Title
Essays on Information and Prices

Author
Heinzman, Andrew

Publication Date
2023

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Information and Prices

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Economics

by

Stephen Andrew Heinzman

2023

ABSTRACT OF THE DISSERTATION

Essays on Information and Prices

by

Stephen Andrew Heinzman
Doctor of Philosophy in Economics
University of California, Los Angeles, 2023
Professor John W. Asker, Chair

I study how information is incorporated into prices. In the first two chapters, I approach this problem by studying the pricing of new products. Retailers face uncertainty about demand for new products, making it challenging to set prices. To learn about demand, retailers rely on information from similar, existing products and demand in similar markets. Thus, retailers have the least knowledge about demand for new products that are not present in any markets and have few substitutes. I show that to address uncertainty retailers launch new products into a subset of markets to learn about demand before introducing a product widely. As retailers gain additional information from similar existing products or prior introductions at other stores, they set prices closer to the optimal level. Additionally, after introducing a new product, retailers continue to learn about demand and refine their pricing strategy. Retailers learn, in part, by sharing information across their network of stores. Information sharing across different geographic regions helps retailers learn about demand if consumers are similar across locations. I find that while information sharing typically leads to prices that are closer to the optimal level, there are cases where information from different geographies has a negative impact on pricing.

In my third chapter, I take these themes in a different direction by studying pricing in financial markets from a theoretical perspective. In modern financial markets, investors seeking to profit from new information compete with high frequency traders (HFTs) who invest in the ability to trade quickly. Using a dynamic model, I find that while HFTs' speed advantage improves market liquidity (as indicated by bid-ask spreads), it also slows down the speed at which new information is incorporated into prices. HFTs decrease the profitability of new information by trading ahead of slower investors, prompting investors to trade less to avoid revealing valuable information. As a result, new information takes longer to be incorporated into prices. However, HFTs also improve market liquidity by ensuring that prices better reflect widely known information.

The dissertation of Stephen Andrew Heinzman is approved.

Hugo A. Hopenhayn

Elisabeth Honka

Martin B. Hackmann

John W. Asker, Committee Chair

University of California, Los Angeles

2023

To my wife Chelsea, without you this would not have been possible.

To my parents Kelly and Steve as well as my grandparents Linda and Floyd, for inspiring me and encouraging me every step of the way.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I owe a great thanks to my committee John Asker, Martin Hackmann, Elisabeth Honka, and Hugo Hopenhayn for their valuable advice, meetings, and support these past few years.

I want to thank my friends and peers Daniel Perez, Domenico Fabrizi, and Chandni Raja for keeping me on track and learning with me week after week.

VITA

2016	B.A. in Economics and Statistics, University of Virginia.
2016–2018	Senior Analyst, Cornerstone Research.
2020	M.A. in Economics, University of California, Los Angeles.

CHAPTER 1

A Descriptive Approach to New Product Introductions, Retailer Learning, and Pricing

This chapter examines how retailers learn about demand and set prices for new products across their network of stores. Retailers face uncertainty about demand for new products, making it challenging to set prices. To learn about demand retailers rely on information from similar, existing products and demand in similar markets. Thus, retailers have the least knowledge about demand for new products that are not present in any markets and have few substitutes. I document that to address this uncertainty, retailers use a staggered rollout strategy, launching new products in a smaller subset of markets to learn about demand before introducing the product widely. As retailers learn about demand they update prices less frequently, reducing price volatility over time. This chapter provides a descriptive analysis of retailers' product introduction and pricing choices, which then informs the development of a structural model in Chapter 2. [1]

1.1 Introduction

Every year, as 30,000 new products are added to store shelves, retailers need to set prices.[2] Typically prices are set as cost plus a markup. Costs are determined by upstream manufacturers and wholesalers and are generally known to retailers. The optimal markup, however, is potentially unknown. Product

[1] Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

[2] https://nielseniq.com/global/en/insights/analysis/2019/bursting-with-new-products-theres-never-been-a-better-time-for-breakthrough-innovation/

markups are determined by two key factors: how sensitive consumers are to higher prices, and how much demand there is for the product. Generally, consumers are consistent in how sensitive they are to prices, especially for products in the same category. As a result, retailers know consumers' price sensitivity from their past experience. However, for new or novel products, it can be difficult for retailers to know what demand will be. This uncertainty makes it challenging for retailers to set the right price, and there may be a significant delay between a product's launch and the time when the retailer determines the optimal price to charge.

To learn about demand a product, retailers usually rely on past sales data. However, this can be a challenge for new products that have no sales history. As a result, retailers need to learn about demand from other sources of information. Two options are to use data from existing products that are similar to the new product, or—for retailers with multiple stores—to use data from previous launches of the new product in similar locations. Retailers with multiple stores can take advantage of their network of stores to improve their pricing strategy. A common approach is to first launch a new product in a limited number of markets to learn about demand before expanding to more stores. Staggering the rollout of new products across stores can help retailers minimize the number of stores with sub-optimal prices while eventually offering the new product to all consumers. After learning about demand in a small number of stores, retailers can apply the insights they gained to set prices across all stores. This allows the retailer to set prices based on customer demand, while also achieving greater efficiency in their pricing strategy by using their entire network of stores.

The usefulness of a staggered rollout strategy for learning about demand is dependent on the similarity of consumers' preferences across locations. If there are significant differences in consumer preferences between the initial market and subsequent markets, then the information gained from the initial market won't be useful for setting prices elsewhere. If preferences vary significantly between markets then staggering the rollout of new products won't be helpful because the retailer will need to learn about demand in each market independently. Previously, Y. Huang, Ellickson, and Lovett (2022) showed that differences in consumer preferences can be a challenge for retailers expanding to new markets. Y. Huang, Ellickson, and Lovett (2022) showed that grocery retailers selling liquor in Washington state for the first time had to relearn about consumer demand—even though they had previously sold the same products in other US states. This is because consumers in other states had significantly different preferences than consumers in Washington, so the same pricing strategy that was optimal in other states was not optimal in Washington.

This chapter builds on Y. Huang, Ellickson, and Lovett (2022) in two ways. First, it demonstrates the impact of learning on a retailer's product rollout strategy. When retailers have limited sources of information about demand, they use a staggered rollout strategy and introduce new products into fewer stores. Conversely, if retailers can learn about demand from other sources, such as similar existing products, they rely less on staggered rollouts and introduce new products more widely. Second, this chapter contributes by providing evidence of learning through observing the retailer's pricing strategy. When retailers are uncertain about demand, they should update prices as they learn, resulting in more volatile prices. The chapter finds that in the initial years after a product's launch, prices are consistently volatile due to learning, but this volatility decreases over time as the retailers gain a better understanding of consumer preferences.

I study learning in the grocery market for dried pasta between 2010 and 2019. During this period, a large number of new pasta products were introduced. New products included variations of existing wheat-based pastas and new varieties made from chickpeas, lentils, rice, or corn that cater to gluten-free consumers. While retailers had access to information about demand for the wheat-based pastas due to the similarities with existing products, they lacked information for the gluten-free pastas as there were few similar, existing products. As a result, while retailers were able to introduce wheat-based products widely they used a staggered rollout strategy and introduced new gluten-free products in a smaller subset of markets to learn about demand. This contrast between wheat and gluten-free pastas provides an excellent opportunity to examine how retailers learn by using their network of stores.

I find that retailers rollout new products more slowly when they lack information about demand from existing products. One common source of information is sales from existing products of the same brand. For new gluten-free pastas, I observe that each additional existing product of the same brand leads to the retailer introducing the new product into 2.8 additional stores. For gluten-free pastas additional existing same-brand products provide more information about demand, allowing the retailer to launch the new product more broadly. In contrast, for wheat pastas, which were widely available before 2010–2019, the relationship between the number of existing same-brand products and the number of stores where a new product was introduced is not significant. This indicates that retailers already had a good understanding of demand for wheat pastas, and so the introduction of additional same-brand products did not provide any new information.

In addition to the impact of information on product rollouts, I observe that retailers are able to learn over time, as evidenced by changes in their pricing strategy. When retailers don't know demand they update prices as they learn, resulting in large fluctuations from month to month. I find that during the first three years following a product's launch, price volatility remains consistent. However, after four years, there is a significant reduction in price volatility, indicating that the retailer has gained information about demand and is now able to implement a more stable pricing strategy. These findings suggest that retailers require a substantial period of time, up to four years, to learn about demand, and that once they do, they set significantly more stable prices.

In this chapter, I take a descriptive approach to study the grocery market for dried pasta and examine how retailers learn about demand using their network of stores. Through observing the retailers' product introduction and pricing choices, I find that retailers introduce new products more slowly when they lack information about demand and that they are able to learn over time, resulting in more stable prices. While these descriptive findings offer valuable insights, a model of the retailer's price setting decisions is needed to account for the influence of potentially confounding factors, such as changing costs or the entry and exit of competing products. I provide such a model—building on the insights in this chapter—in Chapter 2.

1.2 Literature Review

This paper is related to two strands in the literature. The first is the literature on how firms learn about demand. A number of theoretical papers including McLennan (1984); Trefler (1993); Raman and Chatterjee (1995); Bergemann and Välimäki (2006); and Handel and Misra (2015) study the pricing decision of a firm who does not know demand. Studying how firms learn is empirically difficult because it is often impossible to know what information firms have available to them. One approach of recent empirical work is to study learning by imposing structure on the learning process or the parameters firms are uncertain about. This has allowed authors to study firm's decisions about investment (Jeon (2022)), pricing (G. Huang, Luo, and Xia (2019)), and if a new product should be launched or scrapped (Hitsch (2006)). However, Doraszelski, Lewis, and Pakes (2018); Y. Huang, Ellickson, and Lovett (2022) and this chapter approach the problem of studying how firms learn by inferring that firms are learning by observing their actions. I take an approach similar to these papers of studying learning without directly modeling the firm's learning process. I instead infer that the firm is learning by looking for evidence in the firm's product introduction and pricing decisions.

Doraszelski, Lewis, and Pakes (2018) study how electricity providers learn to set prices after a change in the market structure. The authors introduce a technique that allows them to study learning without knowing how firms form their beliefs. The authors do so by first splitting their time period into a period where firms are learning and a period where firms have finished learning and have full information. In order to split time into these two periods Doraszelski, Lewis, and Pakes look at patterns within market prices. The authors show that prices are initially volatile and stabilize to a "rest point" over time. I take a similar approach in this chapter to look for evidence of learning. Given assumptions on costs and the competitive environment I show that firms are learning as evidence by a reduction in price volatility.

The approach in Y. Huang, Ellickson, and Lovett (2022) is even more similar to this chapter. Y. Huang, Ellickson, and Lovett (2022) study how grocers learn to set prices after being allowed to sell liquor in Washington state for the first time. Similar to Doraszelski, Lewis, and Pakes (2018), Huang, Ellickson, and Lovett show that over time price volatility declines and prices settle to a steady state. The authors use the existence of a steady state to show that retailers are able to fully learn about demand. Again, similar to Doraszelski, Lewis, and Pakes (2018), the authors use this evidence to divide time into two periods, one in which the retailer is learning and one in which the retailer has reached the steady state. In this chapter I take a similar approach, showing that after four years price volatility declines which is evidence of learning. Y. Huang, Ellickson, and Lovett (2022) is most similar to this chapter because of both the setting, pricing in the grocery market, and the approach of inferring that a retailer is learning by looking for declines in price volatility.

The second portion of the literature that this chapter is related to is about how firms introduce and price products in the beginning of the product lifecycle (see Nair (2019) for a review). A number of papers including Nosko (2011); Eizenberg (2014); Berry, Eizenberg, and Waldfogel (2016); Wollmann

(2018); and Fan and Yang (2020) study the products firms choose to introduce given the competitive environment. However, the focus of many of these papers is on the number and characteristics of new products rather than how firms introduce and set prices for the new products. Firm's decisions about where to launch new products and how to set prices is especially important given that demand may significantly vary between locations (Duan and Mela (2009)). I contribute to this literature on product introductions by considering how firms learn about demand and the influence of learning on firm's decisions about pricing and the speed of product rollouts.

In the literature on product introductions Bronnenberg and Mela (2004) is most closely related to this paper. Bronnenberg and Mela (2004) study how manufacturers rollout products across stores and across retailer networks nationally. The authors find that manufacturers are more likely to rollout products to stores that are geographically close by and to stores within the same retail chain. Additionally, retailers are more likely to add a product if their competitor already carries the product and if the product is already widely available in the local area. The result of these two findings is that a product's speed of adoption is increasing over time. This paper builds upon Bronnenberg and Mela (2004) by looking at learning as a factor that influences the speed of a product's rollout. I find that information from products with similar characteristics allows retailers to introduce products to more stores. This is related to the findings in Bronnenberg and Mela (2004) that geographic distance and experience with a product increase rollout speeds. Geographic distance, experience with a product, and experience with products that share characteristics all increase the information that retailers have about new products. As a result, the influence of learning helps to explain the patterns in Bronnenberg and Mela (2004). Further, as a product's speed of adoption increases over time, the retailer has more information about demand and can set prices closer to the optimum in more markets.

1.3 Setting

1.3.1 Pasta Products

In this chapter I focus on the grocery sales of dried pastas from a single retailer. Pasta products are relatively simple because they can be defined by three characteristics: brand, type, and shape. For example, a commonly sold product is Barilla wheat spaghetti. Here Barilla is the brand, wheat is the type, and spaghetti is the shape. I use Nielsen retail scanner data to identify the set of pasta products sold during 2010–2019. [3] Nielsen provides some product information including brand, but in order to categorize a pasta's type I collect additional nutrition information. I supplement the Nielsen data with nutritional and ingredient information from three sources: Nutritionix, OpenFoodFacts, and the USDA's FoodData Central database.

Using these three sources of nutritional and ingredient information I define a pasta's type. Pasta types are easily defined by a product's ingredient list because the list of ingredients is relatively short and consist of primarily flour and water. I categorize pastas into four types: wheat, wheat plus, gluten-free, and potato. I define wheat pastas as products that contain only non-whole grain wheat flour, typically durum or semolina. I define wheat plus pastas as products that contain whole grain wheat flour or a mix of wheat and a non-wheat flour that is typically added to increase the nutritional content of the pasta. One example is pasta marketed as protein plus that contains a mix of wheat flour and a legume flour for added protein. All wheat plus pastas are defined to contain wheat and are not gluten-free. In contrast, I define gluten-free pastas as products that do not contain wheat flour. Typically, gluten-free pasta is made from some combination of rice, corn, or legume flours. Finally, I define potato pastas as products that primarily contain potatoes. Potato pastas may also be made with some wheat flour and are often marketed as gnocchi. [4]

Table 1.1 shows the number of pasta products of each type as well as the proportion of revenue and units sold for each pasta type by Nielsen's retailer 32. Wheat pastas make up the majority of products and make up a disproportionately large share of units sold and revenue. In contrast, gluten-free products make up a disproportionately small share of units sold and revenue.

---

[3] For details on the data sources used see Appendix 1.6.1.

[4] Potato pastas only make up a small fraction of all pasta products and are included here for completeness.

Table 1.1: Summary of Pasta Types Sold

		Proportion of		
Pasta Type	**Products**	**Products**	**Units Sold**	**Revenue**
Wheat	733	70%	84%	78%
Wheat Plus	129	12%	11%	12%
Gluten Free	152	15%	3%	6%
Potatoes	27	3%	2%	4%

Aside from a pasta's type, products are also defined by their brand and shape. Using Nielsen data I am able to identify the brand of each product. On average a given store offers products from 14 brands. I do not observe a consistent measure of pasta shapes either from Nielsen or the data sources that provide ingredient information. As a result, I do no include shape information in my analyses. However, from manually looking up pasta products most brands offer the same core pasta shapes (examples include spaghetti, linguine, rotini, macaroni, etc.) while larger brands may also offer more specialty shapes.

1.3.2 Retailer 32

In this paper I focus on Nielsen's retailer numbered 32. I choose to focus on a single retailer to better understand how that retailer introduces new products across the stores it operates. I focus on retailer 32 because it is a large national retailer that operates in a number of distinct geographies. Retailer 32's geographic diffusion is interesting because demand is potentially very different across the country. Further, because retailer 32 operates in multiple markets they are able to stagger product rollouts across markets.

In Table 1.2 I present the top 10 Nielsen designated market areas, DMAs, in which retailer 32 operates as well as the number of stores per DMA. In total retailer 32 has at least 10 stores in 14 different DMAs. In Figure 1.1 I show the geographic distribution of retailer 32's stores. Retailer 32 operates in the Mid-Atlantic region around DC, the Mountain West region around Denver, and the

Pacific Northwest region between San Francisco and Seattle.

