2. *Single-Index: $q^*(p;\theta)$ is non-decreasing in θ and strictly increasing for all types θ such that $q^*(p;\theta)>0$.*
3. *Elasticity: The elasticity of demand of a type θ consumer's demand curve around price p is given by $\frac{\partial q^*(p;\theta)}{\partial p}\frac{p}{q^*(p;\theta)}=\frac{u'(q^*(p;\theta))}{q^*(p;\theta)u''(q^*(p;\theta))}$.*

The proposition characterizes consumer behavior under linear prices, but the utility maximization framework also implies how fully rational consumers *ought* to behave when presented with a nonlinear pricing scheme, such as a subscription program. Specifically, let the pair (S,d) describe an offer which allows the consumer to voluntarily purchase a $d\%$ discount for one period in exchange for an $\$S$ upfront fee. Faced with this option, a consumer now makes a 2-part decision: first, she must decide whether to buy the subscription or not; and second, she chooses purchase volume based on either discounted price $p(1-d)$ (if she subscribes) or default price p (if not). Working backwards, a subscriber would consume $q^*(p(1-d);\theta)$, while a non-subscriber consumes $q^*(p;\theta)$. The subscription is only worthwhile if consumption utility, net of the subscription fee, exceeds default utility; i.e.,

$$\underbrace{\theta u(q^*(p(1-d);\theta))-p(1-d)q^*(p(1-d);\theta)-S}_{\text{Optimized utility, conditional on subscribing}} \;-\; \underbrace{\theta u(q^*(p;\theta))-pq^*(p;\theta)}_{\text{Optimal utility under default pricing}} \;\geq 0. \tag{1}$$

Lemma 1. *The LHS of inequality* (1) *is increasing in θ.*

A formal proof is relegated to Appendix A. As a consequence of Lemma 1, there is a unique cutoff type, $\tilde{\theta}_p(S,d)$ which implies (1), such that (under fully rational subscription behavior), all types $\theta<\tilde{\theta}_p(S,d)$ should optimally *decline* the subscription, while all types $\theta>\tilde{\theta}_p(S,d)$ should optimally subscribe. Moreover, we can obtain empirically testable predictions about the cutoff type's consumption level, as the following result illustrates:

Proposition 2. *The cutoff demand type satisfies $q^*(p;\tilde{\theta}_p(S,d))\leq\frac{S}{p\times d}\leq q^*(p(1-d);\tilde{\theta}_p(S,d))$.*

See Appendix A for a proof. The cutoff $q=\frac{S}{p\times d}$ in Proposition 2 has a simple economic interpretation: it is the demand quantity at which the savings from discount d exactly offsets the upfront subscription fee S. Proposition 2 states that the cutoff type $\tilde{\theta}$ demands less than the cutoff quantity at price p, but demands more than the cutoff quantity with a discount.

Corollary 1. *If consumers are fully rational in purchasing subscriptions, then no consumer will buy a subscription and consume less than* $\frac{S}{p\times d}$*. Conversely, no consumer will abstain from buying a subscription and consume more than* $\frac{S}{p\times d}$*.*

This theoretical framework—a standard model of demand in empirical nonlinear pricing studies—produces sharp predictions when agents behave in a fully rational manner, as illustrated by Corollary 1. While various prior studies have relied on some combination of theory and exogenous price variation to achieve identification—e.g., Luo et al. (2018); D'Haultfœuille and Février (2020); Brugues (2020); Attanasio and Pastorino (2020); Kang and Silveira (2021); Hedblom et al. (2022); Cotton et al. (2026)—a unique aspect of the nonlinear pricing implementation in our data is that it allows for the basic rationality assumptions built into the model to be directly tested with raw data. A subscription program facilitates this test by separating demand into a 2-part sequential decision: subscription choice, followed by conditional consumption choice.

3. Data and Reduced-Form Analysis

In this section, we will see that the rationality implications of Corollary 1 are strongly violated in our empirical setting. This suggests that consumers may commit certain "mistakes" when approaching an ex-ante commitment within a nonlinear pricing scheme—e.g., when deciding whether to subscribe—and thus motivates a more general empirical framework (developed below) that can accommodate observed choice patterns.

3.1. Experimental Design.

Our data come from a controlled experiment that the rideshare firm, Lyft, conducted as part of the design process for a subscription program. The experiment randomized roughly 1 million consumers into one of three treatments. First, a *control group* was given default, business-as-usual price p_0. A second group, *treatment 1*, got an offer with the option to purchase a subscription for upfront fee \$$S$ to receive a discount of $d_1 = 15\%$ off of all rides in the subsequent month. The third group, *treatment 2*, got an offer to purchase a discount of $d_2 = 25\%$ for the same \$$S$ fee. For each consumer i in the sample, Lyft recorded an indicator for i's treatment status, $t_i \in \{c, 1, 2\}$. Finally, Lyft also recorded a variable $\upsilon_i \in \{0, 1\}$, which is an indicator equal to 1 if $t_i \in \{1, 2\}$ *and* if the consumer purchased a subscription during the experiment ($\upsilon_i = 0$ otherwise). In what follows, we will frequently refer to this group of consumers as *uptakers*.

3.2. **Reported Consumption Units.** Before proceeding, it is useful to clarify the units of analysis. Because Lyft rides are heterogeneous, we require an aggregate measure of consumption that can be interpreted as q within the model. We define q as the *cost of the ride in the absence of discounts*. For example, an individual i in the control group who takes two rides priced at \$10 and \$12, respectively, is recorded as having $q_i = 22$. Another individual i' who takes the same rides but receives a 10% discount pays $\$19.80 = (1 - 0.1) \cdot (\$10 + \$12)$, yet is likewise recorded as having *non-discount-equivalent* (NDE) consumption $q_{i'} = 22$.

This NDE convention is adopted for two reasons. First, it normalizes the baseline market price to $p_0 = 1$, simplifying a discount of d as corresponding to a price of $(1 - d)$. Second, it facilitates comparison across heterogeneous rides. Two trips taken contemporaneously between distinct origin–destination pairs may differ in distance or duration, while even a fixed route may constitute distinct goods across time due to variation in traffic conditions. Measuring q as total NDE consumption thus provides a consistent basis for comparison across trips and periods. This approach is analogous to the hedonic valuation of product attributes (Rosen (1974)) and closely related to the "bid homogenization" methods employed in the auctions literature (e.g., Haile et al. (2006); Athey and Haile (2007)).[4]

When presenting results in tables or figures, we apply an additional normalization to q to preserve the confidentiality of Lyft's internal cost and profitability information. Specifically, rather than reporting q in absolute units, we divide each observation by the (monthly) 98$^{\text{th}}$ percentile of q, denoted $\overline{q}$. All reported quantities of q therefore represent consumption expressed as a fraction of $\overline{q}$. This normalization does not affect the interpretation of discounts, which continue to correspond to prices multiplied by $(1 - d)$. For notational consistency, we maintain the normalization $p_0 = 1$ throughout. For similar confidentiality reasons, we also do not report the specific experimental value of $\$S$ so that Lyft's internal profits cannot be reverse-engineered by the reader. However, the economic insights from our study (both qualitative and quantitative) are presented so as to be invariant to these normalizations.

3.3. **Descriptive Analyses.** We begin by analyzing how consumption choices respond to the subscription program. The first three rows of Table 1 report mean values of consumption across the three experimental conditions. Here we see direct evidence for the law of demand, a basic assumption of consumer theory: consumers randomized to receive more generous subscription offers purchased more rideshare services from Lyft, on average. Panel A of Figure 1 similarly shows that treatment group 1's consumption distribution first-order

[4]In auction models, bids are often regressed on observable covariates to homogenize them, with the residual interpreted as bidder-specific demand intensity and the regression terms as hedonic utilities of the covariates. In our setting, ride characteristics (proxied by prices) play a similar role, while θ represents demand intensity.

stochastically dominates the demand CDF for the control group, and is in turn dominated by treatment group 2's demand CDF.

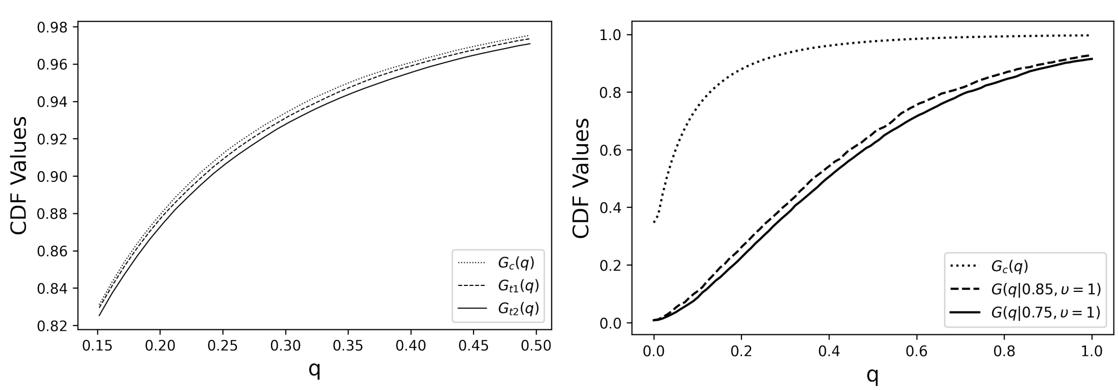
We next examine subscriptions purchasing behavior. In the fourth and fifth rows of Table 1, we report the proportion of consumers in the two treatment groups who ended up buying a subscription. We find fairly low subscription rates (between 1%-1.7%), but treatment group 2, who received the more generous discount offer (for the same fee $\$S$) was 70% more likely to buy the subscription. This is once again consistent with the law of demand, and implies that our consumer base is at least partially responsive to incentives. In Panel B of Figure 1, we also report the consumption behavior of individuals who subscribe. Overall, subscribers purchase considerably more rideshare services than individuals in control.

The evidence presented so far qualitatively aligns with the standard rational-choice theory outlined in Section 2, but we now turn attention to the more detailed predictions of our model regarding subscription purchase decisions. The sixth and seventh rows of Table 1 report the fraction of non-subscribers who nonetheless consumed $q \geq S/(p \times d)$. In the last two rows, we report the probability of consuming $q < S/(p \times d)$ despite buying the subscription. Recall from Corollary 1 that the both sets of probabilities should be 0. We find strong violations of this prediction. Between 19%-29% of consumers who did not buy the subscription would have broken even had they bought the subscription, depending on treatment arm, while 6%-16% of subscribers failed to break even. Moreover, the differences in these error rates across treatment arms are in an intuitive direction: when the offered discount is more generous, the fraction of sub-optimal non-subscribers drops.

Overall, the evidence suggests that while a rational-choice framework for analyzing subscription behavior is a useful start, to fit all features of the data, we must augment the basic model to allow for behavioral frictions influencing uptake behavior.

FIGURE 1. Raw CDFs From Experiemnts

(A) Subscr. Agg. CDFs (B) Subscr. Uptaker CDFs



Notes: Panel (A) compares the demand CDF within the control group to the demand CDF in the treatment group. Panel (B) compares the demand CDF of the control group with the demand CDFs for uptakers within the two treatment arms.

Table 1. Summary Statistics

Treatment Group	Point Estimate	Std. Err.
Mean q		
Control	62.46	0.20
Treatment 1	63.95	0.22
Treatment 2	65.98	0.21
$\mathbf{Pr}[\boldsymbol{v_i=1}]$ (unconditional uptake)		
Treatment 1	0.0098	0.0002
Treatment 2	0.0169	0.0002
$\mathbf{Pr}[\boldsymbol{q \geq S/(p\times d)\vert v_i=0}]$ (high demand, given non-uptake)		
Treatment 1	0.2948	0.0008
Treatment 2	0.1890	0.0007
$\mathbf{Pr}[\boldsymbol{q < S/(p\times d)\vert v_i=1}]$ (low demand, given uptake)		
Treatment 1	0.0644	0.0003
Treatment 2	0.1598	0.0007

Notes: This table reports a number of summary statistics of behavior in the various experimental treatment groups.

4. Structural Demand Model Extension: Behavioral Frictions

The differences between control and uptaker demand CDFs may arise due to either a treatment effect of the discount offered in the subscription, or a selection effect in terms of who subscribes in the first place. One key role of the model is to clarify this distinction and discipline how we disentangle these two phenomena within the data. However, our reduced-form evidence indicates that the baseline model from Section 2 places too many restrictions on the data-generating process to adequately perform this role.

4.1. **Augmented Model.** We now extend the model to allow for mistakes in the discrete choice of subscribing, while maintaining our assumption that consumers are able to choose the optimal quantity $q^*(p;\theta)$ without error.[5] We introduce a *salience* parameter, $\sigma \in (0,1]$, representing the probability that a given (treated) consumer is cognizant of the subscription offer. We also allow for some fraction $\rho \in [0,1)$ of consumers to exhibit *impulsivity* and (conditional on salience) always purchase the subscription without adequately weighing costs and benefits. Finally, we allow for consumers to be imperfect at forecasting their own demand intensity over the period following their subscription decision. Rather than evaluating their uptake decision based on their true demand type θ, they instead choose whether to subscribe based on a noisy estimate $\hat{\theta}_i = \theta_i \exp(-\varepsilon_i)$, where $\varepsilon_i \sim H_\varepsilon$ is independent of θ_i.[6] Conditional on salience and non-impulsiveness, consumer i subscribes if

$$\hat{\theta}_i \geq \tilde{\theta}_p(S,d), \quad \Leftrightarrow \quad \varepsilon_i \leq \log\theta_i - \log\tilde{\theta}_p(S,d). \tag{2}$$

[5]The idea that i makes mistakes only concerning the ex-ante subscription commitment is reasonable if a period (e.g., a month) is composed of D sub-periods (e.g., days) within which i observes a sub-period demand shock ε_{id} before choosing sub-period demand, but i cannot fully anticipate future sub-period shocks.

[6]The model is a static, one-time decision process of whether to purchase a subscription. In a dynamic version where the consumer makes repeated decisions over multiple subscription periods, a similar structure would hold: subscription in period t is based on $\hat{\theta}_{it} = \theta_{it} exp(-\varepsilon_{it})$, where $\theta_{it} = \theta_i + \eta_{it}$ represents realized ex-post demand at the end of period t, η_{it} is cross-period demand fluctuations, and ε_{it} is forecast error.

Throughout the paper, for simplicity we refer to the ε term as *forecast error*, but our identification results and nonlinear pricing framework are not dependent on a narrow interpretation of what it represents. One possibility is that ε reflects consumer beliefs about fluctuating monthly demand. H_ε could then be interpretable as classical rational expectations if it has zero mean. Alternatively, if H_ε has a non-zero mean, indicating skewed consumer views, it could be indicative of a systematic bias in their subjective beliefs about the future, or some degree of loss aversion, or both. Alternatively, ε could reflect other behavioral phenomena: a cutoff rule like (2) could arise if consumers perfectly forecast their demand, but are heterogeneously present-biased in their decision making. As will become clear later, our counterfactual analyses of interest will not hinge crucially on a specific interpretation among these alternatives, so for ease of discussion we simply label ε as "forecast error" in the meantime. We assume the following about the shock CDF:

Assumption 2. H_ε is absolutely continuous, and has strictly positive density on a compact support $[\underline{\varepsilon}, \bar{\varepsilon}]$. Moreover, ε is independent of θ.

