

ESSAYS IN EMPIRICAL INDUSTRIAL ORGANIZATION

By

Andrew Zeyveld

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Economics—Doctor of Philosophy

2025

ABSTRACT

Chapter 1: Steering Consumers' Learning: Evidence from Stockout Substitutions in Curbside Pickup

Items ordered for curbside pickup sometimes go out of stock, obliging the store to choose substitutes on consumers' behalf. Using novel data from a supermarket chain, I show these "stockout substitutions" influence consumers' future purchases through the mechanism of learning. This presents the store with the following opportunity to increase its future profits: if the store selects substitutes from profitable brands that consumers have never tried before, some consumers will learn that they like the brands of their substitutes and purchase these brands' products in the future. However, consumers are less likely to accept such substitutes than they are to accept substitutes from brands they have previously purchased. To quantify the trade-off between steering consumers' learning and maximizing the probability of substitutes' acceptance, I estimate a learning-based model of differentiated products demand. Although steering consumers' learning proves an unprofitable strategy, the store can still increase profits—and consumer welfare—by individualizing substitutions according to consumers' past purchases and demographics.

Chapter 2: Demand Estimation When Consumers' Preferences Vary over Time

This paper shows that workhorse demand systems fail to reproduce important substitution patterns when individual consumers' preferences vary over time. This failure is rooted in the *independence of preferred alternatives (IPA)* properties of conditional and mixed logit, which restrict the relationship between consumers' purchases and their preferences among unpurchased goods. To assess the empirical relevance of the IPA properties, I employ novel data from stockout substitutions in curbside pickup. For the two product categories that I study, I document substitution patterns that are inconsistent with the IPA property of conditional logit. As for mixed logit, its IPA property proves consistent with the substitution patterns in one of the two product categories. To quantify the benefits of relaxing the IPA property of mixed logit, I compare the model's goodness of fit with that of mixed probit (which does not display an IPA property). In keeping with the descriptive evidence, the results of this comparison vary by product category.

ACKNOWLEDGEMENTS

I am grateful for the guidance and mentorship of my committee co-chairs, Mike Conlin and Kyoo il Kim. They have been deeply invested in every aspect of my professional development, from training me to think like an economist to extending my econometric toolkit. I would also like to thank committee members Arijit Mukherjee and Forrest Morgeson for their thoughtful comments.

Additional thanks are extended to the retailer that supplied my data. I am especially grateful to Kevin D., who first introduced me to the problem of stockout substitutions.

I would like to recognize generous financial support from the Michigan State University Department of Economics, College of Social Science, and Graduate School. I am also indebted to dunhumby [sic], which provided supplementary financial support; and to the Institute for Cyber-Enabled Research at Michigan State University, which provided computational resources and services.

Finally, I would like to thank my spouse, Anna Zeyveld Jeffries, for her kindness, compassion, and patience throughout my studies.

TABLE OF CONTENTS

CHAPTER 1	INTRODUCTION	1
CHAPTER 2	STEERING CONSUMERS' LEARNING: EVIDENCE FROM STOCKOUT SUBSTITUTIONS IN CURBSIDE PICKUP	4
2.1	Introduction	4
2.2	Background	8
2.3	Descriptive Evidence	16
2.4	Conceptual Model	26
2.5	Empirical Model and Estimation	30
2.6	Estimation Results	40
2.7	Counterfactual Simulations	44
2.8	Conclusion	55
	BIBLIOGRAPHY	58
APPENDIX 2A	DATA STRUCTURE AND OBSERVABLE CHARACTERISTICS	65
APPENDIX 2B	ADDITIONAL DESCRIPTIVE EVIDENCE	69
APPENDIX 2C	ESTIMATION DETAILS	76
APPENDIX 2D	ESTIMATION RESULTS FOR APPLE SAUCE CUPS	81
APPENDIX 2E	SUPPLEMENTARY COUNTERFACTUAL SIMULATIONS	83
CHAPTER 3	DEMAND ESTIMATION WHEN CONSUMERS' PREFERENCES VARY OVER TIME	86
3.1	Introduction	86
3.2	Relationship to Prior Literature	88
3.3	Theory: Alternate-Choice Data in Demand Systems	94
3.4	Institutional Background and Data	98
3.5	Descriptive Evidence	101
3.6	Structural Evidence	115
3.7	Conclusion	129
	BIBLIOGRAPHY	132
APPENDIX 3A	PROOF OF LEMMA 1	139
APPENDIX 3B	COMPARISON OF THEOREM 1 WITH PRIOR THEORETICAL RESULTS	141
APPENDIX 3C	MONTE CARLO TESTS OF THEOREM 1	142
APPENDIX 3D	CROSS-CHARACTERISTIC CORRELATIONS IN (DIS)SIMILARITY	144
APPENDIX 3E	DETAILS ON THE STRUCTURAL ESTIMATION METHOD	145
APPENDIX 3F	MULTIPLE-UNIT PURCHASES OF INDIVIDUAL PRODUCTS	149
APPENDIX 3G	SUPPLEMENTARY RESULTS FROM STRUCTURAL ESTIMATION	151

CHAPTER 1

INTRODUCTION

People often face discrete choices. For example, someone who is shopping for a pint of ice cream might confront dozens of options which vary in brand, flavor, and price. To understand how people approach such discrete choices, social scientists routinely employ random utility models. These models relate people's choices to their circumstances. Concerning ice cream, for instance, how do consumers respond when the price of their favorite product increases?

To estimate random utility models, researchers need to make assumptions about people's decision-making process. One such assumption is that people possess perfect information about all the available options. Another is that people's preferences remain stable over time. My dissertation explores how these assumptions can be relaxed. How should the researcher model people's discrete choices when they possess imperfect information? Or when their preferences change over time? I take up these questions in the context of grocery shopping, an ideal natural laboratory to study discrete choice. In the first place, I can follow individual consumers' purchases over time (provided they participate in the chain's loyalty program). This allows me to identify heterogeneity in consumers' preferences, as well as shifts in individual consumers' preferences over time. In the next place, grocery chains collect demographic data about the consumers who frequent their stores. This enables me to determine how characteristics like age or income are correlated with consumers' purchases. Neither of these features of grocery shopping data are, of course, unique to this dissertation. In fact, a large literature within empirical industrial organization and quantitative marketing studies grocery shopping for these reasons (such as Nevo (2001); Erdem, Keane, and Sun [2008]; or Backus, Conlon, and Sinkinson [2021]). My dissertation, however, benefits from a recent development within the grocery industry: curbside pickup.

Curbside pickup is a "click-and-collect" form of shopping in which consumers order groceries online and later pick up their groceries from the local supermarket. Importantly, requested items sometimes go out of stock after consumers have already placed their orders, but before the store collects them. This obliges the store to select substitutes on the affected consumers' behalf. Once

the consumers arrive at the store, they are presented with two options: either they can “accept” the substitute chosen by the store, or they can “reject” it and buy no such item.

These “stockout substitutions” sometimes cause consumers to try out new products for the first time. What they learn about these products might, in turn, influence their future purchases. In Chapter 2, I present model-free evidence that is consistent stockout substitutions’ affecting consumers’ future purchases through the mechanism of learning. Further, consumers appear to learn more about their tastes for *brands*—meaning branded product lines, like the Ben & Jerry’s line of ice cream—than about their tastes for other characteristics (like size or flavor). This suggests the store could exploit stockout substitutions to increase its future profits. For, if consumers were offered substitutes from high-margin brands that they have never tried before, some would discover that they like their substitutes’ brands and then purchase these brands’ (profitable) products in the future. Additional descriptive evidence, however, suggests consumers are reluctant to accept stockout substitutes from unfamiliar brands.

To quantify the trade-off between steering consumers’ learning and maximizing the probability of substitutes’ acceptance, I estimate a learning model of differentiated products demand. Counterfactual simulations indicate that the store could not perceptibly increase its profits by steering consumers’ learning. This is partly due to consumers’ disinclination to accept stockout substitutions from unfamiliar brands. In addition, when consumers accept, they do not learn enough about their substitutes’ brands for future profits to meaningfully change.

Although steering consumers’ learning proves an unprofitable strategy, it emerges that store profits—as well as consumer profits—increase when the store individualizes its choice of substitute based on consumers’ original orders, past purchases, and demographics. Moreover, a substantial fraction of these gains can be achieved by individualizing the choice of substitute according to consumers’ original orders alone.

In Chapter 3, I put the stockout data to a different purpose: namely, recovering within-consumer variation in preferences. The idea is that individual consumers’ preferences sometimes change over time. Take the case of coffee: many consumers prefer iced coffee during the summer but hot

coffee during the winter. To what extent do workhorse demand systems accommodate within-consumer preference variation like this? I show that conditional logit imposes independence between consumers' purchases and their preferences among unpurchased goods. As for mixed logit, this more flexible model imposes conditional independence between consumers' purchases and their preferences among unpurchased goods, given the realizations of consumers' random coefficients. In other words, someone's purchase should be uninformative of time-specific factors that influenced both her purchase and her preferences among the goods she did not purchase. I term these the "Independence of Preferred Alternatives" (IPA) properties of conditional and mixed logit, respectively.

These theoretical results raise two empirical questions. First, can data help determine whether consumers' preferences in a given market are consistent with the IPA property of mixed logit? And second, how should demand be estimated when consumers' preferences prove inconsistent with the property?

I answer these questions using the same dataset as in Chapter 2. On these data, the conditional logit IPA imposes independence between consumers' original orders and their willingness to accept specific products as stockout substitutes. This restriction proves inconsistent with the data from both product categories that I study: namely, bottled water and flour. As for mixed logit, its IPA property essentially dictates that individual consumers have the same preferences for stockout substitutes on all of their shopping trips—irrespective of their original order choice. Descriptive evidence suggests that consumers' behavior is consistent with this prediction in only one of the two product categories that I study: bottled water. Concerning flour, by contrast, consumers' preferences for substitutes seem to change across shopping trips (owing perhaps to variation in the planned recipe).

To help quantify the benefits of relaxing the IPA property of mixed logit, I compare the model's goodness of fit with that of mixed probit (which does not display an IPA property). Overall, mixed probit seems to forecast consumers' accept/reject decisions more precisely than mixed logit does. In keeping with the descriptive evidence, this disparity appears to be more pronounced for the product category of flour than for bottled water.

CHAPTER 2

STEERING CONSUMERS' LEARNING: EVIDENCE FROM STOCKOUT SUBSTITUTIONS IN CURBSIDE PICKUP

2.1 Introduction

Consumers often make decisions with imperfect information. This has motivated many studies on government information provision. These studies frequently find that the government could increase consumers' welfare by providing easily accessible information. Take the case of grocery shopping. When the government requires that unhealthy products carry warning labels, consumers reduce their purchases of unhealthy products they had mistakenly believed to be healthy (Barahona, Otero, and Otero 2023). Health insurance is another example: by publishing quality scores, the government can steer consumers towards higher-quality insurance plans (Vatter 2024).¹

Online platforms also steer consumers' learning.² However, online platforms' incentives differ from those of the government: whereas the government steers consumers' learning to increase their welfare, online platforms steer consumers' learning to maximize profits. How, then, are consumers affected when online platforms steer their learning? I take up this question in the context of curbside grocery pickup (hereafter, "curbside pickup"). This is a "click-and-collect" form of shopping in which consumers order groceries online and then pick up their groceries from the local supermarket. Sometimes, however, the store cannot supply an ordered item because it has gone out of stock. This obliges the store to select another item—known as a "stockout substitution"—to serve as a replacement. Once the consumer arrives, she can either purchase this suggested substitute, or reject it and buy no such item.³

Stockout substitutions sometimes cause consumers to try out new products. What they learn

¹Concerning both food and health insurance, government information provision also causes a welfare-increasing response on the supply side. In particular, food manufacturers formulate healthier products (Barahona, Otero, and Otero 2023), while health insurers increase plan quality (Vatter 2024).

²Firms employ many methods to steer consumers' learning. One is advertising, which serves to inform consumers of a firm's product range (Anand and Shachar 2011) as well as to signal its quality level (Akerberg 2003). Strategic pricing is another method of steering consumers' learning. By reducing its prices, a firm *encourages* consumers to try its own products (Osborne 2011) while *discouraging* them from trying its competitors' products (Ching 2010).

³Of course, she could also go into the store to search for a different substitute. However, the data suggest that this is quite rare.

about these products might, in turn, influence their subsequent purchases. Using novel data on curbside pickup at a large regional supermarket chain, I supply model-free evidence that supports this intuition; stockout substitutions do, indeed, affect consumers' future purchases through the mechanism of learning. Moreover, consumers seem to learn more about their preferences for *brands*—meaning branded product lines, like the Häagen-Dazs line of ice cream—than about their preferences for other characteristics (such as size). This suggests that the store could exploit stockout substitutions to steer consumers' learning towards high-margin brands. For, if consumers were offered substitutes from high-margin brands they have never tried before, some consumers would discover that they like their substitutes' brands and then purchase these brands' (profitable) products in the future. However, less profitable outcomes are also possible. Consumers might be inclined to reject substitutes from unfamiliar brands, leaving the store with zero margins on the present transaction. Consistent with this hypothesis, reduced-form evidence suggests that consumers prefer when stockout substitutes belong to brands they have previously purchased.

Given the uncertainty involved, could the store increase profits by steering consumers' learning? And would doing so increase, or decrease, consumer welfare? To answer these questions, I estimate a learning model of differentiated products demand. Counterfactual simulations suggest that the store cannot increase profits by steering consumers' learning. This is because consumers are reluctant to accept stockout substitutes from unfamiliar brands and, when they do accept, tend to learn too little for their subsequent purchases—or the store's future profits—to meaningfully change. However, I find that store profits—as well as consumer welfare—increase when the store individualizes its choice of substitute based on consumers' original orders, past purchases, and demographics. Much of these gains can be realized by personalizing the choice of substitute according to the consumer's original order alone.

From a policy perspective, my findings are significant for the following reason. Regulators worry that online platforms like Amazon favor profitable products in their search rankings and product recommendations (Farronato et al. [2024]). Taking as given that online platforms engage in such behavior, would it be better for consumers if platforms steered demand towards profitable

products with, or without, the benefit of consumer microdata (like purchase histories or household demographics)? Conditional on platforms’ maximizing variable profits (as opposed to revenue or consumer welfare), my results suggest that consumers may benefit when platforms exploit consumer microdata. This result is surprising in the context of curbside pickup, where the “outside option” of procuring the relevant item elsewhere (or going without) is quite unattractive.⁴

The remainder of the paper proceeds as follows. In Section 2.2, I relate my analysis to prior work in industrial organization and quantitative marketing. I also provide details about the purchase environment and data. Briefly, I study a supermarket chain that offers three ways to shop: in-person, home delivery, and curbside pickup (where “stockout substitutions” occur). For each such substitution, I observe the out-of-stock item and the substitute, as well as the consumer’s decision to accept or reject the substitute. Besides the data on stockout substitutions, I also employ “scanner data” that record consumers’ purchases at the store. These scanner data display a household-level panel structure thanks to the chain’s loyalty program, enabling me to compare someone’s purchases before versus after a stockout substitution. I also have household-level demographic data that the store commissioned from a marketing firm.

In Section 2.3, I present descriptive evidence of the trade-offs faced by the store as it chooses stockout substitutes. I begin by characterizing when consumers are willing, or unwilling, to accept stockout substitutes. Probit regressions indicate that the probability of acceptance is increasing in the similarity of the substitute’s observable characteristics—such as brand or size—to those of the out-of-stock product (as well as the consumer’s past purchases). Next, I ask whether stockout substitutions influence consumers’ learning. To provide insight, I examine stockouts where consumers are offered substitutes that feature a brand, flavor, or other attribute they have never tried before. It emerges that these consumers proceed to purchase products with these hitherto-unfamiliar attributes more often in the future than do comparable consumers who successfully picked up before the stockout event. This pattern is consistent with consumers’ learning about

⁴To secure a substitute other than that offered by the store, a consumer who has suffered a stockout in curbside pickup would need to (i) find a new parking spot that is not reserved for curbside pickup, enter the store, and search for an alternative substitute therein; or (ii) add an extra visit to a different store.

their tastes for the substitutes' observable characteristics. Furthermore, consumers seem to learn more about the characteristic of brand than they do about other characteristics. This is intuitive; consumers are unlikely to learn much from, say, purchasing a specific quantity of milk for the first time. Finally, I examine how products' observable characteristics affect their retail margins (meaning the difference between the retail price and the wholesale cost). The characteristic of brand proves a key determinant of retail margins.

These empirical patterns present the store with the following strategic problem. Although stockout substitutions enable the store to steer consumers' learning towards profitable brands, doing so would increase the risk that stockout substitutes are rejected (leaving the store with zero margins on the present transaction). In Section 2.4, I present a conceptual framework that formalizes these trade-offs in a simplified environment. This framework shows that steering consumers' learning has an ambiguous effect on their welfare. The same is true when the store individualizes its choice of substitute based on consumers' past purchases or demographics.

Section 2.5 builds a learning model of differentiated products demand and then explains the estimation procedure. In the model, consumers are unsure of their tastes for a given brand until they purchase one of its products. Consumers' prior beliefs about brands, along with their true tastes, are heterogeneous. Brand aside, I allow for unobserved heterogeneity in consumers' preferences for non-brand characteristics like flavor. Consumers' preferences for these characteristics also vary based on the household-level demographic information observed by the store (like household income).

The estimated model parameters are reported in Section 2.6. With these in hand, I simulate profits and consumer welfare under counterfactual substitution policies. These policies vary along two dimensions. One is the store's objective function. I compare outcomes when the store either (a) maximizes expected present-trip profits alone or (b) maximizes the present-discounted value of expected profits (both present and future). The second dimension along which counterfactual policies vary is the extent of consumer microdata used. I compare policies that exploit (i) none of these data, (ii) just the consumer's original order, or (iii) the consumer's past purchases and

household demographics as well as her original order.

I find that substitution policies designed to maximize present-trip or total profits yield similar outcomes. This pattern, which holds in multiple product categories, suggests that the store cannot profit from steering consumers’ learning. To determine why this is the case, I simulate outcomes under counterfactual changes to the purchase environment or the primitives of consumers’ learning. I find that if consumers were less reluctant to accept stockout substitutes from unfamiliar brands, and if they experienced more learning conditional on acceptance, the store could perceptibly increase its future profits by steering consumers’ learning. As for the store’s use or disuse of the consumer microdata, I find that most gains from individualization can be secured by conditioning the choice of substitute on the out-of-stock product. Concerning super-premium ice cream, for example, 78% (\$0.28) of the per-stockout gains from individualizing the choice of substitute can be achieved if the store individualizes its choice of substitute according to the consumer’s original order.

2.2 Background

2.2.1 Related Literature

An emergent literature studies how online platforms steer consumers’ learning. This literature has so far focused on *search goods*. These are goods whose utility can be determined prior to purchase by inspecting the good or reading a description (see Nelson [1970]). Concerning search goods, platforms steer consumers’ learning by manipulating the set of products that consumers encounter on the platform. Take the case of search rankings, where consumers are more likely to click on—and learn about—products that are highly ranked. This creates an incentive for platforms to assign profitable goods a higher search rank than similarly-popular goods that afford smaller margins. Consistent with this intuition, Farronato, Fradkin, and MacKay (2023) show that Amazon’s search rankings favor its own products (which afford high margins) over the products of third-party sellers (which afford thinner margins). Meanwhile, Reimers and Waldfogel (2023) develop an equilibrium framework to detect bias in search rankings, which they apply to data from Amazon, Expedia, and Spotify. Search rankings aside, platforms also steer consumers’ learning through product recommendations. Chen and Tsai (2024) show that Amazon’s “Frequently Bought

Together” recommendations privilege Amazon’s own products over those of third-party sellers.

Unlike the foregoing studies, I explore how online platforms can steer consumers’ learning about *experience goods*. These are goods for which consumers cannot learn their tastes prior to purchase. Instead, consumers learn their tastes for these goods through usage experiences after purchase. Unlike search goods, I find that online platforms struggle to profit from steering consumers’ learning about experience goods. This suggests that experience goods may require less regulatory scrutiny than search goods with respect to platforms’ (potential) influence over consumers’ learning.

My findings also relate to a literature on the welfare effects of online personalization. Does consumer surplus increase or decrease when platforms exploit consumer microdata? In a field experiment, Donnelly, Kanodia, and Morozov (2024) show that the personalized search algorithm of a large online retailer delivers higher consumer surplus than does a uniform bestseller-based ranking—despite the former’s placing nonzero weight on products’ margins. In another field experiment, Dubé and Misra (2023) show that personalized pricing shrinks total consumer surplus despite benefitting most consumers. Complementary to these studies, I show that when consumers’ preferred products become unavailable, they might benefit when the online platform leverages consumer microdata in suggesting an alternative item—even if the platform exclusively attends to variable profits (as opposed to consumer surplus).

This study also contributes to the literature on incomplete information and consumer learning. This literature spans many environments, from school choice (see Allende, Gallego, and Neilson [2019]) to the demand for household appliances (see Newell and Siikamäki [2014]). Like me, a subset of this literature studies consumers’ learning in connection to online platforms. For instance, Allcott et al. (2025) show that Google’s dominance in web search owes partly to consumers’ imperfect information about competitors. Another subset of the consumer learning literature shares my focus on consumer packaged goods. Using data on consumers’ TV viewing habits and packaged good purchases, Akerberg (2003) show that consumers learn about the quality of yogurt brands from the brands’ TV advertisements.

To the consumer learning literature, I contribute the first empirical characterization of a firm’s optimal strategy to steer consumer learning about experience goods.⁵ The task proves unusually tractable in the context of curbside pickup for the following reasons. First, consumers’ preferences over groceries, along with their learning, can be distilled in a comparatively simple demand model.⁶ And second, general equilibrium effects are negligible so far as stockout substitutions are concerned. That is, the focal store’s optimal substitution policy is not influenced by those of its competitors.⁷

2.2.2 Curbside Grocery Pickup

In curbside pickup, consumers order groceries online and later pick them up from bricks-and-mortar supermarkets. This form of grocery shopping gained traction during the COVID-19 pandemic (Young 2023) and remains popular, with US sales exceeding \$3 billion in February 2024 alone (Brick Meets Click and Mercatus 2024).

To see how curbside pickup works, picture a consumer who wants to purchase two items: ice cream and apple sauce. She begins by visiting the store’s app or website. When she searches for a specific item—such as “ice cream”—she sees a list of relevant products, along with prices, images, and written descriptions. Once she identifies her preferred product—say, Häagen-Dazs vanilla ice cream—she adds it to her virtual “shopping cart.” Having repeated this process for apple sauce—choosing, say, Mott’s Cinnamon—she completes the order by indicating the time when she plans to pick up her groceries (for example, “Between 8 and 9 AM tomorrow morning”).

Once the consumer is ready to pick up her groceries, she drives to the store and parks in a designated “curbside pickup” area. A store worker then brings the groceries out to her car, where she pays for them. Importantly, the store maintains the same prices online as in-store;⁸ our

⁵I am only aware of one other study that empirically characterizes the optimal supply-side strategy to steer consumers’ learning about goods of any description—experience or otherwise. Compiani et al. (2024) consider how online platforms like Expedia should rank products in web searches, given that consumers possess incomplete knowledge of products’ *observable* characteristics. The assumption is that consumers will learn a product’s true utility once they have clicked on its web page, which describes the product’s observable characteristics.

⁶Because packaged foods are highly standardized and have just one usage case (namely, snacking), I adopt a “one-shot” model of learning in which consumers learn their true tastes for products after trying them just once. In addition, grocery shopping is characterized by many fast-paced but low-stakes decisions. I therefore approximate consumers’ behavior as being myopic, as opposed to forward-looking.

⁷As previously mentioned, it seems unlikely that consumers choose where to shop based on grocery stores’ handling of stockout substitutions. See Sections 2.4 and 2.7 for further discussion.

⁸If a consumer places a curbside order such that the sum of the ordered items falls below a specified threshold, she

consumer will pay the same price for a given item as if she had physically entered the store and purchased it there.

Stockout Substitutions.—The store is sometimes unable to supply an ordered item because it has gone out of stock. In that event, the store will offer a similar item to serve as a substitute. To illustrate how stockout substitutions proceed, let us revisit the (hypothetical) consumer who has ordered ice cream and apple sauce. Sometime *after* she has placed her order but *before* her intended pickup time, a store worker will collect the ordered items and set them aside, so that they can be brought out immediately upon her arrival. As he does so, the worker may discover that an ordered item has gone out of stock. Imagine, for instance, that our consumer’s preferred ice cream—namely, Häagen-Dazs vanilla—is unavailable. To ensure that she is not left without ice cream altogether, the worker will choose another product to serve as a substitute—say, Halo Top vanilla ice cream.⁹ Then, when our consumer arrives at the store,¹⁰ she will be presented with two options: 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. If she accepts the substitute, she will pay the substitute’s price (not that of the out-of-stock product).

2.2.3 Data

This study employs data from a regional supermarket chain that offers both in-person and online shopping. Concerning the latter, consumers can choose whether they prefer curbside pickup or home delivery.¹¹ As far as online shopping is concerned, my analysis concentrates on curbside

will pay a fixed fee for curbside pickup.

⁹The store’s website and mobile app allow the consumer to leave item-level instructions for the store. For instance, someone who is ordering strawberries might request “extra-ripe” berries. However, a consumer could also use this feature to request a specific substitute if her preferred product goes out of stock. Although I do not observe whether a consumer makes such a substitution request (or, for that matter, whether she leaves item-level instructions of any kind), the retailer has indicated that, during the time period of my data, consumers almost never left item-level instructions.

¹⁰Since September 2021, the store has also allowed consumers to accept or reject substitutes remotely. When an ordered item goes out of stock, the affected consumer receives a pop-notification or text to that effect, along with information about the substitute (such as the name and price). She can then accept or reject the substitute using her phone or computer. (If she fails to respond electronically, she will be offered the substitute at her car as in the old procedure.)

¹¹Home delivery resembles curbside pickup as far as orders are concerned. Unlike curbside pickup, however, home delivery does not require the shopper to travel to the store. Rather, her groceries are delivered directly to her home. For this convenience, she must pay a fee. (By contrast, curbside pickup is free for sufficiently large orders.)

pickup. This is because in home delivery, consumers select stockout substitutes themselves.¹²

The supermarket data consist of four distinct datasets. These include a “curbside stockout” dataset, which details stockout events in curbside pickup; a “scanner” dataset, which records consumers’ final purchases; a “demographics” dataset, which describes the characteristics of the consumer’s household; and the chain’s product catalog, which describes the products carried by the chain. I will now describe each of these datasets in turn.

Curbside Stockout Data.—The first dataset describes (attempted) stockout substitutions in curbside pickup from February 2020 to March 2022. Each observation includes the universal product code (UPC) of both the out-of-stock product and the substitute. I also see the price of the substitute,¹³ and whether it is accepted or rejected by the consumer.

Importantly, each observation in the data contains the loyalty ID number of the affected consumer,¹⁴ along with the date, time, and store location of pickup. This information enables me to identify the consumer’s past and future purchases within the scanner dataset (as described below).

To see what the curbside stockout data look like in practice, turn to Table 2A.1, which depicts the observations that would result from the stylized example in Section 2.2.2.

Scanner Data.—The second dataset records all purchases at the chain, both online and in-person, between April 2016 and July 2023. Each observation, which consists of a single transaction, includes the UPCs and prices of all the items that were purchased, along with the consumer’s loyalty ID. The data also record the date, time, and store location of the transaction. Finally, I observe the wholesale costs of each item.¹⁵ Hence, by taking the difference between purchase prices and wholesale costs, I can recover the “retail margin” of each item carried by the store.

Where curbside pickup is concerned, the scanner data only include a stockout substitute if it is accepted by the consumer. To illustrate, consider once more the (hypothetical) consumer from

¹²When an item ordered for home deliver becomes unavailable, the store phones the shopper to determine her preferred replacement.

¹³The price of the out-of-stock item is obtained from the scanner data (as I will explain shortly).

¹⁴Participation in the chain’s loyalty program is required to place curbside pickup orders.

¹⁵Prior to 2021, the retailer’s cost measure included some fixed costs in addition to the wholesale cost. There are six months during which both the old cost measure and the new one (i.e., wholesale cost alone) are recorded. For individual products, I observe these two cost measures moving roughly in tandem during this period.

the preceding subsection. Recall that she ordered Häagen-Dazs vanilla ice cream and Mott's apple sauce, but that the former went out of stock. Here, the substitute ice cream (Halo Top Vanilla) would only appear in the data if she accepted the swap. By contrast, the apple sauce would certainly appear in the scanner data, as it is the exact product that she had originally requested. See Table 2A.2 for a comparison of the data entries that would result from acceptance versus rejection.