Table 1.2: Top 10 DMAs by Number of Stores

DMA	Number of Stores
San Francisco-Oakland-San Jose, CA	131
Seattle-Tacoma, WA	105
Phoenix, AZ	77
Portland, OR	74
Denver, CO	73
Washington, DC	72
Sacramento-Stockton-Modesto, CA	49
Baltimore, MD	23
Spokane, WA	23
Tucson, AZ	21

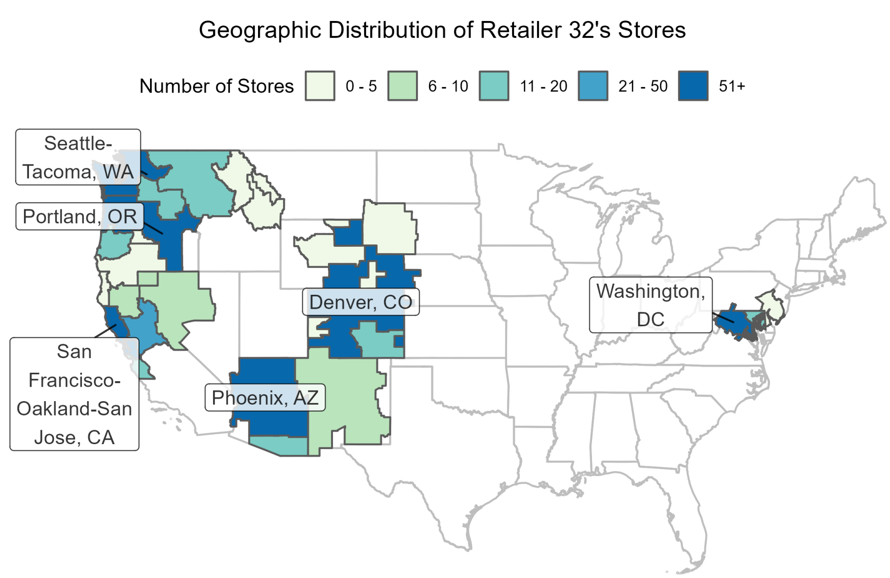


Figure 1.1: Geographic Distribution of Retailer 32's Stores

DMAs with 51+ stores are labeled.

In total from Nielsen scanner data I identify 767 stores belonging to retailer 32 that continuously report sales from 2010 to 2019. [5] [6] On average retailer 32 offers 99 products per store from 14 brands. The number of products per store varies over time as products are added and removed, but the average within-store standard deviation of the number of products offered is 10.7.

1.3.3 New Products

In this paper I focus on the introduction of new products across stores. I define a product's introduction to a store as the first date that I observe a sale. This introduces the potential for measurement error if a product is available, but does not sell for an extended period of time. To account for this I aggregate weekly Nielsen data to the monthly level. Further, to ensure that I can correctly identify the set of existing products I limit the period of my analysis to 2010–2019 although I have data as far back as 2004. [7] By doing so I use the Nielsen data in 2004–2009 to identify existing products. Table 1.3 shows the number of new products added to each store per year. The number of new products shows no large difference between the beginning and end of my sample indicating no major issues in my ability to identify new and existing products using the 2004–2009 data.

---

[5] I identify stores belonging to retailer 32 as those with both parent_code and retailer_code equal to 32. There are an additional 28 stores with the parent_code 32 and a different retailer_code. There are an additional 14 stores with the retailer_code 32 and a different parent_code. I exclude these additional stores.

[6] There are 57 stores who enter the data in or after 2010. I exclude these stores because in the first year data for a store is available I cannot identify if a product is new to that store or if the store did not previously exist in the data. There are 116 stores that I exclude because they exit the sample prior to 2019.

[7] Nielsen scanner data which provides panel data on all products sold within a store is available for 2006–2019. Nielsen Homescan data which provides panel data on all products purchased by a household is available for 2004–2019.

Table 1.3: Summary of New Products Added Per Store Each Year

Year	Min	Median	Average	Max
2010	4	22	22.0	52
2011	3	33	31.5	64
2012	6	25	26.1	58
2013	3	21	21.3	40
2014	9	23	23.8	48
2015	2	10	16.6	47
2016	2	21	25.8	91
2017	5	21	23.5	81
2018	3	28	27.1	89
2019	1	13	14.0	53

The time a given product is introduced to a store varies widely across retailer 32's network of stores. Figure 1.2 shows the difference in time between when a product is introduced to a given store and the time when the product was first introduced by the retailer. Retailer 32 is slower to introduce gluten-free products to all stores compared to wheat and wheat plus products. For wheat and wheat plus products over 81% of product-store introductions occur within one year of the product's first introduction by the retailer. However, for gluten-free products only 73% of product-store introductions occur within one year of the product's first introduction by the retailer. Further, during the second and third years following a gluten-free product's first introduction an additional 10% and 12% of store introductions occur per year, respectively. This difference between wheat pastas and gluten-free pastas may be due to learning if the retailer is better able to predict demand for wheat and wheat plus pastas compared to gluten-free pastas. Learning may also help explain why the retailer waits two to three years to introduce gluten-free products to a large proportion of stores. If the retailer used some stores as a test market to learn about demand then we should expect to see delayed introductions in the non-test markets. The use of introductory markets for products the retailer is uncertain about would explain the difference in rollout speeds between wheat and gluten-free products and why gluten-free products are introduced to waves of stores two and three years after their overall introduction.

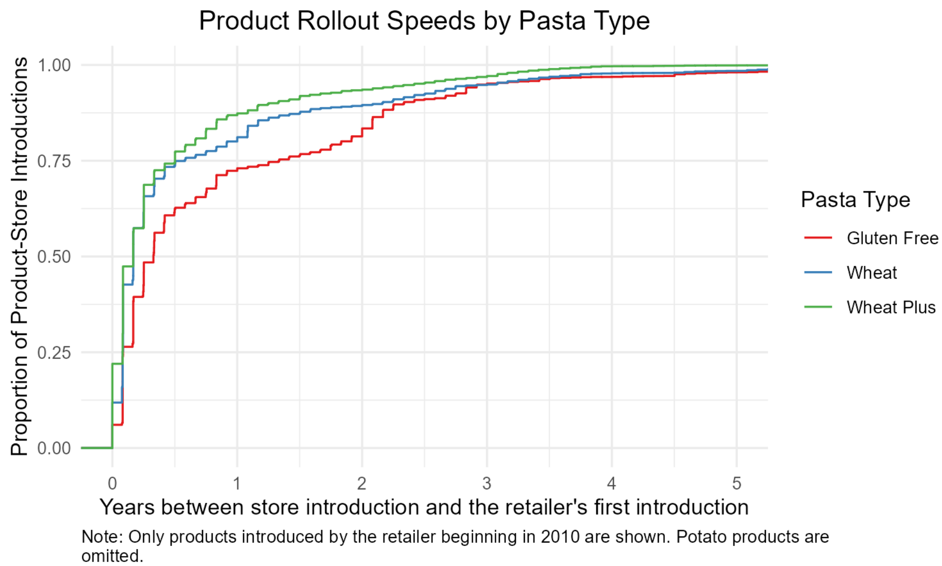


Figure 1.2: Product Introduction Speeds by Pasta Type

1.4 Descriptive Statistics

1.4.1 Differences in Product Rollout Speeds

To explore how the newness of a product impacts the speed at which it is introduced, I look at the impact of the number of previously introduced products that share the same brand and pasta type as the new product. If retailer 32 uses test markets when it is uncertain about demand, then we should expect to see products with few similar, existing substitutes introduced into fewer stores. I test this hypothesis that more similar products increase rollout speeds using the following negative binomial regression, [8]

$$\log(E[\text{Num store intros}_j \mid x_j]) = \beta_0 + \beta_1 \text{Prior brand intros}_j + \beta_2 \text{Prior type intros}_j \times \text{Pasta type}_j$$
$$+ \beta_3 \text{Prior brand intros}_j \times \text{Prior type intros}_j + \text{Pasta type}_j + \epsilon_j \tag{1.1}$$

where j indexes newly introduced products. $\text{Num store intros}_j$ is the number of stores the new product j has been introduced into within the first 12 months. [9] $\text{Prior brand intros}_j$ is the number of existing products the retailer has introduced that share the same brand as the new product j. $\text{Prior type intros}_j$ is the number of existing products the retailer has introduced that share the same pasta type as the new product j. [10]

I show the results of this regression in Table 1.4 as well as Figures 1.3 and 1.4. Figure 1.3 shows the marginal impact of increasing the number of existing same-brand products on the number of stores a new product is introduced into. Increasing the number of same-brand products significantly increases the number of stores a new product is introduced into for both gluten-free and wheat plus pastas. This fits the hypothesis that retailer 32 is learning about demand. When a new product has more existing same-brand products then the retailer has more information about demand and can introduce the new product widely. We can also see in Figure 1.3 that the impact of an additional same-brand product falls as the number of same-type products increases. Additional existing same-type products also provide information about demand, so the addition of an existing same-brand product provides less additional information. This also explains why we do not see a significant impact for additional same-brand wheat pastas. Retailer 32 already sells a large number of wheat pastas and does not gain additional information about demand from additional same-brand products.

The results in Figure 1.3 indicate that the retailer is learning both from existing products of the

[8] I use a negative binomial model because the outcome of interest, the number of store introductions, is a count bounded from below by 1 (because I do not observe products not introduced). I use a negative binomial model rather than a Poisson model because the count of store introductions is over-dispersed.

[9] I consider all products introduced by retailer 32 for the first time between 1/1/10 and 12/31/18. I do not consider products introduced during 2019 because I am unable to observe a full 12 months of data after a product's introduction.

[10] The number of existing products includes all products previously introduced by the retailer that I observe using the Nielsen data back to 2004.

Table 1.4: Impacts on the Speed of a New Product's Store Introductions

	Model 1
(Intercept)	7.559***
	[6.283, 8.842]
Prior brand intros	0.022**
	[0.005, 0.042]
Prior brand intros x Wheat	−0.005***
	[−0.007, −0.003]
Prior brand intros x Gluten-Free	−0.006
	[−0.013, 0.0004]
Prior brand intros x Wheat Plus	−0.043***
	[−0.060, −0.025]
Prior brand intros x Prior type intros x Wheat	−0.000 01
	[−0.000 05, 0.000 02]
Prior brand intros x Prior type intros x Gluten-Free	−0.000 07
	[−0.0002, 0.000 09]
Prior brand intros x Prior type intros x Wheat Plus	0.0002+
	[−0.000 03, 0.0004]
N	436
FE: Pasta type	X

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note:
95% robust confidence intervals are shown. Only products that were sold for at least 12 months are considered.

same brand and of the same type. When the number of same-type products is large (such as for wheat pastas) there is little additional information from increasing the number of same-brand products. As a result, the impact of an additional same-brand product on the number of store introductions declines.

Figure 1.4 shows the marginal impact of increasing the number of same-pasta type products on the number of stores a new product is introduced into. In most instances increasing the number of same-type products has no significant impact on the number of stores a new product is introduced into. I only estimate a significant impact of increasing the number of same-type products for newer brands of wheat pastas. This impact is negative indicating that additional products of the same type are correlated with a new product being introduced into fewer stores. This implies that the retailer is more uncertain about demand for new wheat pastas when there are few existing products of the same brand. One possible explanation is that the product space for wheat pasta is relatively crowded (there

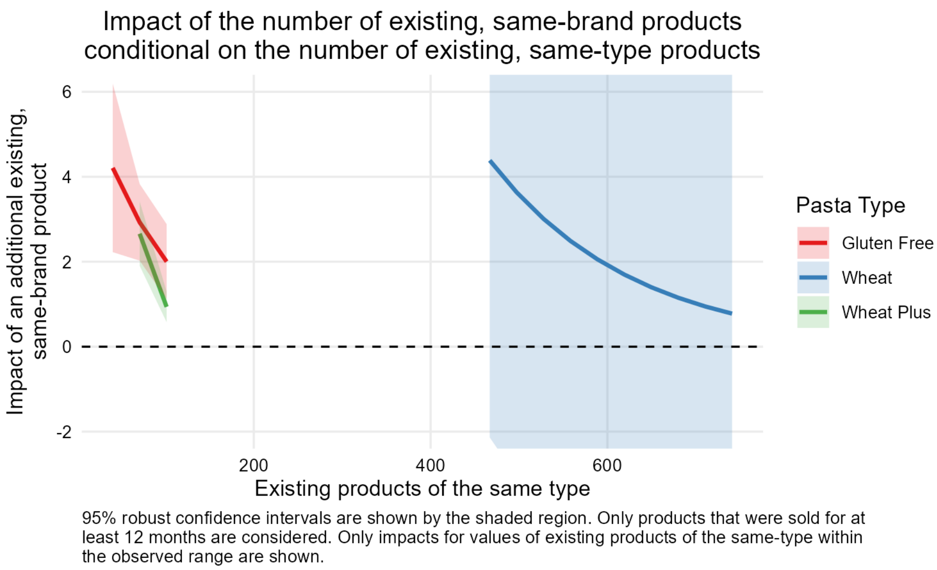


Figure 1.3: Impact of Existing, Same-Brand Products on a New Product's Store Introductions

are a large number of wheat pastas available), so in order for a new product to make it to market it must have some innovative characteristic that the retailer needs to learn about.

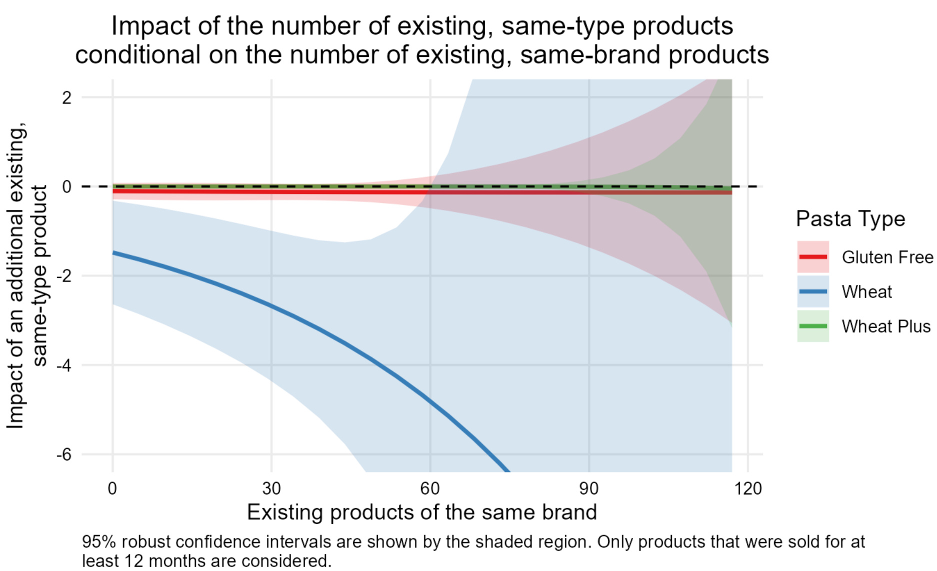


Figure 1.4: Impact of Existing, Same-Type Products on a New Product's Store Introductions

Together, Table 1.4 and Figures 1.3 and 1.4 show that retailer 32 primarily relies on existing products of the same brand rather than existing products of the same type in order to learn about demand for new products. In Appendix 1.6.2 I show that these findings are robust to using the number of DMAs rather than the number of stores a product is introduced into to measure the breadth of a product's rollout.

1.4.2 Evidence of Learning From Prices

We cannot directly observe the retailer's beliefs so it is difficult to determine if the retailer is learning. However, because a change in beliefs about demand should lead to a change in prices we can use prices to find evidence of learning. If retailer 32 is learning about demand for new products then we should see an impact on pricing. As retailer 32's beliefs about demand change prices should change as well. For example, if a product sells better than expected retailer 32 will learn that demand for the product is high and can raise prices. Thus, the variability in prices over time to provides information about beliefs. For example, if the retailer were fully informed about demand prices should be stable and not change month-to-month. However, if the retailer is uncertain about demand and is learning as products sell better or worse than expected then prices should vary.

One way to measure the variability in prices is by looking at how much prices change month-to-month. If we assume costs and the competitive environment are constant over time (a strong assumption that we know is not true) then prices should be constant in the absence of learning. As a result, any change in prices indicates that the retailer is learning about demand. I look at the relationship between (the squared) month-to-month changes in prices and the time since a product has been introduced by the retailer using the regression in Equation (1.2).

$$\begin{aligned}\left(p_{jst} - p_{jst-1}\right)^2 =& \beta_0 + \sum_{\tau=0}^{10} \beta_{1,\tau} \mathbf{1}(\tau \text{ years since retailer intro})_{jt} + \beta_2 \text{Prior brand intros}_j \\ & + FE_s + FE_t + FE_j^{\text{pasta type}} + \epsilon_{jst}\end{aligned} \tag{1.2}$$

In Equation (1.2) the outcome of interest is the squared change in price for product j at store s between month t and the prior month $t-1$. The primary variable of interest in this regression are the indicators $\mathbf{1}(\tau \text{ years since retailer intro})_{jt}$ which are equal to 1 if at time t product j was introduced by the retailer for the first time between τ and $\tau+1$ years ago. I also control for Prior brand intros$_j$, the number of same-brand products previously introduced by the retailer. As shown in the previous section same-brand products provide information about demand and reduce the need for learning. Finally, I also include fixed effects for the store s, time period t, and the pasta type.

In Figure 1.5 I display the estimates of $\beta_{1,\tau}$ where the excluded term is when product j was introduced less than one year ago. As a result, the $\beta_{1,\tau}$ coefficient estimates represent changes to the squared difference in prices compared to the first year after a product is introduced by retailer 32. As shown in Figure 1.5, the coefficient estimates are stable for the first three years and are insignificantly different from zero for two out of those three years. This indicates that before a product's fourth year of availability the retailer is continuing to learn about demand and update prices at the same pace. However, starting in year four the coefficient estimates are significantly different from zero and are declining which indicates that the longer a product has been available from retailer 32 the less prices change month-to-month. Given our assumption that prices are constant aside from the impact of learning, Figure 1.5 indicates that the retailer is learning about demand and this primarily occurs in the first four years of the product's lifecycle.

