Preferences among Unpurchased Goods in Differentiated Products Demand Systems

Andrew Zeyveld*

May 2024

Abstract

Data on consumers' preferences among unpurchased products can help identify demand elasticities. However, workhorse demand systems sometimes fail to replicate important substitution patterns in these data. I consider the following explanation for this shortcoming: both conditional and mixed logit impose a form of independence between consumers' purchases and their preferences among unpurchased goods. Using data on curbside grocery pickup, I document substitution patterns that are inconsistent with both models' independence properties. To quantify the influence of the independence property exhibited by mixed logit, I compare the model's goodness of fit to that of mixed probit (which exhibits no such property).

JEL Classification: C35, D12, L81
Keywords: Discrete choice, differentiated products, micro data, online shopping

*Department of Economics, Michigan State University. E-mail: zeyveld@msu.edu
Special thanks to Kyoo il Kim and Michael Conlin for their advice throughout this project. I also thank Benjamin Bushong and Jeremy Fox for helpful comments. Any remaining errors are my own. Additional thanks are extended to the retailer that supplied the data and provided information about the purchase environment. I am also grateful to dunhummby, which provided financial support, and to the Institute for Cyber-Enabled Research at Michigan State University, which provided computational resources.

1. Introduction

Demand elasticities influence many features of a differentiated products market, such as the magnitude of markups and the impact of policy changes (Berry and Haile 2021). However, it is difficult to estimate elasticities from consumers' purchases alone. It helps to have data concerning the extent to which *unpurchased* goods are substitutable for purchased ones. If someone's preferred good is unavailable, which (if any) of the unpurchased goods would she prefer to the "outside option" of buying nothing? And of these acceptable substitutes, which would she prefer most? Information of this kind—hereafter, *alternate-choice data*—provides direct evidence of product substitutability (Berry, Levinsohn, and Pakes 2004). By requiring a demand system to reproduce the substitution patterns in alternate-choice data, the researcher obtains more precise estimates of the random coefficients that determine demand elasticities.

Workhorse demand systems sometimes fail to replicate important substitution patterns in alternate-choice data. In particular, demand systems tend to underestimate the substitutability of the closest substitutes while overestimating the substitutability of more distant ones (Conlon and Mortimer 2010). This paper studies a possible source of this bias: the restrictions placed on products' substitutability by conditional and mixed logit. I show that conditional logit imposes independence between consumers' purchases and their pairwise preferences among unpurchased goods. In other words, an individual consumer's purchase choice should be uninformative of her preferences among the unpurchased goods—even if some of them are much closer substitutes for the purchased good than others are. As for mixed logit, this (more flexible) model accommodates cross-sectional variation in preferences among unpurchased goods. However, I show that mixed logit still imposes *conditional* independence between consumers' purchases and their pairwise preferences among unpurchased goods, given the realizations of the (consumer-specific) random taste coefficients. Put differently, a consumer's purchase on a particular shopping trip should be uninformative of *trip-specific* factors influencing both her purchase choice and her preferences among the unpurchased goods. Hereafter, I refer to the preceding independence constraints as the *independence of preferred alternatives* (IPA) properties of conditional and mixed logit, respectively.

To what extent do the IPA properties of conditional and mixed logit affect estimated demand elasticities? To provide insight, I analyze novel alternate-choice data from curbside grocery pickup. This is a "click-and-collect" mode of online shopping in which consumers order groceries online and later pick them up from a bricks-and-mortar supermarket. Sometimes, however, an ordered item goes out of stock, in which case the consumer must choose between (i) purchasing a store-selected substitute and (ii) purchasing nothing.[1]

Focusing on the product categories of bottled water and flour, I provide descriptive evidence that consumers' choices to accept (i.e., purchase) or reject (i.e., not purchase) stockout substitutes are inconsistent with the IPA property of conditional logit. As for the IPA property of mixed logit, I find that the choices of bottled water buyers are consistent with the property, whereas those of flour buyers are not. Concerning the latter product category, individual consumers' preferences appear to vary between trips, perhaps due to variation in the type(s) of recipes they plan to bake. However, this within-consumer preference variation is excluded by the mixed logit IPA.

To quantify the bias resulting from the IPA property of mixed logit, I estimate demand using two models: mixed logit and mixed *probit* (a model that does not display an IPA property). Then I compare the models' goodness of fit. Both within- and out-of-sample, mixed probit predicts the acceptance or rejection of stockout substitutes more accurately than mixed logit does. Moreover, the disparity in model fit is larger for flour than for bottled water. This is in keeping with the reduced-form evidence summarized above: namely, that consumers' purchases of bottled water are consistent with the IPA property of mixed logit, whereas their purchases of flour are not.

The remainder of the paper proceeds as follows. Section 2 relates this study to prior literature. Section 3 reviews the standard differentiated products demand model developed by Berry, Levinsohn, and Pakes (1995),[2] and then formalizes the IPA properties of conditional and mixed logit. Both properties stem from an unusual property of logit utilities:[3] given that two goods' utilities are both smaller than some constant K, the conditional probability that one's utility exceeds the other's is identical to the *unconditional* probability of the same.

Section 4 provides institutional details about curbside pickup and introduces the data. As far as stockout substitutions are concerned, I observe the universal product code (UPC) of both the out-of-stock product and the substitute. I also see whether the substitute is accepted (i.e., purchased) or rejected (i.e., not purchased). Importantly, these data can be matched to panel data on the consumer's past and future purchases at the store,

[1] In principle, the consumer could also enter the store in search of a better substitute. However, this is exceedingly rare with respect to the product categories considered in this paper, namely, flour and bottled water. In 0% (0.6%) of cases in which the consumer rejects a stockout substitute for a bottled water (flour) product, she enters the store afterwards to purchase a different bottled water (flour) product.

[2] Unlike Berry, Levinsohn, and Pakes (1995), I abstract away from price endogeneity. I do so for two reasons. First, unobserved "quality" is probably less important for the products studied in this paper—namely, bottled water and flour—than it is for automobiles. And second, my demand specification is much more computationally burdensome than BLP 1995, as I employ a semi-nonparametric estimator. It would be computationally challenging to adopt an IV (or even control function) approach.

[3] That is, utility functions in which the error term is distributed i.i.d. standard Gumbel (a.k.a. type I extreme value).

enabling me to infer which products she tends to like (or dislike) within the relevant grocery category.

Section 5 presents descriptive evidence concerning the extent to which consumer behavior coincides with the IPA properties of conditional and mixed logit. Under the conditional logit IPA, a consumer's original order choice should be independent of her willingness to accept a given product as a substitute. I employ two strategies to test this prediction. The first is a likelihood ratio test of independence. In both product categories, the null hypothesis of independence is strongly rejected (under reasonable assumptions). As for the second strategy, it reframes the conditional logit IPA in terms of products' *observable characteristics* (such as brand or size). I explain that, under the conditional logit IPA, a substitute's probability of acceptance should be independent of whether its observable characteristics resemble those of the out-of-stock product. For instance, a flour substitute should be no likelier to be accepted if it is the same type of flour as the out-of-stock product (say, all-purpose flour) than if it is a different one (say, bread flour). However, this prediction is not supported by the data. Rather, acceptance is likelier if the substitute's observable characteristics resemble those of the out-of-stock product than if they do not.

Turning to mixed logit, the IPA property of this (more flexible) model translates to curbside pickup as follows. The probability that a consumer accepts a stockout substitute should be independent of whether its observable characteristics resemble those of the out-of-stock product, conditional on her (time-invariant) tendency to like or dislike the substitute. Concerning some product categories, such as flour, this is a counterintuitive prediction. To see why, consider a consumer who regularly purchases two types of flour: all-purpose flour, which she uses to bake cupcakes; and bread flour, which she uses to bake bread. Note that her choice of flour on a given trip indicates its intended application: cupcakes, if she has opted for all-purpose flour; or bread, if she has opted for bread flour. Thus, she will be likelier to accept a given all-purpose flour product as a substitute on trips where she has ordered a (different) all-purpose flour than on trips where she has ordered a bread flour. However, the mixed logit IPA excludes this intuition; our baker's order choice should be uninformative of her decision about the substitute.

To test this prediction, I first construct a proxy variable for consumers' (time-invariant) tendencies to like or dislike each version of a product characteristic.[4] (For instance, as far as flour type is concerned, does the consumer tend to like all-purpose flours? Bread flours? Whole wheat flours?) Then I test the mixed logit IPA as follows. Is the probability of acceptance independent of whether the substitute shares the out-of-stock

[4] This proxy variable is based on consumers' individual purchase histories. Specifically, I compute the fraction of the consumer's trips in which the purchased product features a specific version of a characteristic. For instance, in what fraction of her trips does she purchase all-purpose flours? Or bread flours?

product's version of the characteristic, conditional on the proxy for the consumer's (time-invariant) tendency to like or dislike the substitute's version? I find evidence in the affirmative for one product category (bottled water) but not the other (flour). The difference between the categories seems to reflect a disparity in the amount of *within-consumer* preference variation. Whereas the preferences of bottled water buyers tend to persist across trips, those of flour buyers often vary based on the intended recipe (for which a specific type of flour may be ideal).

Section 6 tests whether the mixed logit IPA materially affects demand estimates. I estimate demand for bottled water and flour using mixed probit (a model without an IPA constraint) as well as mixed logit. To allow mixed logit to compete with mixed probit on the best possible footing, I estimate the distributions of random coefficients nonparametrically.[5] This ensures that the random coefficients reflect consumers' (time-invariant) tendencies to like or dislike substitute products as accurately as possible, thereby minimizing the influence of the mixed logit IPA.

With the estimates in hand, I compare the models' goodness of fit within each product category. I separately assess fit for (i) consumers' intended purchases (which comprise orders for curbside pickup or home delivery as well as in-store purchases); and (ii) consumers' decisions to accept or reject stockout substitutes. Notice that the mixed logit IPA applies only to the latter data type (which constitutes alternate-choice data).

I find the models' fit to be comparable with respect to consumers' intended purchases. By contrast, there is a perceptible difference in fit where stockout substitutions are concerned: both within- and out-of-sample, mixed probit fits the data better. Concerning bottled water (flour) substitutes, the average predicted probability of consumers' observed decisions to accept or reject is 0.9 (1.0) percentage points greater for mixed probit than for mixed logit. Out-of-sample, these disparities in fit widen to 1.9 and 2.4 percentage points for bottled water and flour, respectively. That the difference in model fit is somewhat larger for flour than for bottled water is in keeping with the descriptive evidence summarized above: namely, that consumers' purchases of bottled water are consistent with the mixed logit IPA, whereas their purchases of flour are not.

The results of this paper can inform future applied work in which the researcher has access to alternate-choice data. If within-consumer variation in preferences is a second-order concern, mixed logit should be able to reproduce the substitution patterns in the alternate-choice data. But if within-consumer variation in preferences is an important feature of the choice environment, then an alternative model like mixed probit may be preferable.

[5] To the best of the author's knowledge, this is the first study to estimate mixed probit nonparametrically in a multinomial choice context.

2. Relationship to Prior Literature

A growing empirical literature leverages alternate-choice data to estimate demand elasticities. The pioneering work is Berry, Levinsohn, and Pakes's (2004) study of the US automotive market—hereafter, BLP '04. They estimate a mixed logit model of demand using two types of data: aggregated data on products' market shares, and questionnaire data from a representative sample of new-car buyers. The latter indicate buyers' "second choices"—that is, the purchases they would have made if their preferred vehicle were unavailable. By requiring their demand system to match these second-choice substitution patterns, BLP '04 obtains more precise estimates of the parameters that govern product substitutability in their model.

The empirical framework developed in BLP '04 remains the most popular means of incorporating alternate-choice data in demand systems.[6] Of the few studies that do adopt alternative frameworks, most still share the following features with BLP '04:

(i) The consumer's discrete choice problem is modeled with mixed logit.

(ii) The data consist of *cross-sectional* data on consumers' purchases, coupled with *stated-preference* data on consumers' rankings of unpurchased products.[7]

It is these features that mark my point of departure from the existing literature. Regarding (i), I highlight the restrictions imposed by mixed logit on the substitution patterns in alternate choice data. Under the IPA property of mixed logit, the consumer's purchase choice must be independent of her pairwise preferences among unpurchased goods, conditional on her (consumer-specific) taste coefficients. As for (ii), my data differ in important respects from the data employed in earlier studies. Most prior work couples (a) nationally representative, but aggregated, data on market shares with (b) highly detailed, but stated-preference, alternate-choice data. In contrast, my data pair (a) household-level panel data on purchases at a single, regional retailer; with (b) less comprehensive, but revealed-preference,[8] alternate-choice data.

In what follows, I will elaborate on both these points of departure, explaining how they can inform future applied work that uses alternate-choice data. Then I will briefly remark on two other literatures to which my

[6] In addition to a series of studies on the automotive market listed below, Farronato and Fradkin (2022) also adapt the framework of BLP '04 in their study about the welfare effects of Airbnb on the accommodation industry. Other recent examples include Conlon and Gortmaker (2023), who study the soda industry; as well as Montag (2023), who studies the household appliance industry.

[7] These data, which are collected from questionnaires, concern consumers' *hypothetical* preferences over products they did not purchase. For example, "If product A were not available, what would you have purchased instead?"

[8] These data record consumers' decisions to purchase, or not purchase, store-selected substitute products. Consumers therefore have "skin in the game:" if they accept the substitute, they will pay for it.

work relates: the econometric literature on the identifying power of alternate-choice data, and the empirical literature that leverages data on stockout events.

A. Estimating Demand with Alternate-Choice Data

The Model.—In differentiated products demand estimation, the consumer's discrete choice problem is most often represented with mixed logit.[9] However, mixed logit is subject to an IPA property that may be unrealistic in some settings. In the introduction, I used the example of a regular flour buyer to illustrate the kind of behavior that is excluded by the mixed logit IPA. Here I translate this constraint to the automotive market, the subject of the empirical application in BLP '04 as well as several recent studies that integrate alternate choice data in a mixed logit model—Grieco, Murry, and Yurukoglu (2023); Bachmann et al. (2023); and Xing, Leard, and Li (2021). To see the significance of the mixed logit IPA in the automotive market, picture someone who has purchased two cars recently. The first is a large SUV (say, the Chevrolet Suburban); while the second is a small sports car (say, the Chevrolet Camaro). Suppose that she purchased the former about a year before the latter. Under the mixed logit IPA, our consumer's pairwise preferences among the unpurchased automobiles should have been essentially identical when she purchased the SUV as when she purchased the sports car a year later. In other words, she was equally likely to have preferred an unpurchased SUV (say, the Ford Expedition) over an unpurchased sports car (say, the Ford Mustang) on both occasions. But this prediction is counterintuitive, as the uses of an SUV (such as transporting bulky objects or ferrying lots of people) differ from those of a sports car (such as pleasure driving). Thus, when our consumer made her more recent purchase—that of the Camaro—she was probably searching specifically for a small sports car. It seems unlikely that this search would have ended in the purchase of a second large SUV, as such a vehicle would not fulfill the purpose she had in mind. But the mixed logit model might make just such a prediction, because it presumes that her preferences over unpurchased vehicles remained identical between the two shopping occasions (despite the different classes of vehicle purchased).

Even so, the mixed logit IPA remains realistic in many other settings. Take the case of household appliances, for example. An individual consumer is unlikely to purchase a given appliance (such as a furnace or dishwasher) more than a couple of times throughout her lifetime. And even if she does make multiple purchases, her preferences will likely remain quite stable over time. Thus, *within*-consumer preference

[9] Allcott (2013) represents a notable exception. In his research into the accuracy of consumers' beliefs on the savings from fuel efficient vehicles, he employs a nested logit model. Another exception, albeit from outside the field of industrial organization, is provided by Abdulkadiroğlu, Agarwal, and Pathak (2017). Their research into school choice employs the multinomial probit model.

variation is likely minimal in household appliance markets, even if there is considerable *between*-consumer preference variation. In such markets, the mixed logit IPA accurately describes consumers' behavior.

The Data.—The data employed in this study provide a useful complement to the data used in previous work. Within the existing literature, it is customary to couple (i) cross-sectional data on market shares with (ii) detailed, but stated-preference, alternate-choice data. This data combination is ideal for most applications of interest, such as recovering markups or characterizing market responses to counterfactual policy changes. However, it would be challenging to test the IPA property of mixed logit with cross-sectional data of this description. The reason is that the mixed logit IPA imposes a within-panel restriction on product substitutability. To assess the extent to which consumer behavior is consistent with this constraint, it helps to have household-level panel data. My data—which consist of (i) household-level panel data on consumers' purchases and (ii) revealed-preference alternate-choice data—fit this description.

My data display two key limitations. The first concerns external validity: whereas most existing studies employ nationally representative data, mine cover only one (regional) retailer. As for the second limitation, my data provide less detailed information on consumers' preferences over unpurchased products than do the data employed in existing studies. Specifically, my data characterize consumers' revealed preferences between *one* unpurchased good—namely, a store-selected stockout substitute—and the "outside option" of purchasing nothing. By contrast, most existing studies leverage questionnaire data in which consumers either (i) state their second–most-preferred product or (ii) provide a complete ranking of the unpurchased products. Although these data describe hypothetical choices,[10] they are far more detailed than my data, and will thus provide more precise estimates of demand elasticities.

Unlike most previous studies, my objective is not to obtain a nationally-representative model of demand for a specific market. Rather, my task is to evaluate the degree to which the IPA properties of conditional and mixed logit coincide with consumers' observed behavior for various product categories. So far as this task is concerned, the limitations of my data are unlikely to prove a substantial hindrance.

B. The Econometric Literature on Identification with Alternate-Choice Data

An emerging econometric literature documents how alternate-choice data help to identify demand. Conlon and Mortimer (2021) show that, under certain conditions, second-choice data identify the "Average Treatment

[10] See Carlsson and Martinsson (2001) for a discussion in the context of environmental economics; Lusk and Schroeder (2004) for one in agricultural economics; Quaife et al. (2018) for one in health economics; and Brownstone and Small (2005) for one in transportation research.

for the Untreated" (ATUT) which may, in turn, be a good proxy for demand elasticities. In addition, preliminary work by Conlon, Mortimer, and Sarkis (2023) suggests that a pairing of (i) second-choice data and (ii) information on market shares can identify demand *even without data on products' observable characteristics*. Furthermore, nonparametric estimation using such data can sometimes match observed substitution patterns better than BLP '04–style demand systems, despite the latter exploiting additional data on product characteristics. In particular, BLP '04–style demand systems sometimes underpredict diversion to close substitutes and overpredict diversion to more distant ones. This tendency could be partially explained by the IPA property of mixed logit, which rules out within-consumer variation in preferences over product characteristics (such as might arise from variation in purchase circumstances).