Note that within this expanded version of the model, an individual could consume $q > \frac{S}{p \times d}$ and still fail to subscribe, due either to salience failure or if her forecast error ε is sufficiently large (i.e., $\exp(-\varepsilon)$ is sufficiently small). Similarly, one could also subscribe and fail to break even, due to some combination of impulsivity and forecast error. In other words, The augmented model is no longer rejected by the data. In what follows we will show that the available observables are still sufficient to identify model primitives.

Recall that consumers purchase at or below q under default price iff their θ type is smaller than $p_0/u'(q)$. Thus, we have $G_{ctrl}(q) = G(q|p_0) = F_\theta(p_0/u'(q))$. Let (S_t, d_t) be the subscription parameters offered within treatment group $t \in \{1, 2\}$. Corresponding to this offer, we define three quantities: first, the proportion of individuals who bought the subscription,

$$\Pr[\upsilon=1|t] = \sigma\left[\rho + (1-\rho)\int_{\underline{\theta}}^{\bar{\theta}} H_\varepsilon(\log\theta - \log\tilde{\theta}_{p_0}(S_t, d_t)) f_\theta(\theta)\,\mathrm{d}\theta\right]; \tag{3}$$

second, the demand CDF among non-uptakers,

$$G_t(q|p_0, \upsilon=0) = \frac{\int_{\underline{\theta}}^{p_0/u'(q)} \left\{1 - \sigma\left[\rho + (1-\rho)H_\varepsilon(\log\theta - \log\tilde{\theta}_{p_0}(S_t, d_t))\right]\right\} f_\theta(\theta)\,\mathrm{d}\theta}{1 - \Pr[\upsilon=1|t]}; \tag{4}$$

and third, the discounted demand CDF among uptakers,

$$G_t(q|p_0(1-d_t), \upsilon=1) = \frac{\int_{\underline{\theta}}^{p_0(1-d_t)/u'(q)} \sigma\left[\rho + (1-\rho)H_\varepsilon(\log\theta - \log\tilde{\theta}_{p_0}(S, d_t))\right] f_\theta(\theta)\,\mathrm{d}\theta}{\Pr[\upsilon=1|t]}. \tag{5}$$

In what follows, we will use lowercase g_t to denote the corresponding PDFs for the above CDFs. We now show how this system of equations enables recovery of the structural primitives, $(u, F_\theta, \sigma, \rho, H_\varepsilon)$. Our argument proceeds in a series of steps.

4.2. **Identification of Consumer Demand and Behavioral Frictions.**

4.2.1. *Recovering Reduced-Form Treatment Effects.* The first step of our identification argument is to show that our setting can be re-interpreted as a special case of the LATE instrumental variables framework proposed in Imbens and Angrist (1994). To simplify notation, we focus here on identification using only data from control and one of the two treatment arms, say $t = 1$. Passing to that subsample, we define the following random variables: $Y_{1i} \equiv q^*(p_0(1-d); \theta_i)$, $Y_{0i} \equiv q^*(p_0; \theta_i)$, $D_{1i} \equiv \mathbb{1}\{\varepsilon_i \leq \log\theta_i - \tilde{\theta}_{p_0}(S, d)\}$, $D_{0i} = 0$, and $Z_i = \mathbb{1}\{i \text{ offered subscription}\}$. Finally, let Y_i denote individual i's *realized* consumption choice, and let D_i denote individual i's *realized* subscription decision (with $D_i = 0$ if i is in the control group). The experimental design, along with our structural model, implies that $D_i = D_{1i}Z_i + D_{0i}(1-Z_i)$. Similarly, our assumptions about individual behavior imply that $Y_i = Y_{1i}D_i + Y_{0i}(1-D_i)$. Thus, re-interpreting consumption as the *outcome*, presence of discount d as the *treatment*, and experiment group as the *instrument*, behavior in the experiment under our model follows a standard potential outcomes framework with an excluded binary instrument.[7] Moreover, experimental randomization implies $Z_i \perp (Y_{i1}, Y_{i0}, D_{i1}, D_{i0})$, which is the standard exogeneity condition in the IV-LATE literature. Moreover, since $D_{i0} = 0$, the IV-LATE monotonicity condition that $D_{i1} \geq D_{i0}$ is trivially satisfied.

The above mapping from our setting onto a standard potential outcomes framework implies that results from the IV-LATE literature by Imbens and Rubin (1997) apply: the distribution of potential outcomes for compliers is identified. More specifically, recall that we directly observe the conditional CDF $G_t(q|p_0(1-d), \upsilon = 1)$ (equation (5)); i.e., the distribution of consumption *with* discount d among uptakers. We wish to recover $G_t(q|p_0, \upsilon = 1)$, namely, the *counterfactual* distribution of consumption *without* discount among the same group of uptakers. We also directly observe the complementary conditional CDF $G_t(q|p_0, \upsilon = 0)$ (equation (4)), and the CDF of demand within the control group, $G_{ctrl}(q) = G(q|p_0)$.

Let $\tau_t \equiv \Pr[\upsilon = 1|t]$ (equation (3)) denote the observed proportion of consumers within treatment t who subscribe. By randomization (i.e., instrument exogeneity) and by the law of total probability, the consumption distribution in control equals the weighted sum of treated consumption among uptakers and non-uptakers without a discount; that is, $G(q|p_0) =$

[7] The exclusion here is that the instrument Z_i induces variation in D_i, but only affects the outcome Y_i indirectly through its influence on D_i.

$\tau_t G_t(q|p_0, \upsilon=1) + (1-\tau_t)G_t(q|p_0, \upsilon=0)$. Isolating the object of interest gives

$$G_t(q|p_0, \upsilon=1) = \left[G(q|p_0) - (1-\tau_t)G_t(q|p_0, \upsilon=0)\right]/\tau_t. \tag{6}$$

Thus, the counterfactual uptaker distribution without discount is identified.

4.2.2. *Uptake Functions and Their Identification.* We now next extend results from Imbens and Rubin (1997) to show that the the conditional probability $\Upsilon(y) \equiv \Pr[D_{1i}=1|Y_{0i}=y]$, which we refer to as the *uptake function*, is nonparametrically identified from data. We argue that this insight has broader applicability to IV-LATE studies generally. Within our particular setting of nonlinear pricing, $\Upsilon(y)$ represents the probability that an individual purchases a subscription, conditional on baseline consumption. We will show later that the uptake function being identified is key to pinning down behavioral parameters σ and ρ, as well as the nonparametric distribution of forecast errors H_ε.

Returning to the general potential outcomes framework, the proof of nonparametric identification for Υ is a straightforward application of Bayes rule. Let $F_{0c}(y) \equiv \Pr[Y_{0i} \leq y|D_{1i}=1]$ (a generalization of equation (6)) denote the counterfactual outcome distribution of compliers in the absence of treatment, and let $F_0(y) \equiv \Pr[Y_{0i} \leq y]$ (a generalization of $G_{ctrl}(q)$) denote the unconditional outcome distribution in the absence of treatment. Both of these CDFs are known, so their corresponding densities $f_{0c}(y)$ and $f_0(y)$ are also known. In addition, the proportion of compliers $\Pr[D_{1i}=1] = \Pr[D_i=1|Z_i=1]$ (a generalization of τ_t) is also known from data. Therefore, an application of the Radon-Nikodym form of Bayes rule implies that the uptake function is nonparametrically identified:

$$\Upsilon(y) = \Pr[D_{1i}=1|Y_{0i}=y] = \frac{f_{0c}(y)\Pr[D_{1i}=1]}{f_0(y)}. \tag{7}$$

Our derivation of Equation (7) applies to *any* instrumental variables setting with one-sided non-compliance, or settings where treated individuals have more choices than individuals in control (i.e., $Z_i = 0 \Rightarrow D_i = 0$). It may therefore be of interest, even without an explicit structural model to identify: it allows the researcher to gain a deeper understanding of counterfactual scenarios by shedding light on the link between selection patterns and observable outcomes. For example, a well-known paper by Abadie et al. (2002) analyzed the effect of job training programs on worker earnings. Eligibility for training programs (Z) in their sample was assigned randomly, and thus serves as an instrument, yet conditional on eligibility, only about 60% ultimately signed up for training, or $\Pr[D_{1i}=1|Z_i=1] \approx 0.6$. Thus, their setting also exhibits one-sided non-compliance.[8] The authors find that job training

[8]A straightforward alternate form of equation (7) can be derived for application to settings with two-sided non-compliance (i.e., where $\Pr[D_i=0|Z_i=0] < 1$), but with additional tedious notation. While empirical uptake functions may retain some informativeness within such settings, it is limited because the definition of

appears to be particularly beneficial for individuals with low baseline skills, as the quantile treatment effects on earnings (Y) are largest at the lower end of the earnings distribution. In light of this result, one might ask, to what extent do individuals self-select into job training programs based on anticipated treatment effects? The uptake function sheds light on questions like this. If the empirical $\hat{\Upsilon}(y)$ uncovered from Equation (7) is decreasing in earnings y then this would imply that individuals with lower baseline earnings (for whom the treatment is estimated to be most effective) are also more likely to enroll. On the other hand, if $\hat{\Upsilon}(y)$ was flat or increasing in baseline earnings, that would suggest low-wage workers' selection choices reflect lower ex-ante understanding of the program's potential benefits than uptake choices of their higher-wage counterparts.

This discussion hints at an additional informative dimension of uptake functions: model testability. In some settings such as job training programs, there is no *a priori* information from theory which restricts the form of the uptake function, but this is not always the case. Returning to the specific notation of our demand model, $\Upsilon(q) = \Pr[\upsilon_i = 1 | Q_{0i} = q] = \frac{\tau_t g_t(q|p_0, \upsilon=1)}{g_{ctrl}(q)}$. The additional structure here (Sections 2 and 4.1) implies a property that the uptake function must satisfy: $\Upsilon(q)$ should be monotone increasing as a consequence of Assumptions 1 and 2, which together imply that uptake probability implied by the cutoff rule in (2) is monotone increasing in θ. On the other hand, there is nothing about equation (7) that guarantees the empirical uptake function recovered from a given data sample is monotone, so it follows that the structural demand model is testable.

4.2.3. *Recovery of Utility Parameters From Outcome Distributions.* Above we saw that the counterfactual control CDF for uptakers, $G_t(q|p_0, \upsilon=1)$, is uniquely pinned down by data; here, we show that this implies u and F_θ are nonparametrically identified as well. Intuitively, the shape of u influences the effect of discounts on consumer purchasing behavior, so identification relies on isolating the causal effect of a price change on a stable set of consumers. The counterfactual CDF of q for uptakers provides just the comparison we need.

To formalize this idea, we define the *uptaker type distribution* as $\tilde{F}_\theta(\theta) \equiv \Pr[\Theta \leq \theta | \upsilon=1]$. Since $q^*(p;\theta)$ is increasing in θ, the distribution of consumption conditional on price p_0 is related to the distribution of underlying θ types by $G_t(q|p_0, \upsilon=1) = \tilde{F}_\theta(p_0/u'(q))$. In particular, since $p_0/u'(q)$ is a monotone increasing function of q (by concavity), taking inverses of the above implies that for any quantile rank $r \in [0,1]$, the r^{th} quantile of consumption, $G_t^{-1}(r|p_0, \upsilon=1)$ is the endogenous choice of the r^{th} quantile θ, $\tilde{F}_\theta^{-1}(r)$. This is true independent of the underlying price, so in particular, the r^{th} quantile uptaker under price p_0 has the same θ type as the r^{th} quantile uptaker under price $p_0(1-d)$, and her type is the r^{th} quantile

the uptake function must in that case also condition on not being an "always taker," so we focus discussion here on one-sided non-compliance.

of $\tilde{F}_\theta$. To simplify notation for the identification discussion, we will drop the treatment arm subscript, t, and the conditioning on $\upsilon=1$, both of which are taken as given.

The first-order condition of the r^{th} quantile consumer at price p is $\tilde{F}_\theta^{-1}(r)u'(G^{-1}(r|p))=p$, and taking the ratio of this FOC at prices p_0 and $p_0(1-d)$ yields

$$\frac{u'(G^{-1}(r|p_0(1-d)))}{u'(G^{-1}(r|p_0))}=1-d, \tag{8}$$

with the unobserved demand type terms $\tilde{F}_\theta^{-1}(r)$ canceling out. D'Haultfœuille and Février (2020) show that Equation (8) provides the basis for the nonparametric identification of u (subject to scale and location normalizations, as in Assumption 1). Here, we provide a sketch of the identification argument for completeness. Differentiating both sides of (8) with respect to d, and taking limits as $d\rightarrow 0$, we have

$$-\frac{u''(G^{-1}(r|p_0))}{u'(G^{-1}(r|p_0))}\frac{\partial G^{-1}(r|p_0)}{\partial p}p_0=1 \Rightarrow \frac{\partial G^{-1}(r|p_0)}{\partial p}\frac{p_0}{G^{-1}(r|p_0)}=-\frac{u'(G^{-1}(r|p_0))}{u''(G^{-1}(r|p_0))G^{-1}(r|p_0)}. \tag{9}$$

In words, Equation (9) states that the elasticity of the r^{th} demand quantile around price p_0 identifies the *relative risk aversion* of u at the r^{th} quantile evaluated at p_0. As long as demand curves are sufficiently smooth—which is consistent with continuous demand CDFs under the model—then the elasticity of the r^{th} demand quantile is well-approximated by $\frac{G^{-1}(r|p_0(1-d))-G^{-1}(r|p_0)}{d\times G^{-1}(r|p_0)}$ when d is small, which shows that $u'(q)/(u''(q)q)$ can be approximated for any q in the support of the data. But knowing $u'(q)/(u''(q)q)$ at each demand level q pins down a (second-order) differential equation for u, which can be solved up to two boundary conditions, which are supplied by the location and scale normalizations $u(0)=0$ and $u'(0)=1$.

With u identified, θ can inferred directly at the individual level: a consumer observed consuming q units when facing price p must have $\theta = p/u'(q)$. But since the r^{th} quantile of demand is chosen by the r^{th} quantile of θ, the quantile function of demand types is uniquely determined by $F_\theta^{-1}(r) = p/u'(G^{-1}(r|p))$, and thus, F_θ is nonparametrically identified as well.