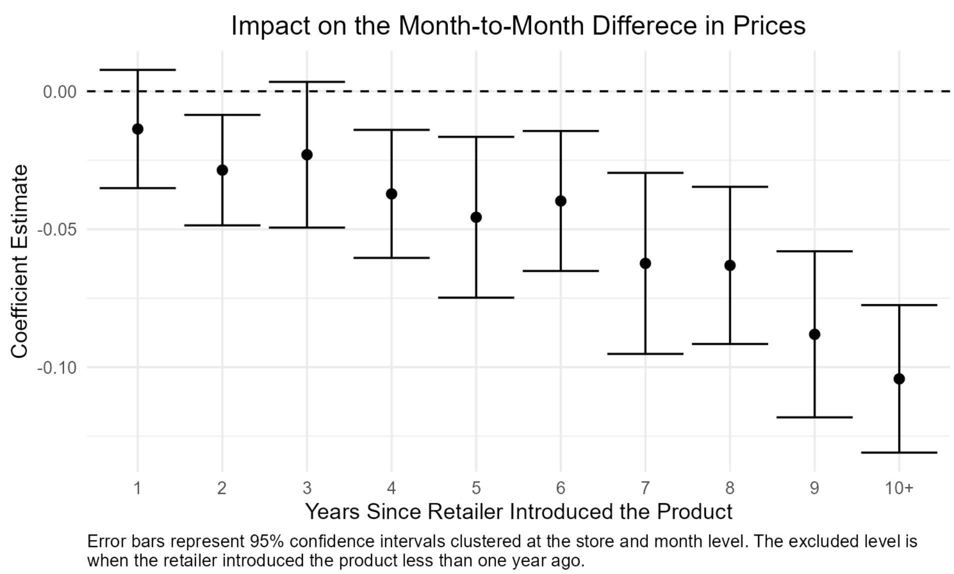


Figure 1.5: Impact of the Time Since Retailer Introduction on Month-to-Month Pricing Errors

One limitation of this analysis is the assumption that the only factor causing prices to change is the retailer learning about demand. Prices might also change if the cost of the product changes, which given the ten year sample seems likely. Additionally prices might change in response to changes in the competitive environment. For example, if during the sample period a close substitute to product j is introduced the price for product j should fall because of the increased competition. The focus of this chapter is about product introductions so we should expect changes to the competitive environment, and changes in prices, throughout the sample period. As a result, we should expect to observe month-to-month price variation throughout a product's lifecycle.

Although the decline in month-to-month price variation shown in Figure 1.5 is indicative that the retailer is learning about demand to fully conclude that learning is causing prices to change we need to control for changes to costs and the competitive environment. Given that cost data is not available, additional modeling assumptions are necessary. A model that includes costs and changes to the competitive environment is provided in Chapter 2. However, the descriptive findings that learning appears to me most concentrated in the first four years will be useful for that analysis.

1.5 Conclusion

In this chapter, I study how retailers use their network of stores to learn about demand and set prices for new products. By examining a retailer's product introduction and pricing choices, I found that existing products with similar product characteristics provide valuable information about demand. This information allowed the retailer to introduce new products widely. However, in the absence of similar products, the retailer faced uncertainty about demand and implemented a staggered rollout strategy, launching the new product in a smaller subset of stores. This approach allowed the retailer to learn about demand in initial markets and use that information when setting prices and launching the new product more widely.

By examining the retailer's pricing strategies, I showed evidence that the retailer learns about demand over time. When retailers learn about demand they update prices to account for new information, leading to fluctuations in prices. I found that during the first three years after introducing a new product, the level of price volatility remained consistent. However, four years into the product's lifecycle, price volatility decreased significantly. This indicates that after four years the retailer has gained a better understanding of demand for the product, and is updating prices less frequently as a result.

Overall, the main contribution of this chapter is to provide insights into how retailers learn about demand for new products and set prices. This chapter takes a descriptive approach to document facts about how retailers introduce new products and provides evidence that the retailer is learning. However, because I cannot directly observe what the retailer knows I am unable to isolate the impact of learning separately from the impact of other factors such as changing costs or changes to the competitive environment from the entry and exist of competing products. As a result, this chapter provides a foundation for Chapter 2 where I introduce a structural model that allows me to control for these factors and more deeply study how the retailer learns about demand.

1.6 Appendix

1.6.1 Data Description

I primarily rely on Nielsen's retail scanner dataset that covers the period 2006–2019. This data provides information about sales, products, and stores. The retailer scanner dataset provides sales information weekly for a national sample of stores. Sales information consists of the weekly units sold and volume weighted average prices for all products (defined by their UPC) sold in a given store. Product information consists of basic product attributes including a product's brand, package size (if the product is a single or multipack), and weight. Product information also includes some rough information about a pasta's shape and type, but I found this information to be too general to categorize products. Store information consists of the store's parent retailer and the store's geographic location at the three-digit zip-code, FIPS county code, and DMA code levels.

In addition to Nielsen's retail scanner data I use Nielsen's Homescan dataset which is a panel dataset of households' grocery purchases during the period 2004–2019. This data provides all of a household's grocery purchases for the period a household is included in Nielsen's panel, typically a multi-year time frame. I use purchases in the Homescan dataset during 2004–2009 to identity products that already existed prior to 2010.

1.6.1.1 Pasta Type Classification

To augment the product characteristics provided by Nielsen I collect nutrition and ingredient data from Nutritionix, OpenFoodFacts, and the USDA's FoodData Central database. All data sources provide information included in the product's nutrition label and the product's ingredient list. Together these datasets provide data on 764 products out of the 1,448 total products I observe. Most notably, Nielsen masks product identifiers (UPCs) for store brand products. This prevents me from collecting nutrition and ingredient information for store brand products from outside sources.

I use ingredient information provided by Nutritionix, OpenFoodFacts, and the USDA's FoodData Central database to classify products into four pasta types (wheat, wheat plus, gluten-free, and potato) based on the types of flour listed as an ingredient. In order to classify products based on their ingredients I collect all ingredient lists where available. I parse these lists and manually categorize individual ingredients based on the flour type or if the ingredient was a spice/flavoring. I then assign types to products based on the categorization of individual ingredients.

Some products have nutrition information but no ingredient information available from Nutritionix, OpenFoodFacts, or the USDA's FoodData Central database. For the products that are missing ingredient information, I interpolate their pasta type by matching on observables. I use a combination of brand, calories, total fat, sodium, and protein to match products with and without ingredient information. In order to classify the remaining products that appear in the Nielsen data for which no nutrition and ingredient information is available I match on the observable product characteristics available from Nielsen. I use a combination of brand and Nielsen product type to match products with and without ingredient and nutrient information.

1.6.2 Robsutness of Differences in Product Rollout Speeds

In this appendix I replicate the analysis in Section 1.4.1. However, instead of measuring the extent of a product's rollout using the number of stores a product is introduced into I use the number of DMAs a product is introduced into. The equivalent version of Equation (1.1) is

$$\begin{aligned}\log(E[\text{Num DMA intros}_j \mid x_j]) = \beta_0 + \beta_1 \text{Prior brand intros}_j + \beta_2 \text{Prior type intros}_j \times \text{Pasta type}_j \\ + \beta_3 \text{Prior brand intros}_j \times \text{Prior type intros}_j + \text{Pasta type}_j + \epsilon_j\end{aligned} \tag{1.3}$$

The only difference between Equation (1.1) and Equation (1.3) is that in Equation (1.1) the outcome of interest is the number of stores a product is introduced to within the first year of its lifecycle. In Equation (1.3) the outcome of interest is the number of DMAs a product is introduced to within the first year of its lifecycle.

The following Figures 1.6 and 1.7 are the equivalent versions of Figures 1.3 and 1.4. Comparing the two sets of figures we can see that the choice of outcome variable is not important. The effect when measured in DMAs is equal in sign, but is scaled down in magnitude compared to the results in Section 1.4.1 that are measured in number of stores. The smaller magnitude of the results in Figures

1.6 and 1.7 is to be expected because there are multiple stores per DMA. Overall, the results in this appendix match the results in Section 1.4.1 and are indicative that the results are robust to how the speed of product's introduction is measured.

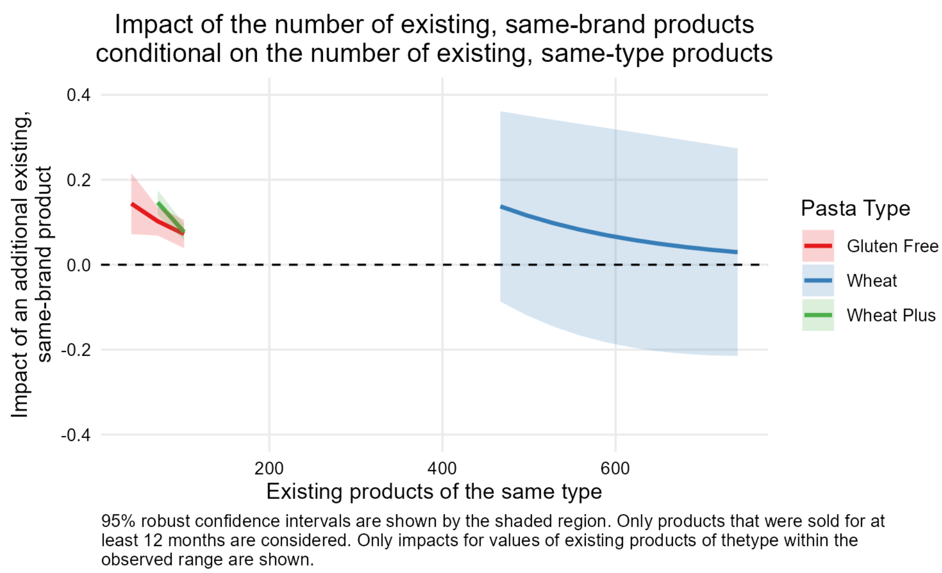


Figure 1.6: Impact of Existing, Same-Brand Products on a New Product's DMA Introductions

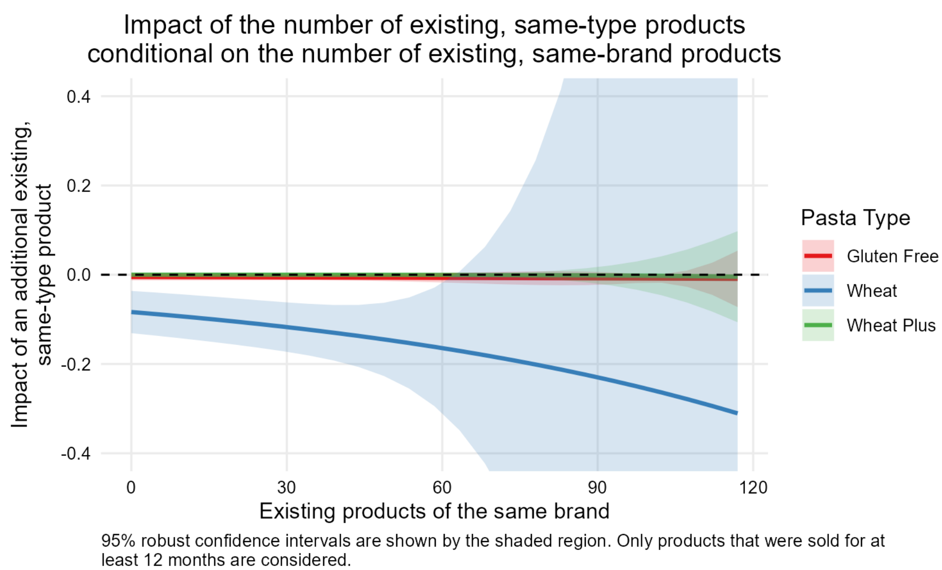


Figure 1.7: Impact of Existing, Same-Type Products on a New Product's DMA Introductions

1.7 References

Bergemann, Dirk, and Välimäki, Juuso, "Dynamic pricing of new experience goods," *Journal of Political Economy*, 114 (2006), 713–743 (The University of Chicago Press).

Berry, Steven, Eizenberg, Alon, and Waldfogel, Joel, "Optimal product variety in radio markets," *The RAND Journal of Economics*, 47 (2016), 463–497 (Wiley Online Library).

Bronnenberg, Bart J., and Mela, Carl F., "Market roll-out and retailer adoption for new brands," *Marketing Science*, 23 (2004), 500–518 (INFORMS).

Doraszelski, Ulrich, Lewis, Gregory, and Pakes, Ariel, "Just starting out: Learning and equilibrium in a new market," *American Economic Review*, 108 (2018), 565–615.

Duan, Jason A, and Mela, Carl F, "The role of spatial demand on outlet location and pricing," *Journal of Marketing Research*, 46 (2009), 260–278 (SAGE Publications Sage CA: Los Angeles, CA).

Eizenberg, Alon, "Upstream innovation and product variety in the us home pc market," *Review of Economic Studies*, 81 (2014), 1003–1045 (Oxford University Press).

Fan, Ying, and Yang, Chenyu, "Competition, product proliferation, and welfare: A study of the US smartphone market," *American Economic Journal: Microeconomics*, 12 (2020), 99–134.

Handel, Benjamin R, and Misra, Kanishka, "Robust new product pricing," *Marketing Science*, 34 (2015), 864–881 (INFORMS).

Hitsch, Günter J, "An empirical model of optimal dynamic product launch and exit under demand uncertainty," *Marketing Science*, 25 (2006), 25–50 (INFORMS).

Huang, Guofang, Luo, Hong, and Xia, Jing, "Invest in information or wing it? A model of dynamic pricing with seller learning," *Management Science*, 65 (2019), 5556–5583 (INFORMS).

Huang, Yufeng, Ellickson, Paul B., and Lovett, Mitchell J., "Learning to set prices," *Journal of Marketing Research*, 59 (2022), 411–434.

Jeon, Jihye, "Learning and investment under demand uncertainty in container shipping," *The RAND Journal of Economics*, 53 (2022), 226–259.

McLennan, Andrew, "Price dispersion and incomplete learning in the long run," *Journal of Eco-*

nomic dynamics and control, 7 (1984), 331–347 (Elsevier).

Nair, Harikesh S., "Diffusion and pricing over the product life cycle," in *Handbook of the economics of marketing, volume 1*, Jean-Pierre Dubé and Peter E. Rossi, eds., Handbook of the economics of marketing (North-Holland, 2019).

Nosko, Chris, "Competition and quality choice in the cpu market," *Manuscript, Harvard University*, (2011).

Raman, Kalyan, and Chatterjee, Rabikar, "Optimal monopolist pricing under demand uncertainty in dynamic markets," *Management Science*, 41 (1995), 144–162 (INFORMS).

Trefler, Daniel, "The ignorant monopolist: Optimal learning with endogenous information," *International Economic Review*, (1993), 565–581 (JSTOR).

Wollmann, Thomas G., "Trucks without bailouts: Equilibrium product characteristics for commercial vehicles," *American Economic Review*, 108 (2018), 1364–1406.

CHAPTER 2

A Structural Approach to New Product Introductions, Retailer Learning, and Pricing

Pricing new products when demand is unknown can be challenging, especially for products with no close competitors or prior sales in other markets. The chapter examines how retailers can use information across stores to learn about consumer demand and set prices. Building on the findings from the previous chapter, I develop a structural model to isolate the impact of learning on pricing. When a retailer has more information from similar existing products or prior introductions at other stores, they set prices closer to the optimal level. Additionally, after introducing a new product, retailers continue to learn about demand and refine their pricing strategy. Sharing information between stores in different geographic regions can help retailers learn about demand if consumers are similar across locations. While information sharing typically leads to prices that are closer to the optimal level, there are cases where information from different geographies has a negative impact on pricing. [1]

2.1 Introduction

As discussed in Chapter 1 the problem of setting prices can be difficult when demand is unknown. Demand is often unknown when a good is first introduced, particularly if it has few similar competitors or has not been introduced in other markets. In Chapter 1, I showed that uncertainty about demand is correlated with a retailer introducing a new product into fewer stores. This pattern can be explained

[1] Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

if the retailers is using a staggered rollout strategy to learn about demand. When retailers operate multiple stores they can share information across stores and learn about demand in a new market using prior data from other markets. A staggered rollout strategy allows the retailer to initially introduce and learn about demand in a small number of stores. The retailer can then use that information when introducing the product and setting prices in new markets.

In this chapter I explore how retailers can share information across stores to learn about demand and set prices. However, information from other stores is only valuable if consumers are similar across stores. Previously, Huang, Ellickson, and Lovett (2022) showed that grocery retailers selling liquor in Washington state for the first time had to relearn about consumer demand because consumers in other states had significantly different preferences than consumers in Washington. In this chapter I study a national retailer to understand to what extent information is useful across geographies. I find that typically information sharing is useful, however this is not uniformly true. In most instances data from other geographic markets allows retailers to set prices closer to the optimal level. However there are some cases where data from other geographic markets results in prices that are further from the optimum. Overall I find that the ability to use information across geographies is valuable, but differences in consumer preferences must be taken into account before information can be used across all stores.

This chapter builds on the findings from Chapter 1 to build a structural model of consumer demand and price setting. Chapter 1 provided a descriptive analysis of the impact of learning on product introductions and pricing, but was unable to control for changes to costs or changes to the competitive environment from the entry and exist of new products. In this chapter I build on my previous findings to incorporate both costs and the competition between products into my model to isolate the impact of learning. As a result, the findings in this chapter can more convincingly be attributed to learning.