C. The Literature on Stockouts and Demand Estimation

There is a large literature in empirical industrial organization and marketing that leverages stockout events to help estimate demand. The intuition is that the substitutability of one good—say, A—for another—say, B—can be inferred from the degree to which A's choice share increases when B goes out of stock. In this literature, the primary points of differentiation are (i) the institutional environment and (ii) the cause of product unavailability. Regarding (i), some of these papers' environments resemble mine, being either supermarkets or convenience stores. These include Musalem et al. (2010) and Bruno and Vilcassim (2008). Another important purchasing environment within this literature is vending machines, the subject of Anupindi, Dada, and Gupta (1998); Conlon and Mortimer (2021); Conlon and Mortimer (2013); and Conlon and Mortimer (2010). As for (ii), most studies rely on endogenous (i.e., naturally occurring) stockouts. Notable exceptions include Conlon and Mortimer's 2021 and 2010 studies, which experimentally manipulate product availability in vending machines.

The key difference between these studies and mine is the data. In my data, stockouts occur *after* the consumer has already made her initial purchase decision. Consequently, I observe two choices per stockout event: the consumer's "first choice" as well her later decision to accept or reject a store-selected substitute (after her first choice has gone out of stock). By contrast, the studies listed above observe only one choice per stockout event: the consumer's purchase from among the available alternatives. It remains unknown what the consumer would have purchased under full availability. Further, the aforementioned studies rely on cross-sectional data, whereas I have panel data. For both these reasons, my data are especially suitable to test the IPA properties of conditional and mixed logit.

One study within this literature may provide suggestive evidence of bias resulting from the mixed logit IPA. Conlon and Mortimer (2010) find that, when a product goes out of stock, the mixed logit model underpredicts the sales increase enjoyed by close substitutes and overpredicts that enjoyed by more distant substitutes. Notice that this is the same pattern identified by Conlon, Mortimer, and Sarkis (2023) in the context of the automotive market (as discussed in Section 2B). Regarding vending machines, Conlon and Mortimer propose several potential explanations for this pattern, such as omitted product characteristics or the absence of price variation in vending machines. However, the IPA property of mixed logit could also be responsible. Under this constraint, an individual consumer cannot be "in the mood" for a certain type of snack on one occasion but a different type on another.[11] So if an individual consumer opts for different categories of snacks on different occasions—such as a savory snack on one occasion and a sweet one on another—then the mixed logit model will assume she is (largely) indifferent between the two categories. But in actual fact, she might have had a strong preference for one category on a given occasion (e.g., "I could really use a salty snack right now") but a strong preference for a different category on another occasion (e.g., "I'm craving something sweet right now").

3. Theory: Alternate-Choice Data in Demand Systems

In this section, I introduce my empirical framework and then formalize the IPA properties of conditional and mixed logit.

Consider a differentiated products market with J goods (or "products"), along with an outside option of no purchase ("good 0"). At time t, each consumer i purchases the good $j \in \mathcal{J} \equiv \{0, 1, \ldots, J\}$ that affords the greatest conditional indirect utility u_{ijt}.[12]

Utility is a linear index of product characteristics (x_j), price (p_{jt}), and an i.i.d. Gumbel error (ε_{ijt}):

$$u_{ijt} = x_j \beta_i - \alpha_i p_{jt} + \varepsilon_{ijt}.$$

[11] To see how this would bias estimates of demand elasticities, picture a consumer who orders a savory snack—say, Lay's potato chips—on one occasion but a sweet one—say, Kit Kat—on another. Under the mixed logit IPA, her relative preferences among the unpurchased snacks must have remained the same on both occasions. Consider the counterfactual where both her first-choice products were out of stock on their respective purchase occasions—that is, Lay's potato chips were out of stock on the first occasion and Kit Kat out of stock on the second. Under the mixed logit IPA, she would have been no likelier to divert to a given savory snack—say, salted peanuts—on the first occasion (when, under full availability, she would have ordered a salty snack) than on the second (when, under full availability, she would have ordered a *sweet* snack).

[12] I assume that $\arg\max_{j \in \mathcal{J}} u_{iAt}$ is a singleton set with probability one. (In other words, there are no "ties.").

Note that the taste coefficients (β_i, α_i) are specific to individual consumers i.

I will show that conditional and mixed logit each impose a form of independence between consumers' purchases and their preferences among unpurchased goods. I begin by proving a lemma about (conditional) logit utilities. Then I use this lemma to derive the IPA properties of conditional and mixed logit.

Lemma 1 (Irrelevance of Identical Upper Bounds on Two Goods' Logit Utilities)**.** *Assume that all consumers share the same taste coefficients, with $(\beta_i, \alpha_i) = (\beta, \alpha)$ for all i. Then, for any two goods $A, B \in \mathcal{J}$ and any constant $K \in \mathbb{R}$,*

$$\Pr\left[u_{iAt} > u_{iBt} \,\middle|\, \max\{u_{iAt}, u_{iBt}\} < K\right] = \Pr[u_{iAt} > u_{iBt}].$$

Proof. See Appendix A. ∎

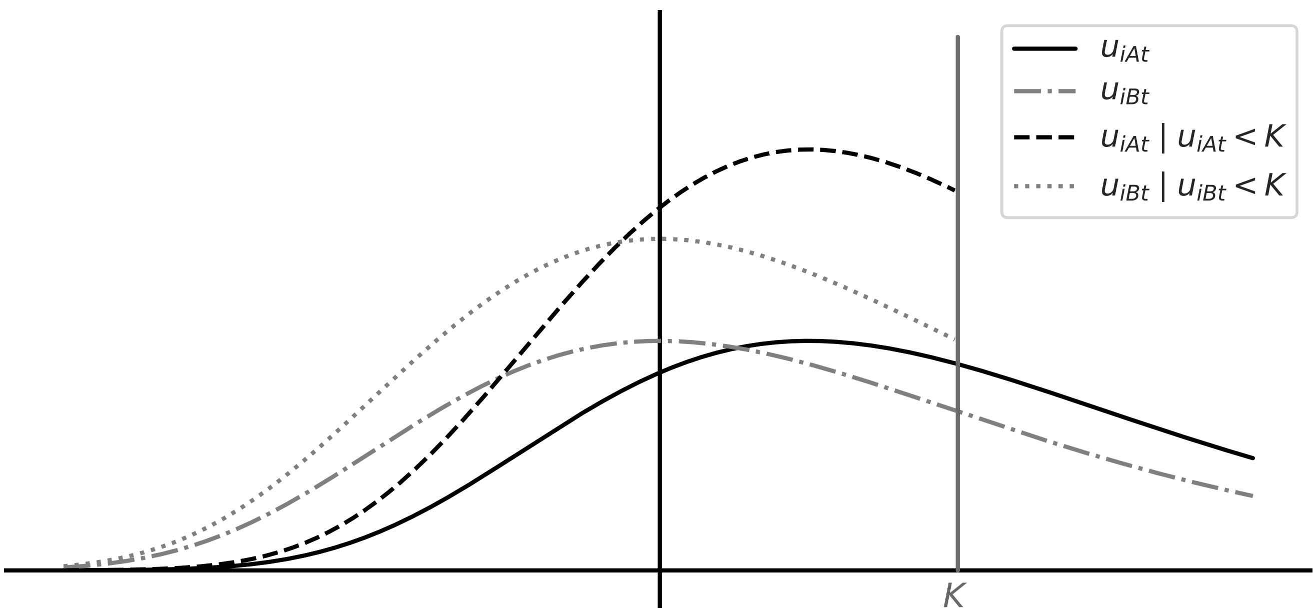

Figure 1 – Irrelevance of Identical Upper Bounds on Two Goods' Logit Utilities

Figure 1 depicts a generic example of Lemma 1. The black solid line and the gray dash-dotted line chart the unconditional PDFs of u_{iAt} and u_{iBt}, respectively; while the conditional PDFs of u_{iAt} and u_{iBt} respectively correspond to the black dashed and gray dotted lines.

Although both unconditional PDFs share the same shape, the unconditional PDF of u_{iAt} is a rightwards location-transformation of u_{iBt}'s. (Evidently, the representative utility of good A exceeds that of good B: $x_A\beta - \alpha p_{At} > x_B\beta - \alpha p_{Bt}$.) It follows that

$$\Pr[u_{iAt} > u_{iBt}] > \frac{1}{2}.$$

Turning to the conditional PDFs, notice that both are bounded above by K. However, they differ in shape, with the conditional PDF of u_{iAt} bunching more tightly around K than does the conditional PDF of u_{iBt}. It follows that

$$\Pr\left[u_{iAt} > u_{iBt} \,\middle|\, \max\{u_{iAt}, u_{iBt}\} < K\right] > \frac{1}{2}.$$

Conditional on being smaller than K, the random variable u_{iAt} is more likely to be "just under" the upper bound K than is the random variable u_{iBt}. Less intuitive, however, is the following result, which is implied by Lemma 1:

$$\Pr\left[u_{iAt} > u_{iBt} \,\middle|\, \max\{u_{iAt}, u_{iBt}\} < K\right] = \Pr[u_{iAt} > u_{iBt}].$$

That is, the probability that u_{iAt} is greater than u_{iBt} remains unchanged after imposing the condition that $\max\{u_{iAt}, u_{iBt}\} < K$. In visual terms, the two distributions will both compress to the left such that the probability of a draw from one distribution exceeding a draw from the other remains unchanged.

Not all distributions display this property. For instance, if the error terms were distributed i.i.d. standard normal (as opposed to i.i.d. Gumbel), then

$$\Pr\left[u_{iAt} > u_{iBt} \,\middle|\, \max\{u_{iAt}, u_{iBt}\} < K\right] \neq \Pr[u_{iAt} > u_{iBt}]$$

in general.[13] So this property represents an unusual feature of the (conditional) logit model and, by extension, of the Gumbel distribution.

I will now employ Lemma 1 to derive the IPA property of conditional logit.

Theorem 1 (Conditional Logit IPA)**.** *Assume that all consumers share the same taste coefficients, with* $(\beta_i, \alpha_i) = (\beta, \alpha)$ *for all* i. *Then, for any three goods* $A, B, C \in \mathcal{J}$,

$$\Pr\left[u_{iBt} > u_{iCt} \,\middle|\, u_{iAt} = \max\nolimits_{j\in\mathcal{J}} u_{ijt}\right] = \Pr[u_{iBt} > u_{iCt}].$$

Proof. By the law of iterated expectations,

$$\begin{aligned}\Pr\left[u_{iBt} > u_{iCt} \,\middle|\, u_{iAt} = \max\nolimits_{j\in\mathcal{J}} u_{ijt}\right] = \mathrm{E}\Big[\Pr\Big[u_{iBt} > u_{iCt} \,\Big|\, u_{iAt} = \max\nolimits_{j\in\mathcal{J}} u_{ijt};\\ (u_{ijt})_{j\in\mathcal{J}\setminus\{B,C\}}\Big] \,\Big|\, u_{iAt} = \max\nolimits_{j\in\mathcal{J}} u_{ijt}\Big].\end{aligned} \tag{1}$$

[13] By way of example, suppose $u_{iAt} = 1 + \varepsilon_{iA}$ and $u_{iBt} = \varepsilon_{iB}$, where the error terms are i.i.d. standard normal. Then $\Pr\left[u_{iAt} > u_{iBt} \,\middle|\, \max\{u_{iAt}, u_{iBt}\} < K\right] \approx 0.66 < 0.76 \approx \Pr[u_{iAt} > u_{iBt}]$.

As far as the inner component of (1) is concerned, only two goods' utilities are random variables: those of B and C. (The remaining goods' utilities are constants.) We can therefore apply Lemma 1 to the inner component of (1), obtaining

$$\Pr\left[u_{iBt} > u_{iCt} \,\middle|\, u_{iAt} = \max_{j\in\mathcal{J}} u_{ijt};\ (u_{ijt})_{j\in\mathcal{J}\setminus\{B,C\}}\right] = \Pr[u_{iBt} > u_{iCt}]. \tag{2}$$

Substituting (2) into (1) yields

$$\begin{aligned}\Pr\left[u_{iBt} > u_{iCt} \,\middle|\, u_{iAt} = \max_{j\in\mathcal{J}} u_{ijt}\right] &= \mathrm{E}\left[\Pr[u_{iBt} > u_{iCt}] \,\middle|\, u_{iAt} = \max_{j\in\mathcal{J}} u_{ijt}\right] \\ &= \Pr[u_{iBt} > u_{iCt}].\end{aligned}$$

■

Importantly, goods A, B, and C need not be "inside goods." Rather, one of them could be the outside option: good 0. Such is the case for the empirical application to curbside pickup in Sections 5 and 6.[14]

I will now show that mixed logit exhibits an analogous IPA property, conditional on the realizations of consumers' random taste coefficients.

Corollary 1 (Mixed Logit IPA)**.** *For any three goods $A, B, C \in \mathcal{J}$,*

$$\Pr\left[u_{iBt} > u_{iCt} \,\middle|\, u_{iAt} = \max_{j\in\mathcal{J}} u_{ijt}; \beta_i, \alpha_i\right] = \Pr[u_{iBt} > u_{iCt} \mid \beta_i, \alpha_i].$$

Proof. Follows immediately from Theorem 1 and the definition of mixed logit. ■

I discuss the practical implications of Theorem 1 and Corollary 1 elsewhere.[15] In addition, Appendix B relates Theorem 1 to prior theoretical results in the literature, while Appendix C presents Monte Carlo tests of Theorem 1.

[14]Briefly: in curbside pickup, a consumer is observed making two related choices: her initial order, and her subsequent decision to accept or reject a stockout substitute. Regarding the former, her decision to order some good j for curbside pickup indicates that she prefers j to the outside option (good 0). As to the latter, she will accept an inside good $j' \in \mathcal{J} \setminus \{0, j\}$ if and only if j' is preferred to the outside option. See Sections 5 and 6 for details.

[15]See Section 5 for an application of Theorem 1 to curbside grocery pickup; and for applications of Corollary 1 to the automotive market and to curbside grocery pickup, see Sections 2 and 5, respectively.

4. Institutional Background and Data

This section introduces the data, which concern curbside grocery pickup at a regional supermarket chain. In what follows, I first provide an overview of curbside grocery pickup and then catalog the contents of the data.

A. *Institutional Background*

Curbside grocery pickup is a form of online shopping in which a consumer orders her groceries online and later picks them up from a bricks-and-mortar supermarket. Her shopping experience proceeds according to the following timeline. First, she uses the supermarket's website or its smartphone app to place her order, indicating which items she wants as well as when she would like to pick them up (e.g., tomorrow morning). Some time later, a supermarket worker gathers the requested items and sets them aside to await pickup. Once the consumer arrives, the worker will bring the items out to her car, where she will pay for them.

Sometimes, however, an item in the consumer's order goes out of stock *after* she has placed the order, but *before* the supermarket worker assembles it. In that event, the worker will choose another product to serve as a substitute. Once the consumer arrives, she will be presented with two choices: either she can accept the substitute that the worker chose earlier on her behalf, or she can reject it and buy no such product at all.

B. *Data*

This study employs three data sets from a regional supermarket chain. The first, hereafter referred to as the "curbside stockout" data set, concerns stockout substitutions in curbside pickup orders. For each stockout event, these data report the universal product code (UPC) of the out-of-stock item as well as that of the substitute offered. I also observe the price of the substitute, as well as whether the substitute is accepted or rejected by the consumer.[16] The data also assign a unique identifier to each transaction, enabling me to match them to the second data set.

The second data set comprises "scanner data," which characterize all purchases at the supermarket chain, irrespective of shopping channel (i.e., in-store, delivery, or curbside pickup). For each purchased item, these data report the UPC and price. I also observe transaction IDs that follow the same system as the curbside stockout data, enabling me to match the two data sets. In addition, the scanner data record the loyalty program ID of the consumer making the purchase, lending the data a panel structure.[17]

[16] The price of the out-of-stock item is obtained from the second data set. See Section 6 for details.

[17] Although participation in the loyalty program is not compulsory in general, it *is* required in order to place curbside pickup

The final data set is the chain's "product catalog," which characterizes all the products sold at the chain. For each product, the catalog reports the UPC and brand, along with the location in the chain's product taxonomy. The catalog also provides a string description of the product, from which I extract information on its observable characteristics (using so-called "regular expressions"). For example, here is the string description for one of the flour products:

"GOLD MEDAL FLOUR HARVEST KING BREAD 5 LB"

This description classifies the product as a bread flour (as opposed to, say, all-purpose or wheat). It also indicates the quantity of flour: five pounds.

Table 1 reports summary statistics for the two product categories studied in Sections 5 and 6: bottled water and flour. These categories were chosen for three reasons. First, I observe many stockout substitutions for products in these categories. Second, product differentiation within each category is fairly uncomplicated. That is to say, a given product's utility depends on only a few observable characteristics (a fact which simplifies the structural analysis in Section 6). And third, the categories display dramatically different levels of variation in consumers' preferences over time. Recall that the mixed logit IPA constrains within-consumer preference variation as follows: each consumer's preferences among unpurchased products should remain constant across all her trips. Thus, if consumers' preferences remain stable over time in a given product category, the IPA property of mixed logit will mirror consumers' true preferences over unpurchased products. On the other hand, if consumers' preferences vary between shopping trips, the mixed logit IPA will be inconsistent with their true preferences over unpurchased products.

To test whether the mixed logit IPA is inconsistent with the behavior of consumers whose preferences differ between trips, I consider a product category with considerable within-consumer preference variation: flour. The reason that flour buyers' preferences vary between trips is that specific flours are suited to specific recipes. If someone plans to bake bread, she would probably prefer bread flour; whereas if she intends to bake cupcakes, she would probably favor all-purpose flour.[18] By way of comparison, I also study a product category whose buyers likely exhibit stable preferences over time: bottled water. Consumers' preferences concerning this category probably persist over time because bottled waters are *functionally* interchangeable.[19] In consequence, a consumer's order choice will largely depend on (i) her subjective assessments of products'

orders. Consequently, I can match the purchases of curbside pickup patrons to their in-store and delivery purchases.

[18] Although any flour can be used in any recipe, using the "wrong" type of flour may require extra work on the baker's part—such as adjusting the recipe—and may also result in an inferior final product.

[19] All bottled waters must satisfy FDA "standard of quality" conditions (U.S. Food & Drug Administration 2022), which regulate the maximum level of contaminants in the product. In addition, most bottled waters share the same size: 16.9 fl oz.

tastes and (b) her price sensitivity. And one would expect both (i) and (ii) to remain fairly constant between trips.