4.2.4. *Recovery of Behavioral Parameters From Uptake Function.* As a final step, we show how one can recover the behavioral primitives of the model using the uptake function, and utility function. Note that the uptake function $\Upsilon_t(q)$ for treatment arm t represents the probability a consumer with baseline demand q (under default linear price p_0) subscribes. This individual has type $\theta = p_0/u'(q)$, and subscribes if and only if either (i) the offer is salient, and the individual acts impulsively, which occurs with probability $\sigma\rho$; or (ii) the offer is salient, the individual does not act impulsively, and forecast error ε satisfies inequality

(2), which occurs with probability $\sigma(1-\rho)H_\varepsilon(\log(\theta)-\log(\tilde{\theta}_{p_0}(S,d_t))$. Thus,

$$\sigma(1-\rho)H_\varepsilon\left(\log(p/u'(q))-\log(\tilde{\theta}_{p_0}(S,d_t))\right)+\sigma\rho=\Upsilon_t(q). \tag{10}$$

Taking limits as $q\to\infty$ yields $\lim_{q\to\infty}\Upsilon_t(q)=\sigma[\lim_{\epsilon\to\infty}\rho+(1-\rho)H_\varepsilon(\epsilon)]=\sigma$. Intuitively, consumers with very high baseline demand will unambiguously know they should subscribe whenever the offer is salient, so their uptake rate pins down σ.

Identification of ρ is in principle similar, but with some additional technical caveats. If negative consumption was possible, then taking limits as $q\to-\infty$, we would have the result $\lim_{q\to-\infty}\Upsilon_t(q)=\sigma\rho$, which would pin down ρ since σ is known. In practice, consumption is lower-bounded by 0; thus, we cannot take this lower limit, but we can identify bounds on ρ. More concretely, we have $\Upsilon_t(0)=\sigma\left[\rho+(1-\rho)H_\varepsilon\left(\log(p_0)-\log(\tilde{\theta}_{p_0}(S,d_t))\right)\right]$, and replacing $H_\varepsilon(\cdot)$ by its lower bound of zero implies that $\Upsilon_t(0)\geq\sigma\rho$. Therefore, the impulsivity parameter must satisfy the following empirically estimable bounds: $\rho\in[0,\Upsilon_t(0)/\sigma]$. The severity of the partial identification problem therefore depends on the mean and variance of ε, but we will see in our rideshare application that the bounds in practice may be fairly tight.

Finally, referring back to Equation (10), since σ, Υ_t, and u are known from previous steps, the conditional distribution $H_\varepsilon\left(\varepsilon\middle|\varepsilon\geq\log(p_0)-\log(\tilde{\theta}_{p_0}(S,d_t))\right)$ is nonparametrically identified up to an affine transformation, where the slope and intercept terms depend on ρ. In summary, there are two partial identification problems, both of which are due to the fact that demand is bounded from below at 0: first, the levels of H_ε depend to some extent on the set-identified impulsivity parameter; and second, the shape of the far lower tail of H_ε is unknown.[9]

4.2.5. *Refining Bounds on ρ and H_ε.* Here we discuss two strategies for mitigating the consequences of partial identification. The first strategy relies on richer instrument variation and an assumption that the salience and impulsivity parameters (σ,ρ) are fixed across multiple treatment arms. If so, then we can obtain tighter bounds on the identified set for ρ as the intersection of the t-specific bounds, or, $\rho\in[0,\min_t\{\Upsilon_t(0)\}/\sigma]$. For some intuition on what type of treatments will lead to tighter bounds on ρ, suppose the true value of ρ is 0, in which case we have $\Upsilon_t(0)=\sigma H_\varepsilon(\log(p_0)-\log(\tilde{\theta}_{p_0}(S,d_t)))$. Holding p_0 fixed, it is clear that $\Upsilon_t(0)$ will be smaller for treatment arms with larger values of the cutoff type $\tilde{\theta}_{p_0}(S,d_t)$, which correspond to smaller discounts d_t (or higher fees S) by Corollary 1. Intuitively, *less* generous

[9]Interestingly, note that if the impulsivity parameter is equal to its upper bound, $\rho=\Upsilon_t(0)/\sigma$, then the forecast error CDF is in fact nonparametrically point identified. For some intuition, recall that the lowest possible demand type is $\underline{\theta}=1/u'(0)$, and note that there are two explanations for why this consumer might still subscribe within the model: either they are impulsive *or* $\varepsilon\leq\log(\underline{\theta})-\log(\tilde{\theta}_p(S_t,d_t))$. When ρ equals its upper bound, the model loads the rationale for $\underline{\theta}$ subscribing solely onto the first explanation, while if $\rho=0$ then the model leans solely on the second explanation.

subscription offers provide more identifying information about ρ because it becomes harder for a forecast error to cause a low-demand individual to nonetheless subscribe.[10]

The second strategy for mitigating partial identification is introducing additional model structure. If one assumes there is no impulsivity, or $\rho = 0$, then the precise levels of H_ε are identified conditional on $\varepsilon \geq \log(p_0) - \log(\tilde{\theta}_{p_0}(S, d_t))$. The only remaining challenge is non-identification in the lower tail, which could be mitigated if the econometrician had a principled method for extrapolation. The following additional assumption provides such extrapolative leverage and has the attractive feature of being testable.

Assumption 3. The forecast error CDF H_ε is symmetric about its median; that is, if $m \equiv H_\varepsilon^{-1}(0.5)$, then $H_\varepsilon(m+\Delta) = 1 - H_\varepsilon(m-\Delta)$.

Since the upper tail of H_ε is point identified, symmetry implies that the lower tail is identified as well, provided that the nonparametric identification cutoff is at least below the median shock, or $\log(p_0) - \log(\tilde{\theta}_{p_0}(S, d_t)) < H_\varepsilon^{-1}(0.5)$. This condition is empirically verifiable, and if it is satisfied one can probe for violations of Assumption 3 by testing the null hypothesis of symmetry on the appropriate set,

$$\mathcal{H}_0 : \sup_{r \in \left[H_\varepsilon\left(\log(p_0) - \log(\tilde{\theta}_{p_0}(S,d_t))\right), 0.5\right]} \left| \left(H_\varepsilon^{-1}(0.5) - H_\varepsilon^{-1}(r)\right) - \left(H_\varepsilon^{-1}(1-r) - H_\varepsilon^{-1}(0.5)\right) \right| = 0.$$

Intuitively, the further apart the median and the identification cutoff ($\log(p_0)$–$\log(\tilde{\theta}_{p_0}(S, d_t))$) are, the more thoroughly one can evaluate the validity of Assumption 3. Thus, by similar logic as above, less generous experimental subscription offers (lower d_t or higher S) and more generous default prices (p_0) facilitate empirical testing of the additional model structure proposed in strategy two for refining partial identification bounds.

5. Estimation and Empirical Results

The constructive identification arguments from the previous section can be used to derive a fairly straightforward procedure for estimating the primitives of consumer demand and behavioral frictions from our experimental data. The estimation proceeds in three stages: first, we extract the relevant counterfactual CDFs; second, we use them to infer the utility function; and third, we recover behavioral parameters.

5.1. Stage 1: Selection Correction.

For treatment arm $t = 1, 2$, we parameterize demand CDFs as cubic B-splines: $\hat{G}(q|1-d_t, \upsilon = 1; \boldsymbol{\omega}_t) \equiv \sum_{k=1}^{K+3} \omega_{tk} \mathcal{B}_k(q)$, and $\hat{G}(q|1, \upsilon = 1; \boldsymbol{\omega}_{0t}) \equiv \sum_{k=1}^{K+3} \omega_{0tk} \mathcal{B}_k(q)$, where basis functions $\boldsymbol{\mathcal{B}} = \{\mathcal{B}_1, \ldots, \mathcal{B}_{K+3}\}$ are determined by the Cox-de

[10]Interestingly for baseline price, lower (i.e., *more* generous) p_0 also delivers more precise bounds on ρ. A lower linear price stimulates non-discounted demand and makes the marginal benefit of buying a subscription smaller; both effects make it less likely that a forecast error will cause low-demand consumers to subscribe.

Boor recursion (see de Boor (2001)) and knot vector $\boldsymbol{\kappa}=\{\kappa_1<\cdots<\kappa_{K+1}\}$, which partitions the support of consumer demand into K uniformly spaced sub-intervals.[11]

Within treatment arm t, we define the following objects: $\tilde{G}_t(q|1-d_t,\upsilon=1)$ is the Kaplan-Meier empirical CDF of consumption among uptakers, $\tilde{G}_t(q|1,\upsilon=0)$ is the empirical CDF of consumption among non-uptakers, $\tilde{\tau}_t$ is the proportion of uptakers, and $\tilde{G}_{ctrl}(q)$ is the empirical CDF of consumption under control (default price). Given a grid of domain points $\{q_1,\dots,q_J\}$, the B-spline forms facilitate a constrained GMM estimator of the B-spline parameters $\boldsymbol{\omega}_t$ and $\boldsymbol{\omega}_{0t}$:

$$\begin{aligned}(\hat{\boldsymbol{\omega}}_t,\hat{\boldsymbol{\omega}}_{0t}) =& \arg\min_{(\boldsymbol{\omega}_t,\boldsymbol{\omega}_{0t})}\Bigg\{\sum_{j=1}^J\left(\hat{G}(q_j|1-d_t,\upsilon=1;\boldsymbol{\omega}_t)-\tilde{G}(q_j|1-d_t,\upsilon=1)\right)^2\\ &+\sum_{j=1}^J\left(\hat{G}(q_j|1,\upsilon=1;\boldsymbol{\omega}_{0t})-\frac{\tilde{G}_{ctrl}(q_j)-(1-\tilde{\tau}_t)\tilde{G}_t(q_j|1,\upsilon=0)}{\tilde{\tau}_t}\right)^2\Bigg\},\\ &\textit{subject to}:\ \omega_{tk}\le\omega_{t,k+1},\ \ \omega_{0tk}\le\omega_{0t,k+1},\ k=1,\dots,K+2;\\ &\omega_{0t1}\ge 0,\ \ \omega_{t1}\ge 0;\ \text{ and }\ \omega_{0t,K+3}=1,\ \ \omega_{t,K+3}=1.\end{aligned} \tag{11}$$

The constraints enforce monotonicity and boundary conditions required to obtain estimates that constitute valid CDFs.

We also specify the empirical uptake function $\hat{\Upsilon}_t(q;\boldsymbol{\omega}_{\upsilon t})\equiv\sum_{k=1}^{K_\upsilon+3}\omega_{\upsilon tk}\mathcal{B}_{\upsilon k}(q)$, for $t=1,2$ as another flexible cubic B-spline, with knot vector $\boldsymbol{\kappa}_\upsilon=\{\kappa_{\upsilon 1}<\cdots<\kappa_{\upsilon,K_\upsilon+1}\}$, which pins down basis functions $\boldsymbol{\mathcal{B}}_\upsilon=\{\mathcal{B}_{\upsilon 1},\dots,\mathcal{B}_{\upsilon K_\upsilon+3}\}$. We similarly estimate its parameters via GMM:

$$\hat{\boldsymbol{\omega}}_{\upsilon t} = \arg\min_{\boldsymbol{\omega}_{\upsilon t}}\sum_{j=1}^J\left(\hat{\Upsilon}_t(q_j;\boldsymbol{\omega}_{\upsilon t})-\frac{\tilde{\tau}_t\hat{g}_t(q_j|1,\upsilon=1;\hat{\boldsymbol{\omega}}_{0t})}{\hat{g}_{ctrl}(q_j;\hat{\boldsymbol{\omega}}_{ctrl})}\right)^2, \tag{12}$$

$$\textit{subject to}:\omega_{\upsilon k}\le\omega_{\upsilon,k+1},\ k=1,\dots,K+2;\ \ \omega_{\upsilon t,1}\ge 0,\omega_{\upsilon t,K_\upsilon+3}\le 1,$$

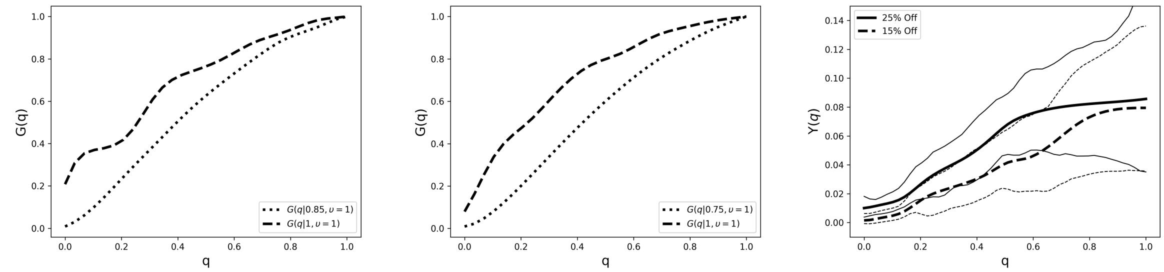
where $\hat{G}_{ctrl}(\cdot;\hat{\boldsymbol{\omega}}_{ctrl})$ is a best-fit B-spline estimate of $\tilde{G}_{ctrl}(q_j)$, and $\hat{g}_{ctrl}(\cdot;\hat{\boldsymbol{\omega}}_{ctrl})$ is its density. The results from this first stage are plotted in Figure 2 below.

5.2. **Stage 2: Utility Estimation.** We specify utility as a quartic B-spline, $\hat{u}(q;\boldsymbol{\omega}_u)=\sum_{k=1}^{K_u+4}\omega_{uk}\mathcal{B}_{uk}(q)$ based on knot vector $\boldsymbol{\kappa}_u$, which ensures that $\hat{u}$ is three times continuously differentiable everywhere on its domain. Recall that the CDF $\hat{G}_t(q|1,\upsilon=1;\hat{\boldsymbol{\omega}}_{0t})$ is a selection-corrected counterpart to the treatment CDF $\hat{G}_t(q|1-d_t,\upsilon=1;\hat{\boldsymbol{\omega}}_t)$, for an identical sub-population of control consumers who would have purchased a subscription, had they received the offer. For treatment arm t, define the *quantile projection at q* by

[11] Previous empirical studies that used a similar B-spline fitting approach have recommended choosing knots to be uniformly spaced in quantile ranks of the raw data (e.g., see Hickman et al. (2017); Bodoh-Creed et al. (2021); Cotton et al. (2026)) so as to balance out the influence of finite data over the various B-spline weights. In our case, the high sample size of roughly a million consumers renders this concern moot.

Figure 2. Uptaker Consumption Distributions

(A) Subscr. Agg. CDFs (B) Subscr. Uptaker CDFs (C) Uptake Functions



Notes: Panel (A) compares consumption CDFs with and without a discount amongst uptakers in treatment 1. Panel (B) compares consumption CDFs with and without a discount amongst uptakers in treatment 2. Panel (C) displays the estimated uptake functions (thick lines) for both treatment arms, with bootstrapped confidence bands (thin lines, 2000 bootstrap samples).

$\tilde{\mathcal{P}}_t(q) \equiv \hat{G}_t^{-1}\left[\hat{G}_t(q|1-d_t, \upsilon=1; \hat{\boldsymbol{\omega}}_t)\middle|1, \upsilon=1; \hat{\boldsymbol{\omega}}_{0t}\right]$. According to our model of demand, this mapping tells us how much an individual consuming q at price $(1-d_t)$ would counterfactually consume if the price was $p=1$ instead. For a given guess of utility parameters $\boldsymbol{\omega}_u$, we can define a model-generated analog of this quantity: $\hat{\mathcal{P}}_t(q;\boldsymbol{\omega}_u) \equiv (u')^{-1}\left(\frac{1}{1-d_t}u'(q;\boldsymbol{\omega}_u);\boldsymbol{\omega}_u\right)$.