I adopt an approach similar to that of Doraszelski, Lewis, and Pakes (2018) and Huang, Ellickson, and Lovett (2022) to study how retailers learn about demand. Building on the results from Chapter 1, I divide the lifecycle of a new product in a store into two distinct periods. During the first period, I make no assumptions about what the retailer knows about demand in order to allow the retailer to learn about it. In contrast, in the second period I assume that the retailer has had sufficient time to learn and therefore knows demand. This division of the product's lifecycle enables me to estimate consumer demand over the full time period and then to recover the cost of offering the product in the second period when the retailer knows demand. I use the cost estimates from the second period along with institutional details to estimate costs in the first period. Using these costs, I generate the optimal prices the retailer would have set in the first period if they knew demand for the new product. Importantly, my measure of optimal prices incorporates both costs and competing products, allowing me to isolate the impact of learning on price setting by comparing optimal prices to the actual prices set by the retailer.

When estimating optimal prices, I make no assumptions about the retailer's knowledge of demand when introducing a new product to a new store. It is possible that the retailer already has information about demand from similar products or other markets, in which case I would expect the actual prices set by the retailer to match the optimal benchmark. Conversely, if the retailer has no prior information, I would expect actual prices to differ from the optimal benchmark. The results I observe align with the descriptive results in Chapter 1. I find that when the retailer has more information from similar existing products or prior introductions at other stores, they are able to set prices closer to the optimal level. For instance, I observe that for every year a product is available at other stores, the retailer can set prices $0.04 closer to the optimal level when launching the product in a new store.[2] This suggests that the retailer learns about demand before introducing a product to a new store by using information from similar products and stores.

In addition to learning about demand before a product's launch, I also find evidence that retailers learn about demand and improve their pricing strategy over time. After a product is introduced the retailer continues to observe sales and learn about consumer demand. In Chapter 1 I found that during the first four years of a product's lifecycle prices were consistently variable, indicating that the retailer was learning and updating prices. After controlling for costs and competing products to construct the optimal price level, I find that during the first four years of a product's lifecycle the retailer is able to set prices $0.03 closer to the optimum level every year on average. This suggests that retailers continue to learn about demand and refine their pricing strategies after a product has been introduced to a new

[2] The average price of a product is $1.87 while the median price is $1.62.

store.

The ability to learn about demand and improve a product's pricing comes from two sources. First, the retailer can use sales data from the relevant market. Second, if consumer preferences are similar, the retailer can use sales data from other markets where the new product is sold. As previously discussed, sales data from other markets improves pricing when a product is first introduced. Similarly, data from other markets may be useful to supplement home market data and improve pricing after the product's introduction. To explore this I use the fact that the retailer I study operates in three distinct geographic regions: the Pacific Northwest, the Mountain West, and the Mid-Atlantic. I use these regions to decompose the impact of an additional year of sales data by region. I find that generally, more data is useful in reducing the difference between actual and optimal prices, regardless of the region the data comes from. However this is not always true. For example, when setting prices in the Mid-Atlantic sales data from both the Mid-Atlantic and Pacific Northwest reduce the difference between actual and optimal prices. Conversely, data from the Mountain West results in prices further from the optimal level. The result is that while information is generally useful across geographic regions, retailers must account for differences in consumer preferences between regions.

In this chapter I validate and extend the descriptive results from Chapter 1 by controlling for costs and the entry and exit of competing products. My findings demonstrate that information sharing between stores is valuable and can be used to reduce initial pricing errors. Retailers learn and improve their pricing strategy over time by using information from similar products and existing markets. Retailers can further improve their pricing strategy by using information from both a store's home market and from other geographically distinct markets. This suggests that there are returns to scale from operating in multiple markets. Large retailers benefit because they can introduce a new product and learn about demand in one market and subsequently use this information to set prices closer to the optimal level in other markets. In contrast, small retailers incur a larger cost to introducing new products because they have less prior information about demand and cannot set prices as close to the optimal level. Overall, the impact of learning is important to how retailers introduce products across their stores and how they set prices during the early stags of a product's lifecycle.

2.2 Literature Review

As Chapter 1 and this chapter study the same topic, the literature review in Chapter 1 is relevant for this chapter as well. However, it is useful to compare the methodological approach in this chapter to the prior literature, especially Doraszelski, Lewis, and Pakes (2018) and Huang, Ellickson, and Lovett (2022). As in these two papers, I take the approach of studying learning without directly modeling the firm's learning process. I instead infer that the firm is learning by comparing the firm's actions to a full-information benchmark. While my approach is similar to both Doraszelski, Lewis, and Pakes (2018) and Huang, Ellickson, and Lovett (2022) I build on these papers and contribute to the larger literature by studying the sources of information firms use to learn about demand.

Within the literature on how firms learn about demand this chapter builds on the methods first used by Doraszelski, Lewis, and Pakes (2018). Doraszelski, Lewis, and Pakes (2018) study how electricity providers learn to set prices. The authors introduce a technique that allows them to study learning without modeling the learning process. The authors do so by first splitting their time period into a learning and informed stage. The important difference between the two time periods is that during the informed stage the authors assume that firms have full information. This assumption allows the authors to estimate demand and then use their demand estimates and the full information assumption to recover costs during the during the informed stage. The authors use their estimate of costs in the informed stage along with institutional details to estimate costs during the learning stage. Finally, the authors use their estimated demand model along with their cost estimates in the learning stage to estimate what prices should have been if firms were informed in learning stage. After recovering counterfactual prices the authors fit different adaptive learning models to the data and compare the price predictions of learning models to prices under a full information assumption. Importantly, the authors did not need to assume a model of learning in the learning stage which allows them to compare the predictions of different learning models. The authors show that prices from adaptive learning models better fit observed prices than prices the result from a full information assumption. The multi-step method for recovering counterfactual, full-information prices without assuming a model of learning is an important innovation and underlies both the structural model in Huang, Ellickson, and Lovett (2022) and this chapter. Huang, Ellickson, and Lovett (2022) and this chapter differ from

Doraszelski, Lewis, and Pakes (2018) by not focusing on determining the correct model of learning, but rater by studying how firms' actions are impacted by the need to learn about demand.

This chapter further differs from the methodology developed by Doraszelski, Lewis, and Pakes (2018) by allowing firms to learn about the same products in different locations and starting at different times. In the setting of both Doraszelski, Lewis, and Pakes (2018) and Huang, Ellickson, and Lovett (2022) there is a single shock that necessitates firms to learn about demand starting at the same time for all firms and all products. In this chapter learning occurs at a number of different locations and may start at different times depending on when new products are introduced. This provides an additional challenge to using the multi-step method introduced by Doraszelski, Lewis, and Pakes (2018). I overcome this challenge by introducing a new assumption about the extent that uncertainty matters for the pricing of different products.

Huang, Ellickson, and Lovett (2022) is the closest to this chapter in the literature on how firms learn about demand. Huang, Ellickson, and Lovett (2022) study how grocers learn to set prices after being allowed to sell liquor for the first time. The authors show that over time prices settle to a steady state. Similar to Doraszelski, Lewis, and Pakes (2018), the authors estimate a multi-step model using two periods, one in which the retailer is learning and one in which the retailer has reached the steady state. Huang, Ellickson, and Lovett (2022) first estimate demand during the steady state to recover demand parameters and estimates of costs. Using steady state cost estimates along with proxies for costs from other states the authors estimate costs prior to the steady state, during the learning phase. Finally, by combining demand estimates from the steady state with cost estimates over the length of their sample the authors estimate the degree of mispricing over time. The authors show that mispricing is largest when the retailer is first able to sell liquor. Mispricing initially declines quickly and then the rate of improvement slows over time. The authors do not model learning parametrically but show that descriptive evidence is consistent with Bayesian learning. This chapter uses a similar estimation approach as Huang, Ellickson, and Lovett (2022) to recover estimates of mispricing over time. However, this chapter goes a step further by examining factors that are related to the size of mispricing errors and factors that change the speed by which mispricing declines.

A third related paper in the literature on how firms learn is Jeon (2022). Rather than taking the approach of Doraszelski, Lewis, and Pakes (2018) and not imposing a structure on the learning process

Jeon specifies a parametric data generating process and allow firms to be uncertain about the true parameters. Although the method of modeling learning is different from this chapter, both Jeon (2022) and this chapter study how learning impacts firm's strategic decisions.

This chapter is also related to the broader literature on how firms price products under uncertainty. Most papers in the literature on pricing under uncertainty (such as Osborne (2011)) focus on consumers' uncertainty and model firms as having full information about demand. While I do include elements in my model to capture consumer hesitancy to try new products, in this chapter I focus on firm's rather than consumers' uncertainty. In the literature on pricing products under uncertainty Ching (2010) also examines firms' problems when they are uncertain about demand. Ching (2010) develops a model of generic drug pricing after patents expire. In their model patients, physicians, and firms are uninformed about the drug's quality and learn through experience. Firms set prices conditional on beliefs about the drug's effectiveness. The findings in Ching (2010) are similar to this chapter. Both this chapter and Ching (2010) find that firms use their experience selling both new and related products to learn about demand.

2.3 Setting

The setting in this chapter is the same as Chapter 1 and readers should refer to Chapter 1 for more details. However, I repeat here important information from Chapter 1 that is relevant for the analysis in this chapter.[3]

2.3.1 Pasta Products

In this chapter I focus on the grocery sales of dried pastas from a single retailer. Pasta products are relatively simple because they can be defined by three characteristics: brand, type, and shape. For example, a commonly sold product is Barilla wheat spaghetti. Here Barilla is the brand, wheat is the type, and spaghetti is the shape. I use Nielsen retail scanner data to identify the set of pasta products sold during 2010–2019.

[3] For additional details on the data sources used see Appendix 2.7.1.

Using three sources of nutritional and ingredient information I define a pasta's type. Pasta types are easily defined by a product's ingredient list because the list of ingredients is relatively short and consist of primarily flour and water. I categorize pastas into four types: wheat, wheat plus, gluten-free, and potato. I define wheat pastas as products that contain only non-whole grain wheat flour, typically durum or semolina. I define wheat plus pastas as products that contain whole grain wheat flour or a mix of wheat and a non-wheat flour that is typically added to increase the nutritional content of the pasta. One example is pasta marketed as protein plus that contains a mix of wheat flour and a legume flour for added protein. All wheat plus pastas are defined to contain wheat and are not gluten-free. In contrast, I define gluten-free pastas as products that do not contain wheat flour. Typically, gluten-free pasta is made from some combination of rice, corn, or legume flours. Finally, I define potato pastas as products that primarily contain potatoes. Potato pastas may also be made with some wheat flour and are often marketed as gnocchi. [4]

Table 2.1 shows the number of pasta products of each type as well as the proportion of revenue and units sold for each pasta type by Nielsen's retailer 32. Wheat pastas make up the majority of products and make up a disproportionately large share of units sold and revenue. In contrast, gluten-free products make up a disproportionately small share of units sold and revenue.

Table 2.1: Summary of Pasta Types Sold

		Proportion of		
Pasta Type	**Products**	**Products**	**Units Sold**	**Revenue**
Wheat	733	70%	84%	78%
Wheat Plus	129	12%	11%	12%
Gluten Free	152	15%	3%	6%
Potatoes	27	3%	2%	4%

Aside from a pasta's type, products are also defined by their brand and shape. Using Nielsen data I am able to identify the brand of each product. On average a given store offers products from 14

[4] Potato pastas only make up a small fraction of all pasta products and are included here for completeness.

brands. I do not observe a consistent measure of pasta shapes either from Nielsen or the data sources that provide ingredient information. As a result, I do no include shape information in my analyses. However, from manually looking up pasta products most brands offer the same core pasta shapes (examples include spaghetti, linguine, rotini, macaroni, etc.) while larger brands may also offer more specialty shapes.

2.3.2 Retailer 32

In this paper I focus on Nielsen's retailer numbered 32. I choose to focus on a single retailer to better understand how that retailer introduces new products across the stores it operates. I focus on retailer 32 because it is a large national retailer that operates in a number of distinct geographies. Retailer 32's geographic diffusion is interesting because demand is potentially very different across the country. Further, because retailer 32 operates in multiple markets they are able to stagger product rollouts across markets.

In Table 2.2 I present the top 10 Nielsen designated market areas, DMAs, in which retailer 32 operates as well as the number of stores per DMA. In total retailer 32 has at least 10 stores in 14 different DMAs. In Figure 2.1 I show the geographic distribution of retailer 32's stores. Retailer 32 operates in the Mid-Atlantic region around DC, the Mountain West region around Denver, and the Pacific Northwest region between San Francisco and Seattle.

Table 2.2: Top 10 DMAs by Number of Stores

DMA	**Number of Stores**
San Francisco-Oakland-San Jose, CA	131
Seattle-Tacoma, WA	105
Phoenix, AZ	77
Portland, OR	74
Denver, CO	73
Washington, DC	72
Sacramento-Stockton-Modesto, CA	49
Baltimore, MD	23
Spokane, WA	23
Tucson, AZ	21

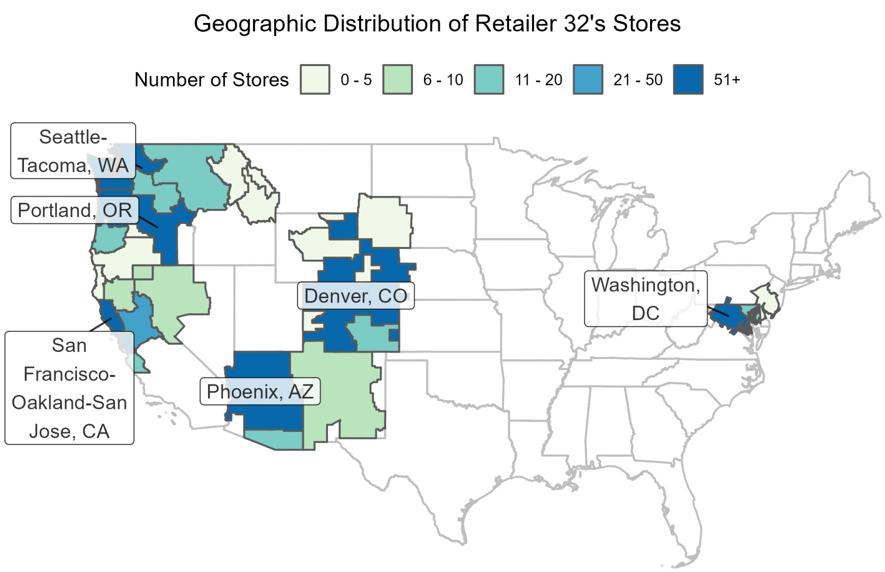


Figure 2.1: Geographic Distribution of Retailer 32's Stores.

DMAs with 51+ stores are labeled.

In total from Nielsen scanner data I identify 767 stores belonging to retailer 32 that continuously report sales from 2010 to 2019. [5] [6] On average retailer 32 offers 99 products per store from 14 brands. The number of products per store varies over time as products are added and removed, but the average within-store standard deviation of the number of products offered is 10.7.

2.3.3 New Products

In this paper I focus on the introduction of new products across stores. I define a product's introduction to a store as the first date that I observe a sale. This introduces the potential for measurement error if a product is available, but does not sell for an extended period of time. To account for this I aggregate weekly Nielsen data to the monthly level. Further, to ensure that I can correctly identify the set of existing products I limit the period of my analysis to 2010–2019 although I have data as far back as 2004. [7] By doing so I use the Nielsen data in 2004–2009 to identify existing products. Table 2.3 shows the number of new products added to each store per year. The number of new products shows no large difference between the beginning and end of my sample indicating no major issues in my ability to identify new and existing products using the 2004–2009 data.

[5] I identify stores belonging to retailer 32 as those with both parent_code and retailer_code equal to 32. There are an additional 28 stores with the parent_code 32 and a different retailer_code. There are an additional 14 stores with the retailer_code 32 and a different parent_code. I exclude these additional stores.

[6] There are 57 stores who enter the data in or after 2010. I exclude these stores because in the first year data for a store is available I cannot identify if a product is new to that store or if the store did not previously exist in the data. There are 116 stores that I exclude because they exit the sample prior to 2019.

[7] Nielsen scanner data which provides panel data on all products sold within a store is available for 2006–2019. Nielsen Homescan data which provides panel data on all products purchased by a household is available for 2004–2019.

Table 2.3: Summary of New Products Added Per Store Each Year

Year	Min	Median	Average	Max
2010	4	22	22.0	52
2011	3	33	31.5	64
2012	6	25	26.1	58
2013	3	21	21.3	40
2014	9	23	23.8	48
2015	2	10	16.6	47
2016	2	21	25.8	91
2017	5	21	23.5	81
2018	3	28	27.1	89
2019	1	13	14.0	53

The time a given product is introduced to a store varies widely across retailer 32's network of stores. Figure 2.2 shows the difference in time between when a product is introduced to a given store and the time when the product was first introduced by the retailer. Retailer 32 is slower to introduce gluten-free products to all stores compared to wheat and wheat plus products. For wheat and wheat plus products over 81% of product-store introductions occur within one year of the product's first introduction by the retailer. However, for gluten-free products only 73% of product-store introductions occur within one year of the product's first introduction by the retailer. Further, during the second and third years following a gluten-free product's first introduction an additional 10% and 12% of store introductions occur per year, respectively. This difference between wheat pastas and gluten-free pastas may be due to learning if the retailer is better able to predict demand for wheat and wheat plus pastas compared to gluten-free pastas. Learning may also help explain why the retailer waits two to three years to introduce gluten-free products to a large proportion of stores. If the retailer used some stores as a test market to learn about demand then we should expect to see delayed introductions in the non-test markets. The use of introductory markets for products the retailer is uncertain about would explain the difference in rollout speeds between wheat and gluten-free products and why gluten-free products are introduced to waves of stores two and three years after their overall introduction.