TABLE 1 – SUMMARY STATISTICS BY PRODUCT CATEGORY

Statistic	Bottled water	Flour
	Panel A. Overview	
No. of households with 1+ substitutions	66,447	22,549
No. of distinct products purchased	39	52
. . . of which ordered for curbside pickup	30	38
No. of distinct brands purchased	9	14
. . . of which ordered for curbside pickup	9	8
	Panel B. Per household with 1+ substitutions	
No. of shopping trips	37.8	11.6
. . . of which curbside pickup	6.9	2.9
. . . of which feature 1+ substitutions	1.6	1.1
No. of distinct products ever purchased	5.6	4.0
. . . of which ordered for curbside pickup	2.4	1.9
No. of distinct brands ever purchased	3.1	2.3
. . . of which ordered for curbside pickup	1.7	1.4
	Panel C. Stockout substitutions	
Prob. of acceptance	0.873	0.920

Notes: All estimates are reported as means or totals. By "brands," I refer to food companies, each of which may sell multiple products in a given category. For instance, the Gold Medal brand sells many types of flour, such as "Whole Wheat" and "All-Purpose Bleached."

Having explained why bottled water and flour form the focus of my empirical analysis, I now return to the summary statistics in Table 1. Panel A presents an overview of these product categories. Notice that almost three times as many households have experienced a stockout substitution for bottled water (66,450) as have experienced one for flour (22,549). The categories also differ, albeit less dramatically, with respect to the number of distinct brands and products carried by the chain. (By "brand," I refer to the name of the food company under which multiple different products may be sold. For instance, the Gold Medal brand sells many different flour products, such as "Whole Wheat" and "All-Purpose Bleached.") Specifically, there are more distinct brands of flour—as well as individual products—than there are of bottled water. In addition, observe that only a proper subset of the chain's offerings in either category are available for curbside pickup.

Turning to the panel dimension of the data, Panel B reports that the average household (who has experienced one or more substitutions) has made more shopping trips that involve bottled water (37.94) than flour (11.78). A modest fraction of these trips are curbside pickup (18% and 25% for bottled water and flour, respectively).

On average, bottled water buyers have experienced slightly more stockout substitutions (1.6) than have their flour counterparts (1.14). Perhaps in consequence of having made more purchases, the average bottled water buyer has purchased more distinct brands and products than has her flour counterpart.

Concerning stockout substitutions, Panel C indicates that flour buyers are likelier to accept the substitute on offer (92.0%) than are their bottled water counterparts (87.2%).

5. Descriptive Evidence

In this section, I provide descriptive evidence concerning the extent to which consumer behavior coincides with the IPA properties of conditional and mixed logit. Because the IPA property of conditional logit is a cross-sectional independence constraint, whereas that of mixed logit is a within-panel constraint, I examine the two properties separately.

A. The Conditional Logit IPA

The IPA property of conditional logit imposes independence between a consumer's purchase and her preferences among the unpurchased products. In the context of curbside pickup, the consumer's "purchase" corresponds to her order choice. Thus, the conditional logit IPA imposes independence between her original order and her preferences among the goods she did *not* order—including the "outside option" of buying nothing.

To see why, consider a consumer i who is placing an order for curbside grocery pickup at time t. She must choose among J_t differentiated goods and the "outside option" of no purchase ("good 0"). She will order whichever good $j \in \mathcal{J}_t = \{0, 1, \ldots, J_t\}$ affords the greatest conditional indirect utility u_{ijt},[20] given by

$$u_{ijt} = x_j\beta - \alpha p_{jt} + \varepsilon_{ijt}.$$

In this equation, x_j is a vector of product characteristics, p_{jt} denotes the price, and ε_{ijt} is an i.i.d. Gumbel

[20] I model "conditional" demand—that is, demand conditional on ordering one of the inside goods. There are two reason for adopting this approach. First, on occasions when someone visits the store but does not purchase a product within a given product category, it is unclear whether (i) she actively considered the store's offerings within the category, but decided the "outside option" of no purchase was preferable; or (ii) she never examined the store's offerings at all, as she had no need for a product in the category. As for the second reason that I model conditional demand, it is that the value of the "outside option" may differ *within a given curbside pickup trip*. When the consumer is assembling her order at home, she may be more (or less) disposed to prefer the outside option than when she has been offered a stockout substitute at the store. (For instance, after she has placed her order, she may be committed to preparing a specific recipe based on the combination of items in her pickup order.)

error. Regarding the outside option, I normalize $u_{i0t} = \varepsilon_{i0t}$.

Suppose that consumer i orders an inside good $j \in \mathcal{J}_t \setminus \{0\}$. This suggests that she prefers j over the other inside goods as well as the outside option: $u_{ijt} = \max_{j' \in \mathcal{J}_T} u_{ij't}$.

Now imagine that our consumer's ordered good j goes out stock. As a result, she faces a binary choice between (i) a stockout substitute $s \in \mathcal{J} \setminus \{0, j\}$ and (ii) the outside option. She will accept the substitute s if and only if $u_{ist} \geqslant u_{i0t}$.[21] Given her original order choice (j), what is the probability that she accepts s? Under Theorem 1,

$$\begin{aligned}
\Pr[i \text{ accepts } s \mid i \text{ ordered } j] &= \Pr[u_{ist} \geqslant u_{i0t} \mid u_{ijt} = \max_{j' \in \mathcal{J}} u_{ij't}] \\
&= \Pr[u_{ist} \geqslant u_{i0t}] \\
&= \Pr[i \text{ accepts } s]
\end{aligned}$$

In other words, the probability of acceptance should be independent of our consumer's original order choice.

This independence constraint can be directly tested in the data by tallying the acceptance probabilities for each ordered/substitute product pairing and then applying a likelihood-ratio test of conditional independence. The null hypothesis is that a given good's probability of being accepted is independent of the consumer's original order.

In taking this test to the data, I entertain two specifications. The first includes all ordered/substitute product pairings observed in the data. However, this specification suffers from (potential) amelioration bias, as most substitute products are only offered as substitutes for a small subset of out-of-stock products. Although I employ the "rule of three" correction,[22] the results should still be interpreted with caution. I therefore prefer a second specification, which focuses on a smaller analysis sample with only the *top ten products* in each product category (in terms of curbside sales among households who experience one or more stockout substitutions).

Under the first specification (which includes all product pairings), the null hypothesis of independence is rejected for bottled water ($p < 10^{-300}$),[23] but not for flour ($p \approx 1$).[24] However, in both these categories, more than three-quarters of the cells in the three-way contingency table are empty. Turning to the second

[21] Without loss of generality, I normalize the utility of the outside option so that $u_{i0t} = \varepsilon_{i0t}$.

[22] See Jovanovic and Levy (1997) or Tuyl, Gerlach, and Mengersen (2009).

[23] The log likelihood ratio test statistic is 4521, with 2186 degrees of freedom. The latter is computed as $((\text{no. of unique out-of-stock products}) - 1)(\text{no. of unique substitute products})$

[24] The p-value is numerically indistinguishable from one, given a log likelihood ratio test statistic of 1615 with 2435 degrees of freedom.

specification, which attends only to pairings of the top ten products in each category, the null hypothesis of conditional independence is rejected with $p < 10^{-300}$ in both categories.[25]

Although the foregoing exercise maps straightforwardly to the conditional logit IPA (as expressed in Section 3), it suffers from two drawbacks. First, there are many products within each category. This makes it difficult to discern why the conditional logit IPA is, or is not, satisfied within a category. And second, the exercise is removed from empirical practice. It is not common practice to estimate consumers' tastes for individual goods (i.e., with product dummies). Rather, it is customary to parameterize utility as a linear index of product characteristics such as brand or size. I therefore emphasize a different descriptive exercise which focuses on product characteristics, as opposed to specific substitute/out-of-stock product pairings. This exercise centers on the following corollary to the conditional logit IPA (Theorem 1).

The conditional logit IPA imposes independence between the following:

(1) The identity of the out-of-stock product

(2) The decision to accept or reject a given substitute

Provided that utility is a linear index of product characteristics, the succeeding pair of factors should also be mutually independent:[26]

(1A) The characteristics of the out-of-stock product

(2A) The decision to accept or reject a substitute with given characteristics

In other words, a substitute is no likelier to be accepted if its characteristics closely resemble those of the out-of-stock product than if they are highly dissimilar. Rather, what matters is the "popularity" of the substitute's characteristics. Are the substitute's characteristics—brand, size, flavor, etc.—ones that feature in a large share of orders? If so, the substitute affords high representative utility and,[27] in consequence, will enjoy a comparatively high acceptance probability. On the other hand, if the substitute's characteristics appear in only a small fraction of orders, then its representative utility must be relatively small, in which case it will suffer a comparatively low acceptance probability. At all events, the extent to which the product is substitutable for the out-of-stock product—as indicated by its (dis)similarity in observable characteristics—is irrelevant.

[25]The likelihood ratio test statistics are 3099 and 515 for the product categories of bottled water and flour, respectively. There are 80 degrees of freedom in each case.

[26]This follows from the definition of conditional independence; factors (1A) and (2A) are more aggregate partitions of the product space than are factors (1) and (2), respectively.

[27]By "representative utility," I mean the modeled portion of utility (as opposed to the error term).

Table 2 presents the results of this test for the product categories of bottled water and flour. For each category, the leftmost column lists the characteristics that differentiate products within the category. (For instance, bottled water is differentiated with respect to four characteristics: brand, the number of bottles in the case, the size of each bottle, and the type of water.) Then the second and third columns catalog possible pairings of the out-of-stock product and substitute's versions of a given characteristic. Where polytomous characteristics are concerned (such as brand or bottle count),[28] there are too many versions of the characteristic to enumerate all possible pairings. I therefore report results solely for the top two versions of each characteristic.[29] (For example, the top two brands of bottled water are Ice Mountain and the store's private label.) Finally, for each pairing of the substitute and out-of-stock products' versions of the characteristic, the remaining columns report the probability of acceptance as well as the number of observations.

According to the conditional logit IPA, the probability of acceptance should depend only on the "popularity" of the substitute's characteristics;[30] whether they match the out-of-stock product's characteristics should be immaterial. However, Table 2 does not support this prediction. To see why, consider a specific characteristic within a product category (such as brand). Notice that the four rows corresponding to the characteristic are ordered on (i) the *substitute's* version of the characteristic and then (ii) the out-of-stock product's version. For under the conditional logit IPA, the probability of acceptance should only depend on the substitute's version of the indicated characteristic, not on the out-of-stock product's version. Hence, among the four rows for a given characteristic, the probability of acceptance should be the same for the first and second rows, as well as for the third and fourth rows. For example, the first and second (third and fourth) rows of panel A both concern stockouts in which the substitute is sold under the private label (Ice Mountain brand). Per the conditional logit IPA, the first and second (third and fourth) rows should thus report identical acceptance probabilities.

In point of fact, the probability of acceptance tends to be greater when the out-of-stock product and the substitute share the same version of the characteristic than when they feature different versions. This is intuitive; one would expect consumers to prefer substitutes that resemble their first-choice products.

There are several apparent departures from this pattern. For example, a comparison of the third and fourth rows in Panel A suggests that an Ice Mountain-branded substitute is likelier to be accepted if the consumer

[28] That is, characteristics with more than two distinct realizations.

[29] Within the analysis sample, comprising purchases by households with 1+ attempted substitutions.

[30] Formally, on the representative utility afforded by the substitute's characteristics.

TABLE 2 – TESTING THE CONDITIONAL LOGIT IPA

Characteristic	Out-of-stock product's version	Substitute's version	Prob. accept	No. of obs.
	Panel A. Bottled water			
Brand	Private label	Private label	0.918	30,918
	Ice Mountain	Private label	0.835	8283
	Ice Mountain	Ice Mountain	0.890	8903
	Private label	Ice Mountain	0.931	11,628
No. of bottles	24	24	0.861	69,823
	40	24	0.930	17,311
	40	40		0
	24	40	0.918	4712
Size of individual bottles	16.9 fl oz	16.9 fl oz	0.878	84,439
	8 fl oz	16.9 fl oz	0.789	1495
	8 fl oz	8 fl oz	0.850	787
	16.9 fl oz	8 fl oz	0.718	840
Water type	Spring	Spring	0.905	33,260
	Purified	Spring	0.845	16,346
	Purified	Purified	0.894	37,955
	Spring	Purified	0.801	19,619
	Panel B. Flour			
Brand	Private label	Private label	0.948	4614
	King Arthur	Private label	0.892	1719
	King Arthur	King Arthur	0.863	3954
	Private label	King Arthur	0.938	838
Quantity	5 lb	5 lb	0.908	17,887
	2 lb	5 lb	0.955	1587
	2 lb	2 lb	0.938	1013
	5 lb	2 lb	0.936	2008
Type of flour	All-purpose flour	All-purpose flour	0.942	19,966
	Bread flour	All-purpose flour	0.825	1778
	Bread flour	Bread flour	0.911	1983
	All-purpose flour	Bread flour	0.840	344
Whether bleached or not	Bleached	Bleached	0.961	9639
	Unbleached	Bleached	0.899	4398
	Unbleached	Unbleached	0.898	8021
	Bleached	Unbleached	0.929	2686

Notes: This table presents the probability of a stockout substitute being accepted, conditional on its own version of a specific characteristic as well as that of the out-of-stock product. If the characteristic in question takes more than two values (as is the case for "brand" in all three product categories), only the top two versions of the characteristic are considered (based on purchases by households with one or more curbside stockouts).

had originally ordered a private-label product than if she had ordered an Ice Mountain product. Results of this kind appear to arise for two reasons. First, where some characteristics are concerned, consumers who have ordered one version of the characteristic are likelier to accept than consumers who have ordered the other—irrespective of the substitute's version. Such is the case for bottled water brands. Whether the substitute is sold under the private label or under the Ice Mountain brand, it is likelier to be accepted if the consumer had originally ordered the private label than if she had originally ordered Ice Mountain. As to the second source of these discrepancies, it concerns the finitude of the product space within a particular product category. Because the store cannot find a substitute that exactly matches the substitute on all characteristics, it will settle for one that matches it in some characteristics but not others. As a result, dissimilarity between the substitute and the out-of-stock product with respect to one characteristic is often associated with similarity with another (see Appendix Table 1 in Appendix D for a correlation matrix). And if the first characteristic is less important to the consumer than the second, the result will be an inverse correlation between the probability of acceptance and the substitute's sharing the first characteristic with the out-of-stock product.

To illustrate, consider a stockout event involving flour. For most consumers, the characteristic of flour type matters more than the characteristic of quantity does. So, given the choice, a consumer would probably prefer a substitute that matches the out-of-stock product's flour type (but not its quantity) over an alternate substitute that matches the out-of-stock product's quantity (but not its flour type). In addition, there is an inverse correlation between (i) being offered a substitute that matches the out-of-stock product's flour type and (ii) being offered a substitute that matches the out-of-stock product's quantity (as reported in Appendix Table 1). The result is an inverse correlation between acceptance and the substitute's sharing the out-of-stock product's quantity.

That the (dis)similarity of the offered substitute's characteristics to those of the out-of-stock product is predictive of acceptance—even conditional on the substitute's characteristics—is inconsistent with the conditional logit IPA. This finding is hardly unexpected. In most differentiated products markets, consumers exhibit heterogeneous preferences over observable characteristics. And, in the context of curbside pickup, an individual consumer's order choice should provide some indication of her tastes (which may differ from the population "average"). The result is a positive correlation between (i) the similarity of the substitute and out-of-stock product and (ii) the probability of acceptance (with a few exceptions due to the finitude of the product space, as described above).

B. The Mixed Logit IPA

Having tested the IPA property of conditional logit, I now turn to its mixed logit counterpart. To see how Corollary 1 relates to curbside pickup, consider the same consumer i as in the preceding subsection. (Recall that she ordered good j at time t and, after j went out of stock, was offered good s as a stockout substitute.)

Unlike conditional logit, mixed logit allows our consumer's random taste coefficients (β_i, α_i) to differ from those of other consumers. How does this affect the probability of acceptance? Per Corollary 1,

$$\begin{aligned}
\Pr[i \text{ accepts } s \mid i \text{ ordered } j; \beta_i, \alpha_i] &= \Pr[u_{ist} \geqslant u_{i0t} \mid u_{ijt} = \max_{j' \in \mathcal{J}} u_{ij't}; \beta_i, \alpha_i] \\
&= \Pr[u_{ist} \geqslant u_{i0t} \mid \beta_i, \alpha_i] \\
&= \Pr[i \text{ accepts } s \mid \beta_i, \alpha_i]
\end{aligned}$$

In other words, our consumer's order choice should be uninformative of her decision to accept or reject the substitute, conditional on her *time-invariant* tendency to like or dislike its observable characteristics.[31]

Table 3 – Stylized Example of the Mixed Logit IPA

	Consumer P			Consumer M		
Trip	Order	Substitute	Prob. accept	Order	Substitute	Prob. accept
1	PL			IM		
2	PL			IM		
3	PL	PL'	p_i	PL	PL'	p_{iii}
4	IM	PL'	p_{ii}	IM	PL'	p_{iv}

Note: Products PL and PL' are sold under the private label, while good IM is sold under the Ice Mountain brand.

To see the significance of this constraint, consider two consumers who regularly order bottled water for curbside pickup. One of them, consumer P, usually purchases the private label;[32] whereas the other, consumer M, typically opts for Ice Mountain. Table 3 summarizes their orders and stockout substitutions. On trips 1 and 2, each consumer orders her customary brand, with consumer P choosing product PL (one of the private label's offerings) and consumer M opting for product IM (one of Ice Mountain's). On trips 3 and 4, by contrast, their orders coincide exactly, with both choosing product PL on trip 3 and product IM on trip 4. However, on trips 3 and 4, both consumers' orders go out of stock, and they are each offered product PL' as

[31] In my differentiated products demand framework, as well as that in Berry, Levinsohn, and Pakes (2004), there is only one source of within-consumer variation in a particular good's representative utility: price changes (for which I include controls in the descriptive exercises below). Although some studies, such as Grieco, Murry, and Yurukoglu (2023), accommodate secular shifts in goods' representative utility over time, they do so at the market level (as opposed to the household level).

[32] That is, the store's eponymous brand of groceries.

a substitute. Assume that PL' shares the same brand as PL—namely, the private label—and is generally a closer substitute for PL than for IM.

How does the probability of acceptance vary across these four (attempted) substitutions? Intuitively, there are two key determinants of acceptance or rejection here: (i) the consumer's time-invariant tendency to like or dislike the characteristics of the substitute, and (ii) trip-specific considerations. To see how these factors figure in our stylized example, let p_i and p_{ii} denote the probability that consumer P accepts PL' on her third and fourth trips, respectively. Likewise, let p_{iii} and p_{iv} denote the probability that consumer M accepts PL' on her third and fourth trips, respectively.