Regarding stockout substitutions, the scanner data enable me to infer the price of the out-of-stock product. To do so, I search the scanner data for purchases of the relevant product on the same day, and at the same store, as the intended pickup—either before or after the stockout event. Provided that I locate at least one such observation, I approximate the out-of-stock product's price as being the mean of the observed purchase prices.¹⁶ If I do *not* observe any purchases of the product on the same day (and at the same store) as the substitution, I instead compute the mean purchase price on the day *before* the substitution.¹⁷ Failing that, I approximate the out-of-stock product's price by taking the average purchase price on the nearest date for which observations appear in the data. If I have still not obtained the out-of-stock product's price, I 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.

Demographic Data.—The third dataset reports demographic information about consumers' households. These data, which are gathered by a third-party consulting firm, report the household's income; the size of the household; the age of the oldest resident male (or, absent male residents, the age of the oldest female); and whether the household owns or rents its home. See Chapter 2A for details.

¹⁶Based on conversations with chain pricing personnel, this procedure should yield a very close approximation of the true price. (The true price may differ slightly from the imputed one due to coupons or other consumer-specific discounts.)

¹⁷Whereas it is possible for a consumer to place an order the day before pickup, it is impossible for her to place the order the day after! Thus, the average purchase price on the day before the pickup is likely more representative of the price that she expected to pay than is the average purchase price on the day after.

Product Catalog.—The fourth dataset describes the products sold by the chain. For each product, the catalog lists the universal product code (UPC) and the brand, as well as the location within the chain’s product taxonomy. I also observe a string description of the product that characterizes its observable characteristics. To illustrate, here is a string description of a package of apple sauce cups:

“MOTTS APPLESAUCE CINNAMON 18/4 OZ”

This description indicates that the apple sauce is sold under the Mott’s brand, that it is cinnamon-flavored, and that the package contains eighteen cups of apple sauce (each measuring 4 oz). I employ so-called “regular expressions” to extract this information. Sometimes, however, a product’s string description omits one or more characteristics of interest. In such cases, I consult either the manufacturer’s website or that of a retailer that carries the product.¹⁸

2.2.4 Summary Statistics

I observe billions of transactions at the relevant supermarket chain. About forty percent of these transactions involve a consumer who participates in the chain’s loyalty program. Because my analysis focuses primarily on curbside pickup (where enrollment in the chain’s loyalty program is required), I will hereafter refer to these individuals as simply “consumers.” The total number of consumers exceeds five million.

Now consider consumers’ choice of shopping channel. Roughly one in five consumers places at least one order for curbside pickup, so the chain fulfills millions of orders for curbside pickup. Regarding stockout substitutions, of the hundred million-plus items ordered for curbside pickup during the two years when I observe stockout substitutions, approximately twelve million undergo stockout substitutions. Because curbside orders contain an average of twenty-three items during this time period, consumers suffer an average of two stockout substitutions per order. Stockout substitutes are accepted in 87.4% of cases.

¹⁸This step proves particularly important for the product category of ice cream, where most products’ characteristics must be constructed completely by hand. Consequently, my analysis of ice cream focuses on the top 97-selling “mainstream” ice cream products and the top 97-selling “super-premium” ice cream products.

In line with the New Empirical IO literature (e.g., Nevo [2001] and Backus, Conlon, and Sinkinson [2021]), I do not model consumers' demand for all packaged foods. Instead, I focus on the following product categories: apple sauce cups, flavored milk, frozen french fries, and ice cream. These product categories were chosen for three reasons. First, each category consists of *experience goods* (as defined in Section 2.1). Second, each category contains many stockout substitutions. This helps me identify the relationship between stockout substitutions and consumers' learning. Finally, the brands within these product categories afford meaningfully different retail margins. If this were not so, the store could not profit from steering consumers' learning towards particular brands through stockout substitutions (which is the key strategy assessed in this paper).

Table 2.1 presents summary statistics for the four product categories studied. The number of households suffering at least one stockout substitution ranges from roughly 7,500 (apple sauce cups) to nearly 57,000 (ice cream). Because some consumers suffer multiple substitutions, the overall number of substitutions per category is somewhat larger. It is the largest for ice cream (about 91,000) and the smallest for apple sauce cups (approximately 9,000). As for consumers' willingness to accept stockout substitutes, the probability of acceptance ranges from 83.9% (ice cream) to 90.6% (frozen french fries).

Table 2.1: Summary Statistics by Product Category

Statistic	Apple sauce cups	Flavored milk	Frozen french fries	Ice cream
<i>Panel A. Overview</i>				
No. of households with 1+ substitutions	7508	13,014	30,588	56,762
Total substitutions	9001	17,484	39,397	91,139
Prob. accept (%)	87.2	88.0	90.6	83.9
<i>Panel B. Per consumer with 1+ substitutions</i>				
No. of shopping trips	20.4	40.7	21.7	33.9
... of which curbside pickup	6.2	9.5	5.7	7.1
... with 1+ substitutions	1.2	1.3	1.3	1.4

Notes: Unless otherwise indicated, estimates are reported as means or totals.

Turning to the panel dimension of the data, Panel B characterizes the purchases of individual

consumers who suffer at least one stockout substitution. Depending on the product category, I observe an average of twenty to forty-one shopping trips per consumer. Six to ten of these shopping trips are curbside pickup (as opposed to in-store shopping or home delivery).

Chapter 2A presents additional summary statistics on state dependence in consumers' purchases, as well as the frequency and duration of stockout events.

2.3 Descriptive Evidence

In this section, I present descriptive evidence of the trade-offs faced by the store as it selects stockout substitutes. I begin by characterizing the circumstances under which consumers are willing to accept stockout substitutes. Next, I present model-free evidence that stockout substitutions influence consumers' learning. Finally, I examine the relationship between products' observable characteristics and their retail margins.

2.3.1 When Do Consumers Accept or Reject Stockout Substitutes?

A consumer's decision to accept or reject a stockout substitute may impact the store's future profits as well as its present ones. Consider first the case where the consumer rejects. Here, the store earns zero margins on the present shopping trip from the (unsuccessful) substitution. As for future profits, the consumer's rejection of the substitute suggests that she may be unhappy with the store's handling of the stockout. This dissatisfaction might, in principle, dent the store's future earnings if the consumer reduces her future patronage. However, it seems unlikely that the store's handling of a single stockout would affect where the consumer shops in the future.¹⁹ Now turn to the case where the substitute is accepted. Regarding the present shopping trip, the store immediately earns the retail margin associated with the substitute. As to future profits,²⁰ the consumer will learn whether she likes or dislikes the substitute product—provided that she has not already purchased it previously, in which case she will already know whether the product is to her taste. This learning

¹⁹Other factors, like the convenience of the store's location or the competitiveness of its prices, probably loom larger in the consumer's choice of grocery store. There is also a cost (in time and effort) associated with enrolling in a rival chain's pickup program and becoming proficient in its interface.

²⁰The consumer might feel unhappy with the store's handling of the stockout despite accepting the substitute (e.g., if she accepts due to the inconvenience of procuring a replacement elsewhere). Although it appears unlikely that a single unsatisfactory substitution could dent the consumer's future patronage of the store (see above), it remains theoretically possible that the store's future profits are related to the consumer's utility from an accepted substitute.

may affect her subsequent purchases (and ultimately the store's future profits).

The task of this subsection is to characterize when consumers are willing to accept stockout substitutes. I focus on two key determinants of acceptance: the substitute's similarity to the out-of-stock product, and the substitute's similarity to products that the consumer has purchased on previous shopping trips. Intuitively, the probability of acceptance should be increasing in the similarity of the substitute's observable characteristics—like its brand or size—to those of the out-of-stock product and those of products that the consumer has purchased on previous shopping trips. Regarding the out-of-stock product, I construct a set of indicator 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 the consumer's past purchases, I include a set of indicator variables for the substitute's sharing a given characteristic k with *any* of the products that the consumer has bought before. Formally, let $\text{ever}_{ik} = 1$ if consumer i is offered a substitute that shares characteristic k with any of the products that she has purchased on past shopping trips, and $\text{ever}_{ik} = 0$ otherwise.

Observable characteristics aside, the prices of the substitute and out-of-stock product should also be informative of acceptance. In particular, consumers may be reluctant to accept stockout substitutes that are perceptibly pricier than the products they had originally ordered. One of my empirical specifications therefore allows 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 ($p_{i,\text{OOS}}$). Another specification incorporates the prices of the substitute and out-of-stock product in a more flexible manner, including each as a separate explanatory variable.

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

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

where

$$a_i^* = \sum_{k=1}^K (\gamma_k \text{same}_{ik} + \zeta_k \text{ever}_{ik}) + w(p_{i,\text{sub}}, p_{i,\text{OOS}}) + u_i. \quad (2.1)$$

and the idiosyncratic error v_i is distributed i.i.d. standard normal. As previously mentioned, I explore several price controls; $w(p_{i,\text{sub}}, p_{i,\text{OOS}}) \in \{0, (p_{i,\text{sub}} - p_{i,\text{OOS}})\eta, \chi p_{i,\text{sub}} + \psi p_{i,\text{OOS}}\}$.

Table 2.2 reports the average marginal effects of the variables in Equation (2.1). For three of the four observable characteristics studied—namely, brand, flavor, the number of mix-ins, and quantity—the probability of acceptance is significantly greater when the substitute shares the relevant characteristic with either (a) the out-of-stock product or (b) at least one previous purchase. As for the characteristic of quantity, the substitute’s similarity in size to the out-of-stock product—or the consumer’s past purchases—is uninformative of the probability of acceptance. This likely reflects the limitations of this reduced-form exercise, not indifference by the consumer as to the substitute’s quantity.²¹

The results for the other product categories studied—which include apple sauce cups, flavored milk, and frozen french fries—are qualitatively similar. In particular, consumers are more likely to accept substitutes whose brands they have previously purchased. (See Chapter 2B for details.)

2.3.2 Stockout Substitutions and Consumers’ Learning about Brands

This subsection supplies model-free evidence that stockout substitutions influence consumers’ learning. Throughout, I adopt the simplifying assumption that consumers learn about their preferences/tastes for products’ observable characteristics, as opposed to their tastes for individual products. This simplifying assumption aligns the descriptive analysis here with the demand model estimated in Sections 2.5 and 2.6.

How might consumers learn about their preferences for products’ observable characteristics? Consider a (hypothetical) consumer who always orders Häagen-Dazs vanilla ice cream. On one occasion, however, her preferred ice cream goes out of stock. In its place, she is offered Halo Top vanilla ice cream as a substitute. If she accepts, she will learn about her tastes for a new brand: Halo Top. However, she won’t necessarily learn anything about her preferences for flavor because the substitute shares the same flavor (namely, vanilla) as her preferred ice cream. What if, instead,

²¹For instance, conditional on the substitute’s differing from the out-of-stock product with respect to a given characteristic, I do not quantify the *degree* of dissimilarity. To more accurately capture the consumer’s underlying choice problem, it helps to estimate a structural model (as I do in Sections 2.5 and 2.6).

Table 2.2: Determinants of Acceptance: Average Marginal Effects from Probit Regressions

	(1)	(2)	(3)
Brand			
Sub shares OOS product's brand	0.015*** [0.004]	0.016*** [0.004]	0.015*** [0.004]
Ever purchased sub's brand before	0.011** [0.003]	0.011** [0.003]	0.011** [0.003]
Flavor(s)			
Sub shares OOS product's flavor(s)	0.041*** [0.003]	0.041*** [0.003]	0.038*** [0.003]
Ever purchased sub's flavor(s) before	0.041*** [0.003]	0.041*** [0.003]	0.041*** [0.003]
No. of mixins			
Sub shares OOS product's no. of mixins ^a	0.035*** [0.003]	0.035*** [0.003]	0.035*** [0.003]
Ever purchased sub's no. of mixins before ^a	0.017*** [0.003]	0.017*** [0.003]	0.017*** [0.003]
Quantity (oz.)			
Sub shares OOS product's quantity (oz.) ^b	-0.011 [0.013]	-0.019 [0.014]	-0.032* [0.014]
Ever purchased sub's quantity (oz.) before ^b	0.000 [0.004]	0.000 [0.004]	-0.002 [0.004]
Sub's price – OOS product's price		-0.009*** [0.002]	
Sub's price			0.004 [0.002]
OOS product's price			-0.016*** [0.002]
Observations	55,270	55,270	55,270
Pseudo R^2	0.024	0.025	0.026

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 brackets. As some consumers suffer multiple stockouts, the standard errors are clustered at the consumer level.

* Significant at the 10 percent level.

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

^a The number of non-ice cream elements mixed into the ice cream (like caramel sauce or chunks of cookie dough).

^b Quantity is discretized as follows: 0 to 5 oz; 5 to 9 oz; 9 to 13 oz; 13 to 25 oz; 13 to 25 oz; 25 to 50 oz; 50 to 100 oz; more than 100 oz.

the store had offered Halo Top *cookies & cream* ice cream as a substitute? Then, if the consumer accepted, she would learn her tastes for a new flavor (i.e., cookies & cream) as well as a new brand (i.e., Halo Top). Notice that the amount of learning will vary by characteristic so far as the consumer holds more accurate prior beliefs about her tastes for some characteristics than others. Intuitively, ice cream buyers are more likely to learn about their preferences for brands or flavors than they are to learn about, say, their preferences for the quantity of ice cream (which will mostly

depend on the size of their freezers and the frequency with which they eat ice cream).

The task of this subsection is, therefore, to determine whether stockout substitutions cause consumers to learn about their preferences for observable product characteristics. I start by identifying stockouts where consumers will learn about their tastes for one of the substitute's observable characteristics if they accept. For example, if I were interested in the characteristic of brand, I would find stockout substitutions where consumers have never purchased the substitute's brand before. Next, I tally how often these consumers' future purchases share the substitute's version of the relevant characteristic. Intuitively, the following empirical pattern should emerge if stockout substitutions affect consumers' learning. Of the consumers who accept the offered substitute—thereby learning their true tastes for its version of the characteristic—some will discover that they like the substitute's version more than they had expected. A disproportionate share of these consumers' future purchases should thus feature the substitute's version of the characteristic, compared to the counterfactual where they never purchased the substitute (and, in consequence, did not learn about the substitute's version of the characteristic). But how can I identify this counterfactual? That's to say, what would these consumers' purchases have looked like if they had never suffered the stockout substitution and, as a result, never learned about the substitute? To approximate this counterfactual, I identify “control consumers” who order the same products as the focal consumers. Unlike the focal consumers, however, these control consumers successfully pick up their preferred products before they go out of stock and, in consequence, do not have the chance to learn about the substitute.

In spelling out my empirical strategy, it helps to focus on just one observable characteristic. I will thus concentrate initially on the characteristic of brand and then explain how my strategy generalizes to other characteristics. With this in mind, consider once more the hypothetical consumer who always buys Häagen-Dazs vanilla ice cream. Now assume that she is offered Halo Top vanilla ice cream as a stockout substitute. If she accepts, she will consume Halo Top-branded ice cream for the first time, thereby learning whether she likes or dislikes the brand. Now assume that our consumer does accept the substitution and that, on her subsequent shopping trips, she begins to purchase Halo Top-branded ice cream. This shift in brands purchased, from Häagen-Dazs to Halo

Top, reflects two factors. One is the consumer's learning about the Halo Top brand. The other is confounding changes in the market environment; perhaps Halo Top has rolled out a new marketing campaign at the same time as the stockout. Or, alternatively, our consumer might be tiring of the taste of Häagen-Dazs ice cream so that, even without the stockout substitution, she would still have switched to a different brand in the near future—like Halo Top.

To isolate the influence of the stockout substitution, I identify a “control consumer” who, like the focal consumer, has never purchased any Halo Top–branded ice cream before. Additionally, the control consumer has ordered the same Häagen-Dazs vanilla ice cream as the focal consumer, from the same store, and on the same day. Unlike the focal consumer, however, the control consumer arrives at the store just before the Häagen-Dazs vanilla ice cream go out of stock. Hence, he does not suffer a stockout substitution and, in consequence, cannot learn his true tastes for the Halo Top brand on the present shopping trip. Any future purchases of the Halo Top brand will, therefore, stem solely from confounding changes in the purchase environment—*not* learning. This enables me to “difference out” confounding changes in the purchase environment: whereas the focal consumer's future purchases reflect both (a) her learning about Halo Top (due to the substitution) and (b) confounding changes in the environment, the control consumer's future purchases reflect only (b). Thus, if the focal consumer proceeds to purchase Halo Top ice cream more often than the control consumer does, the disparity likely stems from the former's learning.

Having sketched the intuition of my strategy, I now spell out the specifics. As suggested by the foregoing thought experiment, I start by identifying stockout substitutions where the consumer has never purchased the substitute's brand before. For each such substitution, I identify all successful curbside pickups of the focal consumer's preferred product *before* it went out of stock.²² Of these successful pickups, I drop those where the purchaser has bought the substitute's brand before. Among the remaining consumers, the “control consumer” is defined as the *last* one to successfully

²²In Chapter 2B, I repeat the same procedure for the *first* consumer to pick up after the stockout ends. However, intuition suggests that stockouts may cause endogenous price changes where the store hikes the prices of products that recently went out of stock. By contrast, purchases *before* the stockout are insulated from such endogenous price adjustments. At all events, the results are quantitatively unchanged by this alternative method of selecting the “control consumer;” see Table 2B.6.

pick up the ordered product before it goes out of stock.²³ Under the null hypothesis that stockout substitutions do not result in consumer learning, the control consumer's future purchases should belong to the substitute's brand just as often as the focal consumer's future purchases do.

The foregoing procedure can be readily adapted to study characteristics other than brand. To do so, I first identify stockout substitutions where the substitute's version of the relevant characteristic is one that the consumer has never purchased before (so that she will learn about the substitute's version if she accepts). Then I single out a "control consumer" from among the population of consumers who have ordered the same product as the focal consumer, and who, like the focal consumer, have never purchased a product with the substitute's version of the relevant characteristic. As with brand, I focus on the last such consumer to successfully pick up before the stockout event.

Table 2.3 presents the results of this descriptive exercise. The results bear an "intent-to-treat" interpretation because I do not distinguish between observations where the substitute is accepted (in which case the consumer learns about the substitute) and observations where the substitute is rejected (in which case the consumer does *not* learn). This is because acceptance is endogenous; consumers who expect to like the substitute's observable characteristics are more likely to accept than are consumers who expect to *dislike* its characteristics.

With this in mind, Table 2.3 is organized as follows. Each panel focuses on a specific observable characteristic. The upper row of the panel corresponds to the "focal consumers," who are offered stockout substitutes with a version of the characteristic that the consumers have never tried. The lower row of the panel, meanwhile, pertains to the "control consumers," who successfully pick up their preferred products (and do not, therefore, learn about the relevant characteristic).

Focus first on the number of purchases made by each type of consumer. For all characteristics studied,²⁴ the "focal consumers" (who suffer stockout substitutions) are observed making more purchases before the stockout than are their control counterparts. The same disparity emerges for

²³To ensure that the purchase environment is comparable to that experienced by the focal consumer, I drop any observations where the "control consumer" picks up the focal consumer's preferred product on a different week than the control consumer. (The store typically updates prices and discounts once per week, on Sundays.)

²⁴I omit results for quantity because there are only seven observations where (i) a "focal consumer" is offered a substitute with an unfamiliar quantity and (ii) a "control consumer" can be identified.

Table 2.3: Model-Free Evidence That Stockout Substitutions Affect Consumers' Learning

Consumer's "treatment"	No. of purchases				Pct. of future purchases with sub's version of characteristic	
	Before stockout		After stockout			
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
<i>Panel A. Characteristic of brand (292 obs.)</i>						
Suffer substitution (focal group)	24.6 [0.1]	35.0 [0.2]	21.3 [0.1]	24.5 [0.1]	6.6 [0.1]	16.9 [0.1]
Successful pickup (control group)	30.1 [0.1]	39.1 [0.2]	26.0 [0.1]	29.7 [0.2]	3.0 [0.0]	7.9 [0.1]
<i>Panel B. Characteristic of flavor (200 obs.)</i>						
Suffer substitution (focal group)	29.8 [0.2]	40.9 [0.3]	19.7 [0.1]	21.8 [0.1]	2.9 [0.0]	10.0 [0.2]
Successful pickup (control group)	38.3 [0.2]	50.3 [0.4]	25.8 [0.1]	27.3 [0.2]	2.0 [0.0]	8.3 [0.2]
<i>Panel C. Characteristic of no. of mix-ins (76 obs.)^a</i>						
Suffer substitution (focal group)	17.0 [0.3]	24.2 [0.5]	17.0 [0.2]	16.1 [0.2]	6.8 [0.2]	11.7 [0.2]
Successful pickup (control group)	24.9 [0.5]	35.8 [0.8]	23.8 [0.3]	23.5 [0.3]	5.7 [0.2]	13.8 [0.5]

Notes: This table presents descriptive evidence that stockout substitutions influence consumers' learning about their preferences for observable product characteristics. Each observation consists of a stockout substitution where the substitute does not share the relevant characteristic with any of the consumer's past purchases (so that, if she accepts, she will learn her tastes for the substitute's version of the characteristic). To capture confounding influences in the environment, results are also reported for "control consumers" who resemble the focal consumers in most respects, but do not suffer stockout (and thus do not learn about the substitute). Each "control consumer" is drawn from the population of consumers who have not yet purchased any products that share the relevant characteristic with the substitute. The control consumer has also ordered the same product as the focal consumer, from the same store, and on the same week. Unlike the focal consumer, however, she successfully picks up her preferred product before it goes out of stock. From the pool of consumers satisfying the foregoing criteria, I select the last one to have successfully picked up before the stockout event. Bootstrapped standard errors are enclosed in brackets. The characteristic of quantity is omitted from this table because there are only seven observations.

^a The number of non-ice cream elements mixed into the ice cream (like caramel sauce or chunks of cookie dough).

the number of purchases *after* the stockout. These discrepancies point to compositional differences between the focal and control consumers; the results of this descriptive exercise should be taken as suggestive, not definitive.

Consider now the percentage of future purchases that share the substitute's version of the characteristic.²⁵ Concerning all three characteristics, this percentage is larger for the "focal" consumers than for their "control" counterparts. That is consistent with a subset of the "focal"

²⁵The corresponding percentage is omitted for past shopping trips as, by construction, none of these consumers have ever purchased a product that shares the relevant characteristic with the substitute.

consumers’ discovering that they like their substitute’s version of the relevant characteristic and, in consequence, purchasing products with that version of the characteristic again in the future. Notice that the disparity between the “focal” and “control” consumers is larger for the characteristic of brand (3.6 percentage points) than for the characteristics of flavor and mix-ins (0.9 and 1.1 percentage points, respectively). This suggests that consumers learn more about their preferences for brands than about their preferences for flavor or the number of mix-ins.

Other Product Categories.—The results for apple sauce cups, flavored milk, and frozen french fries prove qualitatively similar to those for ice cream. Tables 2B.2 to 2B.4 report that consumers who suffer stockouts within these categories proceed to purchase the substitute’s brand 1.9 to 7 percentage points more often than their control counterparts (depending on the product category). And as with ice cream, buyers in these categories generally seem to learn less about non-brand observable characteristics than about brand.

Robustness Checks.—Other mechanisms besides learning could explain these results. One such mechanism is the “buy it again” feature of the store’s app and website, which enables consumers to perform repeat purchases with a single click. Importantly, the “buy-it-again” list includes accepted stockout substitutes. This raises the following question. Do consumers purchase stockout substitutes on subsequent shopping trips because it is convenient, or because they have learned about the substitutes? To adjudicate between these explanations, I modify the foregoing descriptive exercise as follows. Rather than comparing focal and control consumers with respect to all subsequent purchase—both online and offline—I instead focus solely on in-store purchases, which should be unaffected by the “buy-it-again” list. The results (presented in Table 2B.5) prove reassuring. Although sample sizes (and statistical power) shrink dramatically, consumers in two of the four product categories still purchase their substitute’s brand much more frequently on their subsequent in-store shopping trips than do their control counterparts.²⁶

²⁶Concerning apple sauce cups, the focal consumers purchase the substitute’s brand 7.1 percentage points more often than do their control counterparts. The corresponding disparity is 4.3 percentage points for frozen french fries. As for the remaining categories, the focal iced cream buyers proceed to the substitute’s brand slightly more often than do their control counterparts (0.5 percentage points), while the focal flavored milk buyers purchase the substitute’s brand 0.1 percentage points less often.

2.3.3 What Determines Products' Retail Margins?

In this subsection, I study the determinants of products' *retail margins*—that is, the differences between their retail prices and wholesale costs.²⁷ How do observable characteristics like brand, size, or flavor influence a product's profitability?

To provide insight, I estimate the linear regressions of the form

$$p_{jts} - wc_{jts} = x_j\gamma + v_{jts}, \quad (2.2)$$

where p_{jts} and wc_{jts} respectively denote the price and wholesale cost of good j at time t in store s , while x_j denotes the observable characteristics. As the data span more than seven years, I adjust for inflation by converting both prices and margin costs to 2021 dollars.²⁸

Figure 2.1 reports the results. Unlike the foregoing descriptive analysis, here I distinguish between “mainstream” and “super-premium” ice creams. This is because my structural analysis in Sections 2.6 and 2.7 focuses on the latter segment (which is economically important in its own right; see Sullivan [2020]). For a given ice cream segment, the corresponding panel plots the estimated coefficients on key observable characteristics. For discrete characteristics with many values, such as brand or flavor, I assign the top-selling value as the base level and then report the coefficients on the three next-most-popular values.

Consider first the coefficients on the brand dummies. These bear the following interpretation: how do the retail margins of a product sold under the indicated brand differ from those of an otherwise-identical product sold under the omitted brand (which is the top-selling brand in the category)? The estimates suggest that products' brands are, indeed, a key determinant of their respective retail margins. Within both the mainstream and super-premium ice cream segments, the margins of the most profitable brand exceed those of the least profitable brand by fifty cents. Now turn to products' non-brand characteristics. It is immediately evident that the characteristic of flavor is unimportant for margins. Margins do appear to increase with quantity, however—especially where super-premium ice cream is concerned.

²⁷As mentioned previously, the store reported a hybrid cost measure (wholesale cost + some fixed costs) until 2021. For simplicity, these descriptives focus on the time period after 2021, when wholesale costs are directly observed.

²⁸To reduce the influence of brief fluctuations in the CPI, I normalize values using the six-month smoothed CPI.

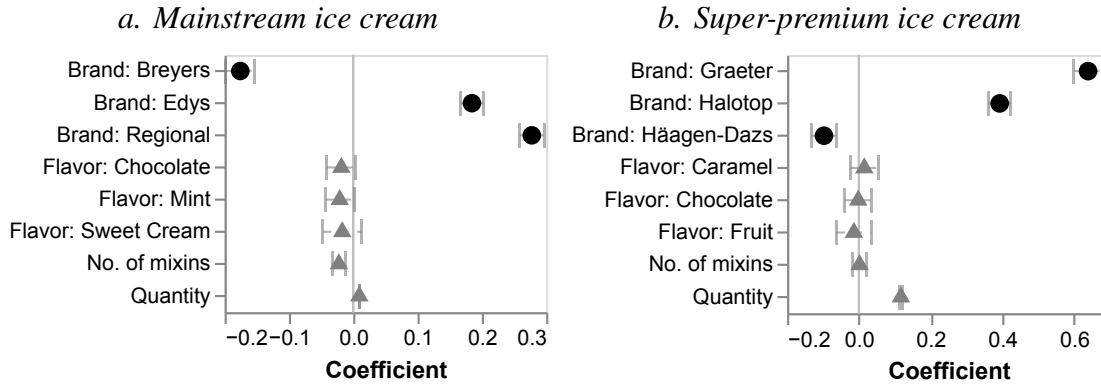


Figure 2.1: Determinants of Retail Margins

Notes: This figure plots estimates of the coefficients (γ) on products' observable characteristics using the specification in Equation (2.2). The horizontal bars provide 95% confidence intervals.

Chapter 2B presents similar descriptive regressions for the other product categories studied. Within each category, the characteristic of brand proves to be one of the key determinants of margins.