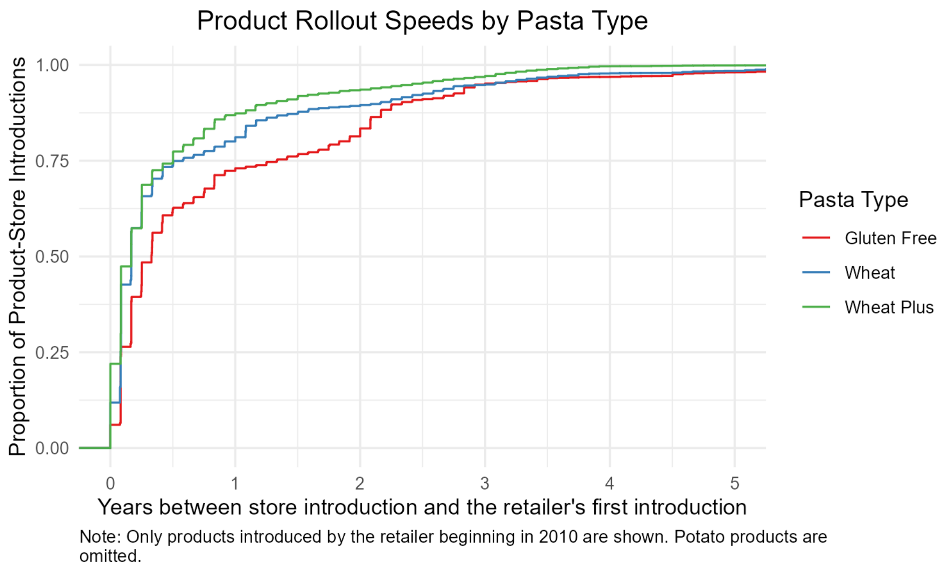


Figure 2.2: Product Introduction Speeds by Pasta Type

2.4 Structural Model

2.4.1 Conceptual Overview

In order to estimate how retailer 32 learns I construct a structural model that allows me to simulate a counterfactual retailer with the same costs and competitive environment, but with full information about demand. By comparing the prices set by the counterfactual, fully-informed retailer to the actual, observed prices I can estimate the degree of mispricing due to learning. If the retailer is fully informed about demand there should be no difference between observed and counterfactual prices because I am holding constant both costs and the competitive environment. As a result, the difference in prices at any point in time is informative about the uncertainty the retailer has about demand. Additionally, changes in the difference between observed and counterfactual prices over time are indicative of the speed at which the retailer learns. For example, if two products have the same amount of initial mispricing and for the first product is takes one year for observed prices to converge to counterfactual prices while for a second product it takes two years for prices to converge. Then from this we can infer that the retailer is learning faster about demand for the first product than the second.

To simulate prices as if the retailer were fully informed I construct a structural model similar to Doraszelski, Lewis, and Pakes (2018) and Huang, Ellickson, and Lovett (2022). This model allows me to estimate mispricing due to learning without imposing structure on how the retailer learns. To gain intuition on how the model works it is useful to think about the components of pricing and what is known by the retailer and by us as researchers.

Prices are set as cost plus a markup. Costs are largely determined by upstream prices set by manufacturers and wholesalers. As a result, costs are known to the retailer but unknown to the researcher because we do not observe upstream prices. Markups are determined by demand as a combination of consumers' price sensitivity and willingness to pay for a product. Consumers' price sensitivity is not product dependent and, as a result, is known by retailers from their experience selling existing products. For new products, retailers may be initially uninformed about consumers' willingness to pay for the product because they lack experience selling the product. However, over time retailers are able to observe sales and learn about consumers' willingness to pay. As a result, retailers may initially lack information about demand, but eventually are able to learn by observing sales. In contrast, we have access to the full sales history of new products and can estimate consumers' demand for all products.

We can exploit the difference between the information we have and the information available to the retailer to simulate what the retailer would have done if they had full knowledge of consumers' willingness to pay when new products are introduced. In order to do so, we need to overcome the hurdle that we don't observe the retailer's costs. Typically in the literature to recover the retailer's costs we assume that retailers set prices according to a full information Nash-Bertrand equilibrium. This assumption allows us to use the relationship that prices are equal to costs plus a markup to recover costs. We observe prices as data and can estimate markups from demand, so using the relationship between prices, markups, and costs we can recover an estimate of costs.

Crucially recovering costs requires the assumption that the retailer sets prices in a full information Nash-Bertrand equilibrium. The onerous part of this assumption is that the retailer has full information about demand. However, we often can't assume that retailer has full information about demand for new products. Thankfully because retailers learn about demand by observing sales there is a point where the full information assumption becomes reasonable. So, if there exists a time period where retailers have had sufficient time to learn about demand, then we can assume retailers set prices according to a full information Nash-Bertrand equilibrium and we can recover costs.

The full information assumption on how retailers set prices allows us to recover costs in time periods when new products have been available for a sufficiently long time, but we care about how retailers set prices when products are still novel. Further, the costs we recovered may be different than the retailer's costs when the products are novel if costs change over time. In order to estimate costs in the time periods when products are novel we can exploit institutional details to impose structure on how costs change over time. For example, manufacturers produce multiple products and their production costs are likely correlated across products. So, if we have information about costs for an existing product from a given manufacturer those costs are informative about costs for a new product from the same manufacturer. Similarly, wholesalers are large and serve multiple stores and regions, but their costs are likely correlated across stores. So, if we have information about costs for existing products in one store those costs are informative about costs for new products in other stores. As a result, we can exploit the correlation in costs between products produced by the same manufacturer and between products offered at different stores to estimate costs during time periods when products are novel and we cannot make a full information assumption.

Finally, given an estimate of costs at every point in time we can construct optimal prices as our estimated costs plus a markup. Recall, we know the optimal markup the retailer should charge because we have access to a full sales history and can estimate demand for all products. This is in contrast to retailers who know costs but may not know the optimal markup to charge because they may not know demand for new products. As a result, the difference between our counterfactual prices and the prices set by retailers can be attributed to differences in markups which comes from differences in information about demand.[8] A large difference between counterfactual prices and actual prices indicates that the retailer has incorrect beliefs about demand. A small difference between counterfactual

[8] There may also be a difference in prices due to out inability to correctly estimate costs. We can approximate the difference due to errors in our cost estimates by examining periods when products are not novel. In these periods the retailer has the same information about demand as we do. So, if we estimate costs as if we could not recover them from the full-information assumption and use these costs to predict prices then any difference between predicted and actual prices is due to misestimating costs.

prices and actual prices indicates that the retailer has correct beliefs about demand. So, by examining the difference between counterfactual prices and actual prices and how this difference changes over time we can understand how the retailer's beliefs change and how the retailer learns about demand.

In the following sections I present the specifics of how I estimate demand, how I recover costs when the retailer is informed about demand, how I estimate costs when the retailer is learning about demand, and how I construct counterfactual, full-information prices.

2.4.2 Demand

I estimate demand using a random coefficients nested logit demand model. I assume the indirect utility of a consumer i conditional on purchasing product j from store s at time t is

$$\begin{aligned} u_{ijst} =& \sigma\nu_i + \left(\alpha + \kappa^{price} y_{it}\right) p_{jst} + \sum_g \mathbf{1}(j \in g)\zeta_{igst} \\ &+ \beta X_{jst} + FE_j + FE_s^{zip3} + \xi_{jst} + (1-\rho_g)\epsilon_{ijst} \end{aligned} \tag{2.1}$$

where ν_i is a household-specific taste shock that is common across all inside goods, is distributed standard normal, and is independent across households. Therefore, σ captures the unobserved dispersion in household's taste for inside goods. y_{it} is the logarithm of household i's income at time t, and κ^{price} captures the effect of household income on the household's taste for price. ζ_{igst} are household-specific taste shocks to pasta type g (wheat, wheat plus, or gluten-free). ζ_{igt} has the unique distribution such that $\zeta_{igt} + (1-\rho_g)\epsilon_{ijst}$ is extreme value (Cardell (1997)). Therefore $0 \leq \rho_g \leq 1$ is the correlation of consumer i's preferences for products belonging to group g. As ρ_g goes to 1 the shocks to utility for products in group g become perfectly correlated. As ρ_g goes to 0 the within-group correlation of the ϵ shocks to utility goes to 0. [9] Covariates X_{jst} are the months since the product j was introduced into store s as well as the months squared. The time since a product has been introduced is important

[9] Berry (1994) shows that nested logit models of demand can be interpreted as a random coefficients model where ζ_{igt} is a random coefficient on a group specific dummy variable. There are two differences between specifying ζ_{igt} as following the distributional assumptions consistent with a nested logit model and specifying ζ_{igt} as a standard normal variable similar to a traditional random coefficient. The first is the distributional assumptions used. Grigolon and Verboven (2014) showed that misspecifying the distribution of ζ_{igt} (ie. specifying ζ_{igt} as consistent with a nested logit when in reality ζ_{igt} is standard normal, or vice versa) has little impact on the predicted substitution patterns between products. The second difference is that specifying ζ_{igt} as a standard normal substantially increase the computational burden of estimating the model. For this reason I choose to specify the model as a nested logit.

if consumers have a preference for new or existing products. I also include fixed effects for products FE_j and fixed effects for the three-digit zip-code where a store is located FE_t^{zip3}. As a result, the error term ξ_{jst} captures transient variations in demand shocks specific to a product and three-digit zip-code combination.

While I don't model what the retailer knows or learns it is useful to think about the components of demand that are unknown to the retailer when a new product is introduced. The coefficient on all inside goods σ and the coefficients on price (α, κ^{price}) do not vary by product and are likely known by the retailer from their long history of selling pasta products. Similarly, retailers likely know ρ_g because all product groups exist prior to the start of my sample period in 2010. Retailers likely also know β and FE_s^{zip3} because they have prior experience selling new products and experience in their geographic location. However, retailers may not know consumers' average preference for new products, FE_j. If a new product is similar to existing products then retailers may be able to estimate consumers' average preference for the new product from existing products. However, if the new product is novel, the retailer may be uncertain about consumers' average preferences and will have to learn FE_j by observing sales.

Additionally, depending on how we interpret ξ_{jst} retailers may know this parameter. ξ_{jst} captures unobserved product characteristics that vary across stores or over time. Examples of unobserved characteristics that retailers know consumers' preferences for include actions taken by the retailer that impact a product's placement on store shelves or promotion in a weekly circular. Retailers move products on their shelves, such as prominently placing a product on an end-cap, or promote a product in a weekly circular because they anticipate it will drive sales. Retailers have a long experience using these types of promotions and likely know their value. ξ_{jst} also captures unobserved product characteristics that retailers may not know consumers' preferences for. Examples include changes to a product's packaging or changes in demand not caused by the retailer such as the appearance of a product in the media because of a manufacturer recall or a beneficial product placement. It may be harder for the retailer to anticipate the impact of these changes on demand because they are relatively rare and potentially short-lived in their impact.

Given a consumer gets utility from purchasing product j as defined by Equation (2.1) the consumer chooses the product that maximizes their utility among all products in the store s at time t, ie.

from the choice set J_{st}. [10] If the consumer does not choose to purchase any product their utility is normalized to $u_{i0st} = \epsilon_{i0st}$. As a result, the market share of a product j within store s at time t is an integral of nested logit choice probabilities over the random coefficients $\{\sigma, \kappa^{price}\}$: [11]

$$s_{jst} = \int s_{ijst} F(\theta_i) = \int \frac{\exp\left[u_{ijst}(p_{jst}, X_{jst}; \theta_i)/(1-\rho_{g(j)})\right]}{\exp\left[V_{ig(j)st}/(1-\rho_{g(j)})\right]} \frac{\exp\left[V_{ig(j)st}\right]}{1+\sum_{g\in G}\exp\left[V_{igst}\right]} F(\theta_i) \tag{2.2}$$

where

$$V_{igst} = (1-\rho_g)\log\sum_{k\in J_{gst}}\exp\left[\frac{u_{ikst}(p_{kst}, X_{kst}; \theta_i)}{1-\rho_g}\right] \tag{2.3}$$

2.4.2.1 Identification

I estimate demand parameters using a set of moments that set the demand shocks ξ_{jst} equal to zero conditional on instruments z_{jst}.

$$E[\xi_{jst} \mid z_{jst}] = 0 \tag{2.4}$$

Endogeneity between ξ_{jst} and shares may arise for a variety of reasons including measurement error (from aggregating prices over weeks) and price setting based on unobserved characteristics ξ_{jst}. I previously argued that ξ_{jst} can include actions taken by the retailer to impact demand such as prominently placing products on an end cap or including products in a weekly circular. Retailers often pair these non-price promotions with a price promotion creating a correlation between ξ_{jst} and prices. In order to instrument for ξ_{jst} I use two sets of instruments. The first are Hausman (1996) style instruments. For each product I construct the average price of the product in all stores in other geographic regions, defined by Nielsen DMAs. Prices in other regions capture common costs shocks due to changes in the manufacturer's or wholesaler's price, but are independent of shocks to demand in

[10] I do not model consumer's choice of stores because consumers generally repeatedly shop at the same store.

[11] Empirical market shares are observed quantities divided by market size. Market size is the population of a store's three-digit zip-code multiplied by the stores yearly market share of all observed sales in that three-digit zip-code. This assumption implies that every person considers purchasing pasta once per week from a store that reports data to Nielsen, not just the retailer considered.

the stores and regions that they instrument for. An alternative interpretation of these instruments comes from Dellavigna and Gentzkow (2019) who show that large retail chains often set prices uniformly across stores. They show that prices across stores are highly correlated and if the timing of retailer level sales is independent of store level demand shocks then these instruments work well.

In addition I also construct differentiation instruments based on Gandhi and Houde (2020). The intuition of these instruments is that they capture the differentiation between a given product and all other products in characteristic space. A product with a high degree of differentiation has few substitutes and as a result the retailer can charge higher markups, increasing prices. Let x_{jlst} be characteristic l for product j in store s at time t. The differentiation instruments are constructed as,

$$z_{jlst} = \sum_{k \in J_{st}\ j} \mathbf{1}(\mid d_{jklst} < SD_l \mid) \tag{2.5}$$

where $d_{jklst} = x_{klst} - x_{jlst}$ is the difference in characteristic l between products k and j, SD_l is the standard deviation of these differences across all markets, and $\mathbf{1}(\mid d_{jklst} < SD_l \mid)$ is an indicator that products j and k are close to each other in characteristic l. I construct differentiation instruments for the time since a product has been introduced to a given store as well as indicators for the product's pasta type. In addition, for the time since a product has been introduced to a given store I use a quadratic version of the differentiation instrument which captures a more continuous measure of the distance between goods.

$$z^{quad}_{jlst} = \sum_{k \in J_{st}\ j} d^2_{jklst} \tag{2.6}$$

As a result, the differentiation instruments capture how similar a given product's pasta type is to other products in the market and how novel a given product is relative to all other products in the market.

To aid identification of the random coefficients I incorporate micro moments following S. Berry, Levinsohn, and Pakes (2004) using the Nielsen Homescan panel data. The Homescan data provides information on individual households purchase decisions over time as well as demographic data. Incorporating moments based on this data helps to identify the random coefficients, especially κ^{price}, the effect of household's income on the household's taste for price. I match model predictions (left) to the empirical counterparts calculated using the Homescan data (right) for the following moments:

- The average price paid by households in different income brackets Inc. [12]

$$\frac{1}{S}\sum_{s}\frac{1}{T}\sum_{t}\frac{1}{|Inc|}\sum_{i\in Inc}\sum_{j\in J_{st}}\frac{s_{ijst}}{(1-s_{i0st})}p_{jst}=\frac{1}{S}\sum_{s}\frac{1}{T}\sum_{t}\frac{1}{|Inc|}\sum_{i\in Inc}\sum_{j\in J_{st}}\mathbf{1}_{ijst}p_{jst} \tag{2.7}$$

- The fraction of purchases in category c by households in different income brackets Inc.

$$\frac{1}{S}\sum_{s}\frac{1}{T}\sum_{t}\frac{1}{|Inc|}\sum_{i\in Inc}\sum_{j\in J_{st}}\frac{s_{ijst}}{(1-s_{i0st})}\mathbf{1}(j\in c)=\frac{1}{S}\sum_{s}\frac{1}{T}\sum_{t}\frac{1}{|Inc|}\sum_{i\in Inc}\sum_{j\in J_{st}}\mathbf{1}_{ijst}\mathbf{1}(j\in c) \tag{2.8}$$

- The difference between the average price of products available at store s and at time t and the price paid by households in different income brackets Inc.

$$\frac{1}{S}\sum_{s}\frac{1}{T}\sum_{t}\frac{1}{|Inc|}\sum_{i\in Inc}\sum_{j\in J_{st}}\frac{s_{ijst}}{(1-s_{i0st})}\overline{p}_{st}=\frac{1}{S}\sum_{s}\frac{1}{T}\sum_{t}\frac{1}{|Inc|}\sum_{i\in Inc}\sum_{j\in J_{st}}\mathbf{1}_{ijst}\overline{p}_{st} \tag{2.9}$$

- The difference between the minimum price of products available at store s and at time t and the price paid by households in different income brackets Inc.

$$\frac{1}{S}\sum_{s}\frac{1}{T}\sum_{t}\frac{1}{|Inc|}\sum_{i\in Inc}\sum_{j\in J_{st}}\frac{s_{ijst}}{(1-s_{i0st})}p_{st}^{(\min)}=\frac{1}{S}\sum_{s}\frac{1}{T}\sum_{t}\frac{1}{|Inc|}\sum_{i\in Inc}\sum_{j\in J_{st}}\mathbf{1}_{ijst}p_{st}^{(\min)} \tag{2.10}$$

- The difference between the maximum price of products available at store s and at time t and the price paid by households in different income brackets Inc.

$$\frac{1}{S}\sum_{s}\frac{1}{T}\sum_{t}\frac{1}{|Inc|}\sum_{i\in Inc}\sum_{j\in J_{st}}\frac{s_{ijst}}{(1-s_{i0st})}p_{st}^{(\max)}=\frac{1}{S}\sum_{s}\frac{1}{T}\sum_{t}\frac{1}{|Inc|}\sum_{i\in Inc}\sum_{j\in J_{st}}\mathbf{1}_{ijst}p_{st}^{(\max)} \tag{2.11}$$

[12] To calculate all moments I divide households into three income brackets: <$40k, $40k–$70k, and >$70k.