Focus first on the consumers' time-invariant tendencies to (dis)like the characteristics of the substitute, PL'. Recall that consumer P tends to favor the private label over Ice Mountain, whereas consumer M exhibits the reverse tendency. Thus, the substitute PL' shares the same brand as consumer P's go-to product, but does not share the brand of consumer M's. As a result, when the two consumers have ordered the same product, consumer P should be likelier to accept PL' as a substitute than is consumer M. In other words, p_i should exceed p_{iii} and p_{ii} should exceed p_{iv}. This intuition is supported by the mixed logit IPA, which allows a given substitute's acceptance probability to vary based on individual consumers' (heterogeneous) time-invariant tendencies to like or dislike the substitute's observable characteristics (here, its brand).

Turning to trip-specific considerations, note that consumers sometimes deviate from their usual order behavior due to unusual circumstances. Take the case of consumer P's order on trip 4, for example. Although consumer P usually prefers the private label, here she departs from this pattern and orders the Ice Mountain brand instead. This departure suggests the presence of trip-specific circumstances that make Ice Mountain more attractive than usual, relative to the private label. Perhaps she is hosting guests who are partial to Ice Mountain, whereas on previous trips she was shopping just for herself (and could therefore purchase the private label, which she prefers). At all events, her decision to pass over the private label in favor of Ice Mountain suggests that she may be less amenable to a private-label substitute than usual. One would therefore expect p_{ii} to be smaller than p_i. By similar logic, consumer M's uncharacteristic decision to order the private label in trip 3, as opposed to her go-to brand (Ice Mountain), suggests that she may be more amenable to a private-label substitute than usual. Consequently, one would expect p_{iii} to exceed p_{iv}. However, neither of these intuitions are consistent with the mixed logit IPA, under which consumers' order choices should be independent of the probability of accepting a given substitute. Here, this means that $p_i = p_{ii}$ and $p_{iii} = p_{iv}$.

Are the foregoing predictions of the mixed logit IPA consistent with the data? To provide insight, I estimate a probit model in which the probability of acceptance depends on (i) the extent to which the substitute's characteristics resemble those of the out-of-stock product, and (ii) the consumer's time-invariant tendency to like or dislike the characteristics of the substitute. Regarding (i), I include a set of indicators variables for the substitute's sharing a given characteristic k (such as brand) with the out-of-stock product. Let $\text{same}_{ik} = 1$ if consumer i is offered a substitute that shares characteristic k with the out-of-stock product, and $\text{same}_{ik} = 0$ otherwise. As for (ii), I proxy for the consumer's time-invariant tendency to like (or dislike) the substitute's characteristics as follows. Leveraging the panel structure of the data, I compute the fraction of the consumer's shopping trips—past, present, and future—in which the purchased product shares the substitute's version of characteristic k.[33] I denote the resulting fraction by frac_{ik}. The intuition is that, if the consumer likes the substitute's version of a given characteristic, a large fraction of her purchases will feature it; whereas if she dislikes it, only a small fraction will. To illustrate, I return to the stylized example about bottled water buyers in Table 3. Minding that this example centers on the product characteristic of brand, consider trip 3. Both consumers' preferred products go out of stock on this trip, and both of them are offered PL' as a substitute. Concentrate first on consumer P. Of the four trips observed in the data, she chooses product PL on three and product IM on one. Only the former product is sold under the same brand as the substitute PL'—namely, the private label—so the proxy variable $\text{frac}_{P,\text{brand}}$ equals three-fourths. Now turn to consumer M, who opts for product IM on three of her four trips and product PL on the remaining one. As the latter (but not the former) shares the brand of the substitute PL', the variable $\text{frac}_{M,\text{brand}}$ equals one-quarter.

Observable characteristics aside, the price of the substitute may also be informative of the decision to accept or reject. In particular, acceptance may be less likely if the substitute is perceptibly pricier than the out-of-stock product. For this reason, I permit the probability of acceptance to depend on the difference between the substitute's price ($p_{i,\text{sub}}$) and that of the out-of-stock product ($p_{i,\text{OOS}}$).[34]

All told, I take the following probit model to the data. Letting $a_i = 1$ if consumer i accepts and $a_i = 0$

[33] Where curbside pickup is concerned, I define the consumer's "purchase" as the product that she originally ordered—even if it goes out of stock and she purchases a substitute. (In that event, her original order choice will be more informative of her preferences than the substitute, which is chosen by the store.)

[34] As discussed in Section 4, I do not observe the out-of-stock product's price. Instead, I search the data for the nearest date on which the out-of-stock product was purchased at the store in question. Then I impute the out-of-stock product's price as being the average purchase price on the date in question. For details on how I impute prices, see Section 6.

otherwise, I estimate:

$$a_i = \begin{cases} 1 & \text{if } a_i^\star \geqslant 0 \\ 0 & \text{if } a_i^\star < 0, \end{cases}$$

where

$$a_i^\star = \sum_{k=1}^{K} (\gamma_k \text{same}_{ik} + \zeta_k \text{frac}_{ik}) + \eta \cdot (p_{i,\text{sub}} - p_{i,\text{OOS}}) + \upsilon_i,$$

and υ_i is distributed i.i.d. standard normal.

Under the mixed logit IPA, whether the substitute matches the out-of-stock product's version of a characteristic k (as captured by the same_{ik} variable) should be uninformative of acceptance, conditional on how often the consumer purchases products with the substitute's version of the characteristic (as given by the frac_{ik} variable). So, if consumers' behavior is consistent with the mixed logit IPA, the ζ_k's should be positive whereas the γ_k's should be indistinguishable from zero.

To illustrate how the mixed logit IPA would manifest in the data, I revisit the stylized example about water bottle buyers in Table 3. Recall that the consumers were offered good PL' as a substitute on two occasions: trip 3, when both consumers had originally ordered good PL, and trip 4, when both consumers had ordered good IM. Now suppose that the consumers' behavior is consistent with the mixed logit IPA—that is, $\gamma_{\text{brand}} = 0$. Although the substitute (PL') shares the same brand as the consumers' preferred good on trip 3 (PL) but not their preferred good on trip 4 (IM), the probability of acceptance should be the same on both trips for each consumer. That is, $p_i = p_{ii}$ and $p_{iii} = p_{iv}$.

Turning to the regression results, Table 4 reports the average marginal effects of the explanatory variables.[35] Notice that there are two variables for each observable characteristic: an indicator for the substitute's sharing the out-of-stock product's version of the characteristic, and a scalar variable for the fraction of the consumer's shopping trips where the purchased product shares the substitute's version of the characteristic. The table is organized so that the coefficients on the former (i.e., the γ_k's) are situated above the coefficients on the latter (i.e., the ζ_k's).

As far as bottled water is concerned, consumers' behavior seems to be consistent with the mixed logit IPA. For all four characteristics, the marginal effect associated with the fraction of purchases that share the

[35] Specifically, I compute the average marginal effect of a change in each variable on the probability of acceptance. (By "average," I mean the following. First, I compute the variables' marginal effects for each individual observation; and second, I take the average across all the observations. An alternative approach, which I do *not* employ, is to compute the marginal effects at the sample means.)

Table 4 – Testing the Mixed Logit IPA: Average Marginal Effects from Probit Regressions

Variable	Product category: Bottled water	Product category: Flour
Brand		
Sub shares OOS product's version	–0.012***	–0.066***
	(0.004)	(0.007)
Frac. of purchases with sub's version	0.148***	0.061***
	(0.006)	(0.011)
No. of bottles		
Sub shares OOS product's version	–0.023***	
	(0.003)	
Frac. of purchases with sub's version	0.037***	
	(0.005)	
Size of each bottle		
Sub shares OOS product's version	0.032***	
	(0.004)	
Frac. of purchases with sub's version	0.051***	
	(0.006)	
Water type		
Sub shares OOS product's version	0.044***	
	(0.003)	
Frac. of purchases with sub's version	0.091***	
	(0.005)	
Flour type		
Sub shares OOS product's version		0.124***
		(0.008)
Frac. of purchases with sub's version		0.026**
		(0.010)
Quantity		
Sub shares OOS product's version		–0.064***
		(0.009)
Frac. of purchases with sub's version		0.016
		(0.011)
Whether bleached or not		
Sub shares OOS product's version		0.002
		(0.007)
Frac. of purchases with sub's version		0.039***
		(0.010)
Sub's price – OOS product's price	0.021***	–0.003
	(0.001)	(0.002)
Observations	81,360	14,181
Pseudo R^2	0.0670	0.0716

Notes: The dependent variable is whether a stockout substitute is accepted (=1) or rejected (=0). The table reports average marginal effects, not coefficients. Standard errors are in parentheses. (Because some households experience multiple stockouts, the standard errors are clustered at the household level.)

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

substitute's version of the characteristic is much larger in magnitude than the marginal effect associated with the substitute's (not) sharing the out-of-stock product's version of the characteristic. This pattern is particularly pronounced where brand and water type are concerned. All else equal, acceptance is 14.8 (9.1) percentage points likelier if the consumer nearly always purchases products with the substitute's brand (water type) than if she virtually never does so.

By contrast, the results for flour are difficult to reconcile with the mixed logit IPA. Whether the substitute matches the out-of-stock product's brand, flour, or quantity is highly predictive of acceptance—even conditional on the frequency with which the consumer purchases the substitute's versions of these characteristics. This is especially true where flour type is concerned; acceptance is 12.4 percentage points likelier if the substitute shares the out-of-stock product's flour type than if it does not. Notice that this marginal effect greatly exceeds that associated with the fraction of trips where the purchased product shares the substitute's flour type; a consumer who almost always purchases the substitute's flour type is only 2.6 percentage points likelier to accept than a consumer who virtually never purchases the substitute's flour type.

Why is the flour type of the out-of-stock product so predictive of the substitute's acceptance or rejection? Recall from Section 1 that specific types of flour are suited to specific recipes—bread flour for bread, all-purpose flour for cupcakes, etc. Hence, if a consumer has a particular recipe in mind when she places her order, she will choose a flour of the corresponding type. She is therefore likely to prefer a substitute of the out-of-stock product's flour type over a substitute of a different flour type—even a flour type that she purchases more frequently—as only the former would enable her to bake the intended recipe (without modification).

In contrast to flour type, the marginal effect of the substitute's sharing the brand or quantity of the out-of-stock product is negative. At face value, this means that the substitute is likelier to be accepted if it differs from the out-of-stock product with respect to these characteristics than if it matches them. However, this counterintuitive result probably reflects the limitations of this reduced-form exercise, which—among other omissions—largely abstracts from the role of price.

Discussion.—These results provide suggestive evidence that consumers' purchases of bottled water are consistent with the mixed logit IPA, whereas their purchases of flour are not. The key difference between the categories is the amount of within-consumer preference variation. Regarding bottled water, individual consumers' preferences largely persist over time. By contrast, individual consumers' preferences for flour appear to vary considerably between trips, perhaps due to variation in intended recipes (for which specific flour types may be optimal).

However, these results also highlight the limitations of reduced-form analysis where the mixed logit IPA is concerned; some determinants of acceptance are difficult to capture without an explicit model of consumer preferences. For this reason, the next section adopts a structural approach to testing the mixed logit IPA.

6. Structural Evidence

In this section, I quantify the benefits of relaxing the mixed logit IPA. To do so, I estimate demand for bottled water and flour using mixed *probit*—which does not suffer from an IPA constraint—as well as mixed logit. Then I compare the models' goodness of fit. The key point of comparison is the models' ability to predict whether a stockout substitute will be accepted or rejected. Whereas mixed logit imposes conditional independence between a consumer's initial order and her decision about the substitute (given the realization of her random taste coefficients), mixed probit does not.

To enable mixed logit to compete with mixed probit on the best possible footing, I nonparametrically estimate the joint distribution of the random coefficients. This ensures that consumers' random taste coefficients provide the most accurate possible representation of their (time-invariant) tendencies to like or dislike substitutes' observable characteristics, thereby minimizing the influence of the mixed logit IPA.[36] My estimation method adapts the fixed grid approach from Fox, Kim, and Yang (2016) and Train (2008). In the case of mixed probit, I employ a novel grid search approach to permit (some) correlation in the error terms.

A. Model

For simplicity, the conceptual framework in Section 5 focused on curbside pickup. Here, I extend this framework to include in-store purchases and home delivery as well as curbside pickup. This provides more observations per consumer, facilitating the identification of the distribution of random taste coefficients.

Consider a consumer i who is shopping at time t. Irrespective of shopping channel (in-person, home delivery, or curbside pickup),[37] she faces a choice between J_t differentiated goods and an outside option of no purchase ("good 0"). She will choose whichever good $j \in \mathcal{J}_t \equiv \{0, 1, \ldots, J_t\}$ maximizes her conditional indirect utility u_{ijt}. As in Section 5B, utility is a consumer-specific function of product characteristics (x_j)

[36] The mixed logit IPA only imposes independence between the order and the accept/reject decision *conditional* on representative utility. Thus, misspecification of representative utility could lead to spurious failures of the mixed logit IPA.

[37] In principle, some goods with a small market share may be solely offered for in-store purchase (as opposed to home delivery or curbside pickup). However, in my empirical estimation, I drop less popular products (because discrete choice models struggle to accommodate alternatives with negligible choice shares). And, unpopular products aside, the online choice set should coincide with its in-store counterpart (e.g., prices should be identical).

and price (p_{jt}):

$$u_{ijt} = x_j\beta_i - \alpha_i p_{jt} + \varepsilon_{ijt}.$$

Unlike in Sections 5A and 5B, the distribution of the error term ε_{ijt} now depends on the model. It is i.i.d. Gumbel in mixed logit, and i.i.d. multivariate normal in mixed probit.

If the consumer has placed an order for curbside pickup, her preferred product j may go out of stock. In that event, the store will offer a substitute $s \in \mathcal{J} \setminus \{0, j\}$. The consumer will accept the substitution if and only if $u_{ist} \geqslant u_{i0t}$, where $u_{i0t} \equiv \varepsilon_{i0t}$ denotes the utility of the outside option.

Identification.—In what follows, I employ a nonparametric mixture estimator for both mixed logit and mixed probit. Do the data afford sufficient variation to support this estimation method? Regarding mixed logit, Fox et al. (2012) prove that the model is nonparametrically identified under fairly minimal data requirements (e.g., local variation in product characteristics). As for mixed probit, Iaria and Wang (2023) show that the model is *semi*-nonparametrically identified. That is, taking as given that the error terms are distributed i.i.d. multivariate normal, the distribution of random coefficients is nonparametrically identified.

B. Estimation Method

I estimate the joint distribution of random coefficients (β_i, α_i) nonparametrically. Following Fox, Kim, and Yang (2016), I approximate the distribution function using a "fixed grid" estimator. In this approach, a fixed grid of heterogeneous coefficients is selected *before* estimation. Then the probability weights on the (pre-specified) grid points are estimated. In what follows, I first derive the likelihood function for these weight parameters. (In so doing, I borrow from the exposition in Heiss, Hetzenecker, and Osterhaus [2022].) Then I explain the expectation-maximization (EM) algorithm employed to maximize the likelihood function, as well as the simulation required for the mixed probit model. To keep this subsection focused, a discussion of the tuning parameters (such as the number and location of the grid points) is relegated to Appendix E. I do the same with respect to the grid-search estimator for correlated errors in the mixed probit model.

The task is to estimate the joint distribution $F(\beta, \alpha)$ of random coefficients. I employ a finite-dimensional sieve approximation that divides the support of (β, α) into a grid of R fixed vectors:

$$\mathcal{B} = \begin{pmatrix} (\beta^1, \alpha^1) \\ \vdots \\ (\beta^R, \alpha^R) \end{pmatrix}$$

Having chosen the grid $\mathcal{B}$, I estimate the probability weights $\theta = (\theta^1, \ldots, \theta^R)$ on each of the coefficient vectors in $\mathcal{B}$. The weight θ^r on a coefficient vector $\beta^r \in \mathcal{B}$ depends on the extent to which it is representative of tastes across the population of consumers. To derive θ^r, focus first on an individual consumer i. Let $\text{choose}_{ijt} = 1$ if good j is her most-preferred product on trip t—that is to say, the ordered product (online) or the purchased product (in-store)—and let $\text{choose}_{ijt} = 0$ otherwise.[38]

Supposing that trip t is curbside pickup, consumer i's ordered good—say, j—may go out of stock before pickup. In that event, she will be offered a substitute good $s \neq j$. To notate stockout substitutions, let $\text{OOS}_{ijt} = 1$ if ordered good j goes out of stock on trip t and $\text{OOS}_{ijt} = 0$ otherwise.[39] And, conditional on ordered good j going out of stock, let $\text{accept}_{ist} = 1$ if the consumer i accepts good s as a substitute on trip t and $\text{accept}_{ist} = 0$ otherwise.

Due to the panel nature of the data, individual consumers are observed making repeated choices over time. Consequently, the likelihood criterion concerns the probability of observing the *entire sequence* of choices made by each consumer (Train 2009). Assuming that β^r represents the true tastes of consumer i, this is given by

$$P_i \mid \beta^r, \alpha^r \equiv \prod_{t \in \mathcal{T}_i} \prod_{j \in \mathcal{J}_t} \Bigg(\big(\Pr[\text{choose } j \mid x_t; \beta^r, \alpha^r]\big)^{\text{choose}_{ijt}} \\ \Bigg(\prod_{s \in \mathcal{J}_t \setminus \{j\}} \big(\Pr[\text{accept } s \mid \text{choose } j; x_t; \beta^r, \alpha^r]\big)^{\text{accept}_{ist}} \Bigg)^{\text{OOS}_{ijt}} \Bigg),$$

where $\mathcal{T}_i$ denotes the set of all her trips.

Of course, consumer i's true tastes are unknown to the researcher. To recover the unconditional probability of her observed sequence of choices, compute the weighted average of the conditional choice probabilities $(P_i \mid \beta^r, \alpha^r)$ associated with each taste vector $(\beta^r, \alpha^r) \in \mathcal{B}$:

$$P_i \equiv \sum_{r=1}^{R} \theta^r (P_i \mid \beta^r, \alpha^r).$$

In this equation, the probability weights θ^r measure the prevalence of tastes (β^r, α^r) across the population of consumers.

[38] In a slight abuse of notation, I now use t to index an individual consumer's trips, as opposed to time.

[39] Where in-store shopping and home delivery are concerned, $\text{OOS}_{ijt} = 0$ for all goods j.