2.4 Conceptual Model

The preceding section highlighted three empirical patterns in the data. First, consumers prefer stockout substitutes whose observable characteristics resemble those of the out-of-stock product (or, at a minimum, consumers' past purchases). Second, stockout substitutions influence consumers' learning about their preferences for observable characteristics—particularly in relation to the characteristic of brand. Last, the characteristic of brand is among the most important determinants of products' retail margins.

I will now distill these empirical patterns in a stylized conceptual model. The task is to formalize the store's strategic problem as it chooses stockout substitutes and to trace the effect on consumer welfare. This conceptual model will inform the empirical model that I take to the data in Sections 2.5 through 2.7.

2.4.1 The Store's Problem

Consider a store that offers three goods for curbside pickup: A , A' , and B . Let p_j and mc_j denote the price and marginal cost, respectively, of good $j \in \{A, A', B\}$. Assume that good B

affords a higher retail margin than do goods A and A' :

$$p_B - mc_B > \max\{p_A - mc_A, p_{A'} - mc_{A'}\}.$$

The store serves a consumer who makes two shopping trips, indexed by $t \in \{1, 2\}$. On each trip, she either (i) purchases one of the three “inside goods” sold by the store; or (ii) chooses the “outside option” of no purchase, indexed by $j = 0$. She values the “inside goods” $j \in \{A, A', B\}$ at $v_j \in \mathbb{R}$ and the “outside option” at zero. The conditional indirect utility of good j is the difference between its valuation and its price:

$$u_j = \begin{cases} v_j - p_j & \text{if } j \in \{A, A', B\} \\ 0 & \text{if } j = 0. \end{cases}$$

Our consumer has imperfect information about her preferences among the three goods: although she knows the valuations of goods A and A' from prior purchase experiences, she does not know her valuation of good B . However, she expects to like good B less than good A' which, in turn, she knows to be less preferable than good A . That is,

$$E[u_B] - p_B =: u_B^E < u_{A'} < u_A. \quad (2.3)$$

In addition, the consumer prefers good A' to the “outside option” of no purchase:

$$u_{A'} > 0. \quad (2.4)$$

Turning to the store, I assume that Equations (2.3) and (2.4) are common knowledge. Only the consumer, however, knows exactly how much utility she expects good B to afford (that is, u_B^E). And neither the store nor the consumer knows the true utility u_B of good B . However, the store has microdata θ on the consumer’s past purchases and household demographics that help it forecast both of these quantities (i.e., u_B^E and u_B).²⁹

Suppose that our consumer orders good A on trip 1, only for the good to go out of stock before pickup. Should the store offer A' or B as a substitute? Its decision depends on several criteria. Two

²⁹Formally, let $H(\cdot \mid \cdot)$ and $H(\cdot)$ denote the conditional and unconditional entropy operators, respectively. Then $H(u_B^E \mid \theta) - H(u_B^E) \in \mathbb{R}_{>0}$ and $H(u_B \mid \theta) - H(u_B) \in \mathbb{R}_{>0}$.

of these criteria concern the first shopping trip. These include (a) the potential substitute's retail margin and (b) the probability of acceptance. Regarding (b), our consumer will accept a substitute $s \in \{A', B\}$ if, and only if, she expects to prefer it to the “outside option” of no purchase (that is, “good 0”).³⁰

The store's choice of substitute also affects its future profits. If the store offers good B as the substitute, the consumer may discover that she likes the good more than she had expected and, in consequence, purchase it on her second trip. That would boost the store's future profits because good B affords greater retail margins than do goods A or A' .

The store's optimal choice of substitute can be formalized as follows. Let δ denote the discount factor for profits on trip 2. Then, given that the consumer originally ordered good A , the optimal substitute should maximize the present-discounted sum of profits on trips 1 and 2;

$$s_A^*(p, mc; \theta) := \arg \max_{s \in \{A', B\}} \{E[\Pi_1 \mid \text{offer } s; \theta] + \delta E[\Pi_2 \mid \text{offer } s; \theta]\}. \quad (2.5)$$

Focus first on trip 1. If the store offers good A' as a substitute, the consumer will certainly accept. Thus, $E[\Pi_1 \mid \text{offer } A'; \theta] = p_A - mc_A$. Concerning good B , by contrast, the store is unsure of acceptance. As a result, $E[\Pi_1 \mid \text{offer } B; \theta] = \Pr[u_B^E > 0 \mid \theta](p_B - mc_B)$.

Turn now to trip 2. Supposing that the store offered A' as a substitute, the consumer cannot have learned anything from the stockout substitution (as she already knew her taste for A'). She will, therefore, order good A on trip 2 just as she did on trip 1. In other words, $E[\Pi_2 \mid \text{accept } A'; \theta] = p_A - mc_A$. What if, however, the store offered good B ? On the one hand, if the consumer rejected it, she will not have learned anything and will, therefore, order good A again on trip 2. Consequently, $E[\Pi_2 \mid \text{reject } B; \theta] = p_A - mc_A$. On the other hand, if the consumer accepted good B , she might have found it preferable to good A . That's to say,

$$E[\Pi_2 \mid \text{offer } B; \theta] = \Pr[u_B > u_A \mid \text{accept } B; \theta](p_B - mc_B) + \Pr[u_B \leq u_A \mid \text{accept } B; \theta](p_A - mc_A).$$

³⁰Here, I implicitly assume that the consumer is myopic, meaning that she overlooks the (expected) value of learning her true tastes for good B . In Section 2.5.1, I explain why this assumption is likely to provide a close approximation of consumers' true behavior in the context of curbside grocery pickup.

Notice that expected trip-2 profits conditional on the store's offering good B strictly exceed those conditional on offering A' .

2.4.2 Substitution Policies and Consumer Welfare

I have hitherto assumed that the store minds the relationship between the stockout substitution and our consumer's learning. I have also supposed that the store exploits the microdata θ to more accurately forecast our consumer's prior-expected and true tastes for brand B (u_B^E and u_B , respectively). In point of fact, accounting for these factors might increase the store's administrative and computational costs.³¹ How, therefore, would our consumer's welfare change if the store determined to disregard her (potential) learning or the consumer microdata?

Suppose first that the store does employ the microdata, but does *not* account for the consumer's (potential) learning. It will then choose a substitute according to the rule $s_A^1(p, mc; \theta) := \arg \max_{s \in \{A', B\}} E[\Pi_1 \mid \text{offer } s; \theta]$. Thus, the store will fail to offer the optimum substitute (as given by s_A^*) if, and only if,

$$E[\Pi_1 \mid \text{offer } B; \theta] < E[\Pi_1 \mid \text{offer } A; \theta] < E[\Pi_1 \mid \text{offer } B; \theta] + \delta E[\Pi_2 \mid \text{offer } B; \theta]. \quad (2.6)$$

When these inequalities hold, which substitution policy is better for the consumer? The answer depends on whether her present-discounted value of expected surplus is larger from the offer of good B or A' . If she is offered the latter, she will accept and enjoy the following present-discounted surplus:

$$E[CS \mid \text{offered } A'] = u_{A'} + \delta(u_A). \quad (2.7)$$

Now consider good B . Given our assumption that the consumer is myopic, she will accept good B if, and only if, $u_B^E \geq 0$. Hence, her present-discounted value of expected surplus is

$$E[CS \mid \text{offered } B] = \begin{cases} u_B^E + \delta \max\{E[u_B \mid u_B > u_A], u_A\} & \text{if } u_B^E \geq 0 \\ 0 + \delta \cdot u_A & \text{if } u_B^E < 0 \end{cases} \quad (2.8)$$

A comparison of Equations (2.7) and (2.8) reveals that the consumer benefits from the store's disregard of her learning whenever she is so pessimistic about good B that she would reject it (i.e.,

³¹The cost associated with forecasting consumer learning and exploiting consumer microdata are probably diminishing with time. Thus, the store might find it more profitable to mind these factors in the future than at present.

$u_B^E < 0$). The welfare effect is ambiguous, however, when the expected utility of good B suffices for her to accept (i.e., $u_B^E \geq 0$). On the one hand, the consumer's expected surplus increases on *trip 1* when the store overlooks her learning (and thus offers A'). This is because $u_{A'} > u_B^E$. On the other hand, her present-discounted value of expected *trip-2* surplus strictly decreases. The reason is that she no longer learns her true taste for good B and will, as a consequence, certainly order good A on trip 2 (even if good B would, in point of fact, afford greater utility).

Consider finally the effect of the microdata on our consumer's welfare. If the store disregards both the microdata θ and the opportunity to steer the consumer's learning, it will select a substitute according to the rule $\tilde{s}_A^1(p, mc) := \arg \max_{s \in \{A', B\}} E[\Pi_1 \mid \text{offer } s]$. Here, expected trip-1 profits conditional on offering good A' remain unchanged because, with or without the microdata, the store is confident that the consumer will accept. In other words, $E[\Pi_1 \mid \text{offer } A'] = E[\Pi_1 \mid \text{offer } A' \mid \theta] = p_{A'} - mc_{A'}$. By contrast, expected trip-1 profits conditional on offering good B' may differ insofar as $\Pr[u_B^E > 0] \neq \Pr[u_B^E > 0; \theta]$. Thus, the store may be more or less likely to offer good B as a substitute depending on the structure of its beliefs, or without, the microdata. As far as the consumer is concerned, the welfare effects of the store's use or disuse of the microdata depend once more on the relative value of the consumer's learning her true taste for good B compared to the greater present-trip utility deliver by good A' .

2.5 Empirical Model and Estimation

In this section, I build a learning model of differentiated products demand. Then I describe the estimation procedure.

2.5.1 The Model

Consider discrete choice among J_t goods/products at “time” t ,³² indexed by $j \in \mathcal{J}_t := \{1, \dots, J_t\}$. These goods are sold under differentiated brands b (such as “Häagen-Dazs” or “Halo Top”). Let $B(j)$ denote the brand of good j .³³

The utility that consumer i derives from good j depends partly on her liking (or “taste”) for its

³²In point of fact, t is defined as the combination of a specific store location and time. For economy of exposition, I focus on the temporal dimension of the index.

³³Formally, the function $B : \bigcup_{t \in \mathcal{T}} \mathcal{J}_t \rightarrow \mathcal{B}$ maps from each good sold to its brand. (Here $\mathcal{T} := \{1, \dots, T\}$ denotes the set of all time periods, while $\mathcal{B}$ denotes the set of all brands.)

brand. This is measured by the scalar $v_{iB(j)} \in \mathbb{R}$. Utility also depends on the good’s non-brand observable characteristics (x_j) and its price (p_{jt}), which is the same online as in-store.³⁴ Besides these observable determinants of demand, there is an unobserved demand factor (ξ_{jt})³⁵ and an i.i.d. Gumbel error (ε_{ijt}). In all,

$$u_{ijt} = v_{iB(j)} + x_j \beta_i - \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt}. \quad (2.9)$$

Of course, the consumer is not obliged to purchase any of the J_t goods on offer. Let $j = 0$ index the “outside option” of purchasing nothing (which provides utility u_{i0t}).³⁶

Learning.—Consumers can, in principle, learn about their tastes for any observable characteristic. However, computational limitations force me to focus on just one characteristic. I choose the characteristic of *brand* for two reasons. First, descriptive evidence suggests that consumers learn more about their tastes for brands than about their tastes for other characteristics (see Section 2.3.2). And second, the characteristic of brand is among the primary determinants of products’ retail margins (see Section 2.3.3). The store may, therefore, profit more from steering consumers’ learning about brands than from steering their learning about other characteristics.

I model consumers’ learning about brands as follows. If consumer i has never purchased brand b , she holds the following (unbiased) beliefs about her tastes for the brand:

$$v_{ib} \sim \text{Normal}(\mu_{ib}, \iota_b^2). \quad (2.10)$$

Once she purchases one of the brand’s products (i.e., some good j such that $B(j) = b$), she will learn her true tastes v_{ib} for the brand. Specifically, v_{ib} will be randomly drawn from Equation (2.10), with the results of the draw determining her tastes for the brand on all future trips.³⁷

³⁴Prices are “exogenous” in the sense that the store cannot individualize prices based on consumers’ decisions to accept or reject substitutes.

³⁵This term captures unobserved store-level promotional activities that temporarily shift demand for the good, such as being featured in a flyer or being placed in a prominent location (such as an “end cap”).

³⁶I normalize $u_{i0t} = \varepsilon_{i0t}$, where ε_{i0t} is an i.i.d. Gumbel error.

³⁷Here, I implicitly assume that a single consumption experience suffices to obtain full knowledge of one’s true tastes for a brand. Although this “one-shot” model of learning is more restrictive than the Bayesian one used in much of the literature (e.g., Erdem and Keane [1996]), it affords two key advantages. First, it accommodates richer heterogeneity in consumers’ underlying tastes than would a more complex model of learning (see Erdem, Keane, and Sun [2008] or Che, Erdem, and Öncü [2015]). And second, “one-shot” learning is likely a close approximation of consumers’ true learning process in this environment. (Intuitively, less experience is required to learn whether one likes a packaged snack or drink than whether one likes a more complex good, such as a car or a computer.)

Consumers hold heterogeneous prior beliefs about their tastes for a given brand. In particular, prior expected tastes for brands (the μ_{ib} 's) are normally distributed across the population of consumers such that

$$\mu_{ib} \sim \text{Normal}(\mu_b, \sigma_b^2)$$

for each brand b . However, all consumers' priors are equally informative about a given brand b (hence the absence of an i subscript on μ_b in equation [2.10]).

In-Store Purchases, Curbside Orders, and Stockout Substitutions.— Whether she is shopping in-store or online, each consumer i purchases one unit of the good with the highest expected utility.³⁸ The source of uncertainty is her tastes for brands. Concerning goods j whose brands $B(j)$ she has never purchased before, the consumer's expected utility depends on her prior-expected tastes for its brand, namely $\mu_{iB(j)}$. As for goods j whose brands she *has* bought before, she knows their exact utilities (u_{ijt}) because she has already learned her true brand tastes ($v_{iB(j)}$) from experience.

Let $\mathcal{I}_{it}$ denote the consumer's prior beliefs and experiential knowledge concerning her preferences for brands at time t . Regarding each brand b that the consumer has *not* yet purchased, the information set contains the parameters μ_{ib} and σ_b^2 that characterize her prior beliefs. As to a brand b that she *has* previously purchased, $\mathcal{I}_{it}$ contains her true tastes v_{ib} .

The expected utility of good $j \in \mathcal{J}_t \setminus \{0\}$ is given by

$$E[u_{ijt} \mid \mathcal{I}_{it}] = E[v_{iB(j)} \mid \mathcal{I}_{it}] + x_j \beta_i - \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt},$$

with

$$E[v_{iB(j)} \mid \mathcal{I}_{it}] = \begin{cases} v_{iB(j)} & \text{if } i \text{ has bought brand } B(j) \text{ before} \\ \mu_{iB(j)} & \text{otherwise.} \end{cases} \quad (2.11)$$

If the consumer is placing an order for curbside pickup, her preferred good—say, $j^\star$ —may go out of stock. She will then be offered a substitute $s \in \mathcal{J}_t \setminus \{0, j^\star\}$, which she will accept if and only if

$$E[u_{ist} \mid \mathcal{I}_{it}] \geq u_{i0t}. \quad (2.12)$$

³⁸I do not model the decision to order a good in the first place. In the data, it is difficult to distinguish between curbside orders where (i) the consumer considered ordering a product from the relevant differentiated-products market, but decided against it; and (ii) the consumer never considered ordering anything from the market in the first place.

Are Consumers Myopic or Forward-Looking?—Consumers’ purchases affect their expected utility on future shopping trips as well as on the present one. The same is true of their decisions to accept or reject stockout substitutes. This is because consumers can learn their true tastes for a brand by either purchasing one of its products or by accepting one of them as a substitute. The resultant learning would enable them to make more informed—and, in expectation, higher-utility—purchases in the future.

Are consumers forward-looking, meaning that they account for the (expected) value of learning? Or are they myopic, meaning that they do not? I assume the latter for two reasons. The first concerns the purchase environment. When shopping for groceries, consumers typically face a multitude of low-stakes decisions. To reduce the cognitive burden, consumers may focus on their present-trip utility, rather than solving the dynamic maximization problem induced by learning’s impact on future utility. Behavioral considerations aside, it is also computationally useful to assume that consumers are myopic. In prior work where consumers are *not* assumed to be myopic, but rather forward-looking, it has usually proved necessary to assume that all consumers share the same underlying preferences among brands.³⁹ By assuming that consumers are myopic, I can accommodate heterogeneous underlying tastes for brands. And, in terms of forecasting consumers’ behavior under counterfactual substitution policies—the ultimate goal of this study—it is arguably more important to capture heterogeneity in consumers’ underlying brand tastes than to model (potentially) forward-looking behavior.⁴⁰

Profits.—As far as this paper is concerned, discounted variable profits correspond to the present-discounted sum of the retail margins associated with consumers’ purchases. With a view to computing variable profits, let $\text{choose}_{ijt} = 1$ if consumer i orders good j online or purchases it in-store; otherwise, let $\text{choose}_{ijt} = 0$. Likewise, define OOS_{jt} as an indicator variable for good j ’s

³⁹Osborne (2011) and Shin, Misra, and Horsky (2012) provide noteworthy exceptions. Both assume that consumers are forward-looking and that they possess heterogeneous underlying preferences. To surmount the resultant computational challenges, however, both studies resort to smaller estimation sample sizes (fewer than 700 households) than would be ideal for this study, where heterogeneity in consumers’ past purchase histories is of direct interest.

⁴⁰Concerning the Norwegian market for new books, Daljord (2022) provides quasi-experimental evidence that consumers evince far greater impatience than the real rate of interest would imply. So, to the extent that consumers are forward-looking while shopping for groceries—arguably, a faster-paced activity (with lower stakes per item purchased) than that of shopping for new books—this feature of their behavior is likely of second-order importance.

undergoing a stockout substitution at time t . Lastly, let accept_{ist} be an indicator for consumer i 's accepting good j as a stockout substitute. Then the discounted variable profits from consumer i at time t are computed as follows:

$$\Pi_i^t = \sum_{t'=t}^{T_i} \prod_{j \in \mathcal{J}_{t'}} \delta^{t'-t} (p_{jt'} - mc_{jt'})^{\text{choose}_{ijt'} - \text{OOS}_{jt'}} \left(\prod_{s \in \mathcal{J}_{t'} \setminus \{j\}} (p_{st'} - mc_{st'})^{\text{accept}_{ist}} \right)^{\text{OOS}_{jt'}},$$

where δ denotes the discount factor.⁴¹

2.5.2 Estimation Method

Several sets of parameters need to be estimated. The first set of parameters pertain to consumers' *prior expected* tastes μ_{ib} for brands b , as well as their *true* tastes v_{ib} . Regarding the latter, two distinct parameters contribute to heterogeneity in consumers true tastes v_{ib} for a given brand b . One is σ_b^2 , which measures heterogeneity in consumers' prior expected tastes for the brand; while the other is ι_b^2 , which gauges the amount of learning when consumers first try the brand. Summing these two parameters yields the standard deviation of consumers' true tastes for a given brand. Specifically,

$$v_{ib} \sim \text{Normal}(\mu_b, \sigma_b^2 + \iota_b^2).$$

This follows immediately from $v_{ib} \sim \text{Normal}(\mu_{ib}, \iota_b^2)$ and $\mu_{ib} \sim \text{Normal}(\mu_b, \sigma_b^2)$. For further details on the brand parameters, see Chapter 2C.

The second set of parameters bears on products' non-brand observable characteristics x_j . Let k index specific characteristics, so that $x_j = (x_{j1}, \dots, x_{jk}, \dots, x_{jK})$ for each good j . Because consumers innately know their tastes β_{ik} for each non-brand characteristic k , the distribution of consumers' taste parameters is recovered with the same procedure as in the familiar mixed logit model (see Arteaga et al. [2022]). For most non-brand characteristics, I assume that tastes β_{ik} follow a normal distribution, conditional on the demographics of consumers' households:

$$\beta_{ik} = \beta_k + \beta_k^D D_i + \sigma_k \omega_{ik},$$

⁴¹Although t indexes both time and location, here I abuse notation by letting $t' - t$ denote the time elapsed between trips t' and t . As to the discount factor itself, I impose a 0.9998 real daily discount rate. This translates to a 0.93 real annual discount rate, which falls between the discount factor of 0.9 used by Ryan (2012) and the discount factor of 0.95 used by Collard-Wexler (2013).

where $\omega_{ik} \sim \text{Normal}(0, 1)$. To reduce the computational burden of simulation, however, I estimate fixed coefficients on a few non-brand characteristics (so that $\beta_{ik} = \beta_k$ for all consumers i).

The third set of parameters governs consumers' price sensitivity. Conditional on household income, I assume that the random price coefficient α_i follows a log-normal distribution with shift parameter α and scale parameter σ_α^2 .

The fourth set of parameters concerns the method by which consumers accept or reject stockout substitutes. Before September 2021, consumers learned of stockouts upon arriving at the store and then accepted or rejected the substitute on the spot. Since September 2021, however, consumers have been able to accept or reject remotely using the store's app or website. Because this new procedure may have lowered the cost of rejecting a substitute,⁴² I allow the utility of rejection to differ before versus after September 2021. In particular, I assume that the consumer will accept a substitute s if and only if

$$E[u_{ist} \mid \mathcal{I}_{it}] \geq u_{i0t} - \gamma \cdot 1[\text{reject in-person}], \quad (2.13)$$

where the parameter γ gauges the added cost of rejecting a substitute in-person (as opposed to remotely).

All the foregoing determinants of demand are observed in the data. However, demand also depends on unobservable factors that vary across space and time. One such factor is store- and time-specific promotional activities, like inclusion in a flyer or placement in a prominent location (like an "end-cap"). In the utility specification, shocks of this description are represented by the term ξ_{jt} .⁴³ To recover ξ_{jt} , I employ the control function approach proposed by Kim and Petrin (2019). This approach proceeds in two steps. In the first, I estimate the reduced-form pricing function. Besides the variables that enter the utility function—namely, the brand $B(j)$ and the non-brand observable characteristics x_j —the pricing function also incorporates a set of instrumental variables that are excluded from demand. I employ products' wholesale costs wc_{jt} as the excluded

⁴²For a start, it may be easier for the consumer to plan a trip to an additional grocery or convenience store if she knows of the stockout in advance. There may also be a psychological dimension; when accepting or rejecting substitutes in-person, consumers might feel social pressure to accept the substitute.

⁴³Although these shocks vary differently over time among consumers i , I omit an i subscript because t indexes combinations of specific store locations and times.

instruments. The intuition is that wholesale costs should be correlated with retail prices, but uncorrelated with store-level promotional activities. All told, the reduced-form pricing function takes the following form:

$$p_{jt} = \eta_{B(j)} + x_j\varphi + \psi \cdot wc_{jt} + \tilde{\xi}_{jt}.$$

I estimate this equation via OLS. Because the store changed its internal cost measure in January 2021,⁴⁴ I perform separate regressions before and after that date. Then, in the second step of the control function procedure, I substitute $\xi_{jt} = \lambda \tilde{\xi}_{jt}$ in the utility function. Here, $\tilde{\xi}_{jt}$ is the residual from the reduced-form price regression and λ is a parameter to be estimated. Due to the change in the store’s internal cost measure during January 2021, I estimate separate coefficients $\lambda_{\text{pre-21}}$ and $\lambda_{\text{post-21}}$ on the control functions before and after that date.

With the control function in hand, the parameters that govern consumers’ utility and learning are obtained via maximum simulated likelihood estimation. My estimation code is adapted from Arteaga et al. (2022). See Chapter 2C for details on the estimation method.

Identification.—Formal identification of the model’s parameters is beyond the scope of this paper. Instead, I will describe how the parameter estimates depend on specific moments of the data. Because previous work has already identified differentiated products demand in the absence of consumer learning (see Berry and Haile [2024, 2021, 2016, 2014]), as well as random-coefficients discrete choice more generally (see Fox et al. [2012] and Iaria and Wang [2024]), my discussion focuses on the parameters that pertain to consumers’ learning.⁴⁵

First consider μ_b . This parameter measures how much the average consumer expects to like brand b before she tries it. Because consumers’ prior beliefs are assumed to be unbiased, μ_b also gauges how much the average consumer would *actually* like the brand if she tried it.⁴⁶ The parameter μ_b is sensitive to the following moment of the data. Are brand b ’s product’s more or

⁴⁴Before January 2021, the store included some fixed costs in its internal cost measure (as well as the wholesale cost).

⁴⁵Although Shin, Misra, and Horsky (2012) identify a Bayesian learning model of demand, the intuition differs from the “one-shot” learning model estimated here. In a Bayesian learning model, the researcher must untangle two distinct learning effects: *bias reduction* and *uncertainty reduction*. In a one-shot model, by contrast, a single consumption experience suffices to eliminate both bias and uncertainty in the consumer’s beliefs.

⁴⁶This is because consumers’ prior beliefs are assumed to be biased.

less popular than would be expected, given their respective (non-brand) observable characteristics, prices, and unobserved demand factors? If they are more popular than expected, brand b must be comparatively well liked. Thus, μ_b should be large. On the other hand, if the brand's products possess smaller market shares than expected, consumers must not like the brand very much. Hence, μ_b should be small.

Now turn to σ_b^2 , which measures heterogeneity in expected tastes for brand b among consumers who have not yet tried the brand. This parameter is sensitive to variation across consumers in the number of shopping trips *before* they purchase the brand for the first time. To see the intuition, suppose first that there is little variation in how long consumers wait before trying one of the brand's products. This suggests that consumers are similarly optimistic about their tastes for the brand, so σ_b^2 is likely small. Now imagine, instead, that there is considerable variation in how long consumers wait before trying the brand; whereas some consumers purchase the brand on one of their earliest shopping trips, others wait a long time before doing so. These two groups of consumers probably differ in their expected tastes for the brand, with the former group being more optimistic than the latter. Thus, σ_b^2 should be large.

Finally, consider ι_b^2 . This parameter measures the amount of learning that consumers experience when they try one of brand b 's products for the first time. To what extent do their true tastes for the brand (v_{ib}) differ from their expected tastes (μ_{ib})? This parameter partly depends on the following moment of the data.⁴⁷ Consider the subset of consumers who try brand b for the first time because of a stockout substitution. How often do these consumers purchase brand b in the future? If they seldom do so, they probably did not learn much about the brand from the substitution. Rather, the experience confirmed their pessimistic prior beliefs about their tastes for the brand. Consequently, ι_b^2 should be small. Now suppose, instead, that many consumers proceed to purchase brand b quite frequently. These consumers likely learned a lot from the substitution, finding brand b more to their tastes than they had expected. Hence, ι_b^2 should be large.

⁴⁷In addition to variation from stockout substitutions, the ι_b^2 estimates also depend on a subtler relationship between (i) the number of purchases before consumers first try out the brand, and (ii) the frequency with which they purchase the brand's products thereafter.

2.5.3 Construction of Estimation Data Set

In this subsection, I describe how I assemble the data set used to estimate the demand model above. As the procedure closely resembles the one used by Zeyveld (2024), much of this subsection is adapted from Section 6 of that paper.

I cannot estimate demand for all the products within a given product category due to computational constraints. For this reason, I exclude slow-selling brands and products from estimation.⁴⁸ Computational constraints also prevent me from including all consumers in estimation. Rather, within each product category, I perform estimation on the following subset of consumers. First, I find consumers who experience stockout substitutions where both the out-of-stock product and the substitute are popular products. (These consumers are used both in estimation and in counterfactuals.) Next, to increase the sample size, I randomly sample additional consumers who have also experienced a stockout substitution—albeit one where either the ordered product or the out-of-stock one is a slow-selling product. (These consumers are included for estimation but excluded from counterfactuals.)

Having sampled consumers for estimation, I need to reconstruct the discrete choice problems that they faced on each shopping trip. What products were available for purchase? And what were their prices? Recall that the scanner data directly record the UPC and price of the item that was purchased. However, these data also enable me to infer the UPCs and prices of goods that the consumer did *not* purchase. To do so, I consult the chain’s product catalog in order to obtain the UPCs of the store’s offerings within the relevant category. Then, turning to the scanner data, I compare these UPCs with those of products sold at the relevant store. If I observe a given product being purchased at the relevant store on the same day as our consumer’s shopping trip, I assume that the product was within her choice menu. Failing that, I presume that the product was available if it was purchased on both the day before *and* the day after our consumer’s trip. Otherwise, I

⁴⁸Regarding super-premium ice cream, I estimate demand for products that are (i) sold under one of the top three brands, (ii) command at least 0.5% market share among consumers who place at least one curbside pickup order, and (iii) are not “limited edition” products (whose non-brand characteristics are not recorded in the product catalog). These products populate 77.1% of purchases by consumers who place at least one curbside pickup order. As for apple sauce cups, I estimate demand for products with at least 0.5% market share among consumers who place at least one curbside pickup order. Such products cover 98.2% of purchases by consumers who place at least one curbside pickup order.

assume that the product was absent from the consumer's choice set (either because it was out of stock, or because the store did not carry it at all).