2.4.2.2 Demand Estimates

In order to allow consumers' preferences to vary by geographic markets I estimate demand separately in each geographic market (Nielsen DMA) for the 15 largest markets measured by the number of stores. To aid estimation I use the pyBLP package and associated numerical and optimization improvements outlined in Conlon and Gortmaker (2020).

Table 2.4: Demand Estimates for San Francisco-Oakland-San Jose, CA

DMA 807 (139 stores)

Coefficient	Estimate	Standard Error
Linear Terms		
Price	2.84	0.56
Months since product-store intro	1.47e-03	4.72e-03
Months since product-store intro, squared	1.18e-05	2.00e-05
Heterogeneous Terms		
σ	1.77	0.53
κ^{price}	-0.33	0.06
Nesting Terms		
Wheat	0.74	0.06
Wheat plus	4.34e-03	2.70
Gluten free	0.26	1.52

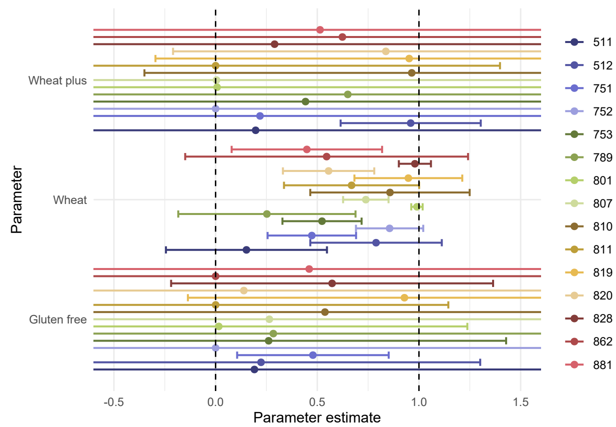
Robust standard errors are clustered at the product and three-digit zip-code levels.

As an example, parameters of the demand model estimated for DMA 807 (San Francisco-Oakland-San Jose, CA), the largest DMA (measured by number of stores), are presented in Table 2.4. Parameters for all DMAs are presented in Figure 2.3. In DMA 807 the estimated linear parameter on price is positive, but the impact of the negative interaction between price and log income, κ^{price}, makes the overall parameter on price negative for all households with incomes greater than $5,198. In DMA 807 the mean and median incomes are $126,083 and $97,000 respectively. As a result, households with incomes at the mean and median have overall price parameters of -1.06 and -0.97 respectively.

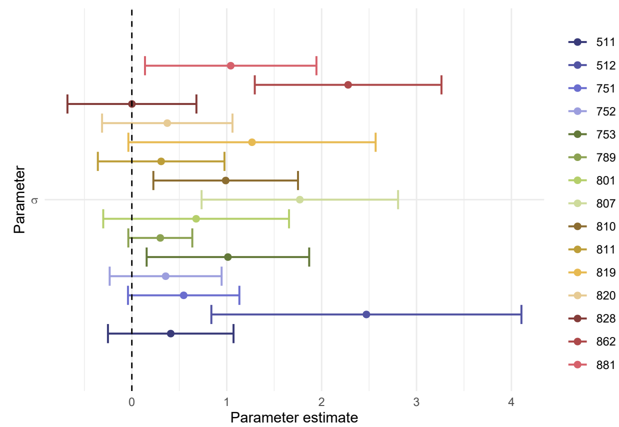
In Figure 2.3 I present demand parameter estimates for all DMAs. Across all DMAs the parameter estimates are largely similar with the exception of the two parameters on price. I estimate a large nesting coefficient on wheat pastas and insignificant coefficients on wheat plus and gluten-free pastas. The nesting coefficient on wheat pasta helps explain why in Table 2.1 wheat pastas make up a disproportionately large share of units sold and revenue. As shown in Figure 2.3(c), in some DMAs the estimated utility from price is increasing in income, while in other DMAs the estimated utility from price is decreasing in income. However, in all DMAs the overall coefficient on price (the combination of the linear term in 2.3(d) and the term interacted with log income in 2.3(c)) for a household at the mean and median income levels is negative.

Figure 2.3(e) shows that the impact of the time a product has been available at a store on consumers' utility is small and insignificant. This indicates that consumers do not have a preference for newer or older products. This zero effect is important because we might be concerned that in addition to the retailer learning about demand consumers may be also be learning or face a cost to switch to a new product. If consumers are learning we should expect to estimate a disutility from new products as consumers purchase products they are certain about over new products that they possibly don't like. Similarly, if consumers face a cost to switch products we should have estimated a disutility for new products as consumers are more likely to purchase existing products than incur a switching cost in order to purchase a new product. Overall, the inability to estimate a significant impact of a product's newness in any DMA should be reassuring that consumers are not hesitant to purchase new products because of either learning or switching costs.

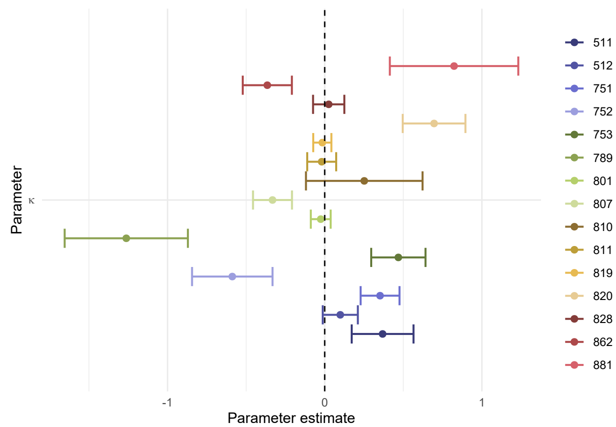
In Figure 2.4 I show own-price elasticities predicted by the estimated demand model across all DMAs. Overall the estimated demand models appear to produce sensible own-price elasticities centered around approximately -1.5. The model does produce some outlier elasticities including both positive and large negative values. However, because most estimated own-price elasticities appear to be reasonable I attribute the outliers to estimation and simulation error.



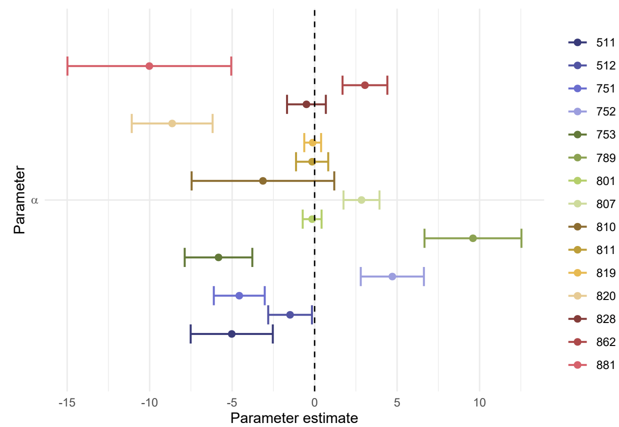
(a) Estimates of the nesting parameters



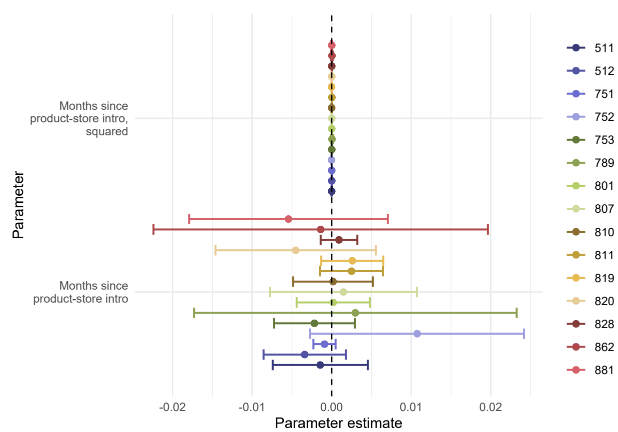
(b) Estimates of the random coefficient on all inside goods



(c) Estimates of the interaction between log income and price



(d) Estimates of the linear price parameters



(e) Estiamtes of the linear parameters on the time a product has been available

Figure 2.3: Demand Estimates

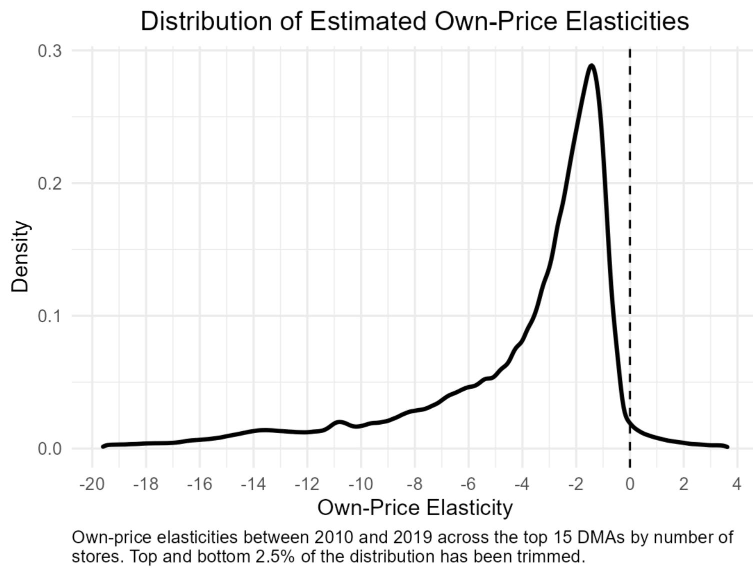


Figure 2.4: Own-Price Elasticity Distribution

2.4.3 Costs

Given estimates of demand I want to recover the retailer's costs. Typically this is done by assuming the retailer is setting prices in a full-information Nash-Bertrand equilibrium. Doing so allows us to take first order conditions of the retailer's profit function in Equation (2.12) where M_{st} is the market size and s_{jst} is the market share of product j consistent with the above demand model summarized in Equation (2.2).

$$\pi_{st} = M_{st} \sum_{j \in J_{st}} s_{jst}(p_{jst} - c_{jst}) \tag{2.12}$$

First order conditions give Equation (2.13), in matrix form, that relates prices, costs, and markups.

$$p - c = \Delta^{-1} s \tag{2.13}$$

Δ is a $J_{st} \times J_{st}$ matrix of demand derivatives $\left(\frac{\partial s}{\partial p}\right)$. Crucially the assumption that the retailer has full information about demand is needed so that the retailers knows Δ. Further, the retailer must know demand for all products in order to know any element of Δ because each element depends on the utility for all products. In order to recover costs I must make some assumptions about the information available to the retailer.

Assumption 1: After a product has been available in a given store for four years the retailer knows demand for that product in that store.

Assumption 1 creates two periods in the lifecycle of a product in a given store. I denote the first four years of a product's lifecycle in a given store as the learning period. During the learning period I make no assumptions about the information available to the retailer in order to allow for the potential that the retailer is learning about demand. I denote the period following the first four years of a product's lifecycle in a given store as the informed period. During the informed period I make Assumption 1, that the retailer has full knowledge of demand for the product in that store. As a result, if all products at a given store are in the informed period then the retailer knows Δ and I can recover costs. However, given that all products need to be in the informed period simultaneously this is a strong requirement and is satisfied 0% of the time. The retailer introduces products throughout the sample period and as a result there are no months when all products are in the informed period simultaneously.

Assumption 2: If a product j is in the informed period and the sum of long-run diversion ratios to all products in the learning period is below a threshold λ then the retailer sets prices for product j independent from all products in the learning period.

Assumption 2 uses the intuition that products with low substitutability are priced independently. Implicitly I already make this assumption by only modeling the dried pasta category. In theory the introduction of a new product in a different product category, such as toothpaste, could impact the demand for pasta. However, I don't model the impact of toothpaste on pasta products because of my perception of the differences between the product categories. Assumption 2 extends this implicit assumption to products within the pasta category. I measure the closeness and substitutability of two products using long-run diversion ratios. The long-run diversion ratio from product j to product k is defined as[13]

$$\bar{\mathcal{D}}_{jkst} = \frac{s_{k(-j)st} - s_{kst}}{s_{jst}} \tag{2.14}$$

where $s_{k(-j)st}$ is the market share of product k if product j were removed from the market. The long-run diversion ratio measures the fraction of consumer who purchase product j that would purchase

[13] See Conlon and Mortimer (2021) for a discussion of diversion ratios.

product k if product j were not available. A diversion ratio of 1 represents that products j and k are close substitutes because all consumers who currently purchase product j would switch to purchasing product k if product j were no longer available. A diversion ratio of 0 represents that products j and k are not substitutes because no consumers who currently purchase product j would switch to purchasing product k if product j were no longer available.

The requirements imposed by Assumptions 1 and 2 for a product to be in the informed period can be summarized by Equation (2.15). Whenever Equation (2.15) is true both Assumptions 1 and 2 are satisfied and I can recover costs using Equation (2.13) for product j at store s and time t.

$$\mathbf{1}\left(j \in \text{learning period}\right) + \sum_{k \in J_{st}} \bar{\mathcal{D}}_{jkst} \mathbf{1}\left(k \in \text{learning period}\right) < \lambda \tag{2.15}$$

In order for the left hand side of Equation (2.15) to fall below $\lambda < 1$ for product j two things must be true. The first is that product j cannot itself be in the learning period. This would violate Assumption 1 and would make the first term $\mathbf{1}\left(j \in \text{learning period}\right) = 1$. The second is that only products that are not substitutes for j can be in the learning period. If a substitute for product j were in the learning period then $\bar{\mathcal{D}}_{jkst}$ would be close to 1 and the second term on the left of (2.15) would cause product j to exceed the threshold λ. As a result, in order for Assumptions 1 and 2 to hold for product j, neither product j nor any close substitutes for product j can be in the learning period.

Using a threshold value of λ = 0.25 I recover costs for 60% of product-store-months. Given these costs I need to estimate costs for the remaining 40% of product-store-months when the retailer is potentially uncertain about demand. To do so I exploit two institutional details of the grocery industry. The first is that pasta products are relatively simple and are defined by the combination of brand, type, and shape. For a given brand and type, different shapes are made by replacing the die used to extrude pasta. As a result, the majority of the production process is identical and costs to produces two different shapes are highly correlated. The same logic extends to different types of pastas made by the same manufacturer, especially for wheat and wheat plus products whose ingredients are similar. The process of producing gluten-free products is largely similar to wheat based products. However, gluten-free products have additional costs because they require dedicated facilities or cleaning of the manufacturing facility to ensure no cross contamination between the production of wheat and gluten-free products. As a result, there is a correlation between the costs to produce different pasta products from the same manufacturer, but costs for gluten-free pastas are scaled up relative to wheat and wheat plus pastas. Correlations between the costs to produce different pastas from the same manufacturer are important because manufacturers typically produce a large number of products. Thus if we are able to recover costs for one product using Assumptions 1 and 2 those costs can be informative about the costs of other products produced by the same manufacturer.

The second institutional detail that is important is that wholesalers serve a large number of stores across a retailer's network. If two stores purchase products from the same wholesaler their costs to purchase products are highly correlated if not identical. We can exploit the correlation in input costs across stores served by the same retailer to estimate costs that we are unable to recover. This is especially true when the retailer uses a staggered rollout strategy. Stores that introduce products earlier learn about demand sooner which allows us to recover costs using Assumptions 1 and 2. Using costs recovered from earlier stores we can estimate costs at later stores because of the correlation generated when stores share a wholesaler.

Based on the correlations of costs from products sharing the same manufacturer and stores sharing the same wholesaler I use these features as inputs to a boosted regression tree. A boosted regression tree groups product-store-month cost observations together by their observable characteristics when those observations have similar costs. For example, a boosted regression tree might group observations that share the same brand if costs within that brand are similar to each other and different from costs of other brands. Among observations that are grouped together the boosted regression tree predicts costs using the average among the group. A boosted regression tree is similar to a fixed effects regression because both methods calculate the average cost conditional on observables. The main difference is that in a fixed effects regression I need to specify exactly how I condition on the observables. In contrast a boosted regression tree takes a set of possible observables as an input and returns conditioning sets that minimize the squared difference between costs and the conditional average.

The advantage of using a boosted regression tree compared to a fixed effects regression in this setting is that I do not observe how the correlations between costs are generated. For example, I know that stores that are served by the same wholesaler will have correlated costs, but I do not observe which set of stores are served by the same wholesaler. If I did observe store-wholesaler relationships I would include them in a fixed effects regression. However, these relationships are unobserved so it is unclear if the best set of fixed effects to capture the retailer-wholesaler relationship are based on smaller geographic regions such as zip-code or larger geographic regions such as Nielsen DMA. Instead, by using a boosted regression tree I am able to allow the model to group stores with similar costs without needing to specify a geographic level for the grouping.[14] The flexibility of a boosted regression tree allows me to capture correlations that I know are important, such as those generated by wholesalers, without observing how the correlations are generated.