At the true utility parameters, the empirical and model-generated projections should (asymptotically) coincide, so we can pin down $\boldsymbol{\omega}_u$ by minimizing the ℓ_2 distance between $\tilde{\mathcal{P}}_t$ and $\hat{\mathcal{P}}_t$. Letting $0=q_1<q_2<\ldots<q_m=q_{max}$ be a uniformly spaced grid of points, we have:

$$\begin{aligned} \hat{\boldsymbol{\omega}}_u =& \arg\min_{\boldsymbol{\omega}_u} \textstyle\sum_{t=1,2}\sum_{j=1}^{m}\left(\hat{\mathcal{P}}_t(q_j;\boldsymbol{\omega}_u) - \tilde{\mathcal{P}}_t(q_j)\right)^2 \\ & \textit{subject to}: \ \ \omega_{u1}=0, \ \ \omega_{u1}=\tfrac{\kappa_{u5}-\kappa_{u2}}{3}, \ \ \omega_{uk}\leq\omega_{u,k+1}-\epsilon, \ k=1,\ldots,K_u+3, \ \epsilon>0, \text{ and} \\ & \tfrac{\omega_{uk}-\omega_{u,k-1}}{\kappa_{u,k+3}-\kappa_{uk}} \leq \tfrac{\omega_{u,k+1}-\omega_{uk}}{\kappa_{u,k+4}-\kappa_{u,k+1}}-\epsilon, \ k=2\ldots,K_u+3, \ \epsilon>0, \end{aligned} \tag{13}$$

where the constraints enforce, in the following order, the utility boundary level and boundary derivative, as well as monotonicity and concavity of the utility function.[12,13]

In our data, we have access to two treatment arms, which generate two separate exogenous shifts in price. Using data from only one of these two exogenous shifts alone suffices to identify utility, which gives rise to an over-identifying test of our model. Specifically, we can re-estimate an analog of (13) twice by only minimizing the GMM objective for one of the two treatments $t\in\{1,2\}$ at a time. We can then compare the two utility estimates with one another to test whether they are the same, modulo finite sampling variance.

5.3. **Stage 3: Behavioral Parameter Estimation.** The last set of structural primitives we must recover are the behavioral parameters. The salience parameter is estimated simply

[12]The boundary derivative normalizes the demand type of the marginal consumer under p to one, or $\theta^*(0,1)=1$. Thus, all estimated θ's are relative to this marginal reference consumer.

[13]Knots were chosen so that $\boldsymbol{\kappa}=\boldsymbol{\kappa}_\upsilon=\boldsymbol{\kappa}_u$ and $K=K_\upsilon=K_u=9$ in order to facilitate comparisons. This afforded high flexibility and additional knots made little difference. The number of objective function (13) evaluations should be $m\geq K_u$; we chose $m=50$. For the tolerance on the shape constraints we chose $\epsilon=10^{-6}$.

by taking the value of the uptake function at the largest consumption basket in the data, $\bar{q}$: $\hat{\sigma}_t = \hat{\Upsilon}_t(\bar{q}; \hat{\boldsymbol{\omega}}_{vt})$. Similarly, we can estimate the upper bound on ρ as $\hat{\bar{\rho}}_t = \hat{\Upsilon}_t(0; \hat{\boldsymbol{\omega}}_{vt})/\hat{\sigma}_t$.

Holding fixed some $\rho \in [0, \hat{\bar{\rho}}_t]$, we can also derive a corresponding estimate of H_ε. We choose a knot vector $\boldsymbol{\kappa}_h = \{-\kappa_{h,K_h/2}, \ldots, -\kappa_{h,1}, \kappa_{h,1}, \ldots, \kappa_{h,K_h/2}\}$ which is centered at and symmetric around 0 (with K_h as an even number) to facilitate a test of symmetry (see Section 5.4.1). Letting $\mathcal{B}_h$ be the quadratic B-spline basis corresponding to this knot vector, we parameterize the forecast error CDF as $\hat{H}_\varepsilon(\epsilon; \boldsymbol{\omega}_h) = \sum_{k=1}^{K_h+2} \omega_{hk} \mathcal{B}_{hk}(\epsilon - \omega_{h0})$, where ω_{h0} is an additional parameter that represents the midpoint of the support, and corresponds to the median shock if H_ε is a symmetric distribution.[14] We estimate H_ε via GMM by matching the left-hand side of (10) to the (pre-estimated) right-hand side as follows:

$$\begin{aligned} \hat{\boldsymbol{\omega}}_h =& \arg\min_{\boldsymbol{\omega}_h} \sum_{t=1,2} \sum_{j=1}^{m} \left(\hat{\sigma}_t (1-\rho) \hat{H}_\varepsilon \left(-\log \left[\hat{u}'(q; \hat{\boldsymbol{\omega}}_u) \tilde{\theta}_1(S_t, d_t; \hat{\boldsymbol{\omega}}_u) \right]; \boldsymbol{\omega}_h \right) + \hat{\sigma}_t \rho - \hat{\Upsilon}_t(q; \hat{\boldsymbol{\omega}}_{vt}) \right)^2 \\ & \textit{subject to}: \quad \omega_{h1} = 0, \omega_{hK_h+2} = 1, \text{ and } \omega_{hk} \le \omega_{h,k+1}, k = 1, \ldots, K_h + 1, \end{aligned} \tag{14}$$

where $\tilde{\theta}_1(S_t, d_t; \hat{\boldsymbol{\omega}}_u)$ is the empirical analog of the cutoff demand type that implies equality in (1), given default price $p=1$, and subscription offer (S_t, d_t).

Treating our B-spline specifications as fixed parametric forms, standard asymptotic theory for GMM estimators applies (see Davidson and MacKinnon (2021), Section 9.5).[15] The estimator $\hat{\boldsymbol{\omega}}_h$ defined above is $\sqrt{N}$-consistent and asymptotically normal under the usual regularity conditions that the model is correctly specified and identified (see Section 4.2.5). In order to assess the role of sampling variability while accounting for the role of estimation error from previous stages, one can estimate confidence bounds on parameters via the bootstrap. Although we have set up the GMM objective (14) to have a common $\boldsymbol{\omega}_h$ vector, we can in principle estimate it separately for each treatment arm in order to facilitate a model test, by comparing behavioral parameter estimates across t, including $\hat{\sigma}_t$, $\hat{\bar{\rho}}_t$, and $\hat{\boldsymbol{\omega}}_{ht}$.

5.4. **Preliminary Model Estimates.** As a preliminary analysis, we begin with a series of model tests, holding the impulsivity parameter fixed at its upper bound. We plot treatment-arm-specific utility estimates (thick lines) and 95% confidence bands (thin lines) in Panel A of Figure 4, with corresponding elasticity functions—$u'(q)/(qu''(q))$—in panel B. In both

[14] Recall from Section 4.2.5 that there is a partial identification problem: the lower bound of the shock support is not known ex-ante. This problem is resolvable if $\rho = 0$ and H_ε is a symmetric distribution; our specification here is designed to facilitate empirical evaluation of these conditions.

[15] One could alternatively view our B-spline specifications as enabling a sieve estimator whose parametric form grows in complexity as the sample size increases. While asymptotic theory exists to formally justify this view for orthogonal polynomial bases (e.g., Gallant and Nychka (1987)), it remains an open econometric problem for B-splines. Since B-splines have advantages over orthogonal polynomials in terms of numerical performance, we adopt the view that our B-spline specifications correspond to a flexible but finite-dimensional parametric form, which we impose for estimator tractability.

TABLE 2. Behavioral Parameters

Parameter	Point Estimate, t_1	95% CI, t_1	Point Estimate, t_2	95% CI, t_2
σ	0.17	[0.07, 0.28]	0.12	[0.07, 0.22]
$\bar{\rho}$	0.02	[0.00, 0.43]	0.27	[0.00, 0.41]
Mean of H_ε	0.90	[0.20, 1.01]	0.45	[0.20, 1.14]
Std. Dev. of H_ε	0.21	[0.07, 0.27]	0.26	[0.16, 0.27]

Notes: This table reports point estimates and bootstrap 95% confidence intervals for behavioral parameters, where the estimates are derived separately for the two treatment groups.

panels, we see that point estimates from one treatment are well within in the confidence bands of the other, so we cannot reject the null hypothesis that a common utility function generated both datasets. In Table 2, where we report estimates and confidence intervals of our behavioral parameters, we see similar overlap in the sampling distributions of $\hat{\sigma}$ and $\hat{\bar{\rho}}$ across treatment arms. We also report mean and standard deviation estimates of the resulting forecast error distributions, and again we cannot reject equality of behavioral parameter estimates across treatment arms. With these various results in mind, our main specification will jointly estimate all parameters using both treatments.

5.4.1. *Empirical Specification Refinements.* Recall that if consumers exhibit the same impulsivity across treatment arms—a hypothesis which is theoretically supported by individual-level randomization, and which we fail to statistically reject in Table 2—then we can refine the upper bound on ρ by taking the minimum of $\hat{\bar{\rho}}_t$ across treatment arms, which is small at 0.02. In other words, this estimate suggests that 2% of observed subscription decisions, at most, are attributable to impulsive buying rather than reasoned weighing of costs and benefits by consumers. Since impulsivity leads to conceptual complications in designing nonlinear pricing policy, and since the estimated upper bound leaves little latitude for it to change our estimate of the role of forecast errors, we henceforth assume a value of $\rho = 0$.[16]

There is a final partial identification challenge to address: $H_\varepsilon(\cdot)$ is only nonparametrically identified for values of $\varepsilon > \log(p_0) - \log(\tilde{\theta}_{p_0}(S,d)) = \underline{\varepsilon}$. Recalling Section 4.2.5, point identification is still possible if the shock CDF is symmetric. To probe for whether this assumption is reasonable within our data, we propose a formal symmetry test in Appendix B. In brief, we define $m \equiv H_\varepsilon^{-1}(0.5)$ as the median of $H_\varepsilon(\cdot)$, and note that symmetry implies $H_\varepsilon^{-1}(r) - m = m - H_\varepsilon^{-1}(1-r)$ should be true for each quantile rank $r \in (H_\varepsilon^{-1}(\underline{\varepsilon}), 0.5)$. We conduct a joint test of the null hypothesis that this system of equalities is satisfied using our B-spline specification of H_ε (see Section 5.3) with $K_h = 7$ sub-intervals, and 5 free parameters after imposing shape and boundary conditions.

[16]To understand the conceptual tension in nonlinear pricing policy with impulsive subscribers, consider an extreme case where $\rho > 0$ and the firm sells a subscription with discount $d = 0$ and upfront cost S. Then no consumers will purchase this subscription *except* for those driven by impulsivity. Such a firm could increase profits by $S\rho\sigma$, which implies incentives to set a high value of S just to exploit impulse subscribers. We view this as a tenuous and short-sighted business model, which we rule out by imposing $\rho = 0$.

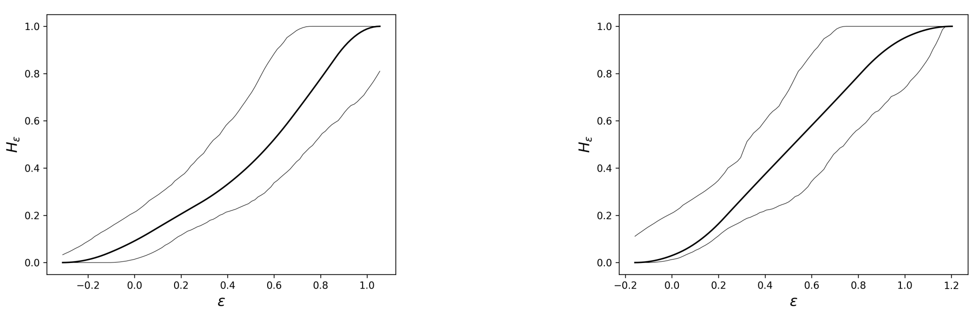
Table 3. Behavioral Parameters

Parameter	Point Estimate	95% CI
Mean of H_ε	0.37	[0.23, 0.67]
Std. Dev. of H_ε	0.35	[0.25, 0.57]
σ	0.14	[0.09, 0.21]

Notes: This table reports point estimates and bootstrap 95% confidence intervals for estimated behavioral parameters.

Full technical details on our procedure are described in Appendix B. We fail to reject the null hypothesis of distributional symmetry, with a p-value of 0.895. Thus, following the logic of Section 4.2.5 our structural model is point identified, and in our main specification we impose an additional set of constraints on (14): for our centered B-spline form, symmetry of $H_\varepsilon(\cdot;\boldsymbol{\omega}_h)$ is equivalent to $\omega_{hk}+\omega_{h,K_h+3-k}=1,\ k=1,\ldots,K_h+1$. Panel (A) of Figure 3 plots the unrestricted H_ε estimate used for the test, and panel (B) plots the estimate with symmetry constraints imposed.

Figure 3. H_ε



Notes: The shock CDF estimate in Panel (A) is the baseline unrestricted B-Spline specification used to test symmetry. The Panel (B) estimate is from a symmetric B-spline specification having 5 free parameters. Thick lines are point estimates and thin lines are 95% confidence bands.

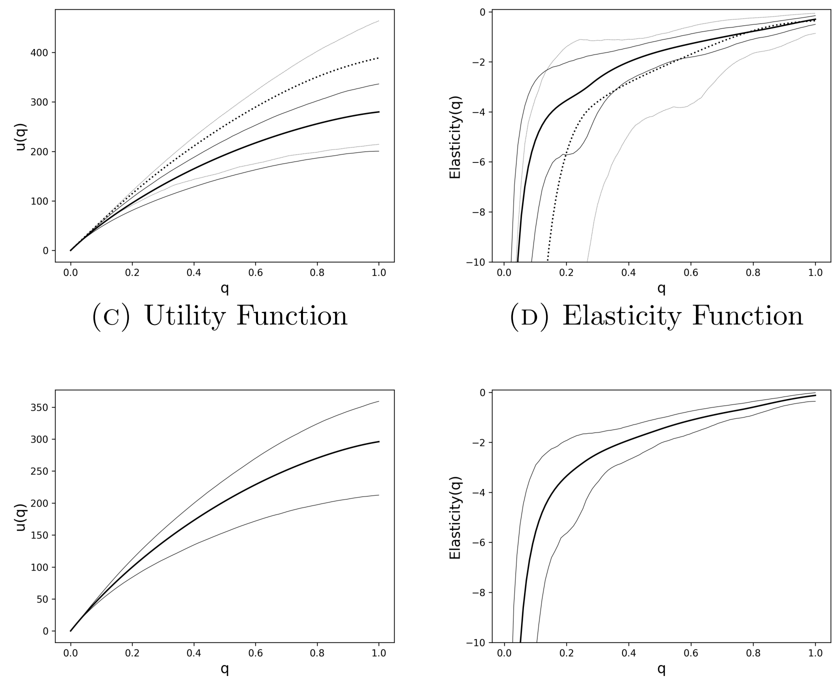
5.5. **Main Specification Results.** Panels C and D of Figure 4 plot the estimated utility and elasticity functions (with 95% confidence bands), using the joint (multi-arm) specification. Consumers across the distribution are highly elastic, with especially high elasticities found for smaller values of q. The average elasticity across the entire distribution of θ types is -3.57. In Table 3 we report mean and standard deviation of forecast error, along with our joint estimate of the salience parameter σ. To contextualize $\hat{H}_\varepsilon$, under multiplicatively separable utility, an $X\%$ increase in the type θ induces the same demand response as an $X\%$ decrease in price. Since ε represents deviations in log-θ space, a consumer with forecast error $\varepsilon=0.1$ forecasts her demand as if prices were 10% larger than they are in practice.