Given that a product appears to be available, I impute its price as being the mean purchase price on the day of the consumer's shopping trip (within the relevant store location).⁴⁹ If no purchases are observed on the precise day of the trip, I instead take the unweighted average of the mean purchase prices on the days immediately before and after.

Consumers' purchases sometimes deviate from the underlying assumptions of my discrete choice model. For a start, consumers sometimes purchase multiple distinct products on a single shopping trip. To illustrate, a consumer shopping for super-premium ice cream might purchase both Häagen-Dazs and Halo Top ice cream on the same trip. I drop all such observations from estimation.⁵⁰ In addition, consumers sometimes purchase multiple units of the same product on a single shopping trips (thereby "stockpiling" the relevant product). For simplicity, I do not model the consumer's choice of quantity. Instead, different quantities of a given product constitute a single option within my discrete choice framework.⁵¹

Initial Conditions Problem.—Some consumers have made purchases at the store before the earliest date recorded in my data (April 24, 2016). This creates an initial conditions problem: When I observe consumers' purchases early in the data, are they experiencing brands for the first time? Or had they purchased them previously, before coverage begins in the data?⁵²

In order to minimize this problem, I drop consumers' four purchases of super-premium ice cream or their first seven purchases of apple sauce cups. These "burn-in" periods are motivated by the following stylized facts. After her first four (nine) shopping trips, three-quarters of super-premium ice cream (apple sauce cups) buyers have purchased all the brands that they will *ever* buy

⁴⁹The chain maintains a policy of uniform prices online and in-store.

⁵⁰This results in the exclusion of 59.6% (25.1%) of transactions involving super-premium ice cream (apple sauce cups).

⁵¹In the product categories of super-premium ice cream and apple sauce cups, consumers with 1+ stockout substitutions purchase multiple units of a single product on 32.3% and 28.7% of shopping trips, respectively.

⁵²A related, but distinct, concerns purchases at *other* supercenter chains. If someone purchases a given brand for the first time at another chain, then her earliest purchase of that brand within the data would not occasion learning. However, most of the behavioral markers that identify the brand parameters are spread over many transactions. This should reduce bias from the misattribution of learning.

at the store.

2.6 Estimation Results

In this section, I report estimates for the demand model developed in Section 2.5 and then evaluate how well the model fits the data.

2.6.1 Parameter Estimates

Table 2.4 presents the demand estimates for super-premium ice cream. Concerning the μ_b estimates, the average consumer narrowly prefers Ben & Jerry's to Häagen Dazs, while Halo Top comes in a distant third. As for the σ_b estimates, consumers display considerable heterogeneity in their prior expected tastes for the all three brands. In fact, $\sigma_{\text{Halo Top}}$ exceeds the difference in mean utility between Halo Top and Ben & Jerry's (that is, $\sigma_{\text{Halo Top}} > \mu_{\text{Ben \& Jerry's}} - \mu_{\text{Halo Top}}$). Finally, consider the ι_b estimates. For each brand b , this parameter is smaller than the corresponding value of σ_b . To see what this means, consider two randomly-selected consumers i and i' . On average, the disparity in our consumers' prior expected tastes for a given brand b —that is, $|\mu_{ib} - \mu_{i'b}|$ —exceeds the amount of learning that consumer i or i' would experience if one of them tried the brand for the first time (i.e., $|v_{ib} - \mu_{ib}|$ or $|v_{i'b} - \mu_{i'b}|$).

Consider next the products' non-brand observables and prices (where my treatment of the former closely follows Sullivan [2020].) According to the β estimates, the average consumer prefers chocolate-flavored ice cream over mint, sweet cream, or vanilla. She also favors a single flavor of ice cream over multiple flavors (particularly when she belongs to a large household). As for mix-ins, such as cookie dough or chocolate chips, the average consumer wants one or two mix-ins (but not three or more, due to the negative coefficient on the quadratic term). The σ_b estimates, meanwhile, point to considerable unobserved heterogeneity in consumers' preferences for both flavor and mix-ins. Turning to the random coefficient on price, recall that α_i is assumed to follow a log-normal distribution with shift parameter α and scale parameter σ_α^2 .⁵³ The former exceeds the latter in absolute value, so there is considerable variation across consumers in the marginal utility of additional income.

⁵³Recall that the price enters the utility function negatively; see Equation (2.9).

Table 2.4: Parameter Estimates

<i>Panel A. Brands</i>					
	Mean exp. tastes (μ_b 's)	Heterogeneity of exp. tastes (σ_b 's)	Amount of learning (ι_b 's)		
Ben & Jerry's	−11.448 [1.210]	2.820 [0.054]	0.091 [0.034]		
Häagen-Dazs	−12.029 [1.062]	3.294 [0.075]	0.903 [0.050]		
Halo Top	−15.808 [1.216]	4.693 [0.123]	1.971 [0.082]		
<i>Panel B. Non-brand observables and prices</i>					
	Means (β 's or α)	Std. devs. (σ_β 's or σ_α)	Demographic interactions (β_k^D 's or α^D)		
			Household income	Household size	Age of oldest HH male ^a
Flavor: chocolate	−0.380 [0.063]	2.110 [0.047]			
Flavor: mint	−2.358 [0.136]	3.297 [0.121]			
Flavor: sweet cream	−1.970 [0.109]	2.121 [0.088]			
Flavor: vanilla	−0.849 [0.067]	2.714 [0.060]			
Multiple flavors	−0.387 [0.089]			−0.082 [0.014]	0.004 [0.002]
No. of mix-ins	0.792 [0.047]	1.357 [0.017]		0.014 [0.006]	−0.002 [0.001]
(No. of mix-ins) ²	−0.279 [0.008]				
Quantity (oz.)	1.052 [0.076]			−0.002 [0.003]	
Price ^b	−0.689 [0.083]	0.697 [0.039]	−0.002 [0.000]		
<i>Panel C. Other explanatory variables</i>					
	Coefficients (λ 's or γ)				
Control function (pre-2021) ^c	0.031 [0.052]				
Control function (post-2021) ^c	0.303 [0.043]				
Reject in-person ^d	1.377 [0.166]				

Notes: estimates are based on 16,827 randomly-sampled observations involving 1024 households. Standard errors (in brackets) do not correct for measurement error in the control function.

^a Or oldest female (if no male in household).

^b Conditional on household income, the random price coefficients α_i are assumed to follow a log-normal distribution.

^c The demand shocks are specified as $\xi_{jt} = \lambda \tilde{\xi}_{jt}$, where $\tilde{\xi}_{jt}$ is the residual from the pricing function and λ is a scaling parameter (reported here). This control function is computed separately before/after January 2021, due to a

change in the store's internal cost measure.

^d Until September 2021, consumers accepted or rejected stockout substitutes upon arrival at the store. Starting September 2021, they could accept or reject substitutes remotely (using the store's app or website).

Other Product Categories.—The estimated parameters for apple sauce cups appear in Table 2D.1. As for flavored milk and frozen french fries, convergence issues prevent me from estimating structural models (or performing counterfactual simulations).

2.6.2 Goodness of Fit

How accurately does the model predict the acceptance or rejection of stockout substitutes? And how well does the model reproduce the dynamics of consumers' subsequent purchases? To answer these questions, I compute the predicted probability of acceptance and then perform forward simulations of consumers' purchases and learning thereafter. In both steps, I compute choice probabilities that reflect consumers' revealed preferences and beliefs. To accomplish this, I do not assign equal weights to all the simulation draws of the random coefficients, but rather compute "conditional weights" that reflect the consumer's observed choices (see Revelt and Train [2000]). To see the intuition, consider a consumer i who never buys chocolate-flavored ice cream. By the logic of revealed preferences, consumer i probably likes chocolate-flavored ice cream less than the average consumer does. This suggests that $\beta_{i,\text{chocolate}} < \beta_{\text{chocolate}}$. The same intuition can be adapted to recover conditional distributions on consumers' prior beliefs and true tastes for brands (the μ_{ib} 's and v_{ib} 's, respectively), along with their marginal utilities of additional income (the α_i 's). In the results discussed below, I employ conditional distributions that reflect the entirety of consumers' purchases in the data—before, during, and after the stockout substitution.⁵⁴

Accept/Reject Decisions.—Regarding the acceptance or rejection of stockout substitutes, the model's fit can be assessed in several ways. The simplest is to compare the predicted and observed rates of acceptance. These prove extremely close; whereas the model predicts that consumers will accept 81% of stockout substitutes, they actually accept 79%. Another measure of fit is the predicted probability assigned to consumers' true decisions. This amounts to the predicted probability of acceptance when the consumer is observed to accept and the predicted probability of rejection

⁵⁴In the counterfactual simulations presented in Section 2.7, I impose that the store cannot foresee consumers' future purchases as it chooses a stockout substitute. Counterfactual substitution policies are, therefore, defined using conditional distributions based on consumers' past purchases only (as these are known to the store at the time of the stockout). Ex post, however, I assess the outcomes that would be realized under a given substitution policy by using conditional distributions that reflect all observed choices in the data.

otherwise. The model assigns consumers' true accept/reject decisions a mean predicted probability of 79%.

Repetition of Brand and Product Choice.—Another important moment concerns the frequency with which consumers purchase the same brand or product on successive shopping trips. There are several reasons why the predicted probability of repetition might be too small. One is the misspecification of demand. Although I accommodate both observed and unobserved heterogeneity in consumers' preferences and beliefs, the deterministic portion of utility (that is, $u_{ijt} - \varepsilon_{ijt}$) imperfectly captures consumers' time-invariant tendencies to like (or dislike) certain products.⁵⁵ To compensate, the model assigns undue importance to the i.i.d. Gumbel errors, which are (by construction) uncorrelated over time. Another challenge is the finite number of shopping trips observed per consumer. Although my predictions employ “conditional” distributions of random coefficients which reflect consumers' observed choices, these conditional distributions remain non-degenerate. This is because many realizations of the random coefficients could, in principle, be consistent with a given consumer's observed choices.

I find that the model closely matches the true frequency of repeat brand purchases in the data. Whereas the predicted probability that consumers purchase the same brand on successive shopping trips is 86%, the true probability is 81%. Regarding individual products, the model performs comparatively worse. The predicted probability that consumers purchase the same product on successive shopping trips is 12%, but the true probability is 27%.

Endogenous Learning.—How often do consumers try out new brands on their own initiative? On the one hand, if consumers frequently experiment with new brands, the store will gain little from introducing a consumer to a profitable new brand through a stockout substitution. For, even if the stockout substitution had never occurred, the consumer would likely have tried the relevant brand soon anyway. On the other hand, if consumers rarely try out new brands, stockout substitutions present an important opportunity for the store to encourage consumers to try out profitable new brands. So, if the model under- (over-) predicts the frequency of endogenous

⁵⁵For one thing, my model excludes behavioral phenomena like inertia or incomplete consideration that may reinforce the tendency to make repeat purchases.

learning, the counterfactual simulations will tend to over- (under-) state the gains from steering consumers' learning through stockout substitutions.

To assess fit in relation to endogenous learning, I start by subtracting the number of brands known to the consumer at the time of the stockout from the total number of brands known at the latest date in the data (July 2023). Then I compare this measure of observed brand experimentation with its predicted counterpart. Whereas consumers are predicted to try an average of 0.43 new brands after the stockout, they actually try out 0.39 new brands.

Learning from Stockout Substitutions.—The profitability of steering consumers' learning depends critically on the following moment of the data. Consider a consumer who has accepted a stockout substitute from brand that she has never purchased before. Is she likely to purchase the substitute's brand on her own initiative in the future? If so, how often? My model predicts that consumers will purchase the brand of the store's chosen substitute on 73% of future shopping trips. The actual proportion in the data is 76%.

2.7 Counterfactual Simulations

How much would profits increase if the store exploited stockout substitutions to steer consumers' learning? And how would this affect consumer welfare? To answer these questions, I conduct counterfactual simulations using the estimated primitives of consumers' beliefs, learning, and tastes.

In what follows, I compare profits and consumer welfare under the store's existing policy with the corresponding outcomes under counterfactual policies. At present, the store's "baseline" substitution policy leaves the choice of substitute to whichever worker happens to be assembling the curbside order. He is asked to exercise his "best judgement" in selecting a suitable replacement for the out-of-stock product. As for the counterfactual substitution policies, these vary along two dimensions. One is the amount of information about consumers that the store employs. Recall that the store knows consumers' past purchases and household demographics, as well as their original order choices. How does the optimal choice of substitute change when the store leverages (i) none of this information; (ii) only its knowledge of consumers' original orders (i.e., the sole information

available to store workers under the baseline policy); or (iii) all the information available to the store? The other dimension along which counterfactual policies vary is the store's objective function. Does the store seek to maximize its expected profits on the present shopping trip alone? Or does it also account for stockout substitutions' influence on consumers' learning (and future purchases)?

2.7.1 Simulation Method

I construct the counterfactual substitution policies using the estimated model from Section 2.5, coupled with plausible assumptions about the future evolution of products' prices and availabilities. Under each counterfactual policy, the store will choose substitutes that maximize (a) expected present-trip profits, (b) the present-discounted value of total profits, or (c) the present-discounted value of future profits alone. Observe that (a) depends on the retail margins and acceptance probabilities of the possible substitutes on the shelf, while (b) additionally depends on the learning that the consumer would experience if she were to accept and (c) depends only on the latter. Here, products' retail margins are directly observed in the data, but the other factors must be simulated.

Focus first on the probability of acceptance. Under counterfactual policies that disregard the consumer microdata, the store will compute this probability based on the joint distribution of tastes and beliefs over the population of consumers. Concerning the other counterfactual policies, by contrast, the store exploits its knowledge of consumers' original orders, purchase histories, and household demographics to calculate more accurate acceptance probabilities. As in Section 2.6.2, I compute "conditional" choice probabilities that reflect consumers' revealed preferences and beliefs. Here, however, I compute conditional probabilities based solely on the data employed by the relevant substitution policy. One of the counterfactual policies, for example, leverages only the microdata on consumers' original orders (as opposed to their past purchases or household demographics). As far as this policy is concerned, the probability of acceptance should be conditioned solely on original orders. Another set of counterfactual policies, meanwhile, exploit all the microdata available to the store at the time of the stockout. Regarding these policies, I compute conditional probabilities of acceptance that reflect consumers' household demographics, original orders, and

past purchases—not their *future* purchases (which the store has not yet observed at the time of the stockout).

Now turn to future profits. How might a consumer’s acceptance (or rejection) of a substitute influence the store’s expected future profits? In principle, the influence of a stockout substitution might extend infinitely into the future. To avoid overstating the returns to steering consumers’ learning, I focus on a short time horizon: one year.

The store faces several sources of uncertainty where future profits are concerned. One is the timing of consumers’ future shopping trips. Here, I assume that the store adopts a simple heuristic: for each consumer i , the frequency of future shopping trips is imputed as being the average frequency of her shopping trips up to (and including) the stockout substitution. The store is also unsure of the future availabilities, prices, and wholesale costs of products within the relevant category. For simplicity, I assume that the store does not possess “insider” knowledge about the evolution of these factors. Instead, the store randomly samples (with replacement) from the choice sets faced by consumer on past shopping trips. (Each such draw consists of the entire choice menu—including availabilities, prices, and wholesale costs—on a single shopping trip.) This allows for persistent variation across consumers in the composition of choice sets. (Such variation might be rooted in the size of the local store, the preferred time of day for shopping, etc.)

This procedure yields a synthetic dataset of future shopping trips. Next, I compute the choice probabilities associated with the future shopping trips within this synthetic dataset. In so doing, I account for the “endogenous learning” that occurs when consumers try new brands on their own initiative. To see why this matters, consider a consumer who has never purchased a given brand b . Even if the store does not offer her one of b ’s products as a substitute, she still might learn her taste for the brand on a future trip if she elects to purchase one of its products. Endogenous learning may, therefore, reduce the potential returns to steering consumers’ learning. Then, having derived the choice probabilities associated with consumers’ subsequent shopping trips, I compute the store’s present-discounted profits using a 0.9998 real daily discount rate.

Of course, this procedure reflects future profits under just one potential future state of the world.

Accordingly, I repeat the entire procedure—synthesizing data and computing choice probabilities—several times in order to “integrate” over possible future states of the world. Finally, I average across these simulation rounds to obtain the present-discounted value of expected future profits associated with the acceptance or rejection of each available substitute. The “steering substitute” is then defined as the product that maximizes the sum of (a) the expected retail margins on the present shopping trip and (b) the present-discounted value of expected future profits.

Evaluating Profits and Consumer Welfare.—Having defined the counterfactual substitution policies, I compare expected profits and consumer welfare under these policies with those under the “baseline” policy. Here, I exploit the entirety of the data—including consumers’ purchases after stockout substitutions.

The profits associated with a stockout substitute depend, once more, on the retail margin, the probability of acceptance, and the present-discounted value of expected profits (conditional on either acceptance or rejection). Regarding the probability of acceptance, I now leverage the entirety of the relevant consumer’s observed choices—before, during, and after the stockout substitution—as I compute the conditional weights on the simulation draws of the random coefficients. As for future profits, I employ a similar heuristic to the one employed to characterize the “steering” substitution policy. Now, however, I impute the frequency of the consumer’s future shopping trips as being the average across the entirety of her shopping trips in the data. Likewise, when simulating products’ future availabilities, prices, and wholesale costs, I sample (with replacement) from the entirety of her shopping trips in the data.

Having computed the choice probabilities associated with future shopping trips, I calculate the expected future profits associated with the substitutes offered under the baseline and steering policies. This entire process is repeated several times (again, with a view to “integrating” over possible future states of the world). Finally, I compare the present-discounted value of expected profits under the baseline and “steering” policies by averaging across the simulations.

2.7.2 Counterfactual Results

Table 2.5 compares profit-related outcomes under the “baseline” and counterfactual substitution policies. Recall that the latter vary along two dimensions. One is the extent of consumer microdata employed. I compare policies that use (a) none of the microdata, (b) only consumers’ original orders, or (c) all available microdata (including consumers’ past purchases and household demographics). As for the second dimension that differentiates the counterfactual policies, I assess three possible store objective functions: (i) maximizing profits on the present shopping trip, (ii) maximizing the present-discounted value of total profits, both present and future; or (iii) maximizing the present-discounted value of future profits alone. Observe that objectives (ii) and (iii) account for stockout substitutions’ influence on consumer learning, whereas objective (i) does not. For brevity, I assume that the store adopts objective (i) if it leverages only some (or none) of the consumer microdata.⁵⁶ As for (iii), this objective serves a purely illustrative function, depicting the outer limits of the store’s ability to increase future profits by steering consumers’ learning.

These criteria translate to the following counterfactual policies. The first, which I refer to as the “one-size-fits-all” substitution policy, exploits no consumer microdata whatsoever. Thus, the store’s choice of substitute only depends on the availabilities, prices, and wholesale costs of possible substitutes. By contrast, the second substitution policy—termed the “individualized-by-order” policy—accounts for consumers’ original orders. To the extent that original orders reveals consumers’ tastes and beliefs, the substitutes offered under this policy should be likelier to be accepted than those offered under the “one-size-fits-all” policy. The remaining counterfactual policies, styled as the “fully individualized” policies, additionally leverage the store’s knowledge of consumers’ past purchases and household demographics. Here, I assess all three possible objective

⁵⁶The following stylized example illustrates why it would be unappealing for the store to steer consumers’ learning without exploiting its consumer microdata. Consider a consumer who has purchased the highest-margin brand—namely, Halo Top—on one past shopping trip. However, none of her intervening purchases are sold under the Halo Top brand. Instead, she has since purchased the brand that affords the second-highest margins: Ben & Jerry’s. Based on her purchase history, the store should probably offer a substitute from the Ben & Jerry’s brand, not Halo top. For a Halo Top-branded substitute would be much likelier to be rejected than a Ben & Jerry’s-branded substitute. And even in the (comparatively unlikely) event that the consumer accepted a substitute from the Halo Top brand, she would not learn anything further about the brand; she already knows that she does not like it very much. Contrary to this intuition, however, a policy that tried to steer all consumers’ learning—irrespective of purchase histories—might still offer our consumer a Halo Top-branded substitute.

functions: namely, maximizing present-trip profits (as in the foregoing policies); maximizing the present-discounted value of total profits; and maximizing the present-discounted value of future profits alone.

Regarding the present shopping trip, notice that the store offers higher-margin substitutes under all the counterfactual policies than under the baseline policy. This is unsurprising. Under the baseline policy, store workers are asked to assess possible substitutes based on their similarity to the out-of-stock product, not their profitability. As to the differences in retail margins among the counterfactual substitution policies, these are partly rooted in the substitutes' brands. Under the "one-size-fits-all" policy, which employs none of the microdata, nearly all the substitutes (96%) are sold under the Ben & Jerry's brand. Here, Ben & Jerry's proves the best substitute brand for the "average" consumer because (a) the average consumer prefers Ben & Jerry's to the other two brands (see Table 2.4); and (b) the brand affords fairly high retail margins (see Figure 2.1). When the store adopts an individualized substitution policy, by contrast, it can select substitutes that reflect individual consumers' brand preferences—which may differ markedly from those of the "average" consumer. Under the "individualized-by-original-order" policy, the substitute shares the same brand as the out-of-stock product even more often than under the baseline policy (97% versus 92%). By comparison, a smaller proportion of the "fully-individualized" substitutes share the brand of the out-of-stock product (specifically, 81% under the policies maximizing present-trip or total discounted profits, and 29% under the policy maximizing discounted future profits).

Turn next to the predicted probability of acceptance. Here, outcomes under the "one-size-fits-all" policy and the baseline are virtually indistinguishable; the former affords acceptance probabilities of 80% and the latter 81%. By comparison, the probability of acceptance rises to 90% under the "individualized-by-original-order" policy, which exploits the store's knowledge of the out-of-stock product (but none of the other microdata). So, under the baseline policy, it seems that the store workers do not fully exploit the information contained in consumers' original orders. This is hardly surprising. Due to tight time constraints, workers are unlikely to exhaustively study all possible substitutes. Instead, they likely focus on products that are situated near the out-of-stock

product on the shelf. (This may explain why the store workers overwhelmingly choose substitutes that share the same brand as the out-of-stock product; ice cream products are typically grouped by brand.) Now consider the “fully-individualized” policies, which additionally leverage the store’s knowledge of consumers’ past purchases and household demographics. Two of these policies—namely, those maximizing present profits or total discounted profits—also deliver 90% predicted acceptance probabilities. Why do the foregoing policies afford such high acceptance probabilities? The explanation does not reside in the substitutes’ brands; as previously discussed, the “fully-individualized” policies match the out-of-stock product’s brand less often than the baseline policy does. Nor are these high acceptance probabilities rooted in the substitutes’ prices (which, under the “fully-individualized” policies, tend to exceed those of the baseline substitutes).⁵⁷ Instead, it seems the individualized substitutes more closely match the non-brand observable characteristics of the out-of-stock product, the consumer’s past purchases, or both. By contrast, the lowest probability of acceptance is associated with the “fully-individualized” policy intended to maximize discounted future profits. The probability of acceptance averages only 54% under this policy. This is partly because two-thirds of the substitutes offered under this policy belong to brands that consumers have never purchased.⁵⁸

Examine now the store’s expected present-trip profits. These correspond to the product of the substitute’s retail margin and its probability of acceptance. It emerges that the baseline and “one-size-fits-all” policies afford identical present-trip profits: \$2.19. Most of the individualized policies, meanwhile, select substitutes that are both higher-margin and likelier to be accepted than are their baseline counterfactuals. The result is higher expected present-trip profits: \$2.47 under the “individualized-by-original-order policy,” and \$2.55 under the “fully-individualized” policies that maximize either present-trip or total discounted profits. However, the policy with the

⁵⁷The mean (median) of the baseline substitutes’ prices come to \$3.82 (\$3.95), while those of the “individualized-by-original-order” policy amount to \$3.74 (\$3.90). As for the “fully-individualized” policies, substitutes’ mean (median) prices when the store maximizes present-trip profits are \$3.87 (\$4.06), while mean (median) prices when the store maximizes total discounted profits are \$3.88 (\$4.06).

⁵⁸In particular, the modal substitution where the predicted probability of acceptance drops by more than average (i.e., 27%) meets the following description: (i) the out-of-stock product is sold under the Häagen-Dazs brand, (ii) all of the consumer’s past purchases are Häagen-Dazs, (iii) the baseline substitute is Häagen-Dazs, and (iv) the “fully-individualized” substitute that maximizes discounted future profits is sold under the Ben & Jerry’s brand.

Table 2.5: Profit-Relevant Outcomes by Substitution Policy

				“Fully individualized” by original order, past purchases, and household demographics		
	Baseline	“One size fits all”	Individualized by original order	Max. present- trip profits	Max. PDV total profits	Max. PDV future profits
Panel A. Present trip						
Retail margin	2.71 (0.33)	2.74 (0.49)	2.77 (0.50)	2.84 (0.50)	2.84 (0.50)	2.83 (0.35)
Prob. accept	0.81 (0.24)	0.80 (0.22)	0.90 (0.14)	0.90 (0.14)	0.90 (0.14)	0.54 (0.30)
Expected present- trip profits	2.19 (0.70)	2.19 (0.63)	2.47 (0.45)	2.55 (0.44)	2.55 (0.44)	1.51 (0.84)
Panel B. Future trips						
PDV future profits	17.16 (25.02)	17.16 (25.02)	17.16 (25.02)	17.16 (25.02)	17.16 (25.02)	17.16 (25.03)
Panel C. Overall						
PDV total profits	19.34 (24.98)	19.34 (25.00)	19.62 (25.02)	19.70 (25.00)	19.70 (25.00)	18.67 (25.00)

Notes: This table compares profit-relevant outcomes under the store’s existing substitution policy (the “baseline”) with outcomes under counterfactual policies. These counterfactual policies exploit the store’s knowledge of the consumer to varying degrees, with the “one-size-fits-all” policy leveraging none of this information; the “individualized-by-original-order” policy employing the store’s knowledge of the consumer’s original order; and the “fully individualized” policies additionally exploiting the store’s knowledge of the consumer’s past purchases and household demographics. Regarding the last, the “fully-individualized” policies are respectively designed to maximize (i) expected profits on the present shopping trip, (ii) the present-discounted value of total profits (both present and future), or (iii) or the present-discounted value of future profits alone. All results are reported as means, with standard deviations appearing in parentheses.

lowest present-trip expected profits is the “fully-individualized” policy intended to maximize future discounted profits. Under this policy, expected present-trip profits shrink to \$1.51 due to the low predicted probability of acceptance.

Consider next the profits from consumers’ future shopping trips. It transpires that the present-discounted value of future profits is identical under all the substitution policies: \$17.16. This is true even of the “fully-individualized” policy that is designed to maximize future profits alone—irrespective of the cost to present-trip profits. I will discuss possible explanations for this result momentarily.

Turn last to the present-discounted value of total profits—both present and future. Here, the “one-size-fits-all” again proves indistinguishable from the store’s baseline policy: each results in total discounted profits of \$19.34. The “individualized-by-original-order” policy increases profits to

\$19.62, while the “fully-individualized” policies maximizing present-trip profits or total discounted profits further increase profits to \$19.70. But the “fully-individualized” policy maximizing solely future profits delivers the lowest total profits of all (\$18.67).

Discussion.—Most of the gains from individualization can be realized by conditioning the choice of substitute on the consumer’s original order. Whereas the “individualized-by-original-order” policy boosts the present-discounted value of total profits by \$0.28 over the store’s baseline policy, the additional gains from conditioning on consumers’ past purchases and household demographics come to \$0.08. This suggests that a single order decision is highly informative of the consumer characteristics that affect the store’s optimal choice of substitutes, such as brand preferences and price sensitivity.

Now consider consumer learning. The counterfactual results suggest that the store cannot perceptibly increase future profits by introducing consumers to new brands through stockout substitutions. This remains true even if the store disregards the dent to present-trip profits associated with offering stockout substitutes from unfamiliar brands (which consumers are quite likely to reject). Admittedly, there some heterogeneity across stockouts in the returns to steering consumers’ learning. For instance, there are eight stockouts where the present-discounted value of expected future profits increases by at least five cents under the “fully-individualized” policy designed to maximize the present-discounted value of total profits, compared to the “fully-individualized” policy tailored to maximize expected present-trip profits alone. But eight stockouts is a small fraction of the analysis sample (2048 stockouts).