One issue with using a boosted regression tree is the potential for overfitting. A boosted regression tree that individually grouped each product-store-month observations would appear to perfectly fit the data, but would preform poorly out-of-sample. In order to prevent overfitting I train the model on a fraction of the recovered costs and use the remaining recovered costs to test the model's performance. To do so I randomly split the costs recovered using Assumptions 1 and 2 into a training and testing set composed of 3/4 and 1/4 of the data respectively. I fit the boosted regression tree on the training set and check the out-of-sample validity of the model by predicting costs in the testing set and comparing the model predictions to recovered values. [15] After training the boosted regression tree on the training set I predict costs and find that on the training set the model fits the data with a root mean squared error (RMSE) of 0.518. This compares to a RMSE of 0.614 on the 1/4 of the data reserved for testing.[16] The similarities between the prediction error on the testing and training sets indicates that the model is not overfit.

One way to examine the results of the boosted regression tree to get a sense of how it predicts costs is to look at what variables were important to the model's predictions. Figure 2.5 shows the relative

[14] Note, the geographic level that predicts cost similarities does not need to be consistent across stores either. A boosted regression tree might group one set of stores together that share the same DMA and group another set of stores together that share the same zip-code.

[15] For more details on the boosted regression tree I used including how I choose the hyperparameters see Appendix 2.7.2.

[16] Note, a small fraction of costs recovered from the demand model appear to be outliers, similar to how some of the predicted own-price elasticity values are outliers. To prevent the outlier costs from skewing the boosted regression tree's predictions I exclude costs that are greater than $10 in absolute value (the 98.5 percentile cost I recover is $9.98). The 0.518 training RMSE and 0.614 testing RMSE values I calculate exclude the outlier costs. If I include the outliers in the testing set the resulting RMSE is 213.52.

importance of each potential variable. The importance of a variable indicates how much on average the cost prediction changes if the variable's values change. The most important variable to predicting a product's cost is the combination of store and month. The combination of store and month captures correlation across time and space that are not well captured by the existing geographic or time variables alone. For example, if a wholesaler served only half of the stores in a DMA during 2010–2015 and then later served all stores in the DMA during 2016–2019 this would be well captured by the store-month variable. All of the observations where the wholesaler provided the product could be grouped together in one node using the store-month variable while more complex interactions of the DMA and month variables would be needed to achieve the same effect. The next three variables, month, Nielsen DMA, and three-digit zip-code, indicate that while there may be some idiosyncratic correlations that can only be achieved by using the store-month variable, there are still important geographic and time factors that influence costs. Finally, it is interesting that all product characteristics have the lowest variable importance. One explanation may be that after accounting for geography and time any variation in costs between products are minimal.

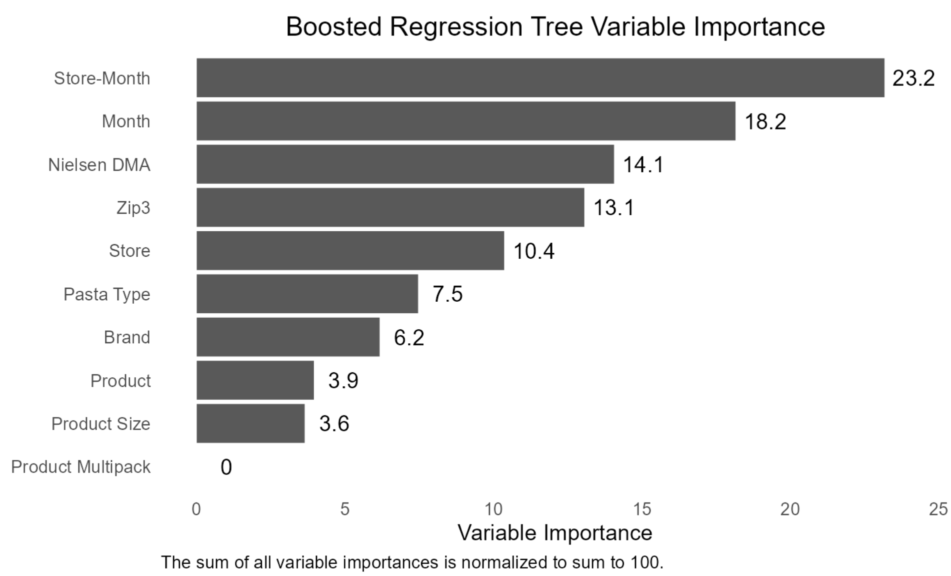


Figure 2.5: Boosted Regression Tree Relative Variable Importance

Given that the boosted regression tree performs well on the testing set and the variables that are important to the model's predictions are sensible I use the boosted regression tree to estimate costs that I was not able to recover. For the 40% of product-store-months that I cannot recover using Assumptions 1 and 2 I estimate costs using the predictions from the boosted regression tree. The process of predicting costs is straightforward and is analogous to predicting costs using a fixed effects model.

2.4.4 Counterfactual Prices

Given cost estimates for every product-store-month I want to estimate what prices would have been if the retailer was fully informed about demand. To do so I assume that the counterfactual retailer knows demand and the demand derivatives Δ. Then using the costs recovered from Assumptions 1 and 2 and the costs predicted from the boosted regression tree I use the FOCs of the retailer's profit in Equation (2.13) to estimate prices. Conceptually this is the reverse of the process used to recover costs from prices under Assumptions 1 and 2. However, the process is complicated because the demand derivatives Δ depend on prices creating an implicit function. Equation (2.13) is not a contraction so to solve for prices I I follow Morrow and Skerlos (2011) who reformulate (2.13) into a contraction mapping.

During periods when all products are informed under Assumption 1 counterfactual prices are equal to observed prices. This is by construction because Assumption 1 guarantees that the retailer is fully informed about demand for all products which is the same requirement that we impose when we estimate counterfactual, full-information prices. In periods where only some products are informed under Assumption 2 counterfactual prices for products in the informed period are close to, but may not be equal to, observed prices. Assumption 2 allows some products to be in the learning period while other products are in the informed period as long as the two sets of products are not close substitutes. As a result, pricing for the two sets of products is largely independent. If pricing were completely independent then, similar to when Assumption 1 holds for all products, counterfactual prices for informed products would be equal to actual prices. However, because there is some small dependence between prices for products in the informed period and products in the learning period counterfactual prices are different from actual prices.

In general there are two reasons why observed and counterfactual prices may differ. The first is that counterfactual prices are constructed under the assumption that the retailer has full information about demand, while in reality the retailer may be uninformed. If the retailer is setting prices based on a misconception of demand this will create a difference from full-information, counterfactual prices. Periods in which the retailer is potential misinformed about demand are when Assumptions 1 and 2 do not hold. This is also when we have to estimate costs using the boosted regression tree. The need to estimate costs introduces a second difference between observed and counterfactual prices, estimation error. To understand how the retailer is learning about demand we want to separate the error coming from misperceptions of demand from the estimation error.

The estimation error we are concerned about comes from costs predicted by the boosted regression tree, which may have some error relative to actual costs. As a result, errors in estimating costs will create an error when costs are used to estimate prices. Estimation error from predicting costs during the learning period is unavoidable, but we can estimate the size of this error. Differences between counterfactual and actual prices during the learning period come from either a lack of information by the retailer or error in predicting costs. This is in contrast to the informed period where we assume the retailer is informed, so errors do not come from a lack of information. If we use predicted costs from the boosted regression tree in the informed period to predict prices we can estimate the error in prices that comes from predicting costs alone. To estimate the size of the estimation error I generate two versions of prices when costs can be recovered from Assumptions 1 and 2. The first set of prices are generated using recovered costs. The second set of prices are generated using costs predicted by the boosted regression tree. The difference between these sets of prices is informative about the error introduced by predicting costs. Further, 1/4 of the costs in the informed period were reserved as a testing set and were not used to fit the boosted regression tree. This testing set provides an ideal sample to estimate the error generated by predicting costs because the retailer is assumed to be informed about demand and these costs were not used to train the boosted regression tree, providing a better estimate of out-of-sample performance. Among observations in the training set I find that the RMSE between observed and predicted prices is 0.96 while the mean absolute error (MAE) is 0.4. Using costs predicted by the boosted regression tree in the testing set I find that the RMSE between observed and predicted prices is 0.94 while the MAE is 0.39. [17] The difference between these measures provide an

[17] Similar to how there are outliers in the estimated own-price elasticities and costs there are outliers in the estimated prices. The RMSE and MAE presented are excluding observations where the counterfactual price is below the 10th percentile or above the 96th percentile. If these counterfactual prices are included the RMSE on the training set is 8365 and the MAE is 31. If these counterfactual prices are included the RMSE on the testing set is 6443 and the MAE is 26.

estimate of the error from predicting costs outside the training sample and are a useful comparison when examining differences between counterfactual prices and actual prices in the learning period.

2.5 Evidence of Learning

2.5.1 Initial Uncertainty

Similar to how I provided evidence that the retailer is learning about demand in Chapter 1, I now want to verify that after controlling for costs and the competitive environment that the evidence still holds. In Chapter 1 I showed that novel products were initially introduced to fewer stores as evidence that the retailer was uncertain about demand. Now given the results of the structural model developed in Section 2.4 I am able to do a similar analysis that shows that the initial difference between actual prices and counterfactual, full-information prices is larger for novel products.

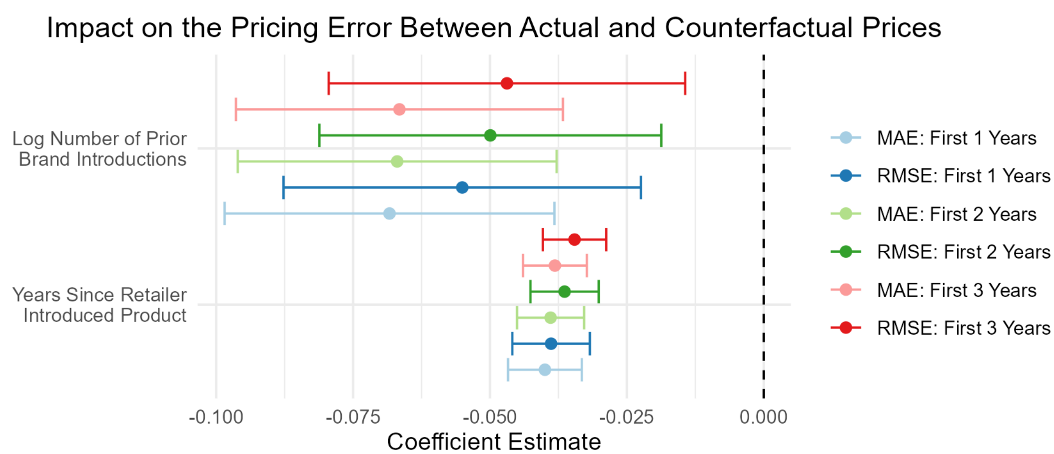
When the retailer is initially uncertain about demand we should expect to see a larger difference between observed and counterfactual prices. In Chapter 1 I showed that the retailer was more uncertain about demand for novel products. To replicate these results, I run the following regression to understand the relationship between the size of the pricing error and the novelty of a new product.

$$\begin{aligned}\text{Pricing error}_{js} =& \beta_0 + \beta_1 \log(\text{Prior brand intros}_{js}) + \beta_2 \text{Years since retailer intro}_{js} \\ &+ FE_s + FE_{js}^{\text{intro year}} + FE_j^{\text{pasta type}} + \epsilon_{js}\end{aligned} \tag{2.16}$$

The outcome of interest for each observation is the mean pricing error after product j is introduced to store s. Years since retailer intro$_{js}$ is the number of years since the retailer introduced product j to any store. Prior brand intros$_{js}$ is the number of products with the same brand as product j that are introduced by the retailer (to any store) before product j is introduced to store s. I include fixed effects for the store s, the year product j was introduced to store s, and the pasta type of product j.

The novelty of the new product is captured by two terms, β_1 and β_2. The first term, β_1, captures the effect of the number of prior products from the same brand that the retailer has introduced. Existing products from the same brand give the retailer information about demand for the new product and should result in a smaller pricing error. The second term, β_2, captures the effect of the years since the retailer introduced the new product to any store. Although a product may be new to a given store s the retailer may have experience selling the product in other markets. More experience with a new product in other markets provides the retailer with information about demand and should result in a smaller pricing error when the product is introduced to store s.

The results of the regression in Equation (2.16) are shown in Figure 2.6. I use six different measures of the pricing error to illustrate the robustness of the results. I measure the pricing error using either the RMSE or the MAE. I also calculate the mean pricing error over either the first one, two, or three years after a product is introduced to a store. Regardless of which measure of the pricing error I use the results are consistent. [18] A lower pricing error is associated with products that come from brands with more existing offerings and products that have been sold by the retailer for a longer time period in other stores. Consistent with Chapter 1, these results indicate that the retailer is more uncertain about demand for novel products and less able to set prices close to the full-information level.



Error bars represent 95% confidence intervals clustered at the store and introduction year level. The pricing error is calcluated as either the mean absolute error (MAE) or the root mean squared error (RMSE) during the first one, two, or three years after a product is introduced to a store. Observations where the counterfactual price is below the 10th percentile or above the 96th percentile are excluded.

Figure 2.6: Impact of Product Novelty on Initial Mispricing

[18] I further show the robustness of the results to different sets of fixed effects in Appendix 2.7.3.

2.5.2 Learning Over Time

In Chapter 1 I showed that prices vary the most month-to-month in the first few years after a retailer introduces a product. Now, given the results of the structural model developed in Section 2.4 I show that—after accounting for changes in costs and the competitive environment—reductions in the variability of prices is driven by learning. If the difference between counterfactual, full-information prices and observed prices is due to misperceptions of demand, then reductions in the pricing error are evidence that the retailer is learning. To understand how the difference between counterfactual, full-information prices and observed prices change over time I run the regression in Equation (2.17). The outcome of interest is the absolute difference between the counterfactual, full-information price and the observed price of product j in store s at time t.

$$\begin{aligned} \mid p_{jst}^{\text{full-info}} - p_{jst} \mid =& \beta_1 \text{Obs. Set}_{jst} + \beta_2 \text{Years since store intro}_{jst} \times \text{Obs. Set}_{jst} \\ &+ \beta_3 \text{Years since retailer intro}_{jst} \times \text{Obs. Set}_{jst} + FE_s + FE_j^{\text{pasta type}} + FE_t + \epsilon_{jst} \end{aligned} \tag{2.17}$$

Obs. Set_{jst} are indicators for if the observation is in one of the three following sets: 1) Product j at store s is in the informed period at time t and the associated recovered cost observation was used to train the cost model. 2) Product j at store s is in the informed period at time t and the associated recovered cost observation was used to test the cost model. 3) Product j at store s is in the learning period at time t and costs were predicted by the cost model. I make a distinction between costs used to train the cost model and costs used to test the model to validate the size of any error resulting from predicting costs on out-of-sample observations not used to train the cost model. The difference between the pricing error in sets 1 and 2 serve as a proxy for errors introduced by the boosted regression tree used to predict prices. The results of the regression in Equation (2.17) are shown in Table 2.5.

In Table 2.5 I present the results of three specifications with an increasing number of controls. In all three specifications the excluded Obs. Set_{jst} are products in the informed period that are used to train the cost model. Across all three specifications I find a small, approximately \$0.008, difference between the pricing error for observations in the informed period that are in the training set versus the testing set for the cost model. This indicates that the out-of-sample accuracy of the cost model does not have an impact on the pricing error. I do find a significant, positive difference between the pricing error for products in the learning period and products in the informed period. The difference between observed and counterfactual prices in the learning period is between $0.264 and $0.32 larger than the pricing error in the informed period. Consistent with Figure 2.6, this indicates that the retailer has difficulty pricing products that are new.

In column 1 I use controls for the length of time a product has been available at a given store. From the results in column 1, the length of time a product has been available in a store decreases the difference between observed and counterfactual prices when the product is in the learning period. This indicates that wile a product is new the retailer is learning about demand and updating prices to reflect the true demand for the product. However, once a product has been available for four years (Assumption 1) and no close substitutes are new (Assumption 2) I observe no significant change in the pricing error indicating that the rate of learning has declined to an extent that the impact on pricing is unobservable.

In columns 2 and 3 I add additional controls for the length of time a product has been available from the retailer (at any store). The purpose of this is to test if learning is local at the store level or if learning occurs at a broader geographic level as information is shared across stores. The results in columns 2 and 3 appear to support the theory that learning occurs at a broad geographic level as information is shared across stores. In column 2 the reduction in the pricing error from information at the retailer level (-$0.012) has approximately the same magnitude as the impact of store level information during the learning period (-$0.014). However in column 3 when the impact of information is differentiated by time period, the reduction in the pricing error from information at the retailer level is approximately 12 times as large as the reduction from store level information (-$0.025 vs. -$0.002). This indicates that reductions in the pricing error, and learning, are the result of retailer level information that is shared across stores.

In column 2 the effect of additional store level information is positive and is significant at the 0.1% level for products in the informed period. These coefficients alone would appear to indicate that the pricing error grows over time. However, the effect of retailer level information reduces the pricing error by approximately 2 times as much. Overall changes to the pricing error from retailer and store level information occur at the same rate as time progresses, so the decline in the pricing error from retailer level information dominates. [19] This means that the retailer is learning about demand over time and setting prices closer to the counterfactual, full-information level across all stores. The relative size and sign of the coefficients indicates that information at the retailer level from all stores is more important than store specific information in reducing the difference between observed and counterfactual prices.