Finally, compute the log-likelihood criterion by summing P_i over the population of consumers:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\log(P_i).$$

Computing the Conditional Choice Probabilities.—So far, I have abstracted away from the calculation of conditional choice probabilities. In the case of mixed logit, they take a closed form. The probability that i orders (purchases) good j while shopping online (in-store) is given by

$$\Pr[\text{order } j \mid x_t; \beta^r, \alpha^r] = \frac{\exp(x_j\beta^r - \alpha^r p_{jt})}{\sum_{j' \in \mathcal{J}} \exp(x_{j'}\beta^r - \alpha^r p_{j't})}.$$

If trip t is curbside pickup, the probability that she accepts a substitute $s \in \mathcal{J}_t \setminus \{j\}$ is

$$\begin{aligned}\Pr[\text{accept } s \mid \text{order } j; x_t; \beta^r, \alpha^r] &= \Pr[\text{accept } s \mid x_t; \beta^r, \alpha^r] \\ &= \frac{\exp(x_s\beta^r - \alpha^r p_{st})}{1 + \exp(x_s\beta^r - \alpha^r p_{st})}.\end{aligned}$$

The former equality follows from the mixed logit IPA, under which a consumer's initial order is uninformative of her accept/reject decision about the substitute (conditional on her time-invariant tastes).

Where mixed probit is concerned, the conditional choice probabilities lack closed forms and must be simulated. To improve the accuracy of the simulated probabilities, I do not draw the simulated error terms directly from a multivariate normal distribution. Rather, I draw the error terms from a scrambled "Sobol' sequence" (Sobol' 1967).[40] For a given number of draws, this quasi-Monte Carlo method should more closely approximate the underlying distribution than would pseudo-random draws from the corresponding multivariate normal distribution.[41]

Simulation proceeds as follows. I take Q quasi-Monte Carlo draws, indexed by $q \in \mathcal{Q} \equiv \{1, \ldots, Q\}$. For each draw q, I draw a vector of low-discrepancy multivariate normal errors $(\varepsilon_{ijt}^q)_{j \in \mathcal{J}_t}$ for each consumer i and trip t.[42] Then the probability of consumer i ordering (purchasing) good j while shopping online (in-store),

[40]Concerning a related simulation problem—namely, computing parametric mixed logit choice probabilities—recent work by Czajkowski and Budziński (2019) suggests that scrambled Sobol' sequences are more efficient than alternative simulation methods, such as scrambled Halton sequences and modified Latin hypercube sampling.

[41]To preserve the balance properties of this quadrature rule, it is necessary that the total number of random draws—that is, the product of (i) the number of orders and (ii) the number of simulations—be a power of two (Virtanen et al. 2020). Throughout, I choose the number of simulations (as well as the number of orders modeled) so that this condition is satisfied.

[42]Precisely speaking, these draws are not from a multivariate normal distribution as such. Rather, they are based on a low-discrepancy Sobol' approximation (as described above).

conditional on having tastes (β^r, α^r), is approximated by the fraction of draws q in which j maximizes her conditional indirect utility u_{ijt}. That is,

$$\hat{\Pr}[\text{order } j \mid x_t; \beta^r, \alpha^r] = \frac{1}{Q} \sum_{q \in Q} 1 \left[u_{ijt}^{rq} = \max_{j' \in \mathcal{J}_t} \{u_{ij't}^{rq}\} \right],$$

where

$$u_{ijt}^{rq} \equiv x_j \beta^r - \alpha^r p_{jt} + \varepsilon_{ijt}^q.$$

Where curbside pickup is concerned, the probability of accepting a substitute depends on the consumer's original order. To see why, suppose that consumer i orders good j on trip t, only for j to go out stock. Conditional on having true tastes β^r, the researcher knows that the error terms $(\varepsilon_{ijt})_{j \in \mathcal{J}_t}$ satisfy

$$u_{ijt}^r = \max_{j' \in \mathcal{J}_t} \{u_{ij't}^r\}.$$

In other words, the consumer's decision to order j is informative of the error terms for the *other* goods $j' \neq j$.

How does this association between order and substitution choices affect estimation? Supposing that consumer i has ordered good j, any draws q such that

$$u_{ijt}^{rq} < \max_{j' \in \mathcal{J}_t} \{u_{ij't}^{rq}\}$$

can be discarded. If β^r represents consumer i's true tastes, draws of the foregoing description would result in her placing a different order than the one observed in the data. Hence, I approximate the probability that i accepts a substitute $s \in \mathcal{J}_t \setminus \{j\}$ as the fraction of the remaining draws

$$\boldsymbol{Q}^\star(j) \equiv \{q \in \boldsymbol{Q} : u_{ijt}^{rq} = \max_{j' \in \mathcal{J}_t} \{u_{ist}^{rq}\}\}$$

for which u_{ist}^{rq} exceeds the (simulated) utility of the outside option, u_{i0t}^{rq}. That is,

$$\hat{\Pr}[\text{accept } s \mid \text{order } j; x_t; \beta^r, \alpha^r] = \frac{1}{|\boldsymbol{Q}^\star(j)|} \sum_{q \in \boldsymbol{Q}^\star(j)} 1[u_{ist}^{rq} > u_{i0t}^{rq}].$$

Because many draws q will be discarded, estimation requires a large total number of draws Q. To avoid unacceptably long run times, I employ the JAX Python library (Bradbury et al. 2018) to spread computation across multiple GPUs as well as to optimize the code.

Expectation-Maximization (EM) Algorithm.—The fixed grid estimator suffers from a curse of dimensionality rooted in the number of random coefficients. To be specific, I compute probability weights on 78,125 fixed grid points in the estimates that follow. This multiplicity of parameters poses a problem for gradient-based optimization. Inversion of the Hessian can fail, and optimization may become "stuck" in regions where the likelihood function is inadequately approximated by a quadratic (Train [2009]).

To surmount this computational difficulty, I employ an "expectation-maximization" (EM) algorithm. Rather than maximizing the likelihood function directly, an EM algorithm instead maximizes (conditional) expectations of the likelihood while holding various parameters constant by turns. See Train (2008), Section 6 for a detailed discussion of the EM algorithm used in this paper.[43]

C. Data Details

Whenever a consumer purchases something (whether in the store, through curbside pickup, or via home delivery), the data report the item's UPC and price. But what about the rest of the consumer's choice menu? Which alternatives did she pass over in favor of her preferred product, and what were their prices?

To reconstruct the consumer's choice menu, I first consult the chain's product catalog to see the UPCs of products in the relevant category. Then I match the resulting list against the UPCs of products sold at the relevant store according to the scanner data. Regarding availability, I assume that a product was in the consumer's choice menu if a *different* consumer purchased it on the same day, at the same store. Failing that, I check if the product was purchased on both the day before *and* the day after (not necessarily by the same consumer). If neither of these conditions is satisfied, I assume that the product was not in the consumer's choice set (either because the product was out of stock, or because the store did not carry it at all).

I employ a slightly different procedure with respect to products that were ordered for curbside pickup but later went out of stock. Observe that a product of this description was likely on the shelf at the time that the consumer placed her order.[44] That, in turn, suggests that the product was either available (i) the day of the attempted stockout substitution or (ii) the day before. Accordingly, I impute the out-of-stock product's price as being the average purchase price on the day of the substitution or, failing that, the average purchase price on the day before. If I do not observe any sales on either day, I impute the price as being the average

[43] In the case of mixed probit, a slight adjustment is necessary: probit kernels, not logit kernels, are computed for each agent.

[44] Unless a stockout was directly caused by an order for curbside pickup, there may be a delay before the store's website indicates that a given item is out of stock.

purchase price on the nearest date for which observations appear in the data (up to seven days before or after the stockout event).[45]

Due to computational constraints, I cannot model demand for all forty bottled-water products or all sixty-one flour products. Rather, I exclude slow-selling or unusual products within each category, leaving me with six bottled water products and ten flour products.[46] For the same reason, I do not perform estimation on all the available data. Instead, I focus on a random sample of households within each product category.[47]

Multi-Product and Multi-Unit Purchases.—In the data, consumers' purchases depart from standard discrete choice frameworks in two ways. First, consumers may purchase multiple *distinct products* on a single shopping trip. For instance, someone might purchase 24-packs of both Ice Mountain and Aquafina on one trip. As for the second departure from discrete choice, consumers may purchase multiple units of a *single product* on one shopping trip. For example, someone might purchase two 24-packs of Ice Mountain bottled water on one trip. Multiple purchases of this kind might be motivated by "stockpiling" to take advantage of discounts.

Within the product categories of bottled water and flour, purchases of more than one product on a single shopping trip are fairly uncommon. Among analysis households, three (seven) percent of shopping trips feature purchases of more than one bottled water (flour) product. I exclude all such transactions from my structural estimation. By comparison, purchases of multiple units of a *single product* are much more common. Roughly 25% (12%) of analysis households' purchases of bottled water (flour) involve multiple units.

Because standard discrete choice models (such as mixed logit and mixed probit) do not accommodate multi-unit purchases, the result may be biased predictions of consumers' choices. And, where mixed logit is concerned, these biased predictions could result in apparent violations of the model's IPA property that do *not* reflect within-consumer preference variation, but rather misspecification of the underlying choice

[45]The structural estimation here focuses on top-selling products, whose prices are comparatively easy to infer. By contrast, the reduced-form regression in Section 5B also includes low-volume products, which may sell infrequently at a given store. If the procedure defined in the main text fails to recover the price of a low-volume out-of-stock product, I instead compute the average purchase price for stores in the same (narrowly-defined) geographic area on the nearest date with observations in the data (once more, up to seven days before or after the stockout event). The assumption is that stores in the same geographic area will coordinate on discounts (which might be advertised through mass mailings or billboards). To group stores by location, I rely on the most granular geographic designation in the chain's internal system. At all events, the results in Section 5B are robust to the inclusion or exclusion of observations whose prices are imputed in this fashion.

[46]For bottled water, I estimate demand solely for the top six products. Together, these products command a 75% market share among "analysis households" (i.e., households that experience one or more curbside stockout substitutions). As for flour, I restrict attention to the top three brands (the private label, King Arthur, and Gold Medal) as well as the top two types of flour (all-purpose and bread). I further exclude products with less than 1.75% market share, along with the one organic flour with nontrivial sales. (To include that organic product, which represents 2% of analysis households' purchases, I would need to add another explanatory variable: an "organic" dummy.) This leaves me with ten products, which together represent 75% of purchases by analysis households.

[47]To ensure the balance properties of the Sobol' sequence, it is necessary that the product of the number of sampled purchases (here, 4096) and the number of simulated error draws (here, 16,384) be a property of two. For the number of purchases to be exactly 4096, I may drop some of the later purchases made by at most one sampled household.

problem. To avoid such an outcome, it is important to minimize the influence of multi-unit purchases on demand estimation. I do so by estimating demand for a subset of households who are especially unlikely to make multi-unit purchases. In particular, I identify households with (i) zero purchases involving multiple units of a single product and (i) ten or more purchases in total. For an additional discussion of multi-unit purchases (including a summary of results when households with multi-unit purchases are *not* dropped), see Appendix F.

D. Results: Mixed Probit versus Mixed Logit

The task is to compare mixed logit's goodness of fit with that of mixed probit, especially in regard to the alternate-choice data on stockout substitutions. I proceed as follows. First, I draw a random sample from the set of households with ten or more purchases (and zero multi-unit purchases) in the data. (Recall from the preceding subsection that I cannot include the universe of households due to memory constraints.) Then I estimate demand using both mixed logit and mixed probit. Finally, I compare the two models with respect to the predicted probabilities assigned to the choices of the *same* random subset of households that I earlier used in estimation.

In addition to the "within-sample" comparison that I have just described, it is also instructive to perform an "out-of-sample" comparison. How accurately do the models forecast the choices of a "holdout" sample of households, whose data were *not* used in estimation? I will briefly discuss the motivations for this alternative procedure—as well as its results—later in this subsection.

Within-Sample Goodness of Fit.—Table 5 compares the within-sample fit of mixed logit and mixed probit. Panel A pertains to in-store purchases, home delivery purchases, and curbside pickup orders; while Panel B attends to stockout substitutions in curbside pickup (i.e., the alternate-choice data). For each portion of the data, I assess model fit based on the average predicted probability assigned to consumers' observed choices.[48] In computing this metric, I leverage the panel structure of the data to derive predicted probabilities that reflect the posterior probabilities of individual consumers' observed choices.[49]

The relative performance of mixed logit and mixed probit varies by data type. Focus first on in-store

[48] In Appendix G, I report results for an alternative measure of fit: the fraction of observations in which consumers' observed choices are assigned the highest predicted probability of any alternative (sometimes termed the "hit rate").

[49] When a consumer is observed making multiple decisions, it may become apparent that she likes or dislikes certain kinds of products. For instance, if a frequent flour buyer always opts for bread flour (as opposed to all-purpose), she probably likes bread flours more than the "average" consumer does. This intuition can be harnessed to situate the taste coefficients (β_i, α_i) of an individual consumer i within the population distribution of random coefficients (Train 2009). To do so in the context of a fixed-grid model, I follow the steps prescribed by Train (2008).

TABLE 5 – GOODNESS OF FIT: MIXED LOGIT VERSUS MIXED PROBIT

Statistic	Bottled water: Mixed logit	Bottled water: Mixed probit	Flour: Mixed logit	Flour: Mixed probit
	Panel A. In-store purchases and online orders			
No. of households	220	220	294	294
No. of purchases[a]	4096	4096	4096	4096
Purchased product's predicted probability	**0.272**	**0.273**	**0.195**	**0.197**
	(0.310)	(0.322)	(0.278)	(0.285)
No. of available alternatives[b]	4.56	4.56	7.58	7.58
	(1.10)	(1.10)	(1.62)	(1.62)
	Panel B. Stockout substitutions			
No. of (attempted) stockout substitutions	316	316	324	324
. . . of which accepted	294	294	300	300
True decision's predicted probability	**0.955**	**0.964**	**0.936**	**0.946**
	(0.152)	(0.148)	(0.201)	(0.208)
	Panel C. Empirical specification			
No. of random coefficients	7	7	7	7
No. of grid points	78,125	78,125	78,125	78,125
No. of simulated error draws[a]		16,384		16,384

Notes: This table compares the within-sample fit of mixed probit and mixed logit models for the product categories of bottled water and flour (see Sections 6B and 6D for details). Where relevant, standard deviations appear in parentheses.

[a] The number of purchases and the number of draws are jointly chosen to maintain the balance properties of the Sobol' sequence.

[b] Excluding the "outside option" of no purchase.

purchases and online orders (Panel A). The models' fit is comparable here, with consumers' chosen bottled water (flour) products being assigned a 0.1 (0.2) percentage point greater probability by mixed probit than by mixed logit. By contrast, there is a perceptible difference in model fit where stockout substitutions are concerned (Panel B). On average, consumers' observed decisions to accept or reject bottled water (flour) substitutes are assigned a 0.9 (1.0) percentage point greater probability by mixed probit than by mixed logit.

Notice that the disparity in model fit is only slightly larger for flour than bottled water. On the face of it, this suggests that the mixed logit IPA does *not* drive the difference in model fit concerning stockout substitutions. Recall that the descriptive evidence in Section 5B suggests that consumers' purchases of bottled water are consistent with the mixed logit IPA, whereas their purchases of flour are not. So, if the mixed logit IPA drives the disparity in fit between mixed logit and mixed probit, one would expect this disparity to be markedly larger for flour than for bottled water.

Overfitting may have biased this model selection exercise, however. Due to the nonparametric estimation approach, the mixed probit and mixed logit models feature nearly eighty thousand parameters each. This complexity enables the models to closely match random noise in the data as well as underlying economic factors.[50] Hence, to the extent that within-sample differences in fit reflect the models' ability to reproduce random noise (as opposed to systematic determinants of demand), the results of a within-sample comparison may be biased. I adopt an out-of-sample approach to address this potential source of bias.

Out-of-Sample Validation.—In contrast to within-sample methods of model selection, out-of-sample methods assess models' ability to forecast the choices of a "holdout sample" of households whose data were *not* used in estimation. The intuition is as follows. To the extent that an estimated model captures statistical noise, as opposed to systematic determinants of demand, it will (incorrectly) project this random noise onto the households in the holdout sample. As a result, the model's accuracy in predicting the choices of the holdout sample depends solely on the extent to which the model has captured generalizable (and economically meaningful) determinants of consumers' choices.[51]

Out-of-sample validation proceeds as follows. First, I randomly draw a "holdout sample" of households whose data were not used to estimate the models above. Second, I compute the posterior distributions of these holdout households' random coefficients based on the empirical CDF of random coefficients from the estimation results above. In computing these posterior distributions, I exclude the holdout households' decisions to accept or reject stockout substitutes, leaving only their original orders (as well as their in-store purchases and their orders for home delivery). This ensures that, so far as stockout substitutions are concerned, the validation exercise is predictive in nature. Finally, I compute predicted probabilities of acceptance or rejection based on these household-level posterior probability distributions.

Table 6 reports the results of out-of-sample validation. Concerning in-store purchases and online orders, the models' relative performance remains unchanged from the within-sample comparison. Consumers' choices of bottled water (flour) are assigned a 0.1 (0.2) percentage point higher probability by mixed probit than by mixed logit. By contrast, where stockout substitutions are concerned, the disparity in model fit is

[50]With a sufficient number of parameters, models can reproduce idiosyncrasies in consumer behavior that should be attributed to the error term. To see the intuition, picture a consumer who is placing a curbside order for bottled water. She intends to order a 24-pack of Ice Mountain bottled water, but mistakenly clicks on a 6-pack of Aquafina instead (and does not spot her error). Although this mistake should be attributed to the error term, a sufficiently complicated model might nevertheless assign our consumer's mistake a fairly high predicted probability.

[51]For an accessible introduction to out-of-sample validation, see Parady, Ory, and Walker (2021); while Zhang and Yang (2015) provide a more technical discussion. As far as applications are concerned, this approach has been employed in variety of economic subdisciplines, including health economics (Deb and Trivedi 2002) and agricultural economics (Haener, Boxall, and Adamowicz 2001), as well as industrial organization (Bajari and Benkard 2005).

TABLE 6 – OUT-OF-SAMPLE VALIDATION: MIXED LOGIT VERSUS MIXED PROBIT

Statistic	Bottled water Mixed logit	Bottled water Mixed probit	Flour Mixed logit	Flour Mixed probit
	Panel A. In-store purchases and online orders			
No. of households	204	204	285	285
No. of purchases[a]	4096	4096	4096	4096
Purchased product's predicted probability	**0.256**	**0.257**	**0.192**	**0.194**
	(0.306)	(0.315)	(0.277)	(0.284)
No. of available alternatives[b]	4.63	4.63	7.48	7.48
	(1.05)	(1.05)	(1.64)	(1.64)
	Panel B. Stockout substitutions			
No. of (attempted) stockout substitutions	293	293	304	304
. . . of which accepted	276	276	283	283
True decision's predicted probability	**0.928**	**0.947**	**0.911**	**0.935**
	(0.213)	(0.215)	(0.246)	(0.244)
	Panel C. Empirical specification			
No. of random coefficients	7	7	7	7
No. of grid points	78,125	78,125	78,125	78,125
No. of simulated error draws[a]		16,384[b]		16,384[b]

Notes: This table compares the fit of mixed probit and mixed logit models on a holdout sample (see Sections 6B and 6D for details). Where relevant, standard deviations appear in parentheses.