Why is the store unable to increase profits by steering consumers’ learning? One possible explanation is endogenous learning: absent the stockout, consumers would have still tried the more profitable brands soon. However, counterfactual simulations presented in Table 2E.1 suggest otherwise. Even if consumers never learned about new brands after stockout substitutions, the store’s future profits under the optimal policy would remain unchanged from the baseline. Another potential explanation for the null results is the assumption of imperfect foresight. What if the store precisely knew goods’ future prices, wholesale costs, and availabilities? Simulations in Chapter 2E

show that perfect foresight would not increase the returns to steering consumers' learning either. Instead, the unprofitability of steering consumers' learning stems from (i) consumers' reluctance to accept stockout substitutes from unfamiliar brands and (ii) the small amount of learning that they experience when they do accept. Regarding (i), recall that the probability of acceptance dips to barely one-half under the policy designed to maximize discounted future profits alone. The reason is that two-thirds of these substitutes are sold under brands that consumers have never tried before. As for (ii), the demand estimates in Section 2.6 indicate that *across-consumer* variation in prior beliefs exceeds the amount of learning that *individual consumers* experience when they try brands for the first time. It is thus unlikely that a consumer will discover that she prefers a hitherto-unfamiliar brand to those she has previously purchased. This stylized fact is consistent with the descriptive evidence presented in Section 2.3. Although consumers who try out new brands as a result of stockout substitutions proceed to purchase those brands more frequently than do otherwise-comparable consumers who do not suffer stockouts, the magnitude of this disparity is modest (3.2 percentage points).

Chapter 2E supplies suggestive evidence that the gains from steering consumers' learning stem, instead, from (a) consumers' reluctance to accept substitutes from new brands and (b) their fairly accurate prior beliefs. There, I perform counterfactual simulations in which consumers are forced to accept stockout substitutions, or consumers learn three times as much as they do in actual fact, or both. Taken in isolation, neither forcing consumers to accept nor tripling the amount of learning translates to increased future profits. But when both occur simultaneously, the average present-discounted value of future profits increases by a cent.

2.7.3 Consumer Welfare

How are consumers affected when the substitution policy is individualized according to their original orders, past purchases, and household demographics? And are consumers better off when the policy accounts for substitutions' influence on consumer learning? To provide insight, Table 2.6 compares consumer welfare under the baseline and counterfactual substitution policies. Focus first on the present shopping trip. The results in Table 2.6 suggest the "one-size-fits-

Table 2.6: Changes in Consumer Welfare (Compared to Baseline Policy)

	“One size fits all”	Individualized by original order	“Fully individualized” by original order, past purchases, and household demographics		
			Max. present-trip profits	Max. PDV total profits	Max. PDV future profits
Expected present-trip consumer surplus (\$)	−0.84 (6.72)	2.15 (5.22)	2.56 (5.59)	2.56 (5.58)	−5.18 (7.77)
PDV future consumer surplus (\$)	0.00 (0.17)	0.00 (0.08)	0.00 (0.12)	0.00 (0.12)	0.12 (0.52)
PDV total consumer surplus (\$)	−0.84 (6.73)	2.16 (5.22)	2.56 (5.59)	2.56 (5.58)	−5.12 (7.79)

Notes: This table reports changes in consumer welfare when the store adopts various counterfactual substitution policies. See notes to Table 2.5 for descriptions of these policies. All results are reported as means (with standard deviations in parentheses).

all” policy diminishes expected present-trip consumer surplus by \$0.84 compared to the store’s baseline policy. As explained above, this disparity is probably not rooted in the substitutes’ prices. Rather, the baseline substitutes are likelier to share the out-of-stock product’s brand or non-brand characteristics than are the “one-size-fits-all” substitutes. The “individualized-by-original-order” policy, by contrast, increases consumers’ expected present-trip surplus by \$2.15 compared to the baseline. And two of the three “fully-individualized policies”—namely, those designed to maximize present-trip profits or the present-discounted value of total profits—secure even larger gains in present-trip surplus (\$2.15). This is likely because the counterfactual policies select substitutes that better match consumers’ preferences for non-brand characteristics. On the other hand, the worst policy for present-trip consumer welfare is the “fully-individualized” policy that maximizes future discounted profits. This policy diminishes present-trip expected surplus by more than \$5 compared to the baseline. As previously discussed in relation to this policy’s low acceptance rate, the problem is that the policy tends to offer stockout substitutes from unfamiliar but profitable brands (about which consumers tend to hold pessimistic prior beliefs).

Now consider future shopping trips. Under all but one of the counterfactual policies, consumers’ present-discounted value of future surplus remains unchanged from the baseline. The exception is the “fully-individualized” policy that maximize the store’s discounted future profits. This policy, which attempts to introduce two-thirds of consumers to a new brand, increases consumers’

discounted future surplus by \$0.12 over the baseline. And this average conceals considerable heterogeneity. Conditional on accepting the stockout substitute, sixty-seven consumers would enjoy increases of a dollar or more in their present-discounted value of expected future surplus.

Overall, the “one-size-fits-all” policy diminishes the present-discounted value of total consumer surplus by \$0.84 compared to the baseline policy. By contrast, the “individualized-by-original-order” policy increases total discounted surplus by \$2.16. Still larger gains are afforded by the “fully-individualized” policies that maximize the store’s expected present-trip or total discounted profits: \$2.56 in both cases. As for the “fully-individualized” policy designed to maximize discounted future profits, the modest increase in consumers’ discounted future surplus is overwhelmed by the slump in expected present-trip surplus. The net result is a drop of \$5.12 in total (discounted) consumer surplus relative to the baseline.

2.7.4 Counterfactual Results: Apple Sauce Cups

In this subsection, I briefly summarize the counterfactual results for apple sauce cups.⁵⁹ The results prove qualitatively similar to those for super-premium ice cream as regards both profits and consumer welfare. Concerning the former, Table 2E.2 shows that discounted future profits remain unchanged from the baseline under the various counterfactual policies studied. On the present shopping trip, however, expected profits increase from the baseline by \$1.48 under the “fully-individualized” policies maximizing either present-trip profits or discounted total profits. Most of these gains—namely, \$1.45 (98%)—can be achieved under the “one-size-fits-all” policy.

As for consumer welfare, Table 2E.3 shows that average consumer surplus is higher under the baseline policy than under any of the counterfactual policies. Among the counterfactual policies, though, consumer surplus is higher under the “fully-individualized” policies than under the “individualized-by-original-order” policy or the “one-size-fits-all” policy.

2.8 Conclusion

This paper shows that stockout substitutions in curbside grocery pickup enable the store to steer consumers’ learning towards high-margin brands. However, consumers are less likely to

⁵⁹Recall that structural models were not estimated for flavored milk or frozen french fries due to non-convergence.

accept substitutes from unfamiliar brands than they are to accept substitutes from familiar brands (whose products they've purchased before). To quantify the trade-off between steering consumers' learning and maximizing the probability of acceptance, I estimate a learning model of differentiated products demand. Counterfactual simulations suggest that steering consumers' learning would prove an unprofitable strategy. Even so, the store could increase profits—and consumer welfare—by individualizing substitutions according to consumers' original orders, past purchases, and demographics.

A natural extension to this study concerns cross-category spillovers in consumers' learning. Many brands sell products in multiple categories, such as the store's private label (which competes in nearly every category of packaged food). Concerning such brands, what a consumer learns about the brand in one product category might also be informative of her tastes for the brand's products in other product categories. For instance, imagine that a stockout substitution causes a consumer to learn that she likes private-label ice cream. If she interprets this as a positive signal of her tastes for the private label as a whole, she might decide to try its offerings in other product categories—such as apple sauce cups—on subsequent shopping trips. Thus, to the extent that brands' retail margins are correlated across product categories, cross-category learning spillovers might increase the returns to steering consumers' learning.

More broadly, further research is needed on the extent to which firms can steer consumers' learning online. My findings suggest that supermarkets would struggle to profit from steering consumers' learning via stockout substitutions—a result that is reassuring as far as consumer welfare is concerned. However, the internet affords many other opportunities to direct consumers' learning. Take the case of web browsers, which are used to access important productivity software—word processors, spreadsheets, calendars, etc.—and to casually surf the web (Taivalsaari et al. 2008). Here, Microsoft leverages the popularity of its Windows operating system to encourage consumers to try its own browser, Edge, and to discourage them from experimenting with those of its competitors (Krasnoff 2022; Hollister 2023).⁶⁰ Another example concerns online shopping,

⁶⁰Microsoft sets Edge as the default browser on Windows 11 (Krasnoff 2022), so that web links and certain file types automatically open in Edge (unless consumers manually change the default browser). And when users try to download

where Google exploits its dominance in web search to promote its eponymous shopping service (Raedts and Evans 2024).⁶¹ Many of the affected consumers are, of course, happy with Edge or Google Shopping. Even so, some consumers might learn that they prefer alternatives—like Firefox or Bing Shopping, respectively—were they to try them. Future work could quantify the welfare effects of tech giants’ efforts to steer consumers’ learning about web browsers, online shopping, and other things.⁶²

the rival Chrome browser, they are first presented with a notice that Edge “. . . runs on the same tech as Chrome, with the added trust of Microsoft,” then asked to complete a poll about their reasons for downloading Chrome (Hollister 2023)

⁶¹When consumers make shopping-related searches, Google displays its own shopping service more prominently than those of its competitors (Raedts and Evans 2024).

⁶²Unlike packaged foods, there are adjustment costs associated with trying out new online software/services. (For instance, when a consumer experiments with a new web browser, she needs to determine where important functions are located in the interface.) These adjustment costs affect welfare analysis as follows. If tech firms stopped steering consumers’ learning, consumers might cross-shop online software/services more frequently. This would, in turn, increase the total adjustment costs incurred by consumers.

BIBLIOGRAPHY

- 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.
- Ackerberg, Daniel A. “Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination”. *International Economic Review* 44, no. 3 (2003): 1007–1040.
- 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.
- Allcott, Hunt, et al. *Sources of Market Power in Web Search: Evidence from a Field Experiment*. National Bureau of Economic Research, 2025.
- Allende, Claudia, Francisco Gallego, and Christopher Neilson. “Approximating the Equilibrium Effects of Informed School Choice”. Working paper, 2019. Visited on 10/28/2024.
- Anand, Bharat N., and Ron Shachar. “Advertising, the Matchmaker”. *The RAND Journal of Economics* 42, no. 2 (June 2011): 205–245.
- 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.
- Arteaga, Cristian, et al. “xlogit: An Open-Source Python Package for GPU-Accelerated Estimation of Mixed Logit Models”. *Journal of Choice Modelling* 42 (2022): 100339.
- 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.
- Backus, Matthew, Christopher Conlon, and Michael Sinkinson. *Common Ownership and Competition in the Ready-to-Eat Cereal Industry*. National Bureau of Economic Research, 2021. Visited on 04/02/2025.
- 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.
- Barahona, Nano, Cristóbal Otero, and Sebastián Otero. “Equilibrium Effects of Food Labeling Policies”. *Econometrica* 91, no. 3 (2023): 839–868.
- 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, and Philip Haile. “Identification in Differentiated Products Markets”. *Annual Review of Economics* 8, no. 1 (Oct. 31, 2016): 27–52.
- 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 T., and Philip A. Haile. “Foundations of Demand Estimation”. In *Handbook of Industrial Organization*, ed. by Kate Ho, Ali Hortaçsu, and Alessandro Lizzeri, 4:1–62. 2021.
- . “Identification in Differentiated Products Markets Using Market Level Data”. *Econometrica* 82, no. 5 (2014): 1749–1797.
- . “Nonparametric Identification of Differentiated Products Demand Using Micro Data”. *Econometrica* 92, no. 4 (2024): 1135–1162.
- Bradbury, James, et al. *JAX: Composable Transformations of Python + NumPy Programs*. Version 0.3.13, 2018.
- 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.
- Brick Meets Click and Mercatus. “February U.S. eGrocery Sales Total \$7.9 Billion, Down 10% versus Year Ago”. Brick meets click, Mar. 13, 2024. Press Release.
- 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.
- Che, Hai, Tülin Erdem, and T. Sabri Öncü. “Consumer Learning and Evolution of Consumer Brand Preferences”. *Quantitative Marketing and Economics* 13, no. 3 (Sept. 2015): 173–202.
- Chen, Nan, and Hsin-Tien Tsai. “Steering Via Algorithmic Recommendations”. *The RAND Journal of Economics* 55, no. 4 (Dec. 2024): 501–518.
- Ching, Andrew T. “A Dynamic Oligopoly Structural Model for the Prescription Drug Market After

- Patent Expiration*". *International Economic Review* 51, no. 4 (Nov. 2010): 1175–1207.
- Collard-Wexler, Allan. "Demand Fluctuations in the Ready-Mix Concrete Industry". *Econometrica* 81, no. 3 (2013): 1003–1037.
- Compiani, Giovanni, et al. "Online Search and Optimal Product Rankings: An Empirical Framework". *Marketing Science* 43, no. 3 (May 2024): 615–636.
- 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. 1, 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.
- Daljord, Øystein. "Durable Goods Adoption and the Consumer Discount Factor: A Case Study of the Norwegian Book Market". *Management Science* 68, no. 9 (2022): 6783–6796.
- 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.
- Donnelly, Robert, Ayush Kanodia, and Ilya Morozov. "Welfare Effects of Personalized Rankings". *Marketing Science* 43, no. 1 (Jan. 2024): 92–113.
- Dubé, Jean-Pierre, and Sanjog Misra. "Personalized Pricing and Consumer Welfare". *Journal of Political Economy* 131, no. 1 (2023): 131–189.
- Erdem, Tülin, and Michael P. Keane. "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets". *Marketing Science* 15, no. 1 (1996): 1–20.

- Erdem, Tülin, Michael P. Keane, and Baohong Sun. “A Dynamic Model of Brand Choice When Price and Advertising Signal Product Quality”. *Marketing Science* 27, no. 6 (2008): 1111–1125.
- 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.
- Farronato, Chiara, Andrey Fradkin, and Alexander MacKay. “Self-Preferencing at Amazon: Evidence from Search Rankings”. In *AEA Papers and Proceedings*, 113:239–243. American Economic Association, 2023.
- Farronato, Chiara, et al. “Understanding the Tradeoffs of the Amazon Antitrust Case”. *Harvard Business Review* (Jan. 11, 2024).
- 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).
- 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.
- Hollister, Sean. “Microsoft Now Thirstily Injects a Poll When You Download Google Chrome”. *The Verge*, Oct. 24, 2023.

- Iaria, Alessandro, and Ao Wang. “Real Analytic Discrete Choice Models of Demand: Theory and Implications”. *Econometric Theory* (2024): 1–49.
- Jovanovic, B. D., and P. S. Levy. “A Look at the Rule of Three”. *The American Statistician* 51, no. 2 (May 1997): 137–139.
- Kim, Kyoo il, and Amil Petrin. “Control Function Corrections for Unobserved Factors in Differentiated Product Models”. Working paper, 2019.
- Krasnoff, Barbara. “How to change your default browser in Windows 11”. *The Verge*, Apr. 15, 2022.
- 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.
- Nelson, Phillip. “Information and Consumer Behavior”. *Journal of Political Economy* 78, no. 2 (Mar. 1970): 311–329.
- Nevo, Aviv. “Measuring Market Power in the Ready-to-Eat Cereal Industry”. *Econometrica* 69, no. 2 (Mar. 2001): 307–342.
- Newell, Richard G., and Juha Siikamäki. “Nudging Energy Efficiency Behavior: The Role of Information Labels”. *Journal of the Association of Environmental and Resource Economists* 1, no. 4 (Dec. 2014): 555–598.
- Osborne, Matthew. “Consumer Learning, Switching Costs, and Heterogeneity: A Structural Examination”. *Quantitative Marketing and Economics* 9 (2011): 25–70.
- 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.
- Raedts, Elske, and Simone Evans. “Google Shopping: Self-Preferencing Can Be Abusive”. Stibbe, Feb. 10, 2024.
- Reimers, Imke, and Joel Waldfogel. *A Framework for Detection, Measurement, and Welfare Analysis of Platform Bias*. National Bureau of Economic Research, 2023.
- Revelt, David, and Kenneth Train. “Customer-Specific Taste Parameters and Mixed Logit”, vol. Working Paper No. E00-274, Department of Economics, University of California, Berkeley. 2000.
- Ryan, Stephen P. “The Costs of Environmental Regulation in a Concentrated Industry”. *Econometrica* 80, no. 3 (2012): 1019–1061.
- Shin, Sangwoo, Sanjog Misra, and Dan Horsky. “Disentangling Preferences and Learning in Brand Choice Models”. *Marketing Science* 31, no. 1 (Jan. 2012): 115–137.
- 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.
- Sullivan, Christopher. “The Ice Cream Split: Empirically Distinguishing Price and Product Space Collusion” (2020).
- Taivalsaari, Antero, et al. “Web Browser as an Application Platform”. In *2008 34th Euromicro Conference Software Engineering and Advanced Applications*, 293–302. 2008.
- 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. Bureau of Labor Statistics. *Consumer Price Index for All Urban Consumers (CPI-U)*.
- U.S. Food & Drug Administration. “Bottled Water Everywhere: Keeping it Safe”. Consumer Updates, Apr. 22, 2022.
- Vatter, Benjamin. “Quality Disclosure and Regulation: Scoring Design in Medicare Advantage”.

Working paper, 2024.

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.

Young, Liz. “Never Mind the Delivery, More Online Consumers Are Turning to Store Pickup”. *The Wall Street Journal* (July 14, 2023).

Zeyveld, Andrew. “Demand Estimation When Consumers’ Preferences Vary over Time”. Working Paper (2024).

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

APPENDIX 2A

DATA STRUCTURE AND OBSERVABLE CHARACTERISTICS

Illustrating the Structure of the Data.—In Section 2.2.2, I describe a hypothetical consumer who ordered Mott’s applesauce and Häagen-Dazs, only for the latter to go out of stock. Tables 2A.1 and 2A.2 portray what the curbside stockout data and scanner data would look like in this hypothetical case. Notice that the former lists the UPCs and product catalog descriptions of both the out-of-stock item and the substitute in our stylized example. However, the price of the out-of-stock product is missing (and must be imputed from other sales at the same store before and after the stockout, using the procedure described in Section 2.5.3).

As for the scanner data, Panels A and B of Table 2A.2 compare the contents when the consumer accepts and rejects the substitute eggs, respectively.

Demographic Data Details.—I employ the following procedure to recover consumers’ demographic information. For each transaction, the scanner data report two variables concerning the consumer: her loyalty ID, which serves as the primary panel identifier in my analysis; and her household ID, which maps to the demographic data.¹ When a given loyalty ID maps onto just one household ID, I assume that the consumer belongs to the household in question. Sometimes, however, a loyalty ID maps onto multiple household IDs. In that event, I compute a weighted

Table 2A.1: Curbside Stockout Data (Example)

	Out-of-Stock Product	Offered Substitute
UPC	71373312281	85808900305
Description	“HAAGEN DAZS VANILLA 14Z”	“HALO TOP ICE CREAM VANILLA LIGHT 16 OZ”
<i>Substitute Only</i>		
Price (\$)		3.79
Accepted?		Yes

Note: The (counterfactual) purchase price of the out-of-stock item is not recorded in the data. I impute it using the scanner data.

¹I rely on the chain’s loyalty program to track consumers’ purchases over time, rather than the household ID, because the chain judges the former to be a more reliable identifier of individual shoppers/households.

Table 2A.2: Scanner Data (Example)

<i>Panel A. Substitute is accepted.</i>						
UPC	Product catalog description	Price (\$)	Date	Store ID	Channel	Loyalty ID
1480000023	“MOTTTS APPLESAUCE CINNAMON 6/4 OZ”	3.35	01/01/21	21	Pickup	12345
85808900305	“HALO TOP ICE CREAM VANILLA LIGHT 16 OZ”	3.79	01/01/21	21	Pickup	12345
<i>Panel B. Substitute is rejected.</i>						
1480000023	“MOTTTS APPLESAUCE CINNAMON 6/4 OZ”	3.35	01/01/21	21	Pickup	12345

average of the demographics associated with the household IDs in question (where the weight is given by the number of transactions in the scanner data with the relevant household ID). Finally, because the demographic data were collected in 2014, I lack demographic information on some consumers’ households. Such consumers are excluded from the structural analysis in Sections 2.5 and 2.7.

The chain transitioned to a new household ID system during the time period studied. Before August 2017, the scanner data only report the old household ID; from August 2017 to April 2021, the scanner data indicate both the old and new household IDs; and from May 2021 onwards, the scanner data contain only the new household IDs. Seeing as the demographic data are organized around the old household IDs, I adopt the following procedure to impute the household demographics associated with a given loyalty ID. If the loyalty ID appears in one or more transactions where the “old household IDs” are observed (i.e., before April 2021), I impute the consumer’s demographics as a weighted average of the demographics ascribed to the relevant “old household IDs” (following the procedure in Section 2.2.3). Rarely, a loyalty ID is solely observed in transactions after May 2021—which only contain “new IDs”—and yet the relevant “new ID(s)” themselves appear in (other) transactions that are old enough to also have “old IDs.” In such cases, for each “new ID,” I take a weighted average of the demographics associated with all the “old household IDs” with which the “new ID” appears. (The weights are, once more, based on the number of transactions.) Finally, the demographics associated with the loyalty ID are imputed as being a transaction-weighted average of the (imputed) demographics associated with the relevant “new IDs.”

Table 2A.3: State Dependence in Brand, Product, and Channel Choice

In consecutive trips, prob. of the same. . .	<i>Panel A. Overall</i>			
	Apple sauce cups	Flavored milk	Frozen french fries	Ice cream
Product being purchased	0.612	0.603	0.364	0.271
Brand being purchased	0.805	0.771	0.692	0.542
Shopping channel	0.862	0.857	0.850	0.907
	<i>Panel B. Conditional on present trip being curbside pickup</i>			
Product being purchased	0.600	0.663	0.379	0.331
Brand being purchased	0.775	0.826	0.698	0.626
Shopping channel	0.793	0.738	0.746	0.778

Notes: Estimates are reported as means. In curbside pickup, when there is a stockout substitution, I define the “purchased product” as being the stockout substitute.

State Dependence in Product, Brand, and Channel Choice.—Do consumers tend to purchase the same products in consecutive trips? Or at least products of the same brand? And how often do consumers switch shopping channels (i.e., in-store shopping versus curbside pickup versus home delivery)?

To provide insight, Table 2A.3 reports the probability of repeated product, brand, and shopping channel choices—both overall, and conditional on the present trip being curbside pickup. Focus first on the overall results, which are presented in Panel A. There are meaningful cross-category differences in the probability of purchasing the same product on consecutive trips; whereas there is a 61% probability that a consumer purchases the same apple sauce cups on consecutive shopping trips, there is only a 36.4% (27.1%) that she does the same with respect to flavored french fries (ice cream). However, in all four categories, a consumer is likely to purchase products that are sold under the same brands on consecutive trips, with probabilities ranging from 54.2% (ice cream) to 80.5% (apple sauce cups). Furthermore, these purchases tend to be made through the same shopping channel. Across the three product categories, between 85% and 91% of consumers select the same shopping channel on consecutive trips.

Do consumers display more, or less, state dependence after a curbside pickup order? Panel B suggests that consumers’ behavior evinces a similar degree of state dependence following curbside

Table 2A.4: Summary Statistics by Product Category

Statistic	Apple sauce cups	Flavored milk	Frozen french fries	Ice cream
No. of stockout events	7332	14,710	28,885	66,635
Median upper bound on duration (hours)	130.7	60.4	123.9	148.7

pickup versus in-store shopping or home delivery. The most perceptible difference concerns the choice of shopping channel. If a consumer has placed an order for curbside pickup, the probability that her next shopping trip shares the same channel (namely, curbside pickup) drops to 79% or less across the three product categories (compared to the unconditional probability of repeat channel choices of 85% or more, depending on the product category).

Frequency and Duration of Stockout Events.—When multiple consumers order the same product from the same store at roughly the same time, a single stockout event can result in more than one stockout substitution. How often do stockouts occur, and how long do they last? To answer these questions, I join the curbside stockout data with the scanner data and then sort the combined data set by store, product, and date. For each store-product pairing in the resulting data set, I observe sequences of successful purchases (from the scanner data), interspersed with sequences of stockout substitutions (from the curbside stockout data). Treating the former as evidence that the product is in stock and the latter as evidence of stockout, I identify the *last* successful purchase before each stockout event as well as the *first* successful purchase afterwards. By computing the time elapsed between these two successful purchases, I obtain an upper bound on the duration of the stockout event.

Panel D reports the results of this descriptive exercise. The total number of stockout events varies across product categories, ranging from seven thousand (flavored milk) to sixty-seven thousand (ice cream). The median upper bound on the duration of an individual stockout event is between sixty and one-hundred forty-nine hours.²

²I report the median, not the mean, because some “stockouts” are of such long duration that they are probably not stockouts per se. Rather, the store has likely dropped the product in question for several months and then reintroduced it.

APPENDIX 2B

ADDITIONAL DESCRIPTIVE EVIDENCE

Reduced-Form Evidence on the Acceptance or Rejection of Substitutes.—Here, I characterize the circumstances under which consumers accept or reject substitutes in the product categories of apple sauce cups, flavored milk, and frozen french fries. For each product category, Table 2B.1 reports the average marginal effects from Equation (2.1). Across all the product categories, consumers are much likelier to accept substitutes whose brands they have previously purchased. As for non-brand characteristics, some of these loom larger than others. For instance, consumers are 9 percentage points likelier to accept substitute flavored milks that share the same high protein–status as the out-of-stock product.

Supplementary Evidence of Stockout Substitutions’ Influence on Consumers’ Learning.—The results in Table 2.3 suggest that stockout substitutions sometimes influence consumers’ purchases through the mechanism of learning. This is because the future purchases of the “focal consumers” (who suffer stockout substitutions and, in consequence, can learn about the substitute’s characteristics) differ from the future purchases of the “control consumers” (who order the same products as the focal consumers, but successfully pick up and thus do not learn about the substitute).

That the focal consumers proceed to purchase the substitute’s brand more often in the future than do their “control” counterparts is consistent with the former’s learning about the brand of the substitute. Specifically, some focal consumers may be discovering that they like the substitute’s brand more than they had anticipated and, as a result, purchasing that brand on subsequent shopping trips. However, other factors could also explain the differences between focal and control consumers. One such factor is the “buy it again” feature of the online order system. When consumers visit the store’s website or mobile app, consumers are presented with a list of items that they have purchased on previous shopping trips—any of which can be ordered again with a single click. (By contrast, ordering an item outside this list requires multiple steps; see Section 2.2.2.) To test whether the “buy it again” list is responsible for the disparity between focal and control consumers, I repeat the descriptive exercise with one modification. Rather than comparing focal and control consumers

Table 2B.1: Acceptance: Average Marginal Effects from Probit Regressions

Variable	Product category		
	Apple sauce cups	Flavored milk	Frozen french fries
Brand			
Sub shares OOS product's brand	−0.099*** [0.029]	0.030*** [0.006]	0.012*** [0.003]
Ever purchased sub's brand before	0.059*** [0.012]	0.044*** [0.006]	0.031*** [0.003]
Fruit			
Ever purchased sub's fruit before	0.023 [0.025]		
Seasoning			
Ever purchased sub's seasoning before	−0.002 [0.017]		
No. of cups ^a			
Sub shares OOS product's no. of cups	−0.298*** [0.083]		
Ever purchased sub's no. of cups before	0.005 [0.031]		
Sweetening			
Sub shares OOS product's sweetening	−0.025 [0.029]		
Ever purchased sub's sweetening before	0.031* [0.015]		
Pct. milkfat			
Sub shares OOS product's pct. milkfat		0.047*** [0.006]	
Ever purchased sub's pct. milkfat before		0.031*** [0.006]	
Whether hi-protein			
Sub shares OOS product's whether hi-protein		0.087*** [0.020]	
Ever purchased sub's whether hi-protein before		0.004 [0.020]	
Size ^{and}			
Sub shares OOS product's size		−0.024** [0.008]	0.010 [0.008]
Ever purchased sub's size before		−0.006 [0.006]	0.007 [0.008]
Base vegetable			
Sub shares OOS product's base vegetable			0.125*** [0.013]
Ever purchased sub's base vegetable before			−0.033*** [0.008]
Sub's price	−0.056*** [0.014]	−0.030*** [0.003]	−0.029*** [0.003]
OOS product's price	−0.030*** [0.008]	0.013*** [0.003]	0.006* [0.003]
Observations	2,052	15,667	31,157
Pseudo R^2	0.1076	0.0577	0.0215

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 brackets.