We highlight three striking findings from these results. First, salience of subscriptions amongst consumers is low: only about 14% of the consumer base was fully cognizant of the subscription offer during the experimental period. Second, consumers are on average

Figure 4. Utility and Elasticity

(a) Utility Function (b) Elasticity Function

(c) Utility Function (d) Elasticity Function



Notes: Thick lines are point estimates and thin lines are 95% confidence bands. In Panels A and B, dotted lines are estimates from treatment arm 1 and solid lines are estimates from treatment arm 2.

significantly biased in their ex-ante forecasts of demand over the ensuing month. The estimated mean of ε is 0.37, meaning the average consumer forecasts their demand as if default price was 37% higher than it is in reality. Given the average price elasticity of -3.57, another way to interpret this number is that, when making subscription decisions, an average consumer under-estimates their own ex-post (monthly) rideshare demand by about $100 \times \frac{3.57 \times 0.37}{1+3.57 \times 0.37} \approx 57\%$. Third, consumers are highly variable in their forecasts of monthly consumption, as the standard deviation of ε is 0.35, or nearly as large as its mean.

Overall, these findings indicate that behavioral mistakes tend to mitigate the profitability of nonlinear pricing by depressing demand for subscriptions through multiple mechanisms. Many consumers are not fully aware of the opportunity to subscribe, and even if they are they require a high degree of certainty that they will save money before being willing to pay the subscription fee S. Our results are complementary to empirical findings in a closely related paper, Grubb and Osborne (2015), who study optimal nonlinear pricing under biased consumer beliefs within the context of demand for cellular phone service.

Grubb and Osborne (2015) find that consumers *overestimate* their demand when deciding which cellphone plan is right for them, often choosing a higher tier plan than a rational benchmark would imply. While this tendency points in the opposite direction of our finding that rideshare consumers *underestimate* their future demand, this difference could arise as a consequence of disparities in how price contracts are structured within the two settings. In deciding on rideshare subscriptions, the consequence of buying a lower tier contract than is optimal (including not subscribing) is simply having to pay more for rides in the ensuing month. In contrast, the consequence of buying a sub-optimally low tier phone plan is that the consumer must live with low-quality service for the entire month, arguably a more consequential mistake.[17] This would be equivalent to Lyft offering a subscription plan to its users where they must commit to a fixed (v, q) (uptake-consumption) *pair* at the beginning of the month, rather than just deciding on subscription v first, and optimizing q in real time.

Taken together, our study and Grubb and Osborne (2015) may suggest a subtle insight on contract design in the presence of behavioral phenomena: consumers may be more naturally biased toward over-spending when contractual commitments are all ex-ante, and biased toward under-spending when ex-ante commitments are only partial. Implications for firms in both settings can be substantial: in cellular phone demand with full up-front commitment, behavioral phenomena augment firm profits, while in rideshare demand with partial up-front commitment, they reduce profits. We quantify excess profits from nonlinear pricing (relative to linear pricing) through a series of counterfactuals in the following section, with explorations of behavioral implications, and internal firm strategies to cope with them.

6. Counterfactuals

We now use our estimated model to explore optimal nonlinear pricing, implemented as a subscription program. To evaluate policy, we need to specify both demand and costs, but in order to protect Lyft's internal data confidentiality, we do not use raw production cost data here. Rather, we assume Lyft faces a constant marginal cost of c, and we follow a conventional cost-imputation approach from the demand estimation literature (e.g., Nevo (2000); Ackerberg et al. (2007), see Appendix C for details) based on an assumption of linear price optimality. Thus, marginal costs are set at $c = 1 - \frac{1}{\zeta} = 0.28$ where $\zeta = 3.57$ is the average elasticity of aggregate demand at baseline price $p = 1$.[18]

In Section 6.1, we compute a benchmark where consumers are perfectly rational, and the firm designs a continuum menu of subscriptions to achieve a fully separating equilibrium. Motivated by the fact that real-world firms tend to view such complexity as operationally costly and potentially overwhelming to consumers, Section 6.2 studies a more realistic case

[17]Another salient empirical difference between cellphones and rideshare is that commitments most often are longer lived in the former, with duration over multiple months or years.

[18]Based on conversations with rideshare industry insiders, we view constant marginal costs as a reasonable approximation in our setting. In theory, extremely large shocks to demand may affect marginal costs by causing congestion on the platform, but empirically, such effects tended to be small during our sample period.

where the firm optimally chooses a one- or two-tiered subscription offer instead. Finally, Section 6.3.1 adds the behavioral frictions into the optimal pricing counterfactual analysis.

6.1. **Optimal Continuum Menu of Subscription Offers.** Absent behavioral mistakes, our pricing problem is a special case of the Maskin and Riley (1984) (MR84) framework, so we only sketch the key points here (see Appendix D for additional detail). We first derive a profit-maximizing, continuous menu of subscription offers to produce a fully self-sorting equilibrium by consumer types. The basic idea is that a firm's choice of tailored fee-discount offers $(S(\theta), d(\theta))$ as a function of θ is pinned down by an analog of the inverse-elasticity markup rule for monopoly pricing. The revelation principle implies that an optimal menu can be parameterized by a *marginal price function* $p^*(\theta)$, which describes the per-unit price paid by a consumer of type θ, with $p^*(\underline{\theta})$ as default price.

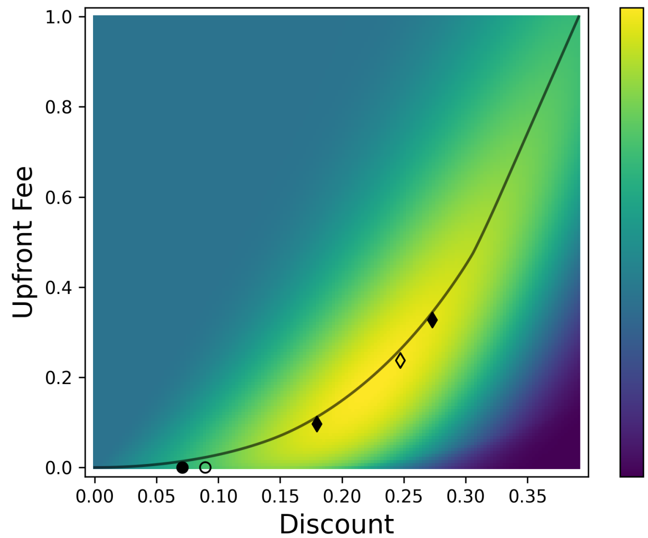
Interpreting the survivor function $1 - F_\theta(\theta)$ as a demand curve, let $\zeta(\theta) = -\frac{1-F_\theta(\theta)}{\theta f_\theta(\theta)}$ be the corresponding elasticity function. Note that incentive compatibility (IC) requires that $p^*(\theta)$ must be decreasing in θ; otherwise higher demand types $\theta' > \theta$ would have incentive to pose as a low-demand θ type. To form an initial candidate solution, suppose the firm could choose optimal price while disregarding IC; then its first-order condition would imply $\tilde{p}(\theta) \equiv \frac{c}{1+1/\zeta(\theta)}$ is optimal.[19] If this function is non-monotone, then the IC constraint binds, and some "ironing" is necessary (e.g., Myerson (1981)), meaning that the optimal price rule $p^*(\theta)$ modifies $\tilde{p}$ to remove the non-monotonicity.[20] IC binds for a range of low θ types in our empirical application, so we develop a numerical method (detailed in Appendix D.1) for computing the optimal menu in the presence of ironing.

Having solved for $p^*(\theta)$, which includes optimizing default price, $p_0^* = p(\underline{\theta})$, we can now derive $(S^*(\theta), d^*(\theta))$. For any $\theta > \underline{\theta}$, $d^*(\theta) = 1 - \frac{p^*(\theta)}{p^*(\underline{\theta})}$ represents the tailored discount value. Then, the optimal subscription fee $S^*(\theta)$ is pinned down by a combination of the participation and IC constraints. Participation implies a boundary condition $S^*(\theta) = 0$ whenever $d^*(\theta) = 0$, while IC implies the ordinary differential equation $S^{*\prime}(\theta) = -p^{*\prime}(\theta) q^*(p^*(\theta), \theta)$. Solving this boundary value problem gives the optimal fee schedule, $S^*(\theta)$.

[19]To see why, we can formally state the profit function by first defining $P(q)$ as a (nonlinear) total payment as a function of a given consumer's q choice. Then, with a slight alteration of our previous notation, we can re-define $q^*(P;\theta) \equiv \text{argmax}_q\, \theta u(q) - P(q)$, from which it follows that the profit function is given by $\int_{\underline{\theta}}^{\bar{\theta}} P(q^*(P;\theta)) - cq^*(P;\theta)\, \mathrm{d}F_\theta(\theta)$. Taking an unconstrained first-order condition to get a candidate optimal $P^*(q)$, the candidate marginal price function above is then given by $\tilde{p}(\theta) = P^{*\prime}(q^*(P^*;\theta))$.

[20]Note that MR84 consider a slightly broader class of nonlinear pricing schemes, whereas we restrict attention to subscriptions, given their popularity as a real-world pricing tool. Model estimates allow one to empirically evaluate *how* restrictive this is. Excess profits would be nearly identical under the fully general MR84 setup, since ironing is only required for a range of low θ types, which produce a tiny fraction of excess profits.

Figure 5. Profits from Subscription Pricing



Notes: The horizontal axis represents discounts (d) while the vertical axis represents the corresponding upfront fee (S). The curve represents the optimal continuum menu. The heatmap represents profits from single-offer (S, d) pairs, holding default price fixed the continuum-optimal $p^*(\underline{\theta})$. The diamonds represent optimal finite menu offers, and circles are optimal finite-menu default prices. open shapes correspond to the single-offer case, and filled shapes are for the two-offer case.

6.1.1. *Continuum Menu Results.* Using our structural estimates, the optimal menu involves increasing the default price $p^*(\underline{\theta})$ by 18.2% relative to the linear-price baseline, and then offering a menu of subscriptions, represented by the curve of (S^*, d^*) pairs plotted in Figure 5. For Lyft's internal data confidentiality, we report fees S^* as a fraction of $S^*(\overline{\theta})$. The optimal menu produces excess profits of 5.71% relative to the linear-price status quo, and it exhibits the "no-distortion-at-the-top" property familiar to mechanism design: the demand type $\overline{\theta}$ receives a subscription offer where marginal price equals marginal cost: $p^*(\overline{\theta}) = c$. Intuitively, the continuum-optimal default price is set *above* the linear-price optimum because nonlinear pricing produces excess profits by inducing high-demand types to self-select into higher-fee offerings. Increasing the default price amplifies the self-selection incentive by increasing the difference between marginal prices paid under different points on the subscription menu.

6.2. **Optimal Finite Menu of Subscription Offers.** While a continuum menu is interesting to market-design researchers, real-world practitioners (e.g., Costco, Lyft, Charge Point, etc.) typically offer one or a small number of subscription options. Firms may prefer simplicity for ease of implementation, or because customers find it difficult to select their optimal choice from a continuum. This raises the question: how much profit is left on the table when a firm forgoes a fully separating equilibrium in favor of a more parsimonious subscription program? To answer this question, it will help to introduce some additional notation: let $\mathcal{M}=\{(S_\alpha,d_\alpha)\}$, $\alpha=0,1,\dots,A$, denote a finite subscription offer menu. To simplify notation later on, $\mathcal{M}$ is assumed to always contain the offer $(S_0,d_0)=(0,0)$, which represents purchasing no subscription and consuming under default price p_0.

For a given (p_0,S,d) triple, let $V(p_0,S,d;\theta)\equiv\max_q\theta u(q)-p_0(1-d)q-S$ be the indirect utility from a type θ consumer who purchases a subscription (S,d) when default price is p_0. A rational type-θ consumer, presented with default price p_0 and menu $\mathcal{M}$ will decide to buy subscription $\alpha^*(\theta;p_0,\mathcal{M})\equiv\arg\max_{\alpha\in\mathcal{M}}V(p_0,S_\alpha,d_\alpha;\theta)$. Upon purchasing this subscription, the consumer will buy $q^*(p_0(1-d_{\alpha^*(\theta;p_0,\mathcal{M})});\theta)$ units of consumption. The profits from offering subscription menu $\mathcal{M}$ and setting default price p_0 is given by

$$\pi\left(p_0,\mathcal{M}\right)=\int_{\underline{\theta}}^{\bar{\theta}}\left\{\left(p_0(1-d_{\alpha^*(\theta;p_0,\mathcal{M})}\right)-c)q^*(p_0(1-d_{\alpha^*(\theta;p_0,\mathcal{M})});\theta)+S_{\alpha^*(\theta;p_0,\mathcal{M})}\right\}dF_\theta(\theta). \quad (15)$$

The above integral can be directly computed for any choice of $(p_0,\mathcal{M})$. We numerically optimize to find the optimal subscription policy, given some cardinality constraint A. We will refer to menus with $A=1$ and $A=2$ as "one-offer" and "two-offer" menus, respectively, with the understanding that default price is freely available to all consumers in both cases.

6.2.1. *Finite Menu Results.* The first row of Table 4 reports *excess profits* (relative to linear pricing) under an optimal single-offer menu: 4.63%, or roughly 82% of excess profits from an optimal continuum menu.[21] For illustrative purposes, and to further elucidate the connection between optimal single-offer and continuum menus, we construct a heatmap of excess profits based on a fine grid of 151 × 151 (S,d) pairs. For each point in the grid, we compute profits based on Equation (15), assuming that default price is set at the continuum-optimal $p^*(\underline{\theta})$. Figure 5 plots the results, with warmer colors representing higher profits, cooler colors representing lower profits, and the shade in the northwest corner—an offer with a high fee and zero discount—corresponding to baseline profits from linear pricing.

The heatmap delivers a number of insights about nonlinear pricing design. First, there is a significant region of single subscription offers that all render high profits, with all being close to one from the optimal continuum menu. Second, the southeast corner of the heatmap shows far lower profits than the northwest corner, which illustrates a decidedly asymmetric loss function: overly generous subscription offers are worse than overly stingy offers. Intuitively, the worse-case for an overly stingy subscription is no uptakers, whereas an overly generous subscription could potentially lead the firm to forfeit *all* profits.

[21]This finding, that a well-chosen low-dimensional subscription offer captures a large fraction of the benefits of a continuum menu, is consistent with results from previous empirical studies of nonlinear pricing by Luo et al. (2018) and Miravete and Röller (2003), though our data allow us to rely on fewer modeling assumptions.