[a] The number of purchases and the number of draws are jointly chosen to maintain the balance properties of the Sobol' sequence.

[b] Excluding the "outside option" of no purchase.

even larger out-of-sample than within-sample. Consumers' decisions to accept or reject bottled water (flour) substitutes are assigned a 1.9 (2.4) percentage point larger probability by mixed probit than by mixed logit. In other words, the disparity in model fit is perceptibly larger for flour than for bottled water. This is in keeping with the descriptive evidence in Section 5B: namely, that consumers' purchases of bottled water are consistent with the mixed logit IPA, whereas their purchases of flour are not.

Notice that the out-of-sample validation magnifies qualitative patterns that were already present within-sample. In particular, the following qualitative patterns are observed both within- and out-of-sample: (i) mixed probit replicates consumers' choices about stockout substitutes more accurately than mixed logit does; and (ii) this disparity in model fit is larger for flour than for bottled water. However, both of these patterns are more pronounced out-of-sample than within-sample. A possible explanation is that the within-sample model comparison is biased by over-fitting. Within-sample, differences in fit between mixed logit and mixed probit reflect both (a) the models' abilities to capture systematic determinants of demand and (b) their capacities to reproduce statistical noise. Out-of-sample, by contrast, differences in model fit solely reflect (a).

7. Conclusion

Data on consumers' preferences among unpurchased goods can help identify demand elasticities. However, workhorse demand systems sometimes fail to replicate important substitution patterns in these "alternate choice" data. This paper highlights one explanation for this shortcoming: the *independence of preferred alternatives* (IPA) properties of logit models. Conditional logit imposes independence between a consumer's purchase and her preferences among unpurchased goods, while mixed logit imposes conditional independence between the same (given the realizations of the random coefficients). To ascertain the extent to which these properties impact demand estimates, I employ novel revealed-preference data on curbside pickup. The data concern consumers' willingness to accept store-selected substitutes when their preferred products go out of stock.

Focusing on the product categories of bottled water and flour, I present descriptive evidence that consumer behavior is sometimes inconsistent with the IPA properties of conditional and mixed logit. Regarding the former, a conditional likelihood ratio test suggests that consumer' purchases of both bottled water and flour are inconsistent with the IPA property of conditional logit; contrary to the property, consumers' original order choices are correlated with their decisions to accept or reject substitutes. By way of an explanation, I present descriptive evidence that consumers prefer substitutes whose observable characteristics resemble those of the out-of-stock product. As for mixed logit, reduced-form analysis suggests that the behavior of bottled water buyers is consistent with the model's IPA property, whereas the behavior of flour buyers is not. These contrasting results appeared to reflect differences between the two product categories in the amount of *within-consumer* preference variation.

To quantify the bias resulting from the IPA property of mixed logit, I estimate demand for bottled water and flour using two models: mixed logit and mixed *probit* (a model without an IPA property). Comparing the models' goodness of fit (both within- and out-of-sample), I find that mixed probit predicts the acceptance or rejection of stockout substitutes more accurately than mixed logit does. Moreover, the disparity in fit is more pronounced for flour than for bottled water. This is in keeping with the reduced-form evidence that consumers' purchases of bottled water are consistent with the IPA property of mixed logit, whereas their purchases of flour are not.

Besides mixed probit, there are many other models that relax the IPA property of mixed logit, such as semi-nonparametric hedonic models (see Bajari and Benkard [2005]) or a pure-characteristics demand model (see Berry and Pakes [2007]). A particularly promising candidate is the "random-coefficients nested logit" model estimated in Brenkers and Verboven (2006) and Grigolon and Verboven (2014). Because its error terms are not Gumbel, but rather Generalized Extreme Value (GEV), the model is unlikely to suffer from an IPA constraint. Furthermore, existing empirical frameworks for alternate-choice data could be adapted to use this more general model in place of mixed logit. In principle, many frameworks are amenable to this adaption, including those proposed by Berry, Levinsohn, and Pakes (2004); Train and Winston (2007); Bachmann et al.; and Grieco et al. (2023). However, the feasibility of incorporating random-coefficients nested logit in these frameworks depends on the conditional choice probabilities of consumers' alternate choices (e.g., second choices or accept/reject decisions in stockout data). Do these probabilities take a parsimonious and predictable form as the number of alternatives and "nests" grows? I leave to future work the question of whether this is true and, if so, the extent to which the resulting model can match the substitution patterns in alternate choice data.

References

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag A. Pathak. "The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match". *American Economic Review* 107, no. 12 (2017): 3635–3689.

Allcott, Hunt. "The Welfare Effects of Misperceived Product Costs: Data and Calibrations from the Automobile Market". *American Economic Journal: Economic Policy* 5, no. 3 (2013): 30–66.

Anupindi, Ravi, Maqbool Dada, and Sachin Gupta. "Estimation of Consumer Demand with Stock-Out Based Substitution: An Application to Vending Machine Products". *Marketing Science* 17, no. 4 (1998): 406–423.

Bachmann, Rüdiger, et al. "Firms and Collective Reputation: A Study of the Volkswagen Emissions Scandal". *Journal of the European Economic Association* 21, no. 2 (2023): 484–525.

Bajari, Patrick, and C. Lanier Benkard. "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach". *Journal of Political Economy* 113, no. 6 (2005): 1239–1276.

Beggs, Steven, Scott Cardell, and Jerry Hausman. “Assessing the Potential Demand for Electric Cars”. *Journal of Econometrics* 17, no. 1 (1981): 1–19.

Berry, Steven, James Levinsohn, and Ariel Pakes. “Automobile Prices in Market Equilibrium”. *Econometrica* 63, no. 4 (1995): 841–890.

– . “Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market”. *Journal of Political Economy* 112, no. 1 (2004): 68–105.

Berry, Steven, and Ariel Pakes. “The Pure Characteristics Demand Model”. *International Economic Review* 48, no. 4 (2007): 1193–1225.

Berry, Steven T., and Philip A. Haile. “Foundations of Demand Estimation”. In *Handbook of Industrial Organization*, 4:1–62. 1. Elsevier, 2021.

Bradbury, James, et al. *JAX: Composable Transformations of Python + NumPy Programs*, 2018. `github.com/google/jax`.

Brenkers, Randy, and Frank Verboven. “Liberalizing a Distribution System: The European Car Market”. *Journal of the European Economic Association* 4, no. 1 (2006): 216–251.

Brownstone, David, and Kenneth A. Small. “Valuing Time and Reliability: Assessing the Evidence from Road Pricing Demonstrations”. *Transportation Research Part A: Policy and Practice* 39, no. 4 (2005): 279–293.

Bruno, Hernán A., and Naufel J. Vilcassim. “Research Note—Structural Demand Estimation with Varying Product Availability”. *Marketing Science* 27, no. 6 (2008): 1126–1131.

Carlsson, Fredrik, and Peter Martinsson. “Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments?: Application to the Valuation of the Environment”. *Journal of Environmental Economics and Management* 41, no. 2 (2001): 179–192.

Conlon, Chris, Julie Mortimer, and Paul Sarkis. “Estimating Preferences and Substitution Patterns from Second Choice Data Alone”. Preliminary and incomplete (2023).

Conlon, Christopher, and Jeff Gortmaker. “Incorporating Micro Data into Differentiated Products Demand Estimation with PyBLP”. Working paper (2023).

Conlon, Christopher, and Julie Holland Mortimer. “Empirical Properties of Diversion Ratios”. *The RAND Journal of Economics* 52, no. 4 (2021): 693–726.

Conlon, Christopher T., and Julie Holland Mortimer. "Demand Estimation under Incomplete Product Availability". *American Economic Journal: Microeconomics* 5, no. 4 (2013): 1–30.

– . "Effects of Product Availability: Experimental Evidence". National Bureau of Economic Research Working Paper 16506 (2010).

– . "Efficiency and Foreclosure Effects of Vertical Rebates: Empirical Evidence". *Journal of Political Economy* 129, no. 12 (Dec. 2021): 3357–3404.

Czajkowski, Mikołaj, and Wiktor Budziński. "Simulation Error in Maximum Likelihood Estimation of Discrete Choice Models". *Journal of Choice Modelling* 31 (2019): 73–85.

Deb, Partha, and Pravin K. Trivedi. "The Structure of Demand for Health Care: Latent Class Versus Two-Part Models". *Journal of health economics* 21, no. 4 (2002): 601–625.

Farronato, Chiara, and Andrey Fradkin. "The Welfare Effects of Peer Entry: The Case of Airbnb and the Accommodation Industry". *American Economic Review* 112, no. 6 (2022): 1782–1817.

Fox, Jeremy T., Kyoo il Kim, and Chenyu Yang. "A Simple Nonparametric Approach to Estimating the Distribution of Random Coefficients in Structural Models". *Journal of Econometrics* 195, no. 2 (2016): 236–254.

Fox, Jeremy T., et al. "The Random Coefficients Logit Model Is Identified". *Journal of Econometrics* 166, no. 2 (2012): 204–212.

Grieco, Paul L.E., et al. "Conformant and Efficient Estimation of Discrete Choice Demand Models". Working Paper (2023). Visited on 05/09/2024. `https://paulgrieco.github.io/files/preprint/like-blp.pdf`.

Grieco, Paul LE, Charles Murry, and Ali Yurukoglu. "The Evolution of Market Power in the US Automobile Industry". *The Quarterly Journal of Economics* (2023).

Grigolon, Laura, and Frank Verboven. "Nested Logit or Random Coefficients Logit? A Comparison of Alternative Discrete Choice Models of Product Differentiation". *Review of Economics and Statistics* 96, no. 5 (2014): 916–935.

Haener, M. K., P. C. Boxall, and W. L. Adamowicz. "Modeling Recreation Site Choice: Do Hypothetical Choices Reflect Actual Behavior?" *American Journal of Agricultural Economics* 83, no. 3 (Aug. 2001): 629–642.

Hausman, Jerry A., and Paul A. Ruud. “Specifying and Testing Econometric Models for Rank-Ordered Data”. *Journal of Econometrics* 34, no. 1 (1987): 83–104.

Heiss, Florian, Stephan Hetzenecker, and Maximilian Osterhaus. “Nonparametric Estimation of the Random Coefficients Model: An Elastic Net Approach”. *Journal of Econometrics* 229, no. 2 (2022): 299–321.

Iaria, Alessandro, and Ao Wang. “Real Analytic Discrete Choice Models of Demand: Theory and Implications”. Working Paper (Apr. 2023). `https://ssrn.com/abstract=4094866`.

Jovanovic, B. D., and P. S. Levy. “A Look at the Rule of Three”. *The American Statistician* 51, no. 2 (May 1997): 137–139.

Lusk, Jayson L., and Ted C. Schroeder. “Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks”. *American Journal of Agricultural Economics* 86, no. 2 (May 2004): 467–482.

Montag, Felix. “Mergers, Foreign Competition, and Jobs: Evidence from the US Appliance Industry”. Working paper (2023).

Musalem, Andrés, et al. “Structural Estimation of the Effect of Out-of-Stocks”. *Management Science* 56, no. 7 (2010): 1180–1197.

Paetz, Friederike, and Winfried J. Steiner. “Utility Independence versus IIA Property in Independent Probit Models”. *Journal of Choice Modelling* 26 (2018): 41–47.

Parady, Giancarlos, David Ory, and Joan Walker. “The Overreliance on Statistical Goodness-of-Fit and Under-Reliance on Model Validation in Discrete Choice Models: A Review of Validation Practices in the Transportation Academic Literature”. *Journal of Choice Modelling* 38 (2021): 100257.

Quaife, Matthew, et al. “How Well Do Discrete Choice Experiments Predict Health Choices? A Systematic Review and Meta-Analysis of External Validity”. *The European Journal of Health Economics* 19, no. 8 (Nov. 2018): 1053–1066.

Sobol’, Il’ya Meerovich. “On the Distribution of Points in a Cube and the Approximate Evaluation of Integrals”. *Zhurnal Vychislitel’noi Matematiki i Matematicheskoi Fiziki* 7, no. 4 (1967): 784–802.

Train, Kenneth E. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2009.

– . “EM Algorithms for Nonparametric Estimation of Mixing Distributions”. *Journal of Choice Modelling* 1, no. 1 (2008): 40–69.

Train, Kenneth E., and Clifford Winston. “Vehicle Choice Behavior and the Declining Market Share of Us Automakers”. *International Economic Review* 48, no. 4 (Nov. 2007): 1469–1496.

Tuyl, Frank, Richard Gerlach, and Kerrie Mengersen. “The Rule of Three, its Variants and Extensions”. *International Statistical Review* 77, no. 2 (Aug. 2009): 266–275.

U.S. Food & Drug Administration. *Bottled Water Everywhere: Keeping it Safe*, Apr. 2022. `https://www.fda.gov/consumers/consumer-updates/bottled-water-everywhere-keeping-it-safe`.

Virtanen, Pauli, et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. *Nature Methods* 17, no. 3 (2020): 261–272.

Xing, Jianwei, Benjamin Leard, and Shanjun Li. “What Does an Electric Vehicle Replace?” *Journal of Environmental Economics and Management* 107 (2021): 102432.

Zhang, Yongli, and Yuhong Yang. “Cross-Validation for Selecting a Model Selection Procedure”. *Journal of Econometrics* 187, no. 1 (2015): 95–112.

Supplementary Appendix

A. Proof of Lemma 1

Denote the representative utility of good $j \in \{A, B\}$ by $v_j \equiv x_j\beta - \alpha p_{jt}$ and, without loss of generality, normalize $v_B = 0$.[52] Then

$$\begin{aligned} P_A &\equiv \Pr\left[u_{iAt} > u_{iBt} \,\middle|\, \max\{u_{iAt}, u_{iBt}\} < K\right] \\ &= \Pr\left[v_A + \varepsilon_{iA} > \varepsilon_{iB} \,\middle|\, \max\{v_A + \varepsilon_{iA}, \varepsilon_{iB}\} < K\right] \\ &= \mathrm{E}_{\varepsilon_{iA}}\left[\Pr\left[v_A + \varepsilon_{iA} > \varepsilon_{iB} \,\middle|\, \max\{v_A + \varepsilon_{iA}, \varepsilon_{iB}\} < K; \varepsilon_{iA}\right] \,\middle|\, v_A + \varepsilon_{iA} < K\right]. \end{aligned} \tag{3}$$

where the last equality follows from the law of iterated expectations.

Consider the inner component of (3), namely, the conditional probability

$$\begin{aligned} P_A \mid \varepsilon_{iA} &\equiv \Pr\left[v_A + \varepsilon_{iA} > \varepsilon_{iB} \,\middle|\, \max\{v_A + \varepsilon_{iA}, \varepsilon_{iB}\} < K; \varepsilon_{iA}\right] \\ &= \Pr\left[\varepsilon_{iB} < v_A + \varepsilon_{iA} \,\middle|\, \max\{v_A + \varepsilon_{iA}, \varepsilon_{iB}\} < K; \varepsilon_{iA}\right]. \end{aligned} \tag{4}$$

Because $\max\{v_A + \varepsilon_{iA}, \varepsilon_{iB}\} < K$, the random variable ε_{iB} possesses the support $(-\infty, K)$. Eq. (4) can thus be expressed as the fraction

$$P_A \mid \varepsilon_{iA} = \frac{F_\varepsilon(v_A + \varepsilon_{iA})}{F_\varepsilon(K)}, \tag{5}$$

where $F_\varepsilon(\varepsilon') \equiv \exp\left(-e^{-\varepsilon'}\right)$ denotes the cumulative distribution function (CDF) of the Gumbel distribution.

---

[52] To see why this assumption is without loss of generality, decompose both goods' utilities into their respective representative utility and error terms;

$$\Pr\left[u_{iA} > u_{iB} \,\middle|\, \max\{u_{iA}, u_{iB}\} < K\right] = \Pr\left[v_A + \varepsilon_{iA} > v_B + \varepsilon_{iB} \,\middle|\, \max\{v_A + \varepsilon_{iA}, v_B + \varepsilon_{iB}\} < K\right].$$

Then subtract v_B from each quantity on the right-hand side to obtain

$$\begin{aligned} \Pr\left[u_{iA} > u_{iB} \,\middle|\, \max\{u_{iA}, u_{iB}\} < K\right] &= \Pr\left[v_A + \varepsilon_{iA} - v_{iB} > \varepsilon_{iB} \,\middle|\, \max\{v_A + \varepsilon_{iA} - v_{iB}, \varepsilon_{iB}\} < K - v_{iB}\right] \\ &\equiv \Pr\left[v'_A + \varepsilon_{iA} > \varepsilon_{iB} \,\middle|\, \max\{v'_A + \varepsilon_{iA}, \varepsilon_{iB}\} < K'\right], \end{aligned}$$

where $v'_A \equiv v_A - v_B$ and $K' \equiv K - v_B$.