^a Discretized.

* Significant at the 10 percent level.

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

Table 2B.2: Model-Free Evidence of Learning: Apple Sauce Cups

Consumer's "treatment"	No. of purchases				Pct. of future purchases with sub's version of characteristic	
	Before stockout		After stockout			
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
<i>Panel A. Characteristic of being (un)sweetened (80 obs.)</i>						
Suffer substitution (focal group)	11.9	18.4	10.4	16.2	11.3	25.3
	[0.2]	[0.4]	[0.2]	[0.3]	[0.3]	[0.4]
Successful pickup (control group)	15.8	26.5	10.6	14.8	4.3	10.8
	[0.3]	[0.6]	[0.2]	[0.4]	[0.1]	[0.2]
<i>Panel B. Characteristic of brand (95 obs.)</i>						
Suffer substitution (focal group)	13.8	25.2	11.3	19.2	13.7	27.4
	[0.3]	[0.6]	[0.2]	[0.4]	[0.3]	[0.4]
Successful pickup (control group)	13.3	28.3	12.0	22.9	11.1	22.8
	[0.3]	[0.6]	[0.2]	[0.5]	[0.2]	[0.3]
<i>Panel C. Characteristic of fruit (25 obs.)</i>						
Suffer substitution (focal group)	24.3	35.1	15.1	21.4	3.5	11.9
	[1.4]	[2.0]	[0.8]	[1.3]	[0.5]	[1.0]
Successful pickup (control group)	23.1	38.4	16.1	19.8	5.3	12.5
	[1.5]	[2.7]	[0.7]	[1.2]	[0.5]	[0.7]

Table 2B.3: Model-Free Evidence of Learning: Flavored milk

Consumer's "treatment"	No. of purchases				Pct. of future purchases with sub's version of characteristic	
	Before stockout		After stockout			
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
<i>Panel A. Characteristic of brand (572 obs.)</i>						
Suffer substitution (focal group)	23.6	41.8	19.8	24.4	5.6	17.3
	[0.1]	[0.2]	[0.0]	[0.1]	[0.0]	[0.1]
Successful pickup (control group)	24.7	42.2	20.5	26.1	3.7	10.6
	[0.1]	[0.1]	[0.0]	[0.1]	[0.0]	[0.0]
<i>Panel B. Characteristic of pct. milkfat (195 obs.)</i>						
Suffer substitution (focal group)	16.3	28.7	14.2	15.9	14.1	25.8
	[0.1]	[0.3]	[0.1]	[0.1]	[0.1]	[0.2]
Successful pickup (control group)	16.6	31.5	16.2	22.6	8.1	17.0
	[0.2]	[0.4]	[0.1]	[0.2]	[0.1]	[0.1]
<i>Panel C. Characteristic of size (150 obs.)</i>						
Suffer substitution (focal group)	10.8	20.4	15.4	19.6	15.6	25.2
	[0.1]	[0.4]	[0.1]	[0.2]	[0.2]	[0.2]
Successful pickup (control group)	13.6	30.3	18.9	23.7	12.1	22.2
	[0.2]	[0.6]	[0.2]	[0.2]	[0.1]	[0.2]

Table 2B.4: Model-Free Evidence of Learning: Flavored milk

Consumer's "treatment"	No. of purchases				Pct. of future purchases with sub's version of characteristic	
	Before stockout		After stockout			
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
<i>Panel A. Characteristic of brand (525 obs.)</i>						
Suffer substitution (focal group)	17.5 [0.0]	21.9 [0.1]	11.1 [0.0]	12.4 [0.0]	8.8 [0.0]	21.0 [0.1]
Successful pickup (control group)	18.1 [0.0]	25.7 [0.1]	10.3 [0.0]	11.0 [0.0]	4.9 [0.0]	15.4 [0.1]
<i>Panel B. Characteristic of flavor (74 obs.)</i>						
Suffer substitution (focal group)	12.5 [0.3]	22.0 [0.6]	9.5 [0.2]	15.2 [0.5]	15.4 [0.4]	27.4 [0.4]
Successful pickup (control group)	11.8 [0.2]	16.6 [0.4]	13.3 [0.2]	17.7 [0.4]	19.1 [0.4]	32.4 [0.4]
<i>Panel C. Characteristic of size (75 obs.)</i>						
Suffer substitution (focal group)	18.1 [0.3]	24.2 [0.7]	10.6 [0.1]	10.7 [0.1]	2.4 [0.1]	10.4 [0.5]
Successful pickup (control group)	25.4 [0.5]	39.3 [1.5]	11.7 [0.2]	14.2 [0.3]	0.5 [0.0]	2.2 [0.1]

with respect to all subsequent purchase—both online and offline—I instead focus solely on in-store purchases. If the disparity between focal and control consumers is entirely driven by the “buy it again” list (as opposed to learning), the disparity should disappear once analysis is confined to in-store purchases (where the “buy it again list” is irrelevant). Table 2B.5 presents the results of this robustness check (where, for brevity, I only report results for the characteristic of brand). Although the sample sizes shrink dramatically, the focal consumers still purchase the substitute’s brand more frequently than do their control counterparts

There may also be underlying differences between the focal and control consumers. In particular, the focal consumers have, by construction, arrived at the store later than their control counterparts (as the stockout occurred in the interim). Could the pickup time be correlated with differential trends in future purchases? Such a correlation might arise if, for instance, the pickup time were associated with consumers’ inclination to try out new products. To test for the presence of any such compositional differences between focal and control consumers, I repeat the descriptive exercise above with one modification: I now define the control consumer as the *first* consumer to successfully pick up the focal consumer’s preferred product after it goes out of stock (from among the subset

Table 2B.5: Model-Free Evidence of Learning About Brands: A Comparison of Future In-Store Purchases

Consumer's "treatment"	No. of purchases				Pct. of future purchases with sub's version of characteristic	
	Before stockout		After stockout			
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
<i>Panel A. Apple sauce cups</i>						
Suffer substitution (focal group)	13.1 [0.6]	16.6 [0.9]	5.6 [0.3]	7.3 [0.4]	20.8 [1.3]	38.9 [1.1]
Successful pickup (control group)	17.0 [0.7]	20.5 [0.8]	3.8 [0.1]	3.1 [0.1]	13.7 [1.0]	28.3 [1.2]
<i>Panel B. Flavored milk</i>						
Suffer substitution (focal group)	27.6 [0.2]	43.5 [0.2]	12.9 [0.1]	17.9 [0.1]	5.2 [0.1]	15.4 [0.1]
Successful pickup (control group)	30.0 [0.2]	48.7 [0.5]	13.0 [0.1]	17.3 [0.1]	5.3 [0.1]	16.4 [0.2]
<i>Panel C. Frozen french fries</i>						
Suffer substitution (focal group)	22.3 [0.1]	25.3 [0.1]	7.6 [0.0]	9.2 [0.1]	10.0 [0.1]	23.5 [0.2]
Successful pickup (control group)	19.0 [0.1]	23.9 [0.2]	5.4 [0.0]	6.2 [0.0]	5.7 [0.1]	16.7 [0.1]
<i>Panel D. Ice cream</i>						
Suffer substitution (focal group)	25.5 [0.2]	32.4 [0.4]	13.1 [0.1]	15.4 [0.2]	5.5 [0.1]	16.0 [0.3]
Successful pickup (control group)	33.2 [0.3]	38.6 [0.3]	16.4 [0.2]	22.2 [0.3]	5.0 [0.1]	14.4 [0.2]

Notes: This table checks whether the results in Tables 2B.2 and 2.3 are robust to focusing only on consumers' future *in-store* purchases. (Unlike order for curbside pickup, in-store purchases are not directly affected by the "buy-it-again" feature of the store's app and website.)

of consumers who, like the focal consumer, have never purchased the substitute's version of the relevant characteristic before).¹ Thus, the focal consumer's order must have been assembled *before* the control consumer's, so that either (a) the focal consumer placed her order earlier than did the control consumer or (b) the focal consumer's stated pickup time was earlier than the control consumer's. As a result, any compositional differences between focal and control consumers that are rooted in order or pickup times should be reversed. Reassuringly, the results—which are presented in Table 2B.6—prove qualitatively similar to the ones above.

Determinants of Retail Margins (Additional Categories).—Figure 2B.1 summarizes the results

¹In principle, this robustness check (unlike the main descriptive exercise above) is vulnerable to endogenous price changes. Specifically, the store might respond to a product's going out of stock by raising the price. This could cause the control consumer to face a different price from the focal consumer.

Table 2B.6: Model-Free Evidence of Learning About Brands: Robustness Check (“First After”)

Consumer’s “treatment”	No. of purchases				Pct. of future purchases with sub’s version of characteristic	
	Before stockout		After stockout			
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
<i>Panel A. Apple sauce cups</i>						
Suffer substitution (focal group)	8.8 [0.0]	16.2 [0.0]	8.9 [0.0]	12.7 [0.0]	18.9 [0.0]	31.1 [0.0]
Successful pickup (control group)	11.0 [0.0]	20.7 [0.1]	8.9 [0.0]	12.4 [0.0]	13.4 [0.0]	25.9 [0.0]
<i>Panel B. Flavored milk</i>						
Suffer substitution (focal group)	20.1 [0.0]	36.1 [0.0]	17.2 [0.0]	22.6 [0.0]	9.1 [0.0]	22.3 [0.0]
Successful pickup (control group)	22.3 [0.0]	41.6 [0.0]	18.0 [0.0]	24.6 [0.0]	6.1 [0.0]	17.0 [0.0]
<i>Panel C. Frozen french fries</i>						
Suffer substitution (focal group)	15.0 [0.0]	22.7 [0.0]	9.7 [0.0]	12.1 [0.0]	11.0 [0.0]	23.4 [0.0]
Successful pickup (control group)	15.9 [0.0]	25.0 [0.0]	10.1 [0.0]	12.9 [0.0]	7.1 [0.0]	18.7 [0.0]
<i>Panel D. Ice cream</i>						
Suffer substitution (focal group)	22.2 [0.0]	36.2 [0.0]	18.3 [0.0]	23.6 [0.0]	6.2 [0.0]	15.8 [0.0]
Successful pickup (control group)	28.4 [0.0]	45.8 [0.1]	21.1 [0.0]	27.8 [0.0]	4.3 [0.0]	13.2 [0.0]

Notes: This table examines whether the results in Table 2.3 are robust to considering a different population of “control consumers.” Although the control consumer is drawn from the same pool of potential control consumers as in Table 2.3, here I select the first consumer to successfully pick up *after* the stockout event.

^a Binned (small/medium/large)

^b Binned (less than 100 cal; between 100 and 200 cal; more than 200 cal)

of descriptive regressions concerning retail margins in the product categories of apple sauce cups, flavored milk, and frozen french fries. As with the product category of ice cream, the characteristic of brand proves to be a key determinant of retail margins.

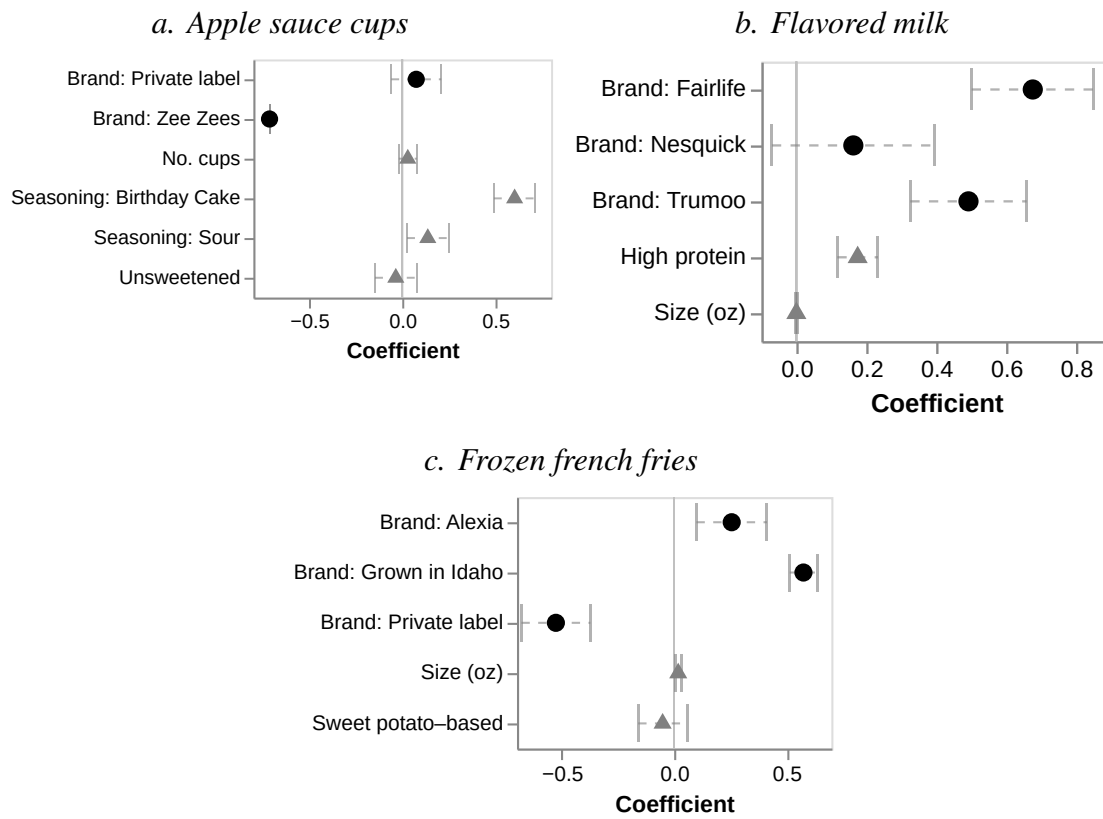


Figure 2B.1: Determinants of Retail Margins

Notes: This figure plots estimates of the coefficients (γ) on products' observable characteristics using the specification in Equation (2.2). The horizontal bars provide 95% confidence intervals.

APPENDIX 2C

ESTIMATION DETAILS

Simulated Likelihood Function.—I employ maximum simulated likelihood estimation to recover the parameters. The likelihood function is based on the probability of the consumer's ordering a particular good, as well as the probability of her accepting a specific substitute. Both those probabilities, in turn, depend on the goods' expected utilities at time t . However, the explanatory variables used in this learning model differ somewhat from those in a traditional mixed (or “random coefficients”) logit model. Thus, I begin my derivation of the likelihood by showing how to compute the goods' expected utilities as a function of (a) the parameters indexing the distributions of consumer tastes and learning, as discussed above; and (b) consumers' observed choices in the data.

Equation (2.10) gives the consumer's expected utility of good j at time t , conditional on the set $\mathcal{I}_{it}$ of brands for which she fully knows her taste. All quantities in equation (2.10) are fully known to the consumer, with the possible exception of her time- t expected taste for good j 's brand. This can be written as

$$E[v_{iB(j)} \mid \mathcal{I}_{it}] = \underbrace{\mu_{iB(j)}}_{\text{prior expected taste}} + \underbrace{(v_{iB(j)} - \mu_{iB(j)}) 1[B(j) \in \mathcal{I}_{it}]}_{\text{learning “correction” (if brand was previously purchased)}} \quad (2C.1)$$

Here the indicator variable $1[B(j) \in \mathcal{I}_{it}]$ equals one if (and only if) the consumer knows her taste for brand $B(j)$ at time t . Until she purchases the brand for the first time, she does *not* fully know her taste for it and must, instead, rely on her prior expected taste $\mu_{iB(j)}$. But upon her first purchase of the brand, she learns the degree to which her true taste $v_{iB(j)}$ differs from her prior expected taste $\mu_{iB(j)}$.

In order to take equation (2C.1) to the data, observe that prior expected tastes $\mu_{iB(j)}$ can be computed as the product of

- (i) a $1 \times B$ vector of brand dummy variables, $(1[B(j) = 1], \dots, 1[B(j) = B])^\top$; and
- (ii) a $B \times 1$ vector of prior expected brand tastes, $(\mu_{i1}, \dots, \mu_{iB})$.

This is true because

$$\begin{aligned}\mu_{iB(j)} &= \sum_{b=1}^B 1[B(j) = b] \cdot \mu_{ib} \\ &= \begin{pmatrix} 1[B(j) = 1] & \cdots & 1[B(j) = B] \end{pmatrix} \cdot \begin{pmatrix} \mu_{i1} \\ \vdots \\ \mu_{iB} \end{pmatrix}\end{aligned}\quad (2C.2)$$

The “learning correction” $(v_{iB(j)} - \mu_{iB(j)})$ can be calculated similarly. Here, the explanatory variables must account for the fact that the learning correction remains latent until the consumer buys the brand for the first time (formally, until $B(j) \in \mathcal{I}_{it}$). I therefore compute the learning correction as

- (i) a $1 \times B$ vector of indicator variables, $(1[B(j) = 1 \text{ and } 1 \in \mathcal{I}_{it}], \dots, 1[B(j) = B \text{ and } B \in \mathcal{I}_{it}])^\top$, such that entry b equals one if b is j ’s brand and also b is a brand the consumer has previously purchased (i.e., $b \in \mathcal{I}_{it}$); and
- (ii) a $B \times 1$ vector of the consumer’s “learning shocks,” $(v_{i1} - \mu_{i1}, \dots, v_{iB} - \mu_{iB})^\top$.

This representation is accurate because

$$\begin{aligned}v_{ib} - \mu_{ib} &= \sum_{b=1}^B 1[B(j) = b \text{ and } b \in \mathcal{I}_{it}] (v_{ib} - \mu_{ib}) \\ &= \begin{pmatrix} 1[B(j) = 1 \text{ and } 1 \in \mathcal{I}_{it}] & \cdots & 1[B(j) = B \text{ and } B \in \mathcal{I}_{it}] \end{pmatrix} \begin{pmatrix} v_{i1} - \mu_{i1} \\ \vdots \\ v_{iB} - \mu_{iB} \end{pmatrix}\end{aligned}\quad (2C.3)$$

Importantly, the learning correction $(v_{ib} - \mu_{ib})$ has a mean of zero for all brands b . This follows from the fact that the consumer’s prior expectation μ_{ib} on her taste for b is unbiased. (Recall that her true taste v_{ib} is drawn directly from her prior, which is normally distributed with mean μ_{ib} .) As a result, there is only one parameter to be estimated in connected with the learning correction: its standard deviation ι_b^2 .

Unlike the random coefficients pertaining to brands, the remaining ones can be recovered with usual procedure employed in mixed (or “random-coefficients”) logit, with x_j , p_{jt} and ξ_{jt} as explanatory variables.

The complete set of explanatory variables for good j can be represented by the vector

$$w_{jt} := \begin{pmatrix} \left(1[B(j) = b] \right)_{b=1}^B \\ \left(1[B(j) = 1 \text{ and } 1 \in \mathcal{I}_{it}], \dots, 1[B(j) = B \text{ and } B \in \mathcal{I}_{it}] \right) \\ x_j \\ p_{jt} \\ 1[\text{before Jan. 2021}] \cdot \tilde{\xi}_{jt} \\ 1[\text{after Jan. 2021}] \cdot \tilde{\xi}_{jt} \\ 1[j = 0] \cdot 1[\text{reject in-person}] \end{pmatrix}$$

while the complete set of parameters can be written as

$$\chi_i := \begin{pmatrix} (\mu_b)_{b=1}^B \\ (v_b - \mu_b)_{b=1}^B \\ (\beta, \sigma_\beta^2) \\ (\alpha, \sigma_\alpha^2) \\ \lambda_{\text{pre-21}} \\ \lambda_{\text{post-21}} \\ \gamma \end{pmatrix}$$

Having written the expected utility of each good j as a function of the parameters to be estimated, as well as the data, I can now derive a parsimonious expression of the (simulated) likelihood function used in estimation. My estimation code borrows from Arteaga et al. (2022); while my exposition here borrows from the same, along with Train (2009). Before elaborating on the mechanics of estimation, I will introduce additional notation concerning an individual consumer’s orders, substitutions, and learning. In reference to orders, let y_{ijt} equal one if consumer i orders good j in trip t , and zero otherwise. Likewise, in reference to substitutions, let a_{ijt} equal one if either (a) consumer i accepts good j as a substitute at time t , or (b) she is not offered j as a substitute

at time t .¹ If neither (a) nor (b) hold—in other words, if the consumer has, in fact, been offered j' as a substitute and proceeded to reject it—then $a_{ij't}$ equals zero.

Take as given that consumer i has taste and learning parameters χ . Then, according to the familiar conditional logit formula, the probability that she orders good j at time t is

$$\begin{aligned} P_{ijt} \mid \chi &:= \Pr \left[j = \arg \max_{j \in \mathcal{J}_t} E[u_{ijt}] \mid w_t; \chi \right] \\ &= \frac{\exp(w_{jt}\chi)}{\sum_{j' \in \mathcal{J}_t} \exp(w_{j't}\chi)} \end{aligned}$$

while her probability of accepting the good as a substitute is given by

$$\begin{aligned} P_{ijt}^A \mid \chi &:= \Pr \left[E[u_{ijt}] > u_{i0t} \mid w_t; \chi \right] \\ &= \frac{\exp(w_{jt}\chi)}{1 + \exp(w_{jt}\chi)} \end{aligned}$$

However, due to the panel structure of the data, the consumer may make a sequence of multiple orders and substitution decisions. The probability of observing a given sequence takes the form

$$P_i \mid \chi := \prod_{t \in \mathcal{T}} \prod_{j \in \mathcal{J}_t} (P_{ijt} \mid \chi)^{y_{ijt}} (P_{ijt}^A \mid \chi)^{a_{ijt}}$$

In reality, though, the consumer's individual taste coefficients are not observed by the econometrician. The unconditional choice-sequence probability P_i is obtained by integrating over the distribution of tastes across the population of consumers:

$$P_i := \int (P_i \mid \chi) f_\chi(\chi) d\chi \tag{2C.4}$$

Here $f_\chi(\cdot)$ denotes the probability density function (PDF) of the parameters χ . (Recall that these include the consumer's prior expected brand tastes [the μ_{ib} 's], her learning shocks [the $(v_{ib} - \mu_{ib})$'s], etc.)

As I previously mentioned, equation Equation (2C.4) does not possess a closed form, and must therefore be simulated. I do this with R random draws, indexed $r \in \{1, \dots, R\}$. For each draw r , I

¹Either because she successfully picks up her original order (whether j or some other good), or because she is offered some other good j' as a substitute.

draw a vector χ_r from $f_\chi(\chi)$ and then compute the choice probabilities conditional on χ_r , denoted $P_i \mid \chi_r$.

After conducting R draws and computing the resulting conditional choice probabilities, the simulated *unconditional* choice-sequence probability $\check{P}_i$ is computed as the average of the conditional choice probabilities:

$$\check{P}_i = \frac{1}{R} \sum_{r=1}^R (P_i \mid \chi^r) \quad (2C.5)$$

For computational efficiency, this simulation is conducted simultaneously for all consumers i . The likelihood function is then computed as the product of the consumers' respective choice probabilities;

$$\check{\mathcal{L}} = \prod_{i \in \mathcal{N}} \check{P}_i$$

Calculating Consumer Surplus.—Suppose that consumer i has been offered good s as a stockout substitute for her preferred good $j^\star$ at time t . Her expected present-trip surplus comes to

$$E[CS_{it}] = E \left[\frac{1}{\alpha_i} \log \left(\exp(r_{ist}^E(\mathcal{I}_{it})) + 1 \right) \mid \text{order } j^\star; \mathcal{H}_{it}, \mathcal{D}_i \right] + C.$$

In this equation, $r_{ist}^E := E[u_{ist} \mid \mathcal{I}_{it}] - \varepsilon_{ist}$ denotes the expected representative utility of the substitute s at time t , while $\mathcal{H}_{it}$ and $\mathcal{D}_i$ respectively denote the consumer's purchase history and household demographics. Finally, C is an unknown constant emphasizing that the absolute magnitude of utility is not identified (Train [2009]). This probability is simulated using a similar approach to that employed during estimation.

Now turn to future shopping trips. Conditional on acceptance, the present-discounted value of expected future surplus is given by

$$V(\text{accept } s) := \sum_{t'=t+1}^T E \left[\frac{1}{\alpha_i} \log \left(\sum_{j' \in \mathcal{J}_t} \exp(r_{ij't'}^E(\mathcal{I}_{it'})) \right) \mid \text{order } j^\star; \mathcal{H}_{it}, \mathcal{D}_i \right] + C.$$

This probability, too, must be simulated.

APPENDIX 2D

ESTIMATION RESULTS FOR APPLE SAUCE CUPS

Table 2D.1 reports the parameter estimates for the product categories of apple sauce cups. There are fewer demographic interactions and random coefficients than for ice cream due to challenges with convergence.

Table 2D.1: Parameter Estimates

<i>Panel A. Brands</i>					
	Mean exp. tastes (μ_b 's)	Heterogeneity of exp. tastes (σ_b 's)	Amount of learning (ι_b 's)		
Mott's	6.150 [0.088]	0.816 [0.029]	0.399 [0.016]		
Private label	5.701 [0.088]	2.578 [0.025]	0.107 [0.014]		
<i>Panel B. Non-brand observables and prices</i>					
	Means (β 's or α)	Std. devs. (σ_β 's or σ_α)	Interactions with demographics		
			Household income	Household size	Age of oldest HH male ^a
Fruit: blueberry	−5.650 [0.165]	4.037 [0.098]			
Fruit: cherry	−6.361 [0.278]	4.416 [0.149]			
Fruit: mixed fruit	−4.686 [0.114]	2.803 [0.062]			
Fruit: peach & mango	−6.050 [0.227]	3.178 [0.101]			
Fruit: strawberry & kiwi	−3.458 [0.166]	2.479 [0.109]			
Unsweetened	−0.276 [0.010]				
No. of cups	0.179 [0.007]		0.002 [0.001]		
Price ^b	0.171 [0.022]	0.595 [0.017]			
<i>Panel C. Other explanatory variables</i>					
	Coefficients (λ 's or γ)				
Control function (pre-2021) ^c	0.852 [0.050]				
Control function (post-2021) ^c	0.820 [0.043]				
Reject in-person ^d	0.792 [0.125]				

Notes: estimates are based on 57,811 randomly-sampled observations involving 2048 households. Standard errors (in brackets) do not correct for measurement error in the control function.

^a The random price coefficients α_i are assumed to follow a log-normal distribution.

^b The demand shocks are specified as $\xi_{jt} = \lambda \tilde{\xi}_{jt}$, where $\tilde{\xi}_{jt}$ is the residual from the pricing function and λ is a scaling parameter (reported here). This control function is computed separately before/after January 2021, due to a

change in the store's internal cost measure.

^c Until September 2021, consumers accepted or rejected stockout substitutes upon arrival at the store. Starting September 2021, they could accept or reject substitutes remotely (using the store's app or website).

APPENDIX 2E

SUPPLEMENTARY COUNTERFACTUAL SIMULATIONS

Explaining the Negligible Returns to Steering Consumers' Learning.—To help explain the unprofitability of steering consumers' learning, Table 2E.1 reports the present-discounted value of future and total profits under counterfactual changes to the purchase environment or the primitives of consumers' learning. Panels A and B show that future profits would remain identical across all stockout substitution policies if there were no endogenous learning or if the store possessed perfect foresight about products' future prices, wholesale costs, and availabilities. Panel C reveals that even if consumers were willing to accept whatever substitute the store offered, the store's expected future profits would remain unchanged from the baseline. Panel D then considers outcomes if consumers experienced three times as much learning as they do in reality.¹ Future profits drop overall, likely because consumers who formerly purchased Ben & Jerry's or Halo top "discovered" that they actually preferred the comparatively low-margin—and inexpensive—Häagen-Dazs brand. But the present-discounted value of future profits remains identical across the baseline policy and the two reasonable counterfactual policies (although future profits do increase by a cent under the purely-illustrative policy tailored to maximize future profits alone). Finally, Panel E indicates that if consumers were guaranteed to accept and if they experienced three times more learning when they tried new brands, the present-discounted value of expected future profits would increase by a cent under the counterfactual policies designed to maximize present-trip profits or total discounted profits compared to the baseline.