When the firm is constrained to offer only a single subscription, it will choose to set a default price p_0^* that is 8.9% *lower* than $p^*(\underline{\theta})$, corresponding to the open dot at location $(S_0^*, d_0^*) = (0, 0.089)$ in Figure 5. By similar intuition as before, one would expect the single-offer optimal default price to lie between $p^*(\underline{\theta})$ and the optimal linear price. This is because constraining the menu to contain only a single subscription reduces the benefits to the seller of inducing selection of high-demand types into higher-fee offers. The optimal single-option offer, represented by the open diamond in the figure, can then be represented as setting discount d_1^* at roughly 17.3% below p_0^*—or roughly 24.7% below $p^*(\underline{\theta})$—and choosing a fee S_1^* just below the corresponding value on the continuum menu.

The optimal two-offer menu—the solid dots/diamonds in Figure 5—involves two fee-discount pairs close to the continuum menu, with one (lexicographically) above and one below the one-offer fee-discount pair. The 2-offer optimal baseline price is naturally closer to the continuum baseline price—$(0, 0)$ in the plot—as this reflects the firm's amplified incentive to induce self-selection when there are more subscription tiers for high-demand consumers to sort into. This two-offer menu achieves excess profits of 5.27%, or roughly 94% of the continuum case, with a more locally-tailored set of options for potential subscribers. The empirical finding that menus with only one or two offers capture nearly the full excess profits of a continuum menu provides a rationale for firms' strong tendency toward simple pricing structures. While recent behavioral work shows that cognitive costs and complexity aversion can substantially affect consumer behavior—e.g., Gabaix (2014); Gabaix and Graeber (2024); Gabaix (2025)—our study offers a complementary insight. Even modest complexity costs can make contract menus with many price options unattractive relative to simple pricing schemes that perform nearly as well when consumers are fully ex-post rational.

6.3. **Optimality and Managerial Coordination Under Behavioral Mistakes.** In all of the settings studied so far, the optimal $(p_0^*, \mathcal{M}^*)$ pair involved choosing a default price higher than the original price, $(p_0^* > 1)$. In practice, the two roles of optimizing p_0 versus optimizing (S, d) are filled by two separate teams within the firm: the *pricing team* and the *promotions team*. Moreover, upon launch of the subscription plan $(p_0^*, \mathcal{M}^*)$, addressing the behavioral issues which tend to limit profitability—salience and customers' tendency to under-estimate their future demand—falls under the purview of a third corporate unit, the *marketing team*. This organizational structure—known as the "M-form" (or "multi-divisional") firm composed of relatively self-contained sub-units in charge of their own specialized tasks—is common in large modern corporations, including Lyft, Uber, Amazon, Walmart, and many others. The idea that the M-form firm creates beneficial intra-firm incentives and information flows at large scale has been documented in the organizational economics literature (Williamson, 1975; Maskin et al., 2000; Mahajan et al., 1988), and is part of standard curriculum in modern business education.

Despite its benefits, the M-form firm has the disadvantage of requiring coordination across different corporate units in scenarios such as optimal pricing. This challenge may arise from intra-firm information frictions, as recently noted by Hortaçsu et al. (2024) in an empirical study of the airline industry. In our setting, the nature of the profit function implies important complementarities between the pricing, promotions, and marketing teams, which arise from demand-side phenomena. We explore this idea in a set of counterfactuals in this section, where we show that nonlinear pricing can be a highly profitable alternative to linear pricing under full intra-firm coordination, but quickly loses its benefits otherwise.

As a preliminary illustration, we re-compute our optimal finite subscription menu while constraining the default price to be at its status-quo value of $p_0 = 1$. In that scenario, the best one-offer menu improves profits by only 3.15%, and the best two-offer menu improves profits by 3.39%. In both cases, this is roughly two thirds of maximal excess profits when p_0 is jointly optimized. Of course, both calculations assume that consumers respond to subscription offerings in a way that is fully ex-post optimal for them. As we will see next, when consumers exhibit behavioral mistakes in their uptake decisions, intra-firm coordination becomes more important still, requiring cooperation with the marketing division, in addition to the pricing and promotions teams.

6.3.1. *Optimal Subscriptions with Behavioral Mistakes.* With the results of Table 3 in mind, it follows that the preceding analysis of optimal pricing represents only a best-case scenario, as behavioral frictions overwhelmingly tend toward sub-optimal non-subscription. First, there is a substantial degree of inattention—uptake remains low even among high-demand consumers. Second, consumers are fairly inaccurate in anticipating their monthly demand, with a negative bias on average. Addressing both limited awareness and mis-calibrated expectations falls squarely within the marketing team's purview. We therefore interpret the behavioral parameters (σ and H_ε) as capturing a short-run information and communication problem—one that is plausibly mitigated through improved targeting, clearer messaging, and consumer learning—highlighting an important role for marketing interventions. Consistent with this interpretation, our results explain why some real-world firms implementing subscription programs actively assist consumers in determining when subscription is privately beneficial.[22] One benefit of our subscription RCT data is that roll-out of the program within the treatment groups involved a single notification email (sent by the promotions team) and

[22]For example, Costco nudges consumers at the checkout when they would strictly benefit from upgrading their membership to a higher-fee tier in order to access larger discounts.

a new menu item that appeared within the Lyft app, but no additional market-level advertising or consumer-awareness efforts. Thus, despite low uptake rates, the experiment cleanly identifies the baseline magnitudes of behavioral issues to be solved by the marketing team.

Following our estimated model, a type θ consumer, given baseline price p_0 and menu $\mathcal{M}$, will purchase no subscription with probability $(1-\sigma)$, and otherwise will choose their optimal subscription based on a mis-perceived type $\hat{\theta}=\theta\exp(-\varepsilon)$, where $\varepsilon \sim H_\varepsilon$. Averaging over the randomness generated by behavioral mistakes, profits from offering subscription menu $\mathcal{M}$ with baseline price p_0 are

$$\begin{aligned}\pi\left(p_0,\mathcal{M};\sigma,H_\varepsilon\right)=\int_\varepsilon\int_{\underline{\theta}}^{\bar{\theta}} \{\sigma[(p_0(1-d_{\alpha^*(\theta\exp(-\varepsilon);p_0,\mathcal{M})})-c)q^*(p_0(1-d_{\alpha^*(\theta\exp(-\varepsilon);p_0,\mathcal{M})});\theta) \\ -S_{\alpha^*(\theta\exp(-\varepsilon);p_0,\mathcal{M})}]+(1-\sigma)(p_0-c)q^*(p_0;\theta)\}\,dF_\theta(\theta)\,dH_\varepsilon(\varepsilon).\end{aligned} \tag{16}$$

Thus, optimal $(p_0^*,\mathcal{M}^*)$ choice in the presence of behavioral frictions is characterized by maximizing (16)—which we do numerically—rather than (15).

As a base case, suppose the pricing and promotions teams work together, but the marketing team is not consulted, leaving behavioral phenomena at their baseline magnitudes. Under this scenario, the best possible subscription plan achieves excess profits of only 0.004% for a one-offer menu, and 0.03% for a two-offer menu. These findings provide a concrete example of the coordination across units required when a company is set up as an M-form organization. Even though the *tasks* that different teams perform are well-siloed within Lyft, the fact that these tasks all ultimately serve the same downstream consumers nonetheless creates demand-induced synergies across corporate units.

6.3.2. *Optimality with Marketing Efforts to Reduce Behavioral Mistakes.* We next model the case where the marketing team takes steps to reduce consumer misperceptions and inattention. For these counterfactuals, we assume the marketing team possesses some capacity for mitigating behavioral mistakes by consumers, though expecting it to *fully* resolve them may be unrealistic. This begs the question, what is the maximal level of residual consumer behavioral frictions that is amenable to reasonable excess profits from nonlinear pricing? To answer this question we consider a range of scenarios where the marketing team achieves various levels of success in reducing salience and forecast error problems.

Specifically, for some $c\in(0,1)$, let $\sigma^c\equiv 1-c(1-\sigma)$ represent the case where non-salience of subscriptions is scaled by a factor of c, and let $H_\varepsilon^c(\epsilon)=H_\varepsilon(c\epsilon)$ represent the distribution of forecast errors where bias and variance are both scaled by a factor of c. Thus, if the marketing team is $(1-c)\times 100\%$ effective at removing behavioral mistakes, the analog of Equation (16) with H_ε^c and σ^c would represent subscription profitability. We report optimal excess profits when the marketing team is 25%, 50%, 75%, and 90% effective (i.e., $c = 0.75, 0.5, 0.25, 0.1$, respectively). Results are reported in Table 4.

Interestingly, subscriptions are not particularly lucrative until the marketing team achieves a relatively high degree of effectiveness. If it can remove 75% of all errors ($c=0.25$), excess profits are about half of the best case for the single-offer menu (2.35%), and just over two thirds of the best case for the two-offer menu (3.66%). To understand this result, it is helpful to further decompose the sources of incremental profits from subscription pricing. Figure 6 provides some visual intuition for the simpler case of a one-offer menu. The horizontal axis represents quantiles of θ, while the vertical axis represents excess profits from nonlinear pricing under different behavioral scenarios. The dotted line assumes all consumers are uptakers, while the dashed line is the case where none subscribe. The dotted line dips below zero at/above the 98^{th} percentile because the firm must settle for lower profits (relative to linear pricing) for the highest demand types in order to maintain incentive compatibility within the optimal one-offer menu. The dashed line dips negative below the 85^{th} percentile because baseline price under the optimal one-offer menu is set above the optimal linear price.

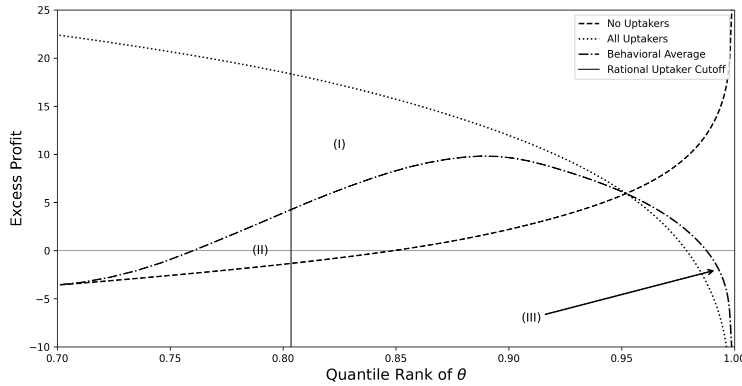
TABLE 4. Optimal Excess Profits from Subscription Menus

Counterfactual	**%Δ From Baseline Profits**	**Counterfactual**	**%Δ From Baseline Profits**
Fully Rational, $p=p^*$			
Continuum Menu	+5.62%	Continuum Menu	+5.62%
One-Offer Menu	+4.63%	Two-Offer Menu	+5.27%
One-Offer Menu, $p=1$		**Two-Offer Menu, $p=1$**	
$c=1$	+0.001%	$c=1$	+0.03%
$c=0.75$	+0.05%	$c=0.75$	+0.45%
$c=0.5$	+0.31%	$c=0.5$	+1.41%
$c=0.25$	+1.21%	$c=0.25$	+2.27%
$c=0.1$	+2.23%	$c=0.1$	+3.26%
$c=0$ (Fully Rational)	+3.15%	$c=0$ (Fully Rational)	+3.39%
One-Offer Menu, $p=p^*$		**Two-Offer Menu, $p=p^*$**	
$c=1$	+0.004%	$c=1$	+0.03%
$c=0.75$	+0.15%	$c=0.75$	+0.48%
$c=0.5$	+0.78%	$c=0.5$	+1.63%
$c=0.25$	+2.35%	$c=0.25$	+3.66%
$c=0.1$	+3.79%	$c=0.1$	+4.95%
$c=0$ (Fully Rational)	+4.63%	$c=0$ (Fully Rational)	+5.27%

Notes: This table reports incremental profits from optimal subscriptions under various counterfactual values of the behavioral parameters and given various assumptions about the constraints that Lyft faces.

The vertical line in Figure 6 represents the fully rational subscriber cutoff, $\tilde{\theta}_{p_0^*}(S^*, d^*)$, at the optimal single-offer subscription. Excess subscription profits from the rational benchmark are therefore the area between 0 and the dashed line *below* $\tilde{\theta}_{p_0^*}(S^*, d^*)$, plus the area between 0 and the dotted line *above* $\tilde{\theta}_{p_0^*}(S^*, d^*)$. When consumers exhibit behavioral deviations from the ex-post rational benchmark, incremental profits are a point-wise (in θ) weighted average of the dotted and dashed curves, represented by the dash-dot line for the case with 75%

FIGURE 6. Subscription Profit Decomposition



Notes: This plot decomposes incremental profits from subscriptions by type. The behavioral average is taken under the assumption that the marketing is 75% effective at removing behavioral phenomena ($c = 0.25$).

marketing team efficacy (i.e., $c = 0.25$).[23] The region labeled (I) in the figure represents lost profits from sub-optimal non-subscription mistakes. Region (II) is additional profits from mistaken subscribership below the rational cutoff, and region (III) is additional profits from mistaken non-subscribership in the far upper tail of demand. As the marketing team approaches full efficacy—i.e., $(1-c) \to 1$—the dash-dot line approaches the dotted line (pointwise) above the cutoff, and it approaches the dashed line (pointwise) below the cutoff, so that regions (I), (II), and (III) vanish in the limit.

From the picture, one can see that consumers who mistakenly fail to subscribe despite reduced mistakes—i.e., the widest part of region (I)—tend to be close to the cutoff. Note that these are also consumer types for whom excess profits are highest when rationally subscribing. Thus, with a small amount of behavioral friction, most people will make the correct subscription decision, but sub-optimal non-uptake behavior is concentrated among consumers who the firm most wishes would subscribe. While this effect is partially offset by consumers who mistakenly subscribe when they should not have (region (II)) and high-demand types that fail to subscribe (region (III)), the negative impact from region (I) strongly dominates. Above 75% efficacy, further improvements tend to reap large gains. As shown in Table 4, at 90% marketing efficacy, the optimal one-offer menu delivers 82% of full excess profits, and the optimal two-offer menu delivers 94%.

6.4. **Strategies for Mitigating Behavioral Frictions.** Thus far, we have examined non-linear pricing profits under scenarios in which the marketing team achieves varying degrees of success in mitigating behavioral frictions. To a large extent, these mitigation efforts will take the form of direct communication strategies aimed at consumers. However, economic

[23]More concretely, the weight on the dotted line (subscribers) is $\sigma^c H^c_\varepsilon(\log(\theta/\tilde{\theta}_{p_0^*}(S^*, d^*)))$ and the weight on the dashed line (non-subscribers) is one minus this number.

theory can also provide guidance on market-design approaches that help blunt the negative impact of behavioral frictions on profits. We briefly discuss two such strategies here.