Substituting (5) into (3) yields

$$
\begin{aligned}
P_A &= \mathrm{E}_{\varepsilon_{iA}}\left[\frac{F_\varepsilon(v_A+\varepsilon_{iA})}{F_\varepsilon(K)} \,\middle|\, v_A+\varepsilon_{iA}<K\right] \\
&= \frac{\mathrm{E}_{\varepsilon_{iA}}\left[F_\varepsilon(v_A+\varepsilon_{iA}) \,\middle|\, v_A+\varepsilon_{iA}<K\right]}{F_\varepsilon(K)}.
\end{aligned} \tag{6}
$$

Now employ the definition of expectation to write (6) as an integral. Notice that $v_A+\varepsilon_{iA}<K$ implies $\varepsilon_{iA}\in(-\infty, K-v_A)$, so the probability density function (PDF) of ε_{iA} is given by $f_\varepsilon(\varepsilon'_{iA})\big/F_\varepsilon(K-v_A)$. (Here, $f_\varepsilon(\varepsilon')\equiv\exp\left(-e^{-\varepsilon'}-\varepsilon'\right)$ denotes the PDF of the Gumbel distribution.) As a result,

$$
\begin{aligned}
P_A &= \frac{1}{F_\varepsilon(K)}\int_{\varepsilon'_{iA}=-\infty}^{K-v_A} F_\varepsilon(v_A+\varepsilon'_{iA})\frac{f_\varepsilon(\varepsilon'_{iA})d(\varepsilon'_{iA})}{F_\varepsilon(K-v_A)} \\
&= \frac{1}{\exp\left(-e^{-K}\right)}\int_{\varepsilon'_{iA}=-\infty}^{K-v_A}\exp\left(-e^{-(v_A+\varepsilon'_{iA})}\right)\frac{\exp\left(-e^{-\varepsilon'_{iA}}-\varepsilon'_{iA}\right)d\varepsilon'_{iA}}{\exp\left(-e^{-(K-v_A)}\right)} \\
&= \exp\left(e^{-K}\right)\exp\left(e^{-(K-v_A)}\right)\int_{\varepsilon'_{iA}=-\infty}^{K-v_A}\exp\left(-e^{-(v_A+\varepsilon'_{iA})}\right)\exp\left(-e^{-\varepsilon'_{iA}}-\varepsilon'_{iA}\right)d\varepsilon'_{iA} \\
&= \exp\left(e^{-K}+e^{-(K-v_A)}\right)\int_{\varepsilon'_{iA}=-\infty}^{K-v_A}\exp\left(-e^{-\varepsilon'_{iA}}e^{-b}\right)\exp\left(-e^{-\varepsilon'_{iA}}-\varepsilon'_{iA}\right)d\varepsilon'_{iA} \\
&= \exp\left(\left(e^{-v_A}+1\right)e^{-(K-v_A)}\right)\int_{\varepsilon'_{iA}=-\infty}^{K-v_A}\left(\exp\left(-e^{-\varepsilon'_{iA}}\right)\right)^{\exp(-v_A)}\exp\left(-e^{-\varepsilon'_{iA}}-\varepsilon'_{iA}\right)d\varepsilon'_{iA}
\end{aligned}
$$

Setting $u\equiv\exp\left(-e^{-\varepsilon'_{iA}}\right)$ and $du\equiv\exp\left(-e^{-\varepsilon'_{iA}}-\varepsilon'_{iA}\right)d\varepsilon'_{iA}$ yields

$$
\begin{aligned}
P_A &= \exp\left(\left(e^{-v_A}+1\right)e^{-(K-v_A)}\right)\int_{u=0}^{\exp(-\exp(-(K-v_A)))}u^{\exp(-v_A)}du \\
&= \exp\left(\left(e^{-v_A}+1\right)e^{-(K-v_A)}\right)\left[\frac{u^{\exp(-v_A)+1}}{e^{-v_A}+1}\right]_{u=0}^{\exp(-\exp(-(K-v_A)))} \\
&= \exp\left(\left(e^{-v_A}+1\right)e^{-(K-v_A)}\right)\frac{\left(\exp\left(-e^{-(K-v_A)}\right)\right)^{\exp(-v_A)+1}}{e^{-v_A}+1} \\
&= \exp\left(\left(e^{-v_A}+1\right)e^{-(K-v_A)}\right)\frac{\exp\left(-\left(e^{-v_A}+1\right)e^{-(K-v_A)}\right)}{e^{-v_A}+1} \\
&= \frac{1}{e^{-v_A}+1} \\
&= \frac{e^{v_A}}{1+e^{v_A}}.
\end{aligned}
$$

■

B. Comparison of Theorem 1 with Prior Theoretical Results

Beggs, Cardell, and Hausman (1981) derive a result that closely resembles Theorem 1. However, their result applies to different types of alternate-choice data. Whereas Theorem 1 pertain to data on consumers' *pairwise* preferences among unpurchased goods, Cardell and Hausman's result applies to *second-choice data* (as well as more comprehensive rankings of the choice set).

Beggs, Cardell, and Hausman's result is as follows. Letting j and j' be any two goods in $\mathcal{J}$, consider the *joint* probability that a consumer both (i) purchases good A and (ii) lists good B as her second-most-preferred good. Cardell and Hausman show that this joint probability equals the product of the *unconditional* probabilities of observing (i) and (ii). Formally,

$$\begin{aligned}\Pr[u_{iAt} = \max\nolimits_{j\in\mathcal{J}} u_{ijt} \text{ and } u_{iBt} = \max\nolimits_{j\in\mathcal{J}\setminus\{A\}} u_{ijt}] = \Pr[u_{iAt} = \max\nolimits_{j\in\mathcal{J}} u_{ijt}]& \\ \cdot \Pr[u_{iBt} = \max\nolimits_{j\in\mathcal{J}\setminus\{A\}} u_{ijt}]&\end{aligned}$$

As to more comprehensive rankings of consumers' preferences, let $\mathcal{S} \subseteq \mathcal{J}$ be any subset of the goods on offer. Then the probability of observing a given ranking of the goods in $\mathcal{S}$ can be written as the product of $|\mathcal{S}| - 1$ logit formulas.[53]

These results indicate that conditional logit restricts consumers' second choices—as well as more comprehensive rankings of the choice set—in a manner that resembles Theorem 1. (Whether Beggs, Cardell, and Hausman's findings imply Theorem 1 is not immediately clear.)

C. Monte Carlo Tests of Theorem 1

In this appendix, I perform Monte Carlo simulations to verify Theorem 1.

Consider a market with J goods, indexed by $j \in \mathcal{J} \equiv \{1, \ldots, J\}$.[54] Utility is specified as

$$u_{ij} = x_j\beta - \alpha p_j + \varepsilon_{ij} \equiv v_j + \varepsilon_{ij},$$

[53]Formally, let $r \equiv (r_1, r_2, \ldots, r_S)$ be any ordinal ranking of the goods in $\mathcal{S}$ such that $u_{ir_1t} > u_{ir_2t} > \cdots > u_{ir_St}$. Then

$$\begin{aligned}\Pr[u_{ir_1t} > u_{ir_2t} > \cdots > u_{ir_St}] = \Pr[u_{ir_1t} = \max\nolimits_{j\in\mathcal{S}} u_{ijt}] \cdot \Pr[u_{ir_2t} = \max\nolimits_{j\in\mathcal{S}\setminus\{r_1\}} u_{ijt}]& \\ \cdot \Pr[u_{ir_3t} = \max\nolimits_{j\in\mathcal{S}\setminus\{r_1,r_2\}} u_{ijt}] \cdots \Pr[u_{ir_{S-1}t} > u_{ir_St}].&\end{aligned}$$

(This notation for preference rankings is borrowed from Hausman and Ruud [1987].)

[54]For simplicity, I abstract from the inside/outside good distinction.

where ε_{ij} is distributed i.i.d. Gumbel. (For simplicity, I abstract from the panel dimension of the data as well as within-product price variation over time.) The task is to ascertain whether

$$\Pr\left[u_{iB} > u_{iC} \,\middle|\, u_{iA} = \max_{j \in \mathcal{J}} u_{ij}\right] = \Pr[u_{iB} > u_{iC}] \tag{7}$$

To do so, I compare (i) the conditional probability of preferring B over C—given A is the most-preferred good—with (ii) the unconditional probability of the same. In computing (i), I do not directly impose the mixed logit IPA (i.e., Theorem 1). Rather, I randomly draw errors from the Gumbel distribution. Then I discard any draws for which A is not the most-preferred good. Finally, I compute the fraction of the remaining draws in which B is preferred to C. This comparison is repeated for S different random draws of the goods' representative utilities.

Each simulation $s \in \mathcal{S} \equiv \{1, \dots S\}$ proceeds as follows. I begin by randomly drawing the representative utility v_{js} of each good $j \in \mathcal{J}$. In so doing, I treat the goods' representative utilities as (mutually independent) random uniform variables with support $[-4.5, 3.5]$.[55] With the representative utility draws in hand, I proceed to compute the probability that B is preferred to C—both unconditionally, and conditional on A being the most-preferred good. The unconditional probability is given by the familiar logit formula:

$$\Pr[u_{iBr} > u_{iCr} \mid (v_{js})_{j \in \mathcal{J}}] = \frac{\exp(v_{Bs})}{\exp(v_{Bs}) + \exp(v_{Cs})}. \tag{8}$$

As for the conditional probability of preferring B to C (given A is the most-preferred good), I simulate it by randomly drawing N different i.i.d. Gumbel errors for each good j, $\{\varepsilon_{ij}\}_{i=1}^{N}$.[56] Then the conditional probability is approximated by:

$$\begin{aligned}\hat{\Pr}[u_{iBt} > u_{iCt} \,|\, u_{iAt} = \max_{j \in \mathcal{J}} u_{ijt}; (v_{js})_{j \in \mathcal{J}}] = \sum_{i=1}^{N} 1 \big[& v_{Br} + \varepsilon_{iB} > v_{Cr} + \varepsilon_{iC} \\ & \text{and } v_{Ar} + \varepsilon_{iA} = \max_{j \in \mathcal{J}}\{v_{jr} + \varepsilon_{ij}\}\big] \\ & \Big/ \sum_{i=1}^{N} 1 \left[v_{Ar} + \varepsilon_{iA} = \max_{j \in \mathcal{J}}\{v_{jr} + \varepsilon_{ij}\}\right]. \end{aligned} \tag{9}$$

[55] This choice of support follows the Monte Carlo experiments in Heiss, Hetzenecker, and Osterhaus (2022).

[56] Regarding the absence of an s subscript: for computational simplicity, I use the same ten million Gumbel draws for all simulations.

With (8) and (9) in hand, I proceed to compute the absolute value of the difference between them:

$$\text{AbsDiff}_s = \left|\hat{\Pr}[u_{iBt} > u_{iCt} \mid u_{iAt} = \max_{j\in\mathcal{J}} u_{ijt}; (v_{js})_{j\in\mathcal{J}}] - \Pr[u_{iBr} > u_{iCr} \mid (v_{js})_{j\in\mathcal{J}}]\right|$$

Having repeated this process for S simulations, I compute the average absolute value of the difference between the conditional and unconditional probabilities: $S^{-1}\sum_{s=1}^{S} \text{AbsDiff}_s$.

Numerical Details and Results.—I perform the steps described above for markets of two different sizes: three goods and four goods. For each market size, I synthesize 100 different representative utility combinations (drawn, as described above, from the uniform distribution with support [-4.5, 3.5]). To approximate the conditional choice probabilities, I take ten million i.i.d. Gumbel draws per good.

The results of this simulation are as follows. For the three-good market, the mean absolute difference between the conditional and unconditional probability is 0.000265 (with a standard deviation of 0.000483). And for the four-good market, the mean absolute difference between the conditional and unconditional probability is 0.000457 (with a standard deviation of 0.000853).

D. Cross-Characteristic Correlations in (Dis)similarity

Appendix Table 1 reports cross-characteristic correlations in the substitutes' similarity or dissimilarity with respect to two characteristics. Letting i index rows and j index columns, cell entry i, j reports the correlation between the substitute's (i) matching the out-of-stock product on characteristic i and (ii) matching the out-of-stock product on characteristic j.

For the most part, similarity between the substitute and the out-of-stock product in one characteristic is inversely correlated with similarity in another. There are only a handful of exceptions. (For instance, a substitute flour is more likely to share the same flour type as the out-of-stock product if it also shares its "bleached" status.)

E. Details on the Structural Estimation Method

This appendix describes two aspects of the structural estimation process. These include (i) the estimation of correlations among the mixed probit error terms and (ii) the choice of tuning parameters (in both mixed logit and mixed probit).

Appendix Table 1 – Correlation Matrices of Similarity in Characteristics between Substitute and Out-Of-Stock Product

Panel A. Bottled water

	Same brand	Similar[a] bottle size	Similar[a] no. of bottles	Same water type
Same brand	1.00			
Similar[a] bottle size	–0.26	1.00		
Similar[a] no. of bottles	–0.31	–0.11	1.00	
Same water type	–0.02	–0.14	–0.09	1.00

Notes: Letting i index rows and j index columns, the entry in cell i, j indicates the correlation between the substitute and out-of-stock product sharing characteristic i and their sharing characteristic j as well. There are 106,484 observations.
[a] Within 10%.

Panel B. Flour

	Same brand	Same "bleached" status	Similar[a] quantity	Same flour type
Same brand	1.00			
Same "bleached" status	–0.05	1.00		
Similar quantity[a]	–0.49	–0.26	1.00	
Same flour type	0.02	0.07	–0.16	1.00

Note: 26,242 observations. (See Panel A for details.) [a] Within 10%.

Grid Search Estimator of Error Correlations in Mixed Probit.—Trip-specific circumstances sometimes shift multiple goods' utilities, causing their error terms to be correlated. To see the intuition, recall the example from Section 1 of a baker who usually bakes bread (for which bread flour is ideal), but who occasionally bakes cupcakes instead (for which all-purpose flour is preferable). On the rare trips when she plans to bake bread, there will be a positive shock to the utilities of all-purpose flours but a negative shock to those of bread flour. Now consider how these trips will figure in a discrete choice model. The positive shocks to all-purpose flours' utilities will appear as positive realization of those products' error terms, whereas the negative shocks to bread flours' utilities will manifest as negative realizations. Thus, the circumstances of a given shopping trip (and, in particular, the planned recipe) cause the error terms of products of a given flour type to be correlated with each other.

The preceding example highlights the following fact. In markets where trip-specific circumstances affect the utilities of multiple goods, a demand system should accommodate correlated errors. This is especially true when alternate choice data are available. In that event, the inclusion of correlated errors should enable the demand system to better match consumers' *observed* preferences over unpurchased products (as reported in the alternate choice data). And, to the extent that preferences over unpurchased products are indicative of product substitutability, the final result is more accurate estimates of demand elasticities.

Unlike mixed logit,[57] the mixed probit model accommodates correlated errors. However, it is challenging to recover the structure of the correlation. One must simulate choice probabilities not only for every point of the fixed grid, but also for each possible correlation structure. It is therefore helpful to minimize the number of potential correlation structures considered. For this reason, I adopt a grid search approach to estimating the correlations between error terms. This method is popular in the machine learning literature, where it is used for a different purpose (namely, tuning so-called "hyperparameters"). Here the method appeals for the same overarching reason: it minimizes the number of times a computationally burdensome procedure must be repeated.

The grid search estimator proceeds as follows. First, I propose a general structure for the correlations among the error terms. I begin by identifying a cluster of products within the category whose error terms are especially likely to be correlated. In so doing, I consult the descriptive evidence in Section 5B concerning within-consumer preference variation across trips. For the product category of flour, I focus on within-consumer preference variation with respect to flour type. The idea is that flours of the type needed for the consumer's intended recipe will enjoy positive utility shocks (which manifest as positive, correlated error terms). As for bottled water, the descriptives provide little guidance regarding which (if any) characteristics experience within-consumer variation in tastes. Resorting to intuition, I opt to model correlation centered on bottle count, the idea being that consumers will sometimes require more water bottles than usual due to trip-specific circumstances (such as preparing for a long road trip).

Having identified a cluster of products whose error terms may be correlated, I compute correlated errors as follows. Assume that the products' error terms are distributed multivariate normal such that (i) all the error terms' variances equal one; (ii) the error terms corresponding to products *within* the "correlation cluster" exhibit a common covariance of σ with one other, but are independent of the error terms of products outside the "correlation cluster;" and (iii) the error terms of products *outside* the "correlation cluster" are independent of both each other and of the error terms of products within the "correlation cluster." To see what the resulting covariance matrix might look like, recall the stylized four-good market from Section 1 in which products A and B are close substitutes for each other, but not for goods C and D (which, in turn, are close substitutes for each other but not for A or B). In relation to this stylized market, I might consider the following covariance

[57] Only generalizations of mixed logit, such as mixed nested logit, can incorporate correlated errors.

matrix:

$$\begin{pmatrix} \mathrm{Var}(\varepsilon_A) & \mathrm{Cov}(\varepsilon_A,\varepsilon_B) & \mathrm{Cov}(\varepsilon_A,\varepsilon_C) & \mathrm{Cov}(\varepsilon_A,\varepsilon_D) \\ \mathrm{Cov}(\varepsilon_B,\varepsilon_A) & \mathrm{Var}(\varepsilon_B) & \mathrm{Cov}(\varepsilon_B,\varepsilon_C) & \mathrm{Cov}(\varepsilon_B,\varepsilon_D) \\ \mathrm{Cov}(\varepsilon_C,\varepsilon_A) & \mathrm{Cov}(\varepsilon_C,\varepsilon_B) & \mathrm{Var}(\varepsilon_C) & \mathrm{Cov}(\varepsilon_C,\varepsilon_D) \\ \mathrm{Cov}(\varepsilon_D,\varepsilon_A) & \mathrm{Cov}(\varepsilon_D,\varepsilon_B) & \mathrm{Cov}(\varepsilon_C,\varepsilon_C) & \mathrm{Var}(\varepsilon_D) \end{pmatrix} = \begin{pmatrix} 1 & \sigma & 0 & 0 \\ \sigma & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Notice that I only model the error correlations within one "cluster" of products within the market: that of goods A and B. Ideally, I would also estimate the correlation between the error terms of goods C and D (the other pair of close substitutes). However, modeling correlations for two clusters, as opposed to one, would exponentially increase the computational burden.[58] Besides, when there are only two product clusters (as is the case here), the key qualitative patterns in the data can be captured by modeling the correlations of just one cluster's errors.[59] Happily, such is case for the product categories of bottled water and flour. Regarding the former category, I model correlations among the error terms of products with twenty-four bottles (as distinct from forty, the other top-selling size). As for the latter category, I model the correlations in the error terms of bread flours (as distinct from all-purpose flours, the other top-selling flour type).

Having identified a cluster of products whose error terms may be correlated, I specify a set $C = \{\sigma_1, \ldots, \sigma_C\}$ of possible covariance parameters. Then I estimate demand separately for each covariance parameter $\sigma_c \in C$. Each time, I follow the steps described above. The only difference between iterations $c = 1, \ldots, C$ concerns the simulated error terms. On iteration c, I assume that the error terms are distributed multivariate normal with the covariance matrix implied by (i) the cluster structure under consideration and (ii) the specific covariance parameter σ_c being evaluated. (Notice that the estimated distribution of the random coefficients (β_i, α_i) will vary across iterations so as to maximize the likelihood function *given the error draws*.)

With the estimates in hand, I identify the covariance parameter that results in the largest log likelihood at convergence. Then I perform estimation a *second time* with that parameter. (Without this step, the log likelihood for the "optimal" covariance parameter may be upwardly biased due to random noise in the

[58]More precisely, the computational burden is squared. For instance, if I considered five different levels of correlation per cluster—0, 0.1, 0.2, 0.3, and 0.4—evaluating the Cartesian product of the candidate correlations would require $5^2 = 25$ rounds of estimation.

[59]To see why, suppose that a consumer has purchased good A, so her second–most-preferred product is probably B. Because the consumer purchased good A, whose errors are correlated with those of good B, the realization of good B's error term is probably positive. Thus, the model would likely predict that good B is the consumer's second–most-preferred product. Now suppose, instead, that the consumer has purchased good C, in which case D is probably her second–most-preferred product. In this case, the realizations of A and B's errors would be disproportionately likely to be negative, thereby increasing the probability that the model assigns good D greater utility than A or B.

simulated probabilities.[60])

Tuning Parameters.—It is necessary to choose both the number and location of the fixed grid points before estimation.