Counterfactual Simulations for Apple Sauce Cups.—Tables 2E.2 and 2E.3 compare profits and consumer welfare, respectively, under the store's baseline substitution policy and several counterfactual ones. See Section 2.7.4 for discussion.

¹Here, I triple the magnitude of the $(v_b - \mu_b)$ "learning correction" parameters before performing the simulation.

Table 2E.1: Profits Under Changes to the Purchase Environment or Model Primitives

	Counterfactual policies (“fully individualized” by original order, past purchases, and household demographics)			
	Baseline	Max. present- trip profits	Max. PDV total profits	Max. PDV future profits
<i>Panel A. No endogenous learning</i>				
PDV future profits (given reject)	17.14 (24.96)	17.14 (24.96)	17.14 (24.96)	17.14 (24.96)
PDV total profits	19.33 (24.93)	19.69 (24.94)	19.69 (24.94)	18.96 (24.95)
<i>Panel B. Perfect foresight</i>				
PDV future profits	17.52 (22.32)	17.52 (22.32)	17.52 (22.32)	17.52 (22.32)
PDV total profits	19.70 (22.29)	20.13 (22.35)	20.13 (22.35)	19.08 (22.29)
<i>Panel C. Guaranteed acceptance</i>				
PDV future profits	17.16 (25.02)	17.16 (25.02)	17.16 (25.03)	17.16 (25.03)
PDV total profits	19.86 (25.01)	20.29 (25.03)	20.29 (25.03)	18.40 (25.07)
<i>Panel D. Three times more learning</i>				
PDV future profits	14.22 (21.08)	14.22 (21.08)	14.22 (21.08)	14.23 (21.08)
PDV total profits	16.41 (21.09)	16.77 (21.09)	16.77 (21.09)	15.72 (21.11)
<i>Panel E. Guaranteed acceptance and three times more learning</i>				
PDV future profits	14.22 (21.08)	14.23 (21.08)	14.23 (21.08)	14.23 (21.09)
PDV total profits	16.93 (21.08)	17.36 (21.08)	17.36 (21.08)	15.50 (21.13)

Notes: This table reports profit-relevant outcomes under counterfactual changes to the purchase environments, under the store’s existing substitution policy (the “baseline”) and counterfactual policies.

Table 2E.2: Profit-Relevant Outcomes by Substitution Policy: Apple Sauce

				“Fully individualized” by original order, past purchases, and household demographics		
	Baseline	“One size fits all”	Individualized by original order	Max. present- trip profits	Max. PDV total profits	Max. PDV future profits
<i>Panel A. Present trip</i>						
Retail margin	1.56 (0.68)	4.66 (2.14)	4.58 (2.07)	4.41 (2.09)	4.41 (2.09)	1.65 (0.83)
Prob. accept	0.92 (0.13)	0.69 (0.27)	0.70 (0.27)	0.73 (0.25)	0.73 (0.25)	0.83 (0.25)
Expected present- trip profits	1.44 (0.69)	2.89 (1.51)	2.89 (1.48)	2.92 (1.47)	2.92 (1.47)	1.35 (0.69)
<i>Panel B. Future trips</i>						
PDV future profits	11.50 (13.71)	11.50 (13.71)	11.50 (13.71)	11.50 (13.71)	11.50 (13.71)	11.50 (13.71)
<i>Panel C. Overall</i>						
PDV total profits	12.94 (13.77)	14.39 (13.92)	14.39 (13.92)	14.42 (13.92)	14.42 (13.92)	12.85 (13.74)

Notes: This table compares profit-relevant outcomes under the store’s existing substitution policy (the “baseline”) with outcomes under counterfactual policies. These counterfactual policies exploit the store’s knowledge of the consumer to varying degrees, with the “one-size-fits-all” policy leveraging none of this information; the “individualized-by-original-order” policy employing the store’s knowledge of the consumer’s original order; and the “fully individualized” policies additionally exploiting the store’s knowledge of the consumer’s past purchases and household demographics. Regarding the last, the “fully-individualized” policies are respectively designed to maximize (i) expected profits on the present shopping trip, (ii) the present-discounted value of total profits (both present and future), or (iii) the present-discounted value of future profits alone. All results are reported as means, with standard deviations appearing in parentheses.

Table 2E.3: Changes in Consumer Welfare Compared to Baseline Policy: Apple Sauce Cups

			“Fully individualized” by original order, past purchases, and household demographics		
	“One size fits all”	Individualized by original order	Max. present- trip profits	Max. PDV total profits	Max. PDV future profits
Expected present- trip surplus (\$)	−0.67 (2.04)	−0.63 (2.04)	−0.57 (2.03)	−0.57 (2.03)	−0.54 (2.56)
PDV future surplus (\$)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.01 (0.04)
PDV total surplus (\$)	−0.67 (2.04)	−0.63 (2.04)	−0.57 (2.03)	−0.57 (2.03)	−0.53 (2.56)

Notes: This table reports changes in consumer welfare when the store adopts various counterfactual substitution policies. See notes to Table 2.5 for descriptions of these policies. All results are reported as means (with standard deviations in parentheses).

CHAPTER 3

DEMAND ESTIMATION WHEN CONSUMERS' PREFERENCES VARY OVER TIME

3.1 Introduction

People's preferences sometimes vary over time. Take the case of coffee, for instance. Many people prefer iced coffee during the summer and hot coffee during the winter. In this paper, I show that workhorse demand systems fail to replicate important substitution patterns in markets where consumers' preferences vary over time. This shortcoming is rooted in the underlying discrete choice model: conditional or mixed logit. I show that conditional logit imposes independence between consumers' purchases and their pairwise preferences among unpurchased goods. As for mixed logit, this more general model imposes *conditional* independence between consumers' purchases and their pairwise preferences among unpurchased goods, given the realizations of the consumers' random coefficients. In other words, what someone purchases on a particular shopping trip should be uninformative of trip-specific factors that influenced both her purchase and her preferences among the goods she did not purchase. Hereafter, I refer to the preceding independence constraints as the *independence of preferred alternatives* (IPA) properties of conditional and mixed logit, respectively.

These theoretical results raise two empirical questions. First, can data help determine whether consumers' preferences in a given market are consistent with the IPA properties of conditional or mixed logit? And second, how should demand be estimated when consumers' preferences prove inconsistent with the IPA property of the (more flexible) mixed logit model? To provide insight, I employ novel data from curbside grocery pickup. This is a "click-and-collect" form of shopping where consumers order groceries online and then pick them up from their local supermarket. Importantly, products ordered for curbside pickup sometimes go out of stock. This obliges the store to select a "stockout substitute" on the affected consumer's behalf. Once she arrives at the store, the consumer is offered two choices: either she can purchase the stockout substitute, or she can purchase nothing.¹ Whether she is willing to purchase this (store-selected) substitute product

¹In principle, the consumer could also enter the store in search of a better substitute. However, this is exceedingly

provides direct evidence of its substitutability for the out-of-stock product.

Focusing on the product categories of bottled water and flour, I provide descriptive evidence that consumers' decisions to accept (i.e., purchase) or reject (i.e., not purchase) stockout substitutes are inconsistent with the IPA property of conditional logit. Contrary to the property, consumers' original orders are informative of their willingness to accept a given stockout substitute. As for mixed logit, I find that the accept/reject decisions of bottled water buyers are consistent with the model's IPA property, whereas those of flour buyers are not. Regarding the latter product category, consumers' preferences for substitute flours vary across trips—perhaps owing to variation in the planned recipe. This kind of within-consumer preference variation is excluded by the mixed logit IPA.

I next turn to an empirical case study. Does the IPA property of mixed logit influence demand estimates? If so, does the extent of this influence vary by product category? To give insight, I estimate demand for bottled water and flour using two models: mixed logit and mixed probit (which does not exhibit an IPA property). Then I compare the models' goodness of fit. As I do so, I focus on the models' fit in relation to the stockout substitution data. On these data, the mixed logit IPA imposes the following restriction: a consumer's original order choice should be conditionally independent of whether she accepts the substitute (given the realization of her random coefficients).

The results of this case study sometimes vary across product categories, model selection strategies (i.e., within- versus out-of-sample), and methods of computing choice probabilities (i.e., “conditional” versus “unconditional”).² But overall, mixed probit seems to forecast consumers' accept/reject decisions more accurately than mixed logit does. Importantly, this disparity tends to be larger for the product category of flour than that of bottled water. This is in keeping with the descriptive evidence summarized above: namely, that consumers' preferences for bottled water are consistent with the IPA property of mixed logit, whereas their preferences for flour are not.

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.

²The “conditional” approach exploits individual consumers' past purchases to supply predictions that reflect their respective choices on past shopping trips (see Train).

My findings can inform future applied work on differentiated products demand. In markets where consumers' preferences are stable across shopping trips, mixed logit should accurately reproduce the underlying substitution patterns. But in markets where consumers' preferences vary over time, an alternative model may be preferable (such as the mixed probit model estimated in this paper).³ Of course, there exist markets where the amount of within-consumer preference variation is not immediately obvious. If the researcher has data on unpurchased goods' substitutability for purchased ones—such as “second choice data” or data on stockout substitutions (as in this paper)—she can adapt the formal tests and informal descriptive analyses developed here to test whether consumers' preferences are consistent with the IPA properties of conditional or mixed logit.

The remainder of the paper proceeds as follows. Section 3.2 relates this study to prior literature. Section 3.3 reviews the canonical differentiated products demand model developed by Berry, Levinsohn, and Pakes (1995),⁴ and then formalizes the IPA properties of conditional and mixed logit. Section 3.4 provides institutional details about curbside pickup and introduces the data. Section 3.5 presents descriptive evidence concerning the extent to which consumer behavior coincides with the IPA properties of conditional and mixed logit. Section 3.6 presents a demand estimation case study, while Section 3.7 concludes.

3.2 Relationship to Prior Literature

An extensive literature within empirical industrial organization employs data on consumers' preferences among unpurchased goods—hereafter, *alternate-choice data*. Both in this existing literature and in my study, alternate-choice data help identify products' substitutability. However, I also use these data for a second purpose: namely, to test whether consumers' preferences are consistent with the IPA properties of conditional and mixed logit.

In what follows, I will elaborate on the relationship between this study and the prior literature

³The mixed probit model is impractical with large datasets. In Section 3.7, I suggest alternative methods of relaxing the mixed logit IPA that impose a smaller computational burden.

⁴Unlike 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.

within empirical industrial organization (“IO”) that employs alternate-choice data. Then I will briefly remark on two other literatures to which my work relates: the econometric literature on the identifying power of alternate-choice data, and the empirical literature that studies stockout events.

3.2.1 Alternate-Choice Data in Empirical IO

A growing empirical literature leverages alternate-choice data to estimate demand elasticities. The pioneering work is Berry, Levinsohn, and Pakes’s Berry, Levinsohn, and Pakes (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.⁵ 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.⁶

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

⁵In 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.

⁶These 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?”

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,⁷ alternate-choice data.

I will now elaborate on both these points of departure, explaining how they can inform future applied work that uses alternate-choice data.

The Model.—In differentiated products demand estimation, the consumer’s discrete choice problem is most often represented with mixed logit.⁸ 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

⁷These 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.

⁸Allcott (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.

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 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,⁹ 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.

3.2.2 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 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).

3.2.3 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,

⁹See 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.

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 3.2.2). 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.¹⁰ 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”).

3.3 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 $j \in \mathcal{J} \equiv \{0, 1, \dots, J\}$ that affords the greatest conditional indirect utility u_{ijt} .¹¹

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

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

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.

¹⁰To 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).

¹¹I assume that $\arg \max_{j \in \mathcal{J}} u_{ijt}$ is a singleton set with probability one. (In other words, there are no “ties.”).

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 $K \in \mathbb{R}$,*

$$\Pr[u_{iAt} > u_{iBt} \mid \max\{u_{iAt}, u_{iBt}\} < K] = \Pr[u_{iAt} > u_{iBt}].$$

Proof. See Chapter 3A. ■

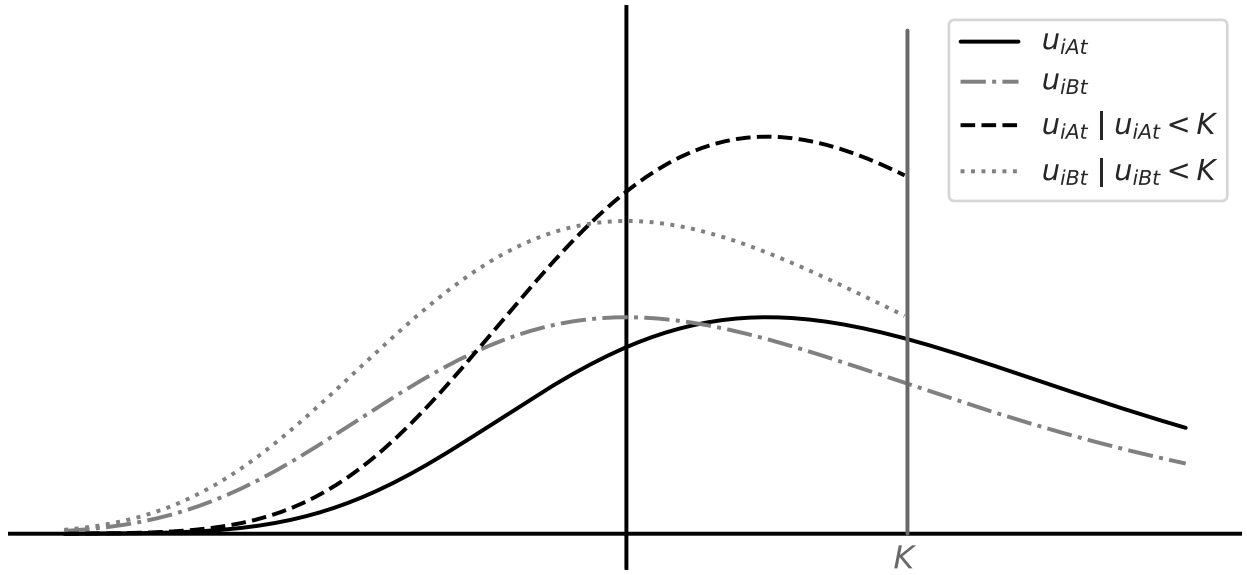


Figure 3.1: Irrelevance of Identical Upper Bounds on Two Goods' Logit Utilities

Figure 3.1 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

$$\Pr [u_{iAt} > u_{iBt} \mid \max\{u_{iAt}, u_{iBt}\} < K] > \frac{1}{2}.$$

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:

$$\Pr [u_{iAt} > u_{iBt} \mid \max\{u_{iAt}, u_{iBt}\} < K] = \Pr[u_{iAt} > u_{iBt}].$$

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

$$\Pr [u_{iAt} > u_{iBt} \mid \max\{u_{iAt}, u_{iBt}\} < K] \neq \Pr[u_{iAt} > u_{iBt}]$$

in general.¹² 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 $A, B, C \in \mathcal{J}$,*

$$\Pr [u_{iBt} > u_{iCt} \mid u_{iAt} = \max_{j \in \mathcal{J}} u_{ijt}] = \Pr[u_{iBt} > u_{iCt}].$$

Proof. By the law of iterated expectations,

$$\begin{aligned} \Pr [u_{iBt} > u_{iCt} \mid u_{iAt} = \max_{j \in \mathcal{J}} u_{ijt}] &= \mathbb{E} \left[\Pr [u_{iBt} > u_{iCt} \mid u_{iAt} = \max_{j \in \mathcal{J}} u_{ijt}; \right. \\ &\quad \left. (u_{ijt})_{j \in \mathcal{J} \setminus \{B, C\}} \mid u_{iAt} = \max_{j \in \mathcal{J}} u_{ijt} \right]. \end{aligned} \quad (3.1)$$

¹²By 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 $\Pr [u_{iAt} > u_{iBt} \mid \max\{u_{iAt}, u_{iBt}\} < 0] \approx 0.64 < 0.76 \approx \Pr[u_{iAt} > u_{iBt}]$.

As far as the inner component of Equation (3.1) 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 Equation (3.1), obtaining

$$\Pr[u_{iBt} > u_{iCt} \mid u_{iAt} = \max_{j \in \mathcal{J}} u_{ijt}; (u_{ijt})_{j \in \mathcal{J} \setminus \{B, C\}}] = \Pr[u_{iBt} > u_{iCt}]. \quad (3.2)$$

Substituting Equation (3.2) into Equation (3.1) yields

$$\begin{aligned} \Pr[u_{iBt} > u_{iCt} \mid u_{iAt} = \max_{j \in \mathcal{J}} u_{ijt}] &= \mathbb{E}[\Pr[u_{iBt} > u_{iCt}] \mid u_{iAt} = \max_{j \in \mathcal{J}} u_{ijt}] \\ &= \Pr[u_{iBt} > u_{iCt}]. \end{aligned}$$

■

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 3.5 and 3.6.¹³

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 $A, B, C \in \mathcal{J}$,*

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

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

I discuss the practical implications of Theorem 1 and Corollary 1 elsewhere.¹⁴ In addition, Chapter 3B relates Theorem 1 to prior theoretical results in the literature, while Chapter 3C presents Monte Carlo tests of Theorem 1.

¹³Briefly: 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 3.5 and 3.6 for details.

¹⁴See Section 3.5 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 3.2 and 3.5, respectively.

3.4 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.

3.4.1 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.¹⁵ 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.

3.4.2 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.

¹⁵The store’s website and mobile app allow the consumer to leave item-level instructions for the store. For instance, someone who is ordering strawberries might request “extra-ripe” berries. However, a consumer could also use this feature to request a specific substitute if her preferred product goes out of stock. Although I do not observe whether a consumer makes such a substitution request (or, for that matter, whether she leaves item-level instructions of any kind), the retailer has indicated that consumers rarely leave item-level instructions.

¹⁶The price of the out-of-stock item is obtained from the second data set. See Section 3.6 for details.

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

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 3.1 reports summary statistics for the two product categories studied in Sections 3.5 and 3.6: 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 3.6). 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

¹⁷Although participation in the loyalty program is not compulsory in general, it *is* required in order to place curbside pickup orders. Consequently, I can match the purchases of curbside pickup patrons to their in-store and delivery purchases.

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

Having explained why bottled water and flour form the focus of my empirical analysis, I now return to the summary statistics in Table 3.1. 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 (66,447) 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 a branded product line under which many distinct products may be sold. For instance, the Gold Medal brand encompasses many distinct 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. Observe also 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

¹⁸Although 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.

¹⁹All 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.

Table 3.1: Summary Statistics by Product Category

Statistic	<i>Panel A. Overview</i>	
	Bottled water	Flour
No. of households with 1+ substitutions	66,447	22,549
No. of distinct products purchased	40	52
... of which ordered for curbside pickup	32	38
No. of distinct brands purchased	9	14
... of which ordered for curbside pickup	9	8
<i>Panel B. Per household with 1+ substitutions</i>		
No. of shopping trips	39.4	12.0
... of which curbside pickup	7.4	3.1
... of which feature 1+ substitutions	1.6	1.1
No. of distinct products ever purchased	5.7	4.1
... of which ordered for curbside pickup	2.5	1.9
No. of distinct brands ever purchased	3.1	2.3
... of which ordered for curbside pickup	1.8	1.4
<i>Panel C. Stockout substitutions</i>		
Prob. accept (%)	87.3	92.0

Notes: All estimates are reported as means or totals. By “brands,” I refer to branded product lines, each of which may include 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.”

has experienced one or more substitutions) has made more shopping trips that involve bottled water (39.4) than flour (12.0). A modest fraction of these trips are curbside pickup (19% and 26% for bottled water and flour, respectively). On average, bottled water buyers have experienced slightly more stockout substitutions (1.6) than have their flour counterparts (1.1). 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 (87.3%).

3.5 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.

3.5.1 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 $j \in \mathcal{J}_t = \{0, 1, \dots, J_t\}$ 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 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 $u_{ist} \geq u_{i0t}$.²¹ Given her original order choice (j), what is the

²⁰I 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.)

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

probability that she accepts s ? Under Theorem 1,

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

In other words, the probability of acceptance should be independent of our consumer's original order choice.²²

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,²³ 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 ($p < 10^{-300}$),²⁴ but not for flour ($p = 0.976$).²⁵ However, in both these categories, more than half of the cells in the three-way contingency table are empty. Turning to the second specification, which attends only to pairings of the top ten products in each

²²Notice that I have *not* conditioned on the fact that the consumer was offered s as a stockout substitute. This is because the store worker who chooses the substitute does not observe the consumer's past purchase history, only the identity of the out-of-stock product. Moreover, it seems unlikely that the worker's choice of substitute reflects “unobservable” product characteristics in the spirit of Berry, Levinsohn, and Pakes (1995), as the worker must choose a substitute quickly (and is probably not an expert about the relevant product category).

²³See Jovanovic and Levy (1997) or Tuyl, Gerlach, and Mengersen (2009).

²⁴The log likelihood ratio test statistic is 4398, with 809 degrees of freedom. The latter is computed as ((no. of unique out-of-stock products) – 1)(no. of unique substitute products)

²⁵The log likelihood ratio test statistic is 1607, with 1721 degrees of freedom.

category, the null hypothesis of conditional independence is rejected with $p < 10^{-300}$ in both categories.²⁶

Although the foregoing exercise maps straightforwardly to the conditional logit IPA (as expressed in Section 3.3), 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:²⁷

- (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

²⁶The 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.

²⁷This 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.

and,²⁸ 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.

Table 3.2 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),²⁹ 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.³⁰ (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;³¹ whether they match the out-of-stock product's characteristics should be immaterial. However, Table 3.2 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

²⁸By “representative utility,” I mean the modeled portion of utility (as opposed to the error term).

²⁹That is, characteristics with more than two distinct realizations.

³⁰Within the analysis sample, comprising purchases by households with 1+ attempted substitutions.

³¹Formally, on the representative utility afforded by the substitute's characteristics.

Table 3.2: Testing the Conditional Logit IPA

Characteristic	<i>Panel A. Bottled water</i>			
	Out-of-stock product's version	Substitute's version	Prob. accept	No. of obs.
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
	<i>Panel B. Flour</i>			
	Out-of-stock product's version	Substitute's version	Prob. accept	No. of obs.
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).

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 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 Table 3D.1 in Chapter 3D 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 Table 3D.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).

3.5.2 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,

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

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.³²

Table 3.3: Stylized Example of the Mixed Logit IPA

Trip	Consumer P			Consumer M		
	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;³³ whereas the other, consumer M, typically opts for Ice Mountain. Table 3.3 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 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

³²In 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).

³³That is, the store’s eponymous brand of groceries.

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 .³⁴ 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 3.3. 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

³⁴Where 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.)

$(p_{i,\text{OOS}})$.³⁵

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

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

where

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

and v_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 3.3. 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 3.4 reports the average marginal effects of the explanatory variables.³⁶ Notice that there are two variables for each observable characteristic: an indicator for

³⁵As discussed in Section 3.4, 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 3.6.

³⁶Specifically, 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;

Table 3.4: Testing the Mixed Logit IPA: Average Marginal Effects from Probit Regressions

Variable	Product category	
	Bottled water	Flour
Brand		
Sub shares OOS product's version	−0.012*** (0.004)	−0.066*** (0.007)
Frac. of purchases with sub's version	0.147*** (0.006)	0.062*** (0.011)
No. of bottles		
Sub shares OOS product's version	−0.023*** (0.003)	
Frac. of purchases with sub's version	0.038*** (0.005)	
Size of each bottle		
Sub shares OOS product's version	0.032*** (0.004)	
Frac. of purchases with sub's version	0.049*** (0.006)	
Water type		
Sub shares OOS product's version	0.043*** (0.003)	
Frac. of purchases with sub's version	0.092*** (0.005)	
Flour type		
Sub shares OOS product's version		0.124*** (0.008)
Frac. of purchases with sub's version		0.027** (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		0.003 (0.007)
Frac. of purchases with sub's version		0.038*** (0.010)
Sub's price – OOS product's price	0.021*** (0.001)	−0.003 (0.002)
Observations	82,001	14,181
Pseudo R^2	0.0672	0.0720

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.

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 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 14.7 (9.2) 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 more likely 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 2.7 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 3.1 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

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.)

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 with respect to testing the IPA property of mixed logit; 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.

3.6 Structural Evidence

In this section, I evaluate the extent to which the mixed logit IPA causes bias in the estimation of demand elasticities. 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. The demand framework includes consumers’ in-store purchases, curbside orders, and decisions to accept or reject stockout substitutes. Then, with the estimated models in hand, I compare mixed probit and mixed logit’s goodness of fit, both within- and out-of-sample. As I do so, I attend especially to the data on consumers’ acceptance or rejection of stockout substitutes. This is because mixed logit, due to its IPA property, imposes conditional independence between a given consumer’s order choice and her subsequent decision about the substitute (given the realization of her random taste coefficients).

Mixed probit, by contrast, does not impose this independence property.

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.³⁷ 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.

3.6.1 Model

For simplicity, the conceptual framework in Section 3.5 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),³⁸ she faces a choice between J_t differentiated goods and an outside option of no purchase ("good 0"). She will choose whichever good $j \in \mathcal{J}_t \equiv \{0, 1, \dots, J_t\}$ maximizes her conditional indirect utility u_{ijt} . As in Section 3.5.2, utility is a consumer-specific function of product characteristics (x_j) and price (p_{jt}):

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

Unlike in Sections 3.5.1 and 3.5.2, 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

³⁷The 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.

³⁸In 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).

substitution if and only if $u_{ist} \geq u_{i0t}$, 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.

3.6.2 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 Chapter 3E. 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 $\theta = (\theta^1, \dots, \theta^R)$ 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.³⁹

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.⁴⁰ 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

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

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.

³⁹In a slight abuse of notation, I now use t to index an individual consumer's trips, as opposed to time.

⁴⁰Where 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).⁴¹ 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.⁴²

Simulation proceeds as follows. I take Q quasi-Monte Carlo draws, indexed by $q \in \mathcal{Q} \equiv \{1, \dots, Q\}$. For each draw q , I draw a vector of low-discrepancy multivariate normal errors

⁴¹Concerning 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.

⁴²To 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.

$(\varepsilon_{ijt}^q)_{j \in \mathcal{J}_t}$ for each consumer i and trip t .⁴³ Then the probability of consumer i ordering (purchasing) good j while shopping online (in-store), 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,

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

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

$$\mathcal{Q}^*(j) \equiv \{q \in \mathcal{Q} : u_{ijt}^{rq} = \max_{j' \in \mathcal{J}_t} \{u_{ij't}^{rq}\}\}$$

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

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

⁴³Precisely 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).

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.⁴⁴

3.6.3 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).

⁴⁴In the case of mixed probit, a slight adjustment is necessary: probit kernels, not logit kernels, are computed for each agent.

Given that a product is (presumably) available to the consumer, I impute its price as being the mean purchase price on the day of the consumer's shopping trip (within the relevant store location). If no purchases are observed on the precise day of the trip, I instead take the unweighted average of the mean purchase prices on the days immediately before and after.

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.⁴⁵ 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 purchase price on the nearest date for which observations appear in the data (up to seven days before or after the stockout event).⁴⁶

All prices are deflated to 2016 dollars using the six-month smoothed CPI (U.S. Bureau of Labor Statistics).

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.⁴⁷ For the same reason, I do not perform estimation on all the available data. Instead, I focus on a random sample of

⁴⁵Unless 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.

⁴⁶The structural estimation here focuses on top-selling products, whose prices are comparatively easy to infer. By contrast, the reduced-form regression in Section 3.5.2 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 3.5.2 are robust to the inclusion or exclusion of observations whose prices are imputed in this fashion.

⁴⁷For 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.

households within each product category.⁴⁸

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 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 (ii) 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 Chapter 3F.

⁴⁸To 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.

3.6.4 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 3.5 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.⁴⁹ I compute these predicted probabilities in two ways. The first, which I refer to as the “unconditional” approach (after Train [2009]), yields posterior probabilities based on the (estimated) population distribution of random coefficients. The second method of computing predicted probabilities, known as the “conditional” approach, leverages the panel structure of the data to derive posterior probabilities conditional on individual consumers' observed choices in the data.⁵⁰ The motivation for reporting both types of predicted probability is as follows. On the one hand, the “unconditional” approach is more commonly used in the literature.

⁴⁹In Chapter 3G, 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”).

⁵⁰When 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).

On the other, the “conditional” approach is closer in spirit to the statement of the mixed logit IPA in Corollary 1 (which conditions on the realizations of consumers’ true random coefficients).