6.4.1. *Multi-Offer Menus.* Note from Table 4 that at various marketing efficacy levels $(1-c)$, the optimal two-offer menu substantially outperforms the optimal one-offer menu, whereas the two are most similar in the fully rational benchmark case $(c=0)$. With the geometric understanding from Figure 6 in mind—that is, sub-optimal non-subscribership tends to concentrate among high-profit prospective subscribers near the rational cutoff—this difference is suggestive of a dual role for richer menu offerings. First, it provides a closer approximation to the continuum menu, as we saw in Section 6.2.1. Second, and perhaps more importantly, when customers exhibit pronounced forecast bias, multi-offer menus also help mitigate sub-optimal uptake choices by providing more locally customized options. Note that the presence of the second subscription tier offer (S_2^*, d_2^*)—upper filled diamond in Figure 5—enables two adjustments to lower tier prices that counteract sub-optimal non-subscribership. First, the lower tier subscription fee S_1^* is set well below the one-offer fee (open diamond), which will deter a smaller fraction of marginal subscribers. Second, the two-offer default price (filled circle) is set higher than the one-offer default (open circle), which makes non-subscribership less appealing to the unconvinced.

6.4.2. *Forecast Errors and Ex-Post Rebates.* Recall that the estimated forecast error distribution H_ε exhibits large variance and downward bias, on average. As discussed previously in Section 4.1, there are various plausible interpretations for ε: it could represent biased subjective beliefs about the future, and/or it could be consistent with some degree of average loss aversion. Under the latter interpretation, consumers' ex-post behavioral mistakes may be an ex-ante attempt at avoiding exploitation by the firm. In other words, consumers may correctly anticipate the future, on average, but may nonetheless require a substantial degree of certainty that they will benefit from a subscription before agreeing to buy one.

Thus, it is conceivable that Lyft's marketing team can only convince consumers to subscribe by promising to make whole those who subscribe but fail to recuperate the upfront fee $\$S$ through discounted consumption. We simulate profits under such a scheme by assuming that consumers who purchase a subscription in error receive a terminal rebate that reverts them to default expenditures (under price p_0) whenever those are lower than their total subscription expenditures (under S and price $p_0(1-d)$). In Figure 6, this corresponds to returning profits from region (II) back to consumers at the end of the month.

We re-compute subscription counterfactuals under this rebate scheme and report the results in Table 5. This guarantee scheme is worthwhile to the firm if it can substantially mitigate sub-optimal non-subscribership. If one assumes that behavioral mistakes could be reduced by 75% ($c = 0.25$) or 90% ($c = 0.1$) with a rebate guarantee, then it would yield similar profits as reducing behavioral mistakes by roughly 63% or 82%, respectively, without a guarantee. In practice, the decision of whether or not a guarantee is worthwhile for the firm turns on the extent to which such a plan truly helps reduce the pessimism of the consumer base in making their subscriptions purchasing decision.

TABLE 5. Optimal Subscription Profits w/Savings Guarantee

One-Offer Menu Counterfactual	**%Δ From Baseline Profits**
$p = p^*$, Savings Guarantee	
$c = 0.75$	+0.22%
$c = 0.5$	+0.51%
$c = 0.25$	+1.47%
$c = 0.1$	+2.97%

Notes: This table reports incremental profits from optimal subscriptions under various counterfactual values of the behavioral parameters and given various assumptions about the constraints that Lyft faces.

7. CONCLUSION

Nonlinear pricing theory predicts that firms can extract substantial surplus by inducing heterogeneous consumers to self-sort across well-chosen contract offers. In the canonical model, consumers correctly evaluate available options and select those that are ex-post optimal. Our evidence from large-scale subscription experiments conducted by Lyft shows that this frictionless sorting assumption may fail in practice. Many high-demand consumers decline subscriptions that would have saved them money, while a share of subscribers fail to break even. These deviations are systematic and quantitatively important.

We develop a structural framework that embeds salience failures, forecast errors, and impulsivity into an otherwise standard intensive-margin demand model. A key insight is that subscription uptake with behavioral frictions can be recast as one-sided noncompliance in a binary-instrument IV framework. This mapping enables a LATE-style identification strategy to recover counterfactual outcome distributions, and to identify a nonparametric "uptake function" linking baseline outcomes to compliance behavior.

Behavioral frictions materially alter the profitability of nonlinear pricing. Once salience and forecast errors are accounted for, profit gains from subscriptions become highly sensitive to consumer awareness and belief formation. Implementing nonlinear pricing is therefore not solely a matter of consumer screening; it also requires shaping the informational environment in which they evaluate price offers. Our counterfactual analyses further show that multi-tiered subscription menus are doubly advantageous. Beyond more closely approximating the fully screening benchmark, additional price offer tiers mitigate the profit consequences of forecast error by offering more locally tailored price–quantity tradeoffs.

Our findings uncover a subtle "reverse-scaling" problem that arises when firms such as Lyft experiment with nonlinear pricing. Because Lyft's user communication occurs mainly at the market level, it is logistically difficult to align experimental treatment with the marketing intensity of a full product launch. As a result, limited salience becomes a substantial impediment to subscription uptake within the experimental setting, despite this being straightforward to address through mass marketing efforts. In the nomenclature of List (2024), this implies that a typical subscription experiment may fail in terms of representativeness of the situation at scale. Thus, naively interpreting experimental average treatment effects can lead to misleading conclusions about nonlinear pricing profitability. In our setting, the estimated average treatment effects of subscription interventions on quantity demanded are statistically significant but economically modest. Our structural approach mitigates the inference aspect of this scaling problem by disentangling underlying preferences from behavioral frictions. Thus, the empirical model can project full potential excess profits from subscription pricing, which turn out to be substantial if the firm can alleviate these frictions.

These findings also have organizational implications. In a frictionless environment, optimal nonlinear pricing requires coordination between pricing and promotions teams. With behavioral frictions present, profitability additionally depends on marketing's effectiveness in increasing salience and improving belief formation. We quantify these complementarities and show that the marginal contribution of any one function depends critically on the performance of the others. In this sense, demand-side frictions generate meaningful intra-firm synergies. More broadly, our findings suggest that nonlinear pricing in practice is best understood as an integrated design problem. Contract structure, baseline price policy, and communication strategy must be jointly optimized. When firms coordinate across these margins—and when menus are structured to accommodate behavioral constraints—subscription pricing can capture a substantial share of the theoretical gains predicted by theory.

References

Abadie, A., Angrist, J., and Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1):91–117.

Ackerberg, D., Benkard, C. L., Berry, S., and Pakes, A. (2007). Econometric tools for analyzing market outcomes. *Handbook of econometrics*, 6:4171–4276.

Alonso, R., Dessein, W., and Matouschek, N. (2008). When does coordination require centralization? *American Economic Review*, 98(1):145–179.

Athey, S. and Haile, P. (2007). Nonparametric approaches to auctions. *Handbook of econometrics*, 6(ch. 60):3847–3965.

Attanasio, O. and Pastorino, E. (2020). Nonlinear pricing in village economies. *Econometrica*, 88(1):207–263.

Bodoh-Creed, A. L., Boehnke, J., and Hickman, B. (2021). How efficient are decentralized auction platforms? *The Review of Economic Studies*, 88(1):91–125.

Bolton, P. and Farrell, J. (1990). Decentralization, duplication, and delay. *Journal of Political Economy*, 98(4):803–826.

Brugues, F. (2020). Take the goods and run: Contracting frictions and market power in supply chains. *Work. Pap., Brown Univ., Providence, RI.*

Chen, C., Jia, Y., Martin, X., and Silveira, B. (2025). Assessing the objective function of the sec against financial misconduct: A structural approach. *Journal of Accounting and Economics*, 80(1).

Cotton, C., Hickman, B., List, J., Price, J., and Roy, S. (2026). Why don't struggling students do their homework? disentangling motivation & study productivity as drivers of adolescent human capital formation. *Journal of Political Economy*, 134(1):86–149.

Davidson, R. and MacKinnon, J. G. (2021). *Econometric Theory and Methods*. New York: Oxford University Press.

de Boor, C. (2001). *A Practical Guide to Splines Revised Edition*. New York: Springer.

DellaVigna, S. and Malmendier, U. (2004). Contract design and self-control: Theory and evidence. *The Quarterly Journal of Economics*, 119(2):353–402.

Dessein, W., Garicano, L., and Gertner, R. (2010). Organizing for synergies. *American Economic Journal: Microeconomics*, 2(4):77–114.

D'Haultfoeuille, X. and Février, P. (2015). Identification of nonseparable triangular models with discrete instruments. *Econometrica*, 83(3):1199–1210.

D'Haultfœuille, X. and Février, P. (2020). The provision of wage incentives: A structural estimation using contracts variation. *Quantitative Economics*, 11(1):349–397.

Esteban, S., Miyagawa, E., and Shum, M. (2007). Nonlinear pricing with self-control preferences. *Journal of Economic theory*, 135(1):306–338.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *Quarterly Journal of Economics*, 129(4):1661–1710.

Gabaix, X. (2025). A theory of complexity aversion. *working paper, Harvard University Economics Department.*

Gabaix, X. and Graeber, T. (2024). The complexity of economic decisions. *NBER working paper series*, 33109.

Gabaix, X. and Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics*, 121(2):505–540.

Gallant, A. R. and Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica*, 55(2):363–390.

Grubb, M. D. (2009). Selling to overconfident consumers. *American Economic Review*, 99(5):1770–1807.

Grubb, M. D. and Osborne, M. (2015). Cellular service demand: Biased beliefs, learning, and bill shock. *American Economic Review*, 105(1):234–271.

Guerre, E., Perrigne, I., and Vuong, Q. (2009). Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. *Econometrica*, 77(4):1193–1227.

Haile, P., Hong, H., and Shum, M. (2006). Nonparametric tests for common values in first-price sealed-bid auctions. *Working Paper, Yale University.*

Hedblom, D., Hickman, B., and List, J. (2022). Toward an understanding of corporate social responsibility: Theory and field experimental evidence. *Working paper, Washington University in St Louis, Olin Business School.*

Hickman, B. R., Hubbard, T. P., and Paarsch, H. J. (2017). Identification and estimation of a bidding model for electronic auctions. *Quantitative Economics*, 8(2):505–551.

Hickman, B. R., List, J. A., Muir, I., and Sun, G. K. (2026). Designing around selection: Identification and inference under multi-dimensional unobserved heterogeneity. *working paper, Washington University in St Louis, Olin Business School.*

Hortaçsu, A., Natan, O. R., Parsley, H., Schwieg, T., and Williams, K. R. (2024). Organizational structure and pricing: Evidence from a large US airline. *The Quarterly Journal of Economics*, 139(2):1149–1199.

Imbens, G. and Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 52(2):467–475.

Imbens, G. and Rubin, D. (1997). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies*, 64(4):555–574.

Kang, K. and Silveira, B. (2021). Understanding disparities in punishment: Regulator preferences and expertise. *Journal of Political Economy*, 129(10:2947–2992.

List, J. A. (2024). Optimally generate policy-based evidence before scaling. *Nature*, 626(7999):491–499.

Luo, Y., Perrigne, I., and Vuong, Q. (2011). Multiproduct nonlinear pricing: Mobile voice service and sms. *Working Paper.*

Luo, Y., Perrigne, I., and Vuong, Q. (2018). Structural analysis of nonlinear pricing. *The Journal of Political Economy*, 126(6):2523–2568.

Mahajan, V., Sharma, S., and Bettis, R. A. (1988). The adoption of the m-form organizational structure: A test of imitation hypothesis. *Management Science*, 34(10):1188–1201.

Maskin, E., Qian, Y., and Xu, C. (2000). Incentives, information, and organizational form. *The review of economic studies*, 67(2):359–378.

Maskin, E. and Riley, J. (1984). Monopoly with incomplete information. *The RAND Journal of Economics*, 15(2):171–196.

Miravete, E. J. (2003). Choosing the wrong calling plan? ignorance and learning. *American Economic Review*, 93(1):297–310.

Miravete, E. J. and Röller, L.-H. (2003). Competitive non-linear pricing in duopoly equilibrium: the early us cellular telephone industry. *Available at SSRN 470844*.

Mussa, M. and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory*, 18:301–317.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of operations research*, 6(1):58–73.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy*, 9(4):513–548.

Oster, S. M. and Scott Morton, F. M. (2005). Behavioral biases meet the market: The case of magazine subscription prices. *The BE Journal of Economic Analysis & Policy*, 5(1):0000101515153806371323.

Perrigne, I. and Vuong, Q. (2011). Nonparametric identification of a contract model with adverse selection and moral hazard. *Econometrica*, 79(5):1499–1539.

Rockafellar, R. T. (1997). *Convex analysis*, volume 28. Princeton university press.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–55.

Sun, G. (2025). What, when, and why can the principal know: Identification in adverse selection models with an application to labor income taxation. *Working Paper, Olin Business School, Washington University in St Louis*.

Torgovitsky, A. (2015). Identification of nonseparable models using instruments with small support. *Econometrica*, 83(3):1185–1197.

Williamson, O. E. (1975). Markets and hierarchies: analysis and antitrust implications: a study in the economics of internal organization.

Appendix A. Proofs

A.1. **Proof of Lemma 1.**

Proof. Define consumer θ's optimized value function under a given linear price p as $V(\theta;p)\equiv \theta u(q^*(p;\theta))-pq^*(p;\theta)$, and note that by Assumption 1, V is differentiable in both θ and p. In other words, V is the total utility experienced by a type θ consumer when prices are p

and consumption is chosen optimally. Note that the LHS of inequality (1) can be written as $V(\theta;p(1-d))-V(\theta;p)-S$, hence, the result follows if $\frac{\partial}{\partial\theta}[V(\theta;p(1-d))-V(\theta;p)]>0$. Note that this inequality is equivalent to the following:

$$\frac{\partial}{\partial\theta}[V(\theta;p(1-d))-V(\theta;p)] = \frac{\partial}{\partial\theta}\int_p^{p(1-d)} \frac{\partial V(\theta;\tilde{p})}{\partial p}\mathrm{d}\tilde{p} = \int_p^{p(1-d)} \frac{\partial^2 V(\theta;\tilde{p})}{\partial\theta\partial p}\,\mathrm{d}\tilde{p}>0,$$

where the first equality follows from the fundamental theorem of calculus, and the second equality follows from Leibniz's Rule. Thus, since the lower limit of integration is greater than the upper limit, the desired result follows if we can show that the integrand is negative. To do this, first note that envelope theorem implies that $\frac{\partial V(\theta;p)}{\partial\theta} = u(q^*(\tilde{p};\theta))$. Differentiating with respect to p, we get $\frac{\partial^2 V(\theta;p)}{\partial\theta\partial p} = u'(q^*(\theta;p))\frac{\partial q^*(\theta;p)}{\partial p}$. Since utility is increasing in q, and q^* is decreasing in p, the RHS of this final expression is negative, as desired. □

A.2. **Proof of Proposition 2.**

Proof. To prove the first inequality, fix (S,d,p), and let $\tilde{\theta}\equiv\tilde{\theta}_p(S,d)$. Let $q_0\equiv q^*(p;\tilde{\theta})$ be the optimal quantity of the cutoff type given no discount, and let $q_1\equiv q^*(p(1-d);\tilde{\theta})$ be the optimal discounted quantity. Since q_1 is optimal with a discount we have

$$\tilde{\theta}u(q_1)-p(1-d)q_1\geq\tilde{\theta}u(q_0)-p(1-d)q_0. \tag{17}$$

Moreover, the definition of $\tilde{\theta}$ implies that $\tilde{\theta}u(q_0)-pq_0=\tilde{\theta}u(q_1)-p(1-d)q_1-S$. Adding S to both sides and combining this with inequality (17) yields $S+\tilde{\theta}u(q_0)-pq_0\geq\tilde{\theta}u(q_0)-p(1-d)q_0$, and after canceling common terms from each side we get $S\geq pdq_0$, as desired.