Regarding the number of grid points, my approach closely resembles that employed by Train (2008). That is to say, I begin by determining the maximum number of grid points that can fit within the memory. Then I divide this total number of grid points evenly among the random coefficients, so that each coefficient's support will be discretized into the same number of values. The final result is that the support of each random coefficient is approximated by five distinct fixed points (which happens to be the same number as one of Train's specifications [2008].)

Having selected the number of distinct fixed grid values per random coefficient, it remains to determine their locations. I follow Heiss, Hetzenecker, and Osterhaus (2022) in basing the grid points' locations on *parametric* mixed logit estimates. Specifically, I center the grid on the mean coefficient estimates from the parametric model. Then, for each coefficient, I place the outermost points two (estimated) standard deviations above and below the mean. In the case of mixed probit, I divide each point by $\sqrt{1.6}$ to adjust for the difference in normalization between multinomial probit and logit models (see Train [2009]).

F. Multiple-Unit Purchases of Individual Products

Contrary to standard discrete choice frameworks, consumers sometimes purchase multiple units of a single product on one shopping trip. This poses a problem for the model selection exercise in Section 6. Recall that the IPA property of mixed logit imposes conditional independence between consumers' orders choices and their decisions to accept or reject the substitute, given their time-invariant tendencies to like or dislike the substitute (based on its observable characteristics). The key assumption is that consumers' preferences do not vary between trips in a fashion that is correlated across products. However, if consumers' choice sets include multiple units of individual products, their observed behavior may be inconsistent with the mixed logit IPA for a different reason: the model misspecifies the underlying choice problem. To see why, consider a consumer who likes to purchase bottled water in large quantities. She might consider the following to be her top two purchase options: (i) a 40-pack of the private label and (ii) *two* 24-packs of Ice Mountain. However, standard discrete choice models exclude option (ii), as they assume that she will purchase only a

[60]The "optimal" covariance parameter σ_c is chosen because it maximizes the log likelihood at convergence. However, the log likelihood is evaluated with error because it is simulated—and sometimes the simulated probabilities of consumers' observed choices exceed the *true* probabilities (perhaps because the error draws spuriously align with consumers' observed choices).

single unit of a given product. In consequence, discrete choice models might underestimate the probability that she purchases Ice Mountain while overestimating the probability that she purchases other brands that offer larger packs.

One possible solution would be to treat different quantities of a product as distinct alternatives. For instance, purchasing one 24-pack of Ice Mountain would be treated as a different alternative from purchase two 24-packs of Ice Mountain. However, the data on stockout substitutions do not report the requested number of units of the out-of-stock product. Although it seems likely that the consumer would be offered a quantity of the substitute such that the total quantity (i.e., size per unit times number of units) would closely match the out-of-stock product's in most situations, it also seems probable that rejection would be especially likely in situations where the substitute's total quantity diverges from the out-of-stock product's. For this reason, I do not attempt to impute the number of units requested of the out-of-stock product based on the substitute's total quantity. Instead, I identify households who are especially *unlikely* to purchase multiple units of a single product. To do so, I find households for whom I observe (i) zero purchases involve multiple units and (i) ten or more purchases in total. (In principle, I could solely drop *transactions* featuring multi-unit purchases, as opposed to entire households. However, because multi-unit purchases are so common [see Section 6C], it seems plausible that a large fraction of households entertained multi-unit purchases during trips where they ultimately purchased a single unit.)

I quantify the importance of excluding households with multi-unit purchases as follows. First, I draw a random sample of households from the *universe* of sample households (as opposed to those with 10+ transactions and 0 multi-unit purchases). And second, I repeat the model selection exercises in Section 6D on this alternative sample. So far as purchases and online orders are concerned, the results resemble those presented in the main text. Concerning stockout substitutions for bottled water, however, mixed logit's performance suffers relative to mixed probit. Within-sample (out-of-sample), the predicted probabilities of consumers' choices about bottled water substitutes are 2.2 (3.6) percentage points higher for mixed probit than for mixed logit. (By comparison, the corresponding disparities in the main text are 0.9 and 1.9 percentage points within- and out-of-sample, respectively.) For flour, the results do not change so dramatically.[61]

[61] Within-sample, the predicted probabilities of consumers' choices about flour substitutes are 1.1 percentage point higher for mixed logit than flour (versus 1.0 percentage points in the main text). Out-of-sample, the corresponding disparity is 1.8 percentage points (versus 2.4 percentage points in the main text).

G. Supplementary Results from Structural Estimation

Verifying the Mixed Logit IPA.—According to Corollary 1, the *conditional* probability of accepting a stockout substitute—given one's original order choice—should be identical to the *unconditional* probability of the same. To verify that this is indeed the case, Appendix Table 2 compares two estimation approaches. The first approach directly imposes the mixed logit IPA, resulting in the closed-form likelihood presented in Section 6B of the main text. By contrast, the second approach simulates the likelihood function without imposing the mixed logit IPA. Simulation proceeds in two steps. First, I compute the order choice probabilities by drawing from the standard Gumbel distribution. And second, I calculate the accept/reject probabilities based solely on the error draws that resulted in "correct" order predictions.

Appendix Table 2 – Verifying the Mixed Logit IPA by Simulation

	Product category	
Statistic	Bottled water	Flour
Frac. of stockouts with same prediction	0.997	0.997
Avg. absolute difference in predicted prob. accept	0.010	0.014
Root mean square difference in predicted prob. accept	0.024	0.038

Notes: This table compares the predictions of two mixed logit estimators: (i) directly imposing the mixed logit IPA (and using the resultant closed-form likelihood), and (ii) simulating the choice probabilities. For (ii), the accept/reject probabilities are solely based on Gumbel error draws that result in the "correct" original online order. (Consequently, most of the 20,000 draws used to compute the order probabilities are discarded for the accept/reject stage.)

Appendix Table 2 reports three measures of the similarity of the two estimation approaches. All these measures pertain to the predicted probability of acceptance. The first measure is the fraction of stockout substitutions in which both models predict the same outcome.[62] As for the second measure, I compute the average absolute difference between the two models' predicted probabilities of acceptance. Letting $s \in \{1, \ldots, S\}$ index (attempted) stockout substitutions, the measure is given by

$$\text{AAD} = \frac{1}{S} \sum_{s=1}^{S} |P_s - \hat{P}_s|$$

where P_s indexes acceptance probabilities derived from the closed-form likelihood and $\hat{P}_s$ denotes their simulated counterparts. The third (and final) measure is the root-mean-square difference in predicted

[62]That is, I compute the fraction of substitutions in which either (i) both models assign a predicted probability of >50% to acceptance or (ii) both assign a predicted probability of <50% to acceptance.

acceptance probabilities:

$$\text{RMSD} = \sqrt{\frac{1}{S}\sum_{s=1}^{S}(P_s - \hat{P}_s)^2}$$

Observe that the second and third measures are similar in spirit; both gauge the average "distance" in acceptance probabilities. However, the average absolute difference employs the L_1 norm whereas the root-mean-square difference employs the L_2 norm.

The results in Appendix Table 2 indicate that the two estimation approaches arrive at very similar predictions. This is especially true where bottled water is concerned.[63]

Random Coefficients.—Appendix Table 3 reports summary statistics for the random coefficients (i.e., the β's) in each product category. To compare the mixed logit coefficients with their mixed probit counterparts, divide the former by $\sqrt{1.6}$. Concerning mixed probit, Appendix Table 3 also indicates the estimated correlation parameter (σ) for the indicated "cluster" of products. This parameter is estimated to be 0.1 for bottled water and 0.2 for flour.[64]

Whether the error terms are correlated or uncorrelated, mixed probit does not display an IPA property.[65] Consequently, the model allow a consumer's initial order to be correlated with her decision to accept or reject a stockout substitute. In principle, this correlation might have no real-world economic content (being a purely mathematical property). However, it is also possible that this correlation reflects real-world consumer behavior. Regarding the latter hypothesis, recall that multinomial probit with uncorrelated errors—hereafter, "independent probit"—does not suffer from the familiar independence of preferred alternatives (IIA) property displayed by conditional logit (Paetz and Steiner 2017). Furthermore, independent probit relaxes the IIA property in a systematic way. Consider a market with two goods: A and B. Without loss of generality, assume that good A commands a larger choice share than does good B. Simulations performed by Paetz and Steiner (2018) suggest that the introduction of a third good—say, C—will cause the choice share of the *less* popular good (B) to shrink more dramatically in percentage terms than the choice share of the *more* popular good (A).

[63]There are two reasons why the results for bottled water are more precise than those for flour. First, a larger fraction of error draws translate to "correct" order predictions in the former category than in the latter (27% versus 20%). This leaves more draws with which to simulate the accept/reject probabilities. And second, the utility specification for bottled water is more flexible than the utility specification for flour. Whereas the former includes dummies for individual products, the latter relies on observable characteristics (brand, flour type, quantity, etc.)

[64]I tested five potential correlation parameters in each category: $\{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$.

[65]By way of example, consider a three-good market with goods A, B, and C. Suppose that $u_{iAt} = \varepsilon_{iA}$, $u_{iBt} = 1 + \varepsilon_{iB}$, and $u_{iCt} = 2 + \varepsilon_{iB}$ (where the error terms are i.i.d. standard normal). Then $\Pr\left[u_{iBt} > u_{iCt} \mid u_{iAt} = \max_{j\in\mathcal{J}} u_{ijt}\right] \approx 0.331 > 0.240 \approx \Pr[u_{iBt} > u_{iCt}]$.

Appendix Table 3 – Summary Statistics on Structural Parameters

	Panel A. Bottled water			
	Mixed logit		Mixed probit	
Variable	Means	Std. devs.	Means	Std. devs.
Price	−0.682	0.675	−0.456	0.562
Aquafina (24 ct.)[a]	6.688	2.709	5.316	1.944
Ice Mtn. (24 ct.)[a]	7.635	2.443	5.966	1.728
Nestle (24 ct.)[a]	7.065	1.925	5.559	1.468
Pvt. lbl. purified water (24 ct.)[a]	7.518	1.447	6.166	1.056
Pvt. lbl. purified water (40 ct.)[a]	7.625	2.114	5.951	1.617
Pvt. lbl. spring water (24 ct.)[a]	8.497	1.980	6.279	1.431
Error correlation cluster			24-packs	
Correlation parameter (σ)			0.1	
	Panel B. Flour			
Price	−1.974	1.060	−1.578	0.799
All-purpose flour	5.372	0.193	4.270	0.169
Bread flour	3.819	4.298	3.365	3.521
Gold Medal brand	1.237	2.972	0.867	1.698
King Arthur brand	1.006	4.123	0.288	3.148
Log quantity	2.134	1.229	1.512	0.977
Unbleached	−0.639	3.375	−0.847	1.899
Error correlation cluster			Bread flours	
Correlation parameter (σ)			0.2	

Notes: This table presents summary statistics for the nonparametrically-estimated distributions of random coefficients. To compare the mixed logit coefficients with the mixed probit ones, divide the former by $\sqrt{1.6}$. The "error correlation clusters" in mixed probit consist of products whose error terms are correlated. See Appendix E for details.

[a] Product-specific dummy.

The "Hit Rate."—In Section 6D, I compare mixed logit and mixed probit's goodness of fit based on the average predicted probability assigned to consumers' observed choices. An alternative measure of fit is the fraction of observations in which consumers' observed choices are assigned the highest predicted probability of any alternative—hereafter, the "hit rate."

The discussion in the main text focuses on the average predicted probabilities of consumers' observed choices—as opposed to the hit rate—for two reasons. First, the predicted probability of the chosen product directly enters the likelihood function, whereas the "hit rate" does not. And second, the predicted probability of the chosen product is more closely related to the product's (estimated) cross-price elasticities than the "hit rate" is.[66]

Appendix Table 4 – "Hit Rate:" Mixed Logit versus Mixed Probit

Data type	Bottled water: Mixed logit	Bottled water: Mixed probit	Flour: Mixed logit	Flour: Mixed probit
Panel A. Within sample				
In-store purchases and online orders	0.657	0.653	0.672	0.661
Stockout substitutions	0.959	0.959	0.951	0.948
Panel B. Out of sample				
In-store purchases and online orders	0.571	0.568	0.652	0.646
Stockout substitutions	0.942	0.942	0.928	0.934

Notes: This table compares the "hit rates" of mixed probit and mixed logit models for the product categories of bottled water and flour. The hit rate is defined as the fraction of choice situations for which the model assigns the consumer's observed choice the highest predicted probability of any alternative. (See Sections 6B and 6D for details on estimation).

Appendix Table 4 compares the hit rates of mixed logit and mixed probit. Within sample, mixed logit delivers a (weakly) higher hit rate for both data types, irrespective of product category. Out of sample, by contrast, the results vary by data type. Regarding in-store purchases and online orders, mixed logit affords a higher hit rate for both product categories. As for stockout substitutions, mixed probit provides an identical hit rate for bottled water and a 0.6 percentage higher hit rate for flour.

Estimated Price Elasticities.—Appendix Tables 5 and 6 report estimated conditional price elasticities for bottled water and flour, respectively.[67] The top entry in each cell corresponds to mixed logit, whereas the bottom one corresponds to mixed probit.

[66] The cross-price elasticity of good j with respect to good j' is defined as $(\partial s_j / \partial p_{j'})(p_{j'}/s_j)$, where s_j denotes the market share of good j. In this equation, s_j is computed as the average predicted probability of j being purchased (across all the observed choice situations), while $(\partial s_j / \partial p_{j'})$ is defined as marginal changes in the same. See Train (2009).

[67] That is, conditional on purchasing an "inside good" (as opposed to the "outside option" of purchasing nothing.)

Appendix Table 5 – Conditional Own- and Cross-Price Elasticities: Bottled Water

	Nestle (24 ct.)	PL spring (24 ct.)	Ice Mtn. (24 ct.)	Aquafina (24 ct.)	PL pfd. (40 ct.)	PL pfd. (24 ct.)	*Model*
Nestle Pure Life (24 ct.)	−0.88	0.18	0.19	0.18	0.17	0.17	*Logit*
	−1.00	0.21	0.22	0.22	0.19	0.20	*Probit*
Pvt. lbl. spring (24 ct.)	0.18	−0.85	0.20	0.16	0.16	0.18	*Logit*
	0.20	−0.98	0.21	0.20	0.18	0.21	*Probit*
Ice Mtn. (24 ct.)	0.19	0.18	−0.92	0.19	0.17	0.16	*Logit*
	0.21	0.19	−1.05	0.24	0.20	0.19	*Probit*
Aquafina (24 ct.)	0.17	0.15	0.19	−0.91	0.20	0.16	*Logit*
	0.19	0.17	0.21	−1.02	0.21	0.19	*Probit*
Pvt. lbl. purified (40 ct.)	0.15	0.15	0.17	0.19	−0.81	0.16	*Logit*
	0.16	0.16	0.18	0.19	−0.87	0.17	*Probit*
Pvt. lbl. purified (24 ct.)	0.17	0.18	0.17	0.17	0.18	−0.84	*Logit*
	0.20	0.22	0.19	0.21	0.21	−1.01	*Probit*

Notes: Letting i index rows and j index columns, cell i, j reports the percent change in product i's market share when product j's price increases by one percent. The top entry in each cell corresponds to mixed logit, whereas the bottom one corresponds to mixed probit.

Appendix Table 6 – Conditional Own- and Cross-Price Elasticities: Flour

	GM APB (5 lb)	PL APB (5 lb)	PL APB (2 lb)	PL APU (5 lb)	GM APU (5 lb)	PL APB (10 lb)	KA UBF (5 lb)	GM HKB (5 lb)	KA APU (5 lb)	KA APU (10 lb)	*Model*
Gold Medal all-purpose bleached (5 lb)	−1.96	0.28	0.21	0.20	0.20	0.21	0.22	0.17	0.21	0.23	*Logit*
	−1.88	0.26	0.21	0.22	0.19	0.19	0.20	0.18	0.20	0.22	*Probit*
Pvt. lbl. all-purpose bleached (5 lb)	0.28	−1.88	0.25	0.22	0.17	0.20	0.18	0.18	0.20	0.24	*Logit*
	0.26	−1.89	0.26	0.21	0.19	0.19	0.16	0.18	0.20	0.23	*Probit*
Pvt. lbl. all-purpose bleached (2 lb)	0.20	0.29	−1.88	0.28	0.21	0.20	0.20	0.17	0.15	0.15	*Logit*
	0.20	0.30	−1.90	0.29	0.21	0.19	0.20	0.17	0.15	0.16	*Probit*
Pvt. lbl. all-purpose unbleached (5 lb)	0.21	0.25	0.33	−1.94	0.26	0.23	0.19	0.22	0.19	0.16	*Logit*
	0.21	0.22	0.32	−1.86	0.27	0.21	0.18	0.18	0.17	0.14	*Probit*
Gold Medal all-purpose unbleached (5 lb)	0.23	0.20	0.27	0.30	−2.08	0.22	0.22	0.19	0.24	0.21	*Logit*
	0.21	0.21	0.25	0.31	−1.94	0.22	0.19	0.20	0.19	0.19	*Probit*
Pvt. lbl. all-purpose bleached (10 lb)	0.22	0.24	0.23	0.25	0.22	−2.19	0.24	0.22	0.19	0.25	*Logit*
	0.21	0.20	0.20	0.23	0.22	−1.93	0.22	0.20	0.17	0.20	*Probit*
King Arthur unbleached bread flour (5 lb)	0.24	0.21	0.22	0.20	0.21	0.26	−2.18	0.26	0.27	0.21	*Logit*
	0.22	0.19	0.22	0.20	0.20	0.25	−2.04	0.28	0.24	0.20	*Probit*
Gold Medal “Harvest King” bread (5 lb)	0.18	0.22	0.20	0.22	0.18	0.21	0.26	−2.04	0.31	0.28	*Logit*
	0.18	0.19	0.18	0.21	0.19	0.20	0.26	−1.93	0.29	0.26	*Probit*
King Arthur all-purpose unbleached (5 lb)	0.22	0.22	0.16	0.18	0.23	0.17	0.23	0.26	−1.91	0.34	*Logit*
	0.21	0.21	0.16	0.18	0.19	0.17	0.21	0.26	−1.87	0.34	*Probit*
King Arthur all-purpose unbleached (10 lb)	0.23	0.27	0.16	0.15	0.18	0.21	0.18	0.23	0.27	−1.88	*Logit*
	0.23	0.26	0.16	0.15	0.18	0.19	0.18	0.24	0.29	−1.87	*Probit*

Notes: See notes for Appendix Table 5