Table 3.5: Goodness of Fit: Mixed Logit versus Mixed Probit

Statistic	<i>Panel A. In-store purchases and online orders</i>			
	Bottled water		Flour	
	Mixed logit	Mixed probit	Mixed logit	Mixed probit
No. of households	121	121	405	405
No. of purchases ^a	4096	4096	4096	4096
No. of available products ^b	4.56 (0.98)	4.56 (0.98)	7.55 (1.67)	7.55 (1.67)
Predicted probability of purchase				
... using the “unconditional” approach ^c	0.284 (0.129)	0.285 (0.130)	0.186 (0.094)	0.186 (0.094)
... using the “conditional” approach ^d	0.632 (0.315)	0.631 (0.312)	0.545 (0.299)	0.530 (0.293)
<i>Panel B. Stockout substitutions</i>				
No. of (attempted) stockout substitutions	147	147	353	353
... of which accepted	125	125	330	330
True decision’s predicted probability				
... using the “unconditional” approach ^c	0.771 (0.291)	0.777 (0.293)	0.878 (0.210)	0.890 (0.219)
... using the “conditional” approach ^d	0.902 (0.178)	0.906 (0.173)	0.946 (0.159)	0.944 (0.166)
<i>Panel C. Empirical specification</i>				
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 3.6.2 and 3.6.4 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.

^c This corresponds to the posterior probability based on the (estimated) population distribution of random coefficients.

^d This yields the posterior probability of the purchase, conditional on the consumer’s observed choices in the data.

The relative performance of mixed logit and mixed probit varies by data type. Focus first on in-store purchases and online orders (Panel A). Regarding the “unconditional” choice probabilities,

the models' fit is comparable; consumers' observed purchases of bottled water (flour) are assigned a 0.0 (0.1) percentage point lower predicted probability by mixed logit than by mixed probit. As for the "conditional" approach, the models' fit is comparable as far as bottled water is concerned; the predicted probabilities of consumers' observed purchases are 0.1 percentage points higher for mixed logit than for mixed probit. Regarding flour, by contrast, there is a perceptible difference in the models' fit. Consumers' observed purchases are assigned a 1.5 percentage points greater predicted probability by mixed logit than by mixed probit.

Turn next to stockout substitutions (Panel B). Under the "unconditional" approach, the predicted probabilities associated with consumers' observed decisions to accept or reject substitute bottled waters are 0.6 percentage points greater for mixed probit than for mixed logit. The disparity in relation to flour is twice as large: 1.2 percentage points. This is consistent with the descriptive evidence in Section 3.5.2. Since consumers' preferences for bottled water seem to be largely consistent with the IPA property of mixed logit, the model should forecast consumers' acceptance or rejection of stockout substitutes almost as accurately as mixed probit does. Regarding flour, by contrast, descriptive evidence suggests that consumers' preferences may be inconsistent with the mixed logit IPA. Consequently, mixed probit (which does not suffer from an IPA constraint) should predict acceptance or rejection more accurately than mixed logit (which does).

It is more difficult to reconcile the results under the "conditional" approach with the descriptive evidence in Section 3.5.2. Here, mixed probit and mixed logit supply predictions of comparable accuracy with respect to consumers' accept/reject decisions. Concerning bottled water, the predicted probabilities associated with consumers' observed accept/reject decisions are 0.4 percentage points greater for mixed probit than for mixed logit. As for flour, the predicted probabilities are nearly identical, being 0.2 percentage points greater for the mixed logit model than for mixed probit.

Over-fitting 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.⁵¹ 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 consumers 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 consumers 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.⁵²

Out-of-sample validation proceeds as follows. First, I randomly draw a “holdout sample” of consumers whose data were not used to estimate the models above. And second, I compute the posterior probabilities using both the “unconditional” and “conditional” approaches. For both approaches, I rely on the empirical CDF of random coefficients from the estimation results above. And regarding the “conditional” method, I report posterior probabilities conditional on consumers' original orders (as well as their in-store purchases and their orders for home delivery), but excluding their decisions to accept or reject stockout substitutes. This ensures that, so far as stockout substitutions are concerned, the validation exercise is predictive in nature.

Table 3.6 reports the results of out-of-sample validation. So far as consumers' in-store purchases and online orders are concerned, the results remain qualitatively unchanged from the within-

⁵¹With 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.

⁵²For 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 fields, including Health Economics (e.g., Deb and Trivedi 2002) and Agricultural Economics (e.g., Haener, Boxall, and Adamowicz 2001), as well as Industrial Organization (e.g., Bajari and Benkard 2005).

Table 3.6: Out-of-Sample Validation: Mixed Logit versus Mixed Probit

Statistic	<i>Panel A. In-store purchases and online orders</i>			
	Bottled water		Flour	
	Mixed logit	Mixed probit	Mixed logit	Mixed probit
No. of households	111	111	427	427
No. of purchases ^a	4096	4096	4096	4096
No. of available products ^b	4.60 (1.01)	4.60 (1.01)	7.53 (1.69)	7.53 (1.69)
Predicted probability of purchase				
... using the “unconditional” approach ^c	0.294 (0.130)	0.294 (0.130)	0.193 (0.099)	0.194 (0.100)
... using the “conditional” approach ^d	0.544 (0.293)	0.537 (0.292)	0.523 (0.287)	0.504 (0.279)
<i>Panel B. Stockout substitutions</i>				
No. of (attempted) stockout substitutions	157	157	356	356
... of which accepted	140	140	329	329
True decision’s predicted probability				
... using the “unconditional” approach ^c	0.809 (0.248)	0.818 (0.249)	0.870 (0.227)	0.880 (0.233)
... using the “conditional” approach ^d	0.862 (0.263)	0.859 (0.277)	0.890 (0.252)	0.896 (0.254)
<i>Panel C. Empirical specification</i>				
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 3.6.2 and 3.6.4 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.

^c This corresponds to the posterior probability based on the (estimated) population distribution of random coefficients.

^d This yields the posterior probability of the purchase, conditional on the consumer’s observed choices in the data (prior to the stockout substitution).

sample comparison. Irrespective of the product category and the method of computing probabilities (i.e., “conditional” versus “unconditional”), mixed logit supplies predictions of comparable—or superior—accuracy to those of mixed probit. Now consider stockout substitutions (Panel B). Using the “unconditional” approach, mixed probit forecasts the acceptance or rejection of stockout

substitutions much more accurately than mixed logit does. (The advantage in acceptance probabilities is 0.9 percentage points for bottled water and 1 percentage point for flour.) Turning to the “conditional” approach, mixed probit delivers *less* accurate predictions than mixed logit where bottled water is concerned (with the latter’s predicted probabilities exceeding the former’s by 0.3 percentage points). The results are more than reversed for flour, however. Here, the predicted probabilities of consumers’ observed accept/reject decisions are 0.6 percentage points greater for mixed probit than mixed logit. This is in keeping with the descriptive evidence in Section 3.5.2: namely, that consumers’ purchases of bottled water are consistent with the mixed logit IPA, whereas their purchases of flour are not.

3.7 Conclusion

This paper shows that workhorse demand systems fail to reproduce important substitution patterns when individual consumers’ preferences vary over time. This shortcoming is rooted in 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 consumer-specific random coefficients). To assess these properties’ influence on 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 both formal tests and informal descriptive evidence that consumers’ preferences are inconsistent with the IPA property of conditional logit. As for mixed logit, descriptive evidence suggests that consumers’ purchases of bottled water are consistent with the model’s IPA property, but not their purchases of flour.

I next present a demand estimation case study. The goal is to quantify what (if any) bias results from the IPA property of mixed logit. To this end, I estimate demand for bottled water and flour using two models: mixed logit and mixed probit (which does not display an IPA property). Then I compare the models’ goodness of fit in relation to the stockout substitution data. The results of this comparison vary by product category, as well as the model selection approach (within-

versus out-of-sample) and the method of computing choice probabilities (“conditional” versus “unconditional”). On balance, however, mixed probit seems to forecast consumers’ accept/reject decisions more accurately than mixed logit does. Importantly, this disparity tends to be larger for the product category of flour than that of bottled water. This is in keeping with the descriptive evidence summarized above: namely, that consumers’ preferences for bottled water are consistent with the IPA property of mixed logit, whereas their preferences for flour are not.

My findings can inform future applied work on differentiated products demand. In markets where consumers’ preferences are stable across shopping trips, mixed logit should accurately reproduce the underlying substitution patterns. But in markets where consumers’ preferences vary over time, an alternative model may be preferable. One such model is mixed probit (as employed in this paper). However, mixed probit is too computationally burdensome in many applications. Another possibility, therefore, is the “random-coefficients nested logit” model estimated in Brenkers and Verboven (2006) as well as 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. Many frameworks should be 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? And, if so, does the resulting model match the true substitution patterns in markets where consumers’ preferences vary over time? These questions are left for future work.

Another method of relaxing the mixed logit IPA is to expressly model within-consumer preference variation across shopping trips. Rather than assuming that each consumer’s preferences are characterized by a single vector of random coefficients, one could instead assume that her

preferences are given by two (or more) distinct vectors of random coefficients based on the circumstances of her shopping trip. To illustrate, consider the problem of modeling demand for flour. An individual consumer i could have one vector of random coefficients for shopping trips in which she plans to bake bread, denoted by β_i^{bread} ; and another vector of random coefficients for shopping trips in which she plans to bake cupcakes (for which all-purpose flour is ideal), denoted by $\beta_i^{\text{all-purpose}}$. Such a model would constitute a discrete mixture of two mixed logit models, where the partition of shopping trips between “bread trips” and “cupcake trips” is latent. I leave to future research the following questions. Is such a model identified? If so, does identification require data on consumers’ preferences among unpurchased goods? Or does it suffice to observe purchases alone? And how might such a model be estimated?⁵³

⁵³One promising direction is to adapt the expectation maximization (EM) algorithm from Section 5 of Train (2008). The key modification would center on the kernel probability.

BIBLIOGRAPHY

- 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.
- Ackerberg, Daniel A. “Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination”. *International Economic Review* 44, no. 3 (2003): 1007–1040.
- 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.
- Allcott, Hunt, et al. *Sources of Market Power in Web Search: Evidence from a Field Experiment*. National Bureau of Economic Research, 2025.
- Allende, Claudia, Francisco Gallego, and Christopher Neilson. “Approximating the Equilibrium Effects of Informed School Choice”. Working paper, 2019. Visited on 10/28/2024.
- Anand, Bharat N., and Ron Shachar. “Advertising, the Matchmaker”. *The RAND Journal of Economics* 42, no. 2 (June 2011): 205–245.
- 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.
- Arteaga, Cristian, et al. “xlogit: An Open-Source Python Package for GPU-Accelerated Estimation of Mixed Logit Models”. *Journal of Choice Modelling* 42 (2022): 100339.
- 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.
- Backus, Matthew, Christopher Conlon, and Michael Sinkinson. *Common Ownership and Competition in the Ready-to-Eat Cereal Industry*. National Bureau of Economic Research, 2021. Visited on 04/02/2025.
- 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.
- Barahona, Nano, Cristóbal Otero, and Sebastián Otero. “Equilibrium Effects of Food Labeling Policies”. *Econometrica* 91, no. 3 (2023): 839–868.
- 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, and Philip Haile. “Identification in Differentiated Products Markets”. *Annual Review of Economics* 8, no. 1 (Oct. 31, 2016): 27–52.
- 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 T., and Philip A. Haile. “Foundations of Demand Estimation”. In *Handbook of Industrial Organization*, ed. by Kate Ho, Ali Hortaçsu, and Alessandro Lizzeri, 4:1–62. 2021.
- . “Identification in Differentiated Products Markets Using Market Level Data”. *Econometrica* 82, no. 5 (2014): 1749–1797.
- . “Nonparametric Identification of Differentiated Products Demand Using Micro Data”. *Econometrica* 92, no. 4 (2024): 1135–1162.
- Bradbury, James, et al. *JAX: Composable Transformations of Python + NumPy Programs*. Version 0.3.13, 2018.
- 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.
- Brick Meets Click and Mercatus. “February U.S. eGrocery Sales Total \$7.9 Billion, Down 10% versus Year Ago”. Brick meets click, Mar. 13, 2024. Press Release.
- 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.
- Che, Hai, Tülin Erdem, and T. Sabri Öncü. “Consumer Learning and Evolution of Consumer Brand Preferences”. *Quantitative Marketing and Economics* 13, no. 3 (Sept. 2015): 173–202.
- Chen, Nan, and Hsin-Tien Tsai. “Steering Via Algorithmic Recommendations”. *The RAND Journal of Economics* 55, no. 4 (Dec. 2024): 501–518.
- Ching, Andrew T. “A Dynamic Oligopoly Structural Model for the Prescription Drug Market After

- Patent Expiration*". *International Economic Review* 51, no. 4 (Nov. 2010): 1175–1207.
- Collard-Wexler, Allan. "Demand Fluctuations in the Ready-Mix Concrete Industry". *Econometrica* 81, no. 3 (2013): 1003–1037.
- Compiani, Giovanni, et al. "Online Search and Optimal Product Rankings: An Empirical Framework". *Marketing Science* 43, no. 3 (May 2024): 615–636.
- 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. 1, 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.
- Daljord, Øystein. "Durable Goods Adoption and the Consumer Discount Factor: A Case Study of the Norwegian Book Market". *Management Science* 68, no. 9 (2022): 6783–6796.
- 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.
- Donnelly, Robert, Ayush Kanodia, and Ilya Morozov. "Welfare Effects of Personalized Rankings". *Marketing Science* 43, no. 1 (Jan. 2024): 92–113.
- Dubé, Jean-Pierre, and Sanjog Misra. "Personalized Pricing and Consumer Welfare". *Journal of Political Economy* 131, no. 1 (2023): 131–189.
- Erdem, Tülin, and Michael P. Keane. "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets". *Marketing Science* 15, no. 1 (1996): 1–20.

- Erdem, Tülin, Michael P. Keane, and Baohong Sun. “A Dynamic Model of Brand Choice When Price and Advertising Signal Product Quality”. *Marketing Science* 27, no. 6 (2008): 1111–1125.
- 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.
- Farronato, Chiara, Andrey Fradkin, and Alexander MacKay. “Self-Preferencing at Amazon: Evidence from Search Rankings”. In *AEA Papers and Proceedings*, 113:239–243. American Economic Association, 2023.
- Farronato, Chiara, et al. “Understanding the Tradeoffs of the Amazon Antitrust Case”. *Harvard Business Review* (Jan. 11, 2024).
- 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).
- 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.
- Hollister, Sean. “Microsoft Now Thirstily Injects a Poll When You Download Google Chrome”. *The Verge*, Oct. 24, 2023.

- Iaria, Alessandro, and Ao Wang. “Real Analytic Discrete Choice Models of Demand: Theory and Implications”. *Econometric Theory* (2024): 1–49.
- Jovanovic, B. D., and P. S. Levy. “A Look at the Rule of Three”. *The American Statistician* 51, no. 2 (May 1997): 137–139.
- Kim, Kyoo il, and Amil Petrin. “Control Function Corrections for Unobserved Factors in Differentiated Product Models”. Working paper, 2019.
- Krasnoff, Barbara. “How to change your default browser in Windows 11”. *The Verge*, Apr. 15, 2022.
- 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.
- Nelson, Phillip. “Information and Consumer Behavior”. *Journal of Political Economy* 78, no. 2 (Mar. 1970): 311–329.
- Nevo, Aviv. “Measuring Market Power in the Ready-to-Eat Cereal Industry”. *Econometrica* 69, no. 2 (Mar. 2001): 307–342.
- Newell, Richard G., and Juha Siikamäki. “Nudging Energy Efficiency Behavior: The Role of Information Labels”. *Journal of the Association of Environmental and Resource Economists* 1, no. 4 (Dec. 2014): 555–598.
- Osborne, Matthew. “Consumer Learning, Switching Costs, and Heterogeneity: A Structural Examination”. *Quantitative Marketing and Economics* 9 (2011): 25–70.
- 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.
- Raedts, Elske, and Simone Evans. “Google Shopping: Self-Preferencing Can Be Abusive”. Stibbe, Feb. 10, 2024.
- Reimers, Imke, and Joel Waldfogel. *A Framework for Detection, Measurement, and Welfare Analysis of Platform Bias*. National Bureau of Economic Research, 2023.
- Revelt, David, and Kenneth Train. “Customer-Specific Taste Parameters and Mixed Logit”, vol. Working Paper No. E00-274, Department of Economics, University of California, Berkeley. 2000.
- Ryan, Stephen P. “The Costs of Environmental Regulation in a Concentrated Industry”. *Econometrica* 80, no. 3 (2012): 1019–1061.
- Shin, Sangwoo, Sanjog Misra, and Dan Horsky. “Disentangling Preferences and Learning in Brand Choice Models”. *Marketing Science* 31, no. 1 (Jan. 2012): 115–137.
- 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.
- Sullivan, Christopher. “The Ice Cream Split: Empirically Distinguishing Price and Product Space Collusion” (2020).
- Taivalsaari, Antero, et al. “Web Browser as an Application Platform”. In *2008 34th Euromicro Conference Software Engineering and Advanced Applications*, 293–302. 2008.
- 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. Bureau of Labor Statistics. *Consumer Price Index for All Urban Consumers (CPI-U)*.
- U.S. Food & Drug Administration. “Bottled Water Everywhere: Keeping it Safe”. Consumer Updates, Apr. 22, 2022.
- Vatter, Benjamin. “Quality Disclosure and Regulation: Scoring Design in Medicare Advantage”.

Working paper, 2024.

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.

Young, Liz. “Never Mind the Delivery, More Online Consumers Are Turning to Store Pickup”. *The Wall Street Journal* (July 14, 2023).

Zeyveld, Andrew. “Demand Estimation When Consumers’ Preferences Vary over Time”. Working Paper (2024).

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

APPENDIX 3A

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 $v_B = 0$.¹ Then

$$\begin{aligned} P_A &\equiv \Pr [u_{iAt} > u_{iBt} \mid \max\{u_{iAt}, u_{iBt}\} < K] \\ &= \Pr [v_A + \varepsilon_{iA} > \varepsilon_{iB} \mid \max\{v_A + \varepsilon_{iA}, \varepsilon_{iB}\} < K] \\ &= E_{\varepsilon_{iA}} \left[\Pr [v_A + \varepsilon_{iA} > \varepsilon_{iB} \mid \max\{v_A + \varepsilon_{iA}, \varepsilon_{iB}\} < K; \varepsilon_{iA}] \mid v_A + \varepsilon_{iA} < K \right]. \end{aligned} \quad (3A.1)$$

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

Consider the inner component of Equation (3A.1), namely, the conditional probability

$$\begin{aligned} P_A \mid \varepsilon_{iA} &\equiv \Pr [v_A + \varepsilon_{iA} > \varepsilon_{iB} \mid \max\{v_A + \varepsilon_{iA}, \varepsilon_{iB}\} < K; \varepsilon_{iA}] \\ &= \Pr [\varepsilon_{iB} < v_A + \varepsilon_{iA} \mid \max\{v_A + \varepsilon_{iA}, \varepsilon_{iB}\} < K; \varepsilon_{iA}]. \end{aligned} \quad (3A.2)$$

Because $\max\{v_A + \varepsilon_{iA}, \varepsilon_{iB}\} < K$, the random variable ε_{iB} possesses the support $(-\infty, K)$. Eq. Equation (3A.2) can thus be expressed as the fraction

$$P_A \mid \varepsilon_{iA} = \frac{F_\varepsilon(v_A + \varepsilon_{iA})}{F_\varepsilon(K)}, \quad (3A.3)$$

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

¹To see why this assumption is without loss of generality, decompose both goods' utilities into their respective representative utility and error terms;

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

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

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

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

Substituting Equation (3A.3) into Equation (3A.1) yields

$$\begin{aligned} P_A &= E_{\varepsilon_{iA}} \left[\frac{F_{\varepsilon}(v_A + \varepsilon_{iA})}{F_{\varepsilon}(K)} \mid v_A + \varepsilon_{iA} < K \right] \\ &= \frac{E_{\varepsilon_{iA}} [F_{\varepsilon}(v_A + \varepsilon_{iA}) \mid v_A + \varepsilon_{iA} < K]}{F_{\varepsilon}(K)}. \end{aligned} \quad (3A.4)$$

Now employ the definition of expectation to write Equation (3A.4) 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 $f_{\varepsilon}(\varepsilon'_{iA}) / F_{\varepsilon}(K - v_A)$. (Here, $f_{\varepsilon}(\varepsilon') \equiv \exp(-e^{-\varepsilon'} - \varepsilon')$ denotes the PDF of the Gumbel distribution.) As a result,

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

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

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

APPENDIX 3B

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 pertains 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,

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

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.¹

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.)

¹Formally, let $r \equiv (r_1, r_2, \dots, r_S)$ be any ordinal ranking of the goods in $\mathcal{S}$ such that $u_{ir_1t} > u_{ir_2t} > \dots > u_{ir_S t}$. Then

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

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

APPENDIX 3C

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 $j \in \mathcal{J} \equiv \{1, \dots, J\}$.¹ Utility is specified as

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

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

$$\Pr[u_{iB} > u_{iC} \mid u_{iA} = \max_{j \in \mathcal{J}} u_{ij}] = \Pr[u_{iB} > u_{iC}] \quad (3C.1)$$

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 $[-4.5, 3.5]$.² 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:

$$\Pr[u_{iBr} > u_{iCr} \mid (v_{js})_{j \in \mathcal{J}}] = \frac{\exp(v_{Bs})}{\exp(v_{Bs}) + \exp(v_{Cs})}. \quad (3C.2)$$

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 , $\{\varepsilon_{ij}\}_{i=1}^N$.³ Then

¹For simplicity, I abstract from the inside/outside good distinction.

²This choice of support follows the Monte Carlo experiments in Heiss, Hetzenecker, and Osterhaus (2022).

³Regarding the absence of an s subscript: for computational simplicity, I use the same ten million Gumbel draws for all simulations.

the conditional probability is approximated by:

$$\hat{\Pr}[u_{iBt} > u_{iCt} \mid u_{iAt} = \max_{j \in \mathcal{J}} u_{ijt}; (v_{js})_{j \in \mathcal{J}}] = \sum_{i=1}^N 1 [v_{Br} + \varepsilon_{iB} > v_{Cr} + \varepsilon_{iC} \\ \text{and } v_{Ar} + \varepsilon_{iA} = \max_{j \in \mathcal{J}} \{v_{jr} + \varepsilon_{ij}\}] \\ / \sum_{i=1}^N 1 [v_{Ar} + \varepsilon_{iA} = \max_{j \in \mathcal{J}} \{v_{jr} + \varepsilon_{ij}\}]. \quad (3C.3)$$

With Equations (3C.2) and (3C.3) in hand, I proceed to compute the absolute value of the difference between them:

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

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).

APPENDIX 3D

CROSS-CHARACTERISTIC CORRELATIONS IN (DIS)SIMILARITY

Table 3D.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.)

Table 3D.1: Correlation Matrices of Similarity in Characteristics between Substitute and Out-Of-Stock Product

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%.

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%.

APPENDIX 3E

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).

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 3.1 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,¹ the mixed probit model accommodates correlated errors. However, it is challenging to recover the structure of the correlation. One must simulate choice probabilities not

¹Only generalizations of mixed logit, such as mixed nested logit, can incorporate correlated errors.

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 3.5.2 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 3.1 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 matrix:

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

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.² 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.³ 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 $C = \{\sigma_1, \dots, \sigma_C\}$ 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 $c = 1, \dots, C$ 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

²More 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.

³To 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.

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 simulated probabilities.⁴)

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]).

⁴The “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).

APPENDIX 3F

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 3.6. 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 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 3.6.3], 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 3.6.4 on this alternative sample. The results qualitatively resemble those presented in the main text (both within- and out-of-sample). The primary difference is that mixed probit *always* delivers more accept/reject predictions than does mixed logit—irrespective of product category, model selection approach, or method of computing predicted choice probabilities. However, the disparity still tends to be larger in relation to flour than in relation to bottled water. This is consistent with the descriptive evidence presented in Section 3.5.2.

APPENDIX 3G

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, Table 3G.1 compares two estimation approaches. The first approach directly imposes the mixed logit IPA, resulting in the closed-form likelihood presented in Section 3.6.2 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.

Table 3G.1: Verifying the Mixed Logit IPA by Simulation

Statistic	Product category	
	Bottled water	Flour
Frac. of stockouts with same prediction	0.996	0.993
Avg. absolute difference in predicted prob. accept	0.003	0.005
Root mean square difference in predicted prob. accept	0.024	0.025

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.)

Table 3G.1 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.¹ As for the second measure, I compute the average absolute difference between the two models’ predicted probabilities of acceptance. Letting $s \in \{1, \dots, S\}$ index (attempted) stockout substitutions, the measure is given

¹That 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.

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 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 Table 3G.1 indicate that the two estimation approaches arrive at very similar predictions. This is especially true where bottled water is concerned.²

Random Coefficients.—Table 3G.2 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, Table 3G.2 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.³

Whether the error terms are correlated or uncorrelated, mixed probit does not display an IPA property.⁴ 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

²There 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.)

³I tested five potential correlation parameters in each category: $\{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$.

⁴By 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 $\Pr[u_{iBt} > u_{iCt} \mid u_{iAt} = \max_{j \in \mathcal{J}} u_{ijt}] \approx 0.331 > 0.240 \approx \Pr[u_{iBt} > u_{iCt}]$.

Table 3G.2: Summary Statistics on Structural Parameters

Variable	<i>Panel A. Bottled water</i>			
	Mixed logit		Mixed probit	
	Means	Std. devs.	Means	Std. devs.
Price	2.174	0.682	1.190	0.405
Aquafina (24 ct.) ^a	8.689	2.699	6.682	1.871
Ice Mtn. (24 ct.) ^a	8.838	2.962	7.538	2.272
Nestle (24 ct.) ^a	8.211	2.155	6.499	1.727
Pvt. lbl. purified water (24 ct.) ^a	8.622	2.252	6.688	1.837
Pvt. lbl. purified water (40 ct.) ^a	10.207	2.957	8.697	2.207
Pvt. lbl. spring water (24 ct.) ^a	8.284	2.540	7.154	1.908
Error correlation cluster	24-packs			
Correlation parameter (σ)	0.0			
	<i>Panel B. Flour</i>			
	Mixed logit		Mixed probit	
	Means	Std. devs.	Means	Std. devs.
Price	2.298	0.475	1.555	0.450
All-purpose flour	6.801	3.045	5.077	2.411
Bread flour	4.781	3.043	3.359	2.653
Gold Medal brand	0.099	2.811	0.615	2.038
King Arthur brand	2.153	5.536	1.777	3.325
Log quantity	1.693	1.046	1.574	0.846
Unbleached	0.085	4.007	−0.663	2.238
Error correlation cluster	Bread flours			
Correlation parameter (σ)	0.0			

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 Chapter 3E for details.

^a Product-specific dummy.

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*).

The “Hit Rate.”—In Section 3.6.4, 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.⁵

Table 3G.3 compares the hit rates of mixed logit and mixed probit. Using the “unconditional” approach, mixed logit and mixed probit deliver extremely similar hit rates—irrespective of the product category or the model selection strategy (i.e., within- versus out-of-sample). Using the “conditional” approach, by contrast, mixed logit performs weakly better than mixed probit. The sole exception is out-of-sample predictions about stockout substitutions within the product category of flour. There, mixed probit’s hit rate exceeds that of mixed logit by 0.3 percentage points.

⁵The 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).

Table 3G.3: “Hit Rate:” Mixed Logit versus Mixed Probit

Data type	<i>Panel A. Within sample</i>			
	Bottled water		Flour	
	Mixed logit	Mixed probit	Mixed logit	Mixed probit
In-store purchases and online orders				
... using “unconditional” approach ^a	0.405	0.401	0.254	0.255
... using “conditional” approach ^b	0.719	0.717	0.686	0.669
Stockout substitutions				
... using “unconditional” approach ^a	0.850	0.850	0.935	0.935
... using “conditional” approach ^b	0.952	0.952	0.966	0.960
<i>Panel B. Out of sample</i>				
In-store purchases and online orders				
... using “unconditional” approach ^a	0.432	0.423	0.305	0.305
... using “conditional” approach ^b	0.646	0.638	0.681	0.659
Stockout substitutions				
... using “unconditional” approach ^a	0.892	0.892	0.924	0.924
... using “conditional” approach ^b	0.898	0.892	0.916	0.919

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 3.6.2 and 3.6.4 for details on estimation).

^a This yields the posterior probability of the purchase, conditional on the consumer’s observed choices in the data.

^b This is the posterior probability based on the (estimated) population distribution of random coefficients.