The second inequality follows by a similar revealed preference argument. Note that

$$\tilde{\theta}u(q_1)-pq_1\leq\tilde{\theta}u(q_0)-pq_0 \tag{18}$$

by definition of q_0 and q_1. Again, the cutoff type satisfies $\tilde{\theta}u(q_0)-pq_0=\tilde{\theta}u(q_1)-p(1-d)q_1-S$, which combined with inequality (18) implies $S\leq pdq_1$. □

Appendix B. Testing the Symmetry Assumption

In this section, we describe a formal statistical test for Assumption 3 that H_ε is symmetric about its median. Let $r_1,r_2,\ldots,r_{K_r}\in(0,1)$ be a grid of quantile ranks and let $H_\varepsilon(\cdot)$ denote the true distribution of forecast errors (versus our estimate $\hat{H}_\varepsilon(\cdot;\hat{\boldsymbol{\omega}}_h)$). Let $\hat{\mathbf{Y}}\equiv(\hat{H}_\varepsilon(r_1;\hat{\boldsymbol{\omega}}_h),\ldots,\hat{H}_\varepsilon(r_{K_r};\hat{\boldsymbol{\omega}}_h),\hat{H}_\varepsilon(1-r_{K_r};\hat{\boldsymbol{\omega}}_h),\ldots,\hat{H}_\varepsilon(1-r_1);\hat{\boldsymbol{\omega}}_h)$ denote the vector of estimated CDF values at the quantile ranks and their complements, with population analog $\boldsymbol{Y}\equiv(H_\varepsilon(r_1),\ldots,H_\varepsilon(r_{K_r}),H_\varepsilon(1-r_{K_r}),\ldots,H_\varepsilon(1-r_1))$. Since $\hat{\boldsymbol{\omega}}_h$ is consistent and asymptotically normal, and since $\hat{\boldsymbol{Y}}$ is a continuously differentiable function of $\hat{\boldsymbol{\omega}}_h$, standard asymptotic theory implies that $\sqrt{n}(\hat{\mathbf{Y}}-\mathbf{Y})\xrightarrow{d}\mathcal{N}(0,\boldsymbol{\Sigma})$, for some asymptotic covariance matrix $\boldsymbol{\Sigma}$, which we estimate by taking the sample covariance of the bootstrapped values of $\hat{\mathbf{Y}}$. The null hypothesis that Assumption 3 holds implies a linear restriction on $\mathbf{Y}$ that $\mathbf{Y}_k + \mathbf{Y}_{K_r-k+1} = 1$. We compute the Wald test statistic for this set of linear restrictions on a grid of five quantile ranks chosen to be uniformly spaced between just above $\underline{\varepsilon}$ and just below the median: $\{r_1 = 0.15, r_2 = 0.2125, r_3 = 0.275, r_4 = 0.3375, r_5 = 0.40\}$. The resulting p-value is 0.895, so we are unable to reject the null hypothesis that symmetry holds in our data.[24]

Appendix C. Imputation of Marginal Costs

Optimal uniform price under constant marginal costs is obtained by choosing p to solve for the firm's first order condition. Formally, if $\theta_m(p)$ is the type who is on the margin between 0 consumption and positive consumption, we have:

$$p = \frac{\zeta(p)}{1+\zeta(p)}c, \quad \zeta(p) = \frac{1}{\int_{\underline{\theta}}^{\bar{\theta}} q^*(p,\theta)\mathrm{d}F(\theta)} \int_{\theta_m(p)}^{\bar{\theta}} \frac{u'(q(p,\theta))\mathrm{d}F(\theta)}{u''(q(p,\theta))} \tag{19}$$

Here, $\zeta(p)$ is aggregate elasticity of demand price p implied by the model with multiplicatively separable utility. 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 $p_0 = 1$ as the optimal uniform price and calibrate c according to Equation (19). 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.

Appendix D. Subscription Menu Optimality

In this appendix, we discuss solving for the optimal subscription program. Our approach follows the seminal work of Maskin and Riley (1984) (MR84), which studies optimal monopoly pricing in the presence of unobservedly heterogeneous consumers. MR84 characterized a nonlinear pricing schedule based on quantity demanded (which is in turn a function of demand type). We first show that this approach can be implemented (with little loss of generality) as a continuum menu of subscription offers, and then we discuss its numerical computation. For the purpose of this exercise, consumers are initially assumed to make

[24]Empirically, the cutoff shock value $\hat{\underline{\varepsilon}}$ is the 14.5^{th} percentile. For robustness, we tried two alternate forms of the test: one with four uniformly spaced quantile ranks between 0.15 and 0.4 (p-value= 0.818), and one with five uniformly spaced between 0.15 and 0.45 (p-value= 0.903); both fail to reject the null hypothesis.

subscription decisions in a fully rational manner. Having established the baseline procedure, we then briefly discuss solutions for an optimal finite menu of subscription offers when consumers exhibit the behavioral phenomena included in the structural model.

Our initial approach begins with the observation (formalized in Lemma 2) that any menu of subscriptions can alternatively be represented by a nonlinear price $P(q)$, which is increasing and concave.[25] Note that P represents price levels as a function of observable consumption q, making it different from (but related to) the marginal-price function $p(\theta)$ which depends on unobserved type θ. We will establish the link between them below.

Lemma 2. *Let $p_0 > 0$ be some baseline price, let $\mathcal{M} = \{(S_\alpha, d_\alpha)\}$ be a menu of subscriptions indexed by α where $d_\alpha \in [0,1)$ for all α. Then there exists a pricing rule $P(q)$, where P is increasing and concave, such that consumer behavior with menu $\mathcal{M}$ and default price p_0 is equivalent to consumer behavior under P, that is:*

$$\max_{q,\alpha} u(q;\theta) - p_0(1-d_\alpha)q - S_\alpha = \max_q u(q;\theta) - P(q), \quad \forall\theta, \tag{20}$$

and the optimal choice of q under both is identical. Conversely, corresponding to any increasing and concave nonlinear pricing rule $P(q)$, there exists some menu $\mathcal{M}$ and default price p_0 such that the equivalence in (20) *holds, with the optimal choices of q being identical.*

Proof. Fix a menu $\mathcal{M}$ and default price p_0. For any quantity q, define pricing rule $P(q) \equiv \min_{\alpha \in \mathcal{M}} p_0(1-d_\alpha)q + S_\alpha$ as the cost-minimizing total price paid, given q and $\mathcal{M}$. This pricing rule is the pointwise minimum of a collection of linear functions (one for each value of the index α), and hence is concave (see, e.g., Rockafellar (1997)). It is increasing since all of the constituient linear functions are themselves increasing. For all θ, we have

$$\max_{q,\alpha} u(q;\theta) - p_0(1-d_\alpha)q - S_\alpha = \max_q u(q;\theta) - \min_{\alpha\in\mathcal{M}} p_0(1-d_\alpha)q + S_\alpha = \max_q u(q;\theta) - P(q).$$

Conversely, let $P(q)$ be a concave, increasing pricing function, and let $\partial P(q)$ denote the *supergradient* of P at q.[26] Let $p_0 \equiv \sup \partial P(0)$. Define $\mathcal{T}$ to be the set of all slope-intercept pairs of the tangent lines of $P(q)$, that is,

$$\mathcal{T} \equiv \bigcup_q \{(b, P(q) - bq) : b \in \partial P(q)\}.$$

[25]The restriction that $P(q)$ must be concave is with some loss of generality relative to the class of pricing rules allowed in MR84. As a result, while incentive compatible implementation of the candidate pricing rule $\tilde{p}(\theta)$ as a subscription menu is only possible when $\zeta(\theta)$ is non-decreasing, MR84's notion of incentive compatibility only implies the weaker requirement that $\theta - \frac{F(\theta)}{f(\theta)}$ is non-decreasing in θ.

[26]In other words, $\partial P(q)$ is the set containing the slopes of all tangent lines of P at q that always lie above P. Since P is concave, standard results in convex analysis (Rockafellar, 1997) imply that the supergradient is always non-empty.

Intuitively, the b term above can be thought of as a (linear) price, and the $P(q) - bq$ term can be thought of as a subscription fee, as it is the difference between total price paid and volumetric price paid. Thus, $\mathcal{T}$ represents the set of price-fee pairs that define tangency lines of the price schedule $P(q)$ (with slope b and intercept $P(q) - bq$). By concavity of P and by the definition of p_0, it is always the case that $b < p_0$ for all elements in the set $\mathcal{T}$. By concavity, $P(q)$ can equivalently be characterized by the pointwise minimum of its tangency lines, $P(q) = \min_{(b,S)\in\mathcal{T}} bq + S$. On the other hand, each tangent line $(b, S) \in \mathcal{T}$ represents exactly how much a consumer pays if they buy a subscription with upfront price S and discount $d \equiv 1 - b/p_0$ when default price is p_0, where $d \in [0, 1)$ since P is increasing and $b \leq p_0$. Hence, defining $\mathcal{M} \equiv \{(1 - b/p_0, S) : (b, S) \in \mathcal{T}\}$, it is clear that Equation (20) holds. $\square$

Next, we argue the equivalency of $P(q)$ and $p(\theta)$. First, note that any nonlinear pricing rule $P(q)$ can be used to pin down the *marginal* price that each θ type pays, that is, $p(\theta) \equiv P'(q^*(P(\cdot); \theta))$. Second, to see why $p(\theta)$ can be used to fully back out $P(q)$, let $U(\theta) \equiv \max_q u(q; \theta) - P(q)$ be the indirect utility function for a type θ consumer, and assume for a moment that $U(\theta)$ is known. If it were known that a type θ consumer chooses marginal price $p(\theta)$, then we also know they purchase $q^*(p(\theta); \theta)$. Thus, knowing this for all values of θ, we can back out $P(\cdot)$ via the identity $P(q^*(p(\theta); \theta)) = \theta u(q^*(p(\theta); \theta)) - U(\theta)$. In order to see why the indirect utility function $U(\theta)$ is known, given knowledge of $p(\theta)$, note the following. Since the lowest-type consumer must get 0 surplus $U(\underline{\theta}) = 0$, and by the envelope theorem we know that $U'(\theta) = u(q^*(p(\theta); \theta))$, then $U(\theta)$ is determined by the unique solution of the resulting boundary value problem. Hence, knowledge of $p(\theta)$ is sufficient to reconstruct $P(q)$ for all q which some consumer type $\theta \in [\underline{\theta}, \overline{\theta}]$ purchases. Finally, note that since $\theta u(q)$ is a supermodular function of θ and q, and since $q^*(P; \theta)$ is increasing in θ, the resulting $P(q)$ is concave if and only if $p(\theta)$ is *non-increasing* in θ.

D.1. **Numerical Optimization: Continuum Menu with Fully Rational Consumers.** The above discussion implies that we can focus on optimizing $p(\theta)$ directly when computing the optimal continuum menu of subscription offers. We parameterize this function as a flexible quartic B-spline in *log* type space, that is, $p(\theta; \boldsymbol{\omega}_p) \equiv \sum_{k=1}^{K+4} \omega_k^p \mathcal{B}_k^p(\log(\theta))$ with knot vector $\boldsymbol{\kappa}_p = \{\kappa_{p1} < \cdots < \kappa_{p,K_p+1}\}$, which partitions $[\log(\underline{\theta}), \log(\bar{\theta}))]$ into K_p uniformly spaced sub-intervals. We constrain $\omega_{p,K+4} = c$ to reflect the "no-distortion at the top" property mentioned above. To reflect the condition that $p(\theta)$ is non-increasing, we also constrain $\omega_{pk} \leq \omega_{p,k+1}$ for $k = 1, \ldots, K+3$.

Given this paraemterization, for a given $\boldsymbol{\omega}_p$ vector, if marginal cost is c, the resulting profit function can be computed via the following procedure:

(1) Define indirect utility and its derivative via $U_p'(\theta;\boldsymbol{\omega}_p) \equiv u^*(q^*(p(\theta;\boldsymbol{\omega}_p);\theta)$, and

$$U_p(\theta;\boldsymbol{\omega}_p) \equiv \int_{\underline{\theta}}^{\theta} U_p'(t;\boldsymbol{\omega}_p)\,\mathrm{d}t. \tag{21}$$

(2) Define the revenue function

$$R(\theta;\boldsymbol{\omega}_p) \equiv \theta u(p(\theta;\boldsymbol{\omega}_p);\theta) - U_p(\theta;\boldsymbol{\omega}_p). \tag{22}$$

(3) Thus, the profit function is

$$\pi(\boldsymbol{\omega}_p) \equiv \int_{\underline{\theta}}^{\bar{\theta}} [R(\theta;\boldsymbol{\omega}_p) - cq^*(p(\theta;\boldsymbol{\omega}_p);\theta)] f_\theta(\theta)\,\mathrm{d}\theta \tag{23}$$

Our final optimization problem takes the form

$$\max_{\boldsymbol{\omega}_p} \pi(\boldsymbol{\omega}_p) \quad \text{s.t.} \quad \omega_{p,K+4} = c,\ \omega_{pk} \leq \omega_{p,k+1}, k = 1, \ldots, K+3. \tag{24}$$

which we solve using a standard nonlinear optimizer. Denote the optimized B-Spline coefficients by $\boldsymbol{\omega}_p^*$. Once we have $\boldsymbol{\omega}_p^*$, we can compute the implied $S(\theta)$ using the fact that $S(\theta;\boldsymbol{\omega}_p^*) = R(\theta;\boldsymbol{\omega}_p^*) - p(\theta;\boldsymbol{\omega}_p^*)q^*(p(\theta;\boldsymbol{\omega}_p^*),\theta)$. The optimal baseline price is then given by $p^* \equiv \inf\{p(\theta) : q^*(p(\theta;\boldsymbol{\omega}_p^*)) > 0\}$, while the menu of subscriptions is given by the set $\mathcal{M}^* \equiv \{(1 - p(\theta;\boldsymbol{\omega}_p^*)/p^*, S(\theta;\boldsymbol{\omega}_p^*)) : \theta \in [\underline{\theta}, \bar{\theta}]\}$.

D.1.1. *Numerical Optimization: Finite Menus and Behavioral Mistakes.* Optimization becomes significantly simpler once the monopolist constrains itself to a finite menu of subscription offers, as discussed in Section 6.2. This simplified setup can then be easily augmented to account for the impacts of imperfect salience and forecast errors on excess profits from nonlinear pricing, as discussed in Section 6.3.1.