[19] Summing the coefficients in column 2 on the years since store introduction and on years since retailer introduction gives overall estimates of -$0.006, -$0.006, and -$0.026 for the informed period (training set), the informed period (testing set), and the learning period respectively. The combined coefficients are all significant at the 0.1% level.

Table 2.5: Impacts on the Pricing Error Over Time

	(1)	(2)	(3)
Informed Period (Testing Set)	−0.008**	−0.008**	−0.007+
	[−0.013, −0.002]	[−0.013, −0.002]	[−0.014, 0.0005]
Learning Period	0.312***	0.264***	0.320***
	[0.290, 0.335]	[0.240, 0.288]	[0.289, 0.350]
Years Since Store Introduction x Informed Period (Training Set)	0.0004	0.006***	0.003***
	[−0.001, 0.002]	[0.005, 0.007]	[0.002, 0.004]
Years Since Store Introduction x Informed Period (Testing Set)	0.0007	0.006***	0.003***
	[−0.0007, 0.002]	[0.005, 0.008]	[0.002, 0.004]
Years Since Store Introduction x Learning Period	−0.024***	−0.014***	−0.002
	[−0.029, −0.020]	[−0.018, −0.009]	[−0.007, 0.002]
Years Since Retailer Introduction		−0.012***	
		[−0.014, −0.011]	
Years Since Retailer Introduction x Informed Period (Training Set)			−0.005***
			[−0.007, −0.003]
Years Since Retailer Introduction x Informed Period (Testing Set)			−0.005***
			[−0.007, −0.003]
Years Since Retailer Introduction x Learning Period			−0.025***
			[−0.028, −0.022]
FE: Stores	X	X	X
FE: Month	X	X	X
FE: Pasta Type	X	X	X
N	7 129 997	7 129 997	7 129 997

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note:
95% confidence intervals clustered at the store and month level are shown. Observations are categorized as either belonging to 1) the set of observations in the informed period used to train the cost model, 2) the set of observations in the informed peroid used to test the cost model, or 3) the set of observations in the learning period. Indicators for these three sets are included in the regressions and the set of observations in the informed period used to train the cost model is the excluded set. Observations where the counterfactual price is below the 10th percentile or above the 96th percentile are excluded.

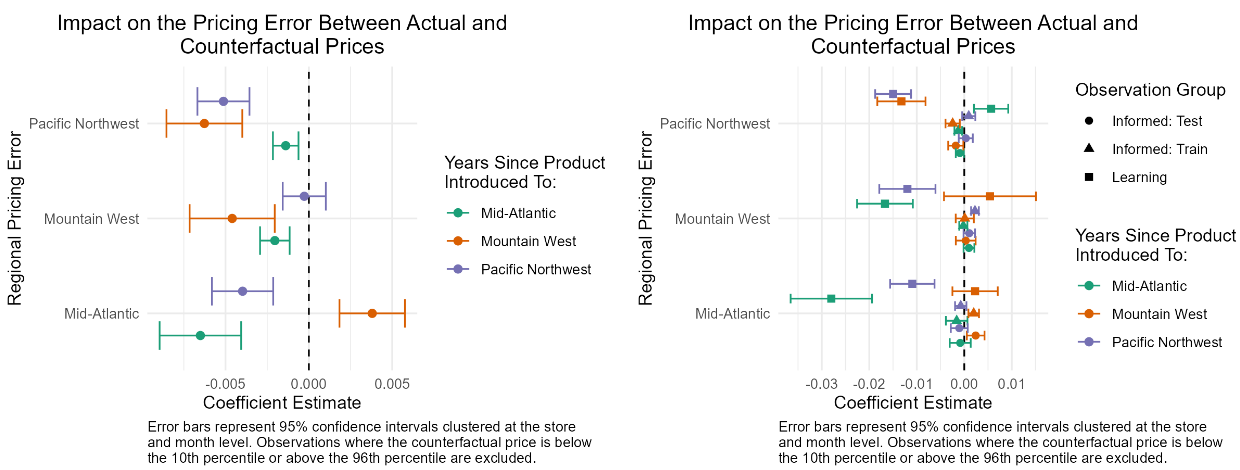
The results from column 2 largely hold for column 3 as well. In column 3 I allow the impact of the time since a product was introduced by the retailer to vary based on if product j was in the learning or informed periods at store s and time t. Regardless of if a product is in the the informed period (training set), the informed period (testing set), or the learning period, the impact of information at the store level is either positive or insignificant while the impact of information at the retailer level the time since the retailer's overall introduction is negative and larger in magnitude. This indicates that the retailer is improving their ability to set prices based on retailer rather than store level information. The combination of the coefficients on the time since a product's introduction by the retailer and to store s and are -\$0.002, -\$0.002, and -\$0.027 for products in the informed period (training set), the informed period (testing set), and the learning period respectively. The combined coefficients during the informed period are both insignificant while the combined coefficient during the learning period is significant at the 0.1% level. This result for the learning period confirms the results in columns 1 and 2 that the retailer is learning about demand and improving pricing during the learning period. The combined coefficients in column 3 for the informed period show no significant change in the pricing error indicating that learning has declined to an extent to be unobservable. This matches the results in column 1, but contradicts the results in column 2. Overall, the results in Table 2.5 indicate that the retailer is learning about demand and improving pricing during the learning period largely as a result of sharing information across all of the retailer's stores.

2.5.3 Learning Across Regions

Although the results in Table 2.5 indicate that learning occurs at the retailer level rather than the store level, not all information may be useful when setting prices in a given store. Retailer 32 operates across the nation so information about demand in one region may not be applicable across all regions. For example, retailer 32 operates in both the Mid-Atlantic and the Mountain West. Demand might be very different across these two locations, so information gained in the Mountain West might not be applicable when setting prices in the Mid-Atlantic. To test the applicability of information across regions I run the regression in Equation (2.18).

$$
\begin{aligned}
\mid p_{jst}^{\text{full-info}} - p_{jst} \mid =& \beta \sum_{r \in R} \text{Years since region intro}_{jr(s)} \times \text{Region}_{r(s)} \times \text{Obs. Set}_{jst} \\
&+ FE_{jst}^{\text{Obs. Set}} + FE_{s} + FE_{j}^{\text{pasta type}} + FE_{t} + \epsilon_{jst}
\end{aligned}
\tag{2.18}
$$

Years since region intro$_{jr(s)}$ is the number of years since product j was introduced into region r. Region$_{r(s)}$ are indicators for the three regions (the Mid-Atlantic, the Mountain West, and the Pacific Northwest). The results of the regression is shown in Figure 2.7.



(a) Basic model parameters

(b) Full model parameters

Figure 2.7: Regional Learning Estimates

Figure 2.7(a) shows the regression results when I do not interact Years since region intro$_{jr(s)}$ with the observation group. From this figure we can see that mispricing declines in the Pacific Northwest the most for products that have been around longer in the Pacific Northwest and the Mountain West. This indicates that the retailer is learning about demand in the Pacific Northwest primarily from sales in that region as well as in the Mountain West. In the Mountain West mispricing declines significantly for products that have been introduced earlier to that region as well as the Mid-Atlantic region. This indicates that the retailer is learning about demand in the Mountain West primarily from sales in that region as well as the Mid-Atlantic. Finally, in the Mid-Atlantic mispricing declines for products that have been introduced earlier in that region as well as the Pacific Northwest (and mispricing increases for products that are introduced earlier to the Mountain West). This indicates that the retailer is learning about demand in the Mid-Atlantic primarily from sales in that region as well as the Pacific

Northwest.

The patterns in Figure 2.7(a) are interesting. They show that the retailer is learning about demand in a given region from sales in that region and sales in one of the other two regions. However, the patterns of which regions impact learning in other regions is not symmetric. For example, the retailer learns about demand in the Pacific Northwest from the Mountain West, but learns about demand in the Mountain West from the Mid-Atlantic. Additionally, Figure 2.7(a) indicates that more information is not always useful. The results from the Mid-Atlantic show that more sales data from the Mountain West is associated with larger pricing errors in the Mid-Atlantic. One rationale for this is that consumer demand is sufficiently different between the Mid-Atlantic and Mountain West, so additional Mountain West data provides no benefit to price setting in the Mid-Atlantic.

In Figure 2.7(b) I allow changes to mispricing to vary based on if the product is in the learning or informed periods. Similar to the results in Table 2.5, the largest changes to pricing errors occur when the product is in the learning period (indicated by the square markers). In general the patterns between Figures 2.7(a) and 2.7(b) are similar with a few exceptions. Most notably, mispricing in the Mountain West no longer declines when products from the same region have been available for longer. Also in Figure 2.7(b), the impact of additional sales data from the Pacific Northwest decreases pricing errors in all three regions during the learning period. One reason that information from the Pacific Northwest might be particularly useful is because retailer 32 operates the most stores in the Pacific Northwest. From Table 2.2 in Section 2.3.2 we can see that both the two largest DMAs in terms of stores and five out of the top ten DMAs are all in the Pacific Northwest. Operating a large number of stores in the Pacific Northwest gives the retailer more information about demand and allows the retailer to better set prices in other regions, even if demand does vary between regions.

The ability for the retailer to use information across regions demonstrates the benefits to operating stores across the nation. The large number of stores retailer 32 operates provide a wealth of information about demand. This information is generally useful both within and across regions contributing to a reduction in the pricing error over time. However, this is not without exception as seen in Figure 2.7(a). Overall, the results in both Table 2.5 and Figure 2.7 indicate that more information results in better price setting and the ability to share information across stores and regions is beneficial in most instances. Retailer 32 benefits from its large scale because it can collect information from all stores and learn to set prices better across its full network of stores.

2.6 Conclusion

This chapter examines how retailers can share information across stores to learn about demand and set prices. Retailers face pricing challenges when they are unsure about demand, particularly when introducing a new product in a new store. I find that retailers set prices further from the optimum when little information is available from similar products and stores. However, over time, I find that retailers learn about demand and improve their pricing strategies. Overall, this indicates that the ability to learn about demand and refine pricing strategies is essential for retailers to successfully introduce and price new products across different markets.

To better understand how retailers learn about demand I study the effectiveness of information sharing across stores and geographic regions. My results suggest that sales data from the same geographic region leads to more optimal pricing. However, the usefulness of data from other regions is more mixed. In most instances data from other geographic regions improve pricing but, I do find cases where data from other regions have a negative impact on pricing. One explanation for this is that consumer preferences may be significantly different across regions. Therefore, while information sharing is useful in most instances, it is necessary for retailers to consider differences in consumer preferences before information can be utilized across all stores.

Large retailers who operate multiple stores or in multiple regions benefit from the ability to share information across stores. Information sharing allows large retailers to gather more sales data and learn about demand faster than small retailers. These returns to scale in information decrease the cost of introducing new products, as prices can be set closer to the optimal level. As a result, information sharing is a benefit that should be considered alongside conventional returns to scale in retailing such as warehousing and transportation costs.

Although this chapter identifies the benefits of information sharing across regions, it does not consider retailers' decisions about where to introduce new products or locate stores. As the value of sales information varies across geographies, the choice of where to launch new products is important because it impacts pricing and potential launch decisions elsewhere. More broadly, the decision of where to locate new stores may also depend on the value of the retailer's information based on their current store locations. While these questions are not addressed in this chapter they are of great interest to the literature and require further investigation in future research.

2.7 Appendix

2.7.1 Data Description

I primarily rely on Nielsen's retail scanner dataset that covers the period 2006–2019. This data provides information about sales, products, and stores. The retailer scanner dataset provides sales information weekly for a national sample of stores. Sales information consists of the weekly units sold and volume weighted average prices for all products (defined by their UPC) sold in a given store. Product information consists of basic product attributes including a product's brand, package size (if the product is a single or multipack), and weight. Product information also includes some rough information about a pasta's shape and type, but I found this information to be too general to categorize products. Store information consists of the store's parent retailer and the store's geographic location at the three-digit zip-code, FIPS county code, and DMA code levels.

In addition to Nielsen's retail scanner data I use Nielsen's Homescan dataset which is a panel dataset of households' grocery purchases during the period 2004–2019. This data provides all of a household's grocery purchases for the period a household is included in Nielsen's panel, typically a multi-year time frame. In addition to the information on households' purchases, the Homescan data provides demographic information for households. I use the Homescan data for two purposes. First, I use purchases during 2004–2009 to identity products that already existed prior to 2010. Second, I use household demographic information and purchase patterns to construct moments that are used to aid identification of my structural model.

2.7.1.1 Pasta Type Classification

To augment the product characteristics provided by Nielsen I collect nutrition and ingredient data from Nutritionix, OpenFoodFacts, and the USDA's FoodData Central database. All data sources provide information included in the product's nutrition label and the product's ingredient list. Together these datasets provide data on 764 products out of the 1,448 total products I observe. Most notably, Nielsen masks product identifiers (UPCs) for store brand products. This prevents me from collecting nutrition and ingredient information for store brand products from outside sources.

I use ingredient information provided by Nutritionix, OpenFoodFacts, and the USDA's FoodData Central database to classify products into four pasta types (wheat, wheat plus, gluten-free, and potato) based on the types of flour listed as an ingredient. In order to classify products based on their ingredients I collect all ingredient lists where available. I parse these lists and manually categorize individual ingredients based on the flour type or if the ingredient was a spice/flavoring. I then assign types to products based on the categorization of individual ingredients.

Some products have nutrition information but no ingredient information available from Nutritionix, OpenFoodFacts, or the USDA's FoodData Central database. For the products that are missing ingredient information, I interpolate their pasta type by matching on observables. I use a combination of brand, calories, total fat, sodium, and protein to match products with and without ingredient information. In order to classify the remaining products that appear in the Nielsen data for which no nutrition and ingredient information is available I match on the observable product characteristics available from Nielsen. I use a combination of brand and Nielsen product type to match products with and without ingredient and nutrient information.

2.7.1.2 Demographic Information

I collect demographic information from the American Community Survey (ACS). I construct demographic information about the consumers shopping at a given store using households sampled in the one-year ACS surveys from 2010–2019. I assign households from the ACS to stores by matching PUMAs (the smallest identifiable geographic unit in the ACS) to a store's three-digit zip-code and county FIPS code.[20] Matching ACS households to stores gives me a demographic profile of each store that I use when constructing moments in my structural model.

In order to infer the total size of a market I use population estimates from the five-year ACS surveys from 2010–2019. I assume the total market for a given store is the yearly population estimate in the store's three-digit zip-code multiplied by the store's yearly market share (by units sold) out of

[20] In order to match PUMAs to three-digit zip-codes and county FIPS codes I use a crosswalk between zip-codes and ZCTAs provided by UDS Mapper (https://udsmapper.org/zip-code-to-zcta-crosswalk/, accessed 2/28/22) and a crosswalk between PUMAs, ZCTAs, and country FIPS codes provided by Missouri Census Data Center (https://mcdc.missouri.edu/geography/ZIP-resources.html, accessed 2/28/22).

all stores in the scanner data in that three-digit zip-code. This implies that all individuals in a three-digit zip-code shop at one of the stores in the Nielsen retailer scanner data once per week. Further the number of individuals who shop at a store is consistent within a year and is proportionate to that store's total pasta sales.

2.7.2 Boosted Regression Tree

In order to predict costs that cannot be recovered from Assumptions 1 and 2 I use a boosted regression tree. In general, a regression tree attempts to predict costs by splitting data based on observable characteristics and then averaging costs among the groups, called leaves. A boosted regression tree improves upon the performance of a single regression tree by combining the outputs of multiple trees. A single regression tree is formed by iteratively splitting the data at each node into two leaves. Nodes are split into leaves until either a node has too few observations or the tree reaches a predetermined depth. Once a tree's structure is determined, model predictions are the average outcome variable in each leaf. A boosted regression tree builds on the concept of regression trees by making predictions using multiple smaller trees. The boosted regression tree algorithm iteratively creates multiple regression trees to predict the residuals from the previous tree. The final predictions of a boosted regression tree are the weighted sums of the predictions of each individual tree.

In order to create a boosted regression tree I need to specify the variables used to predict costs. Cost observations are at the product-store-month level. The characteristics I use are: month identifiers, store identifiers, store-month identifiers, the store's DMA, the store's three-digit zip-code, product identifiers, the product's brand, the product's pasta type, the product's weight, and the number of items if the item is a multi-pack. Given that most of the characteristics are categorical I use the Python CatBoost package which is specifically designed to work with categorical data. The advantage of the CatBoost package in my setting is that the package is design to quickly test different permutations of a categorical variable to determine the optimal grouping to split the data.

To ensure that the boosted regression tree is not overfitting the data and performs well out-of-sample, I split the recovered costs into a training and testing set. The training sample consists of a random 75% sample of product-store-month observations. I use the training sample to fit the boosted regression tree to the data and find that the RMSE of the boosted regression tree on the training sample is 0.518. To check for overfitting I then predict costs on the testing set. The testing set was not used in any way to train the boosted regression tree so the model's performance should be indicative of the out-of-sample model performance. I find that the RMSE of the boosted regression tree on the testing sample is 0.614. [21]

The CatBoost algorithm requires a few hyperparameters to determine the number and size of the regression trees. The hyperparameters I specify are the number of trees, the depth of each tree, the learning rate, and a leaf regularization parameter. The depth of a tree determines the maximum number of splits that occur in each regression tree. The learning rate determines how much weight to put on the predictions of each tree. The leaf regularization parameter controls the information gain required to split a node into leaves. In order to choose values for these hyperparameters I use three fold cross-validation. This process randomly splits the data into thirds and trains three models using the same set of hyperparameters. Each model uses 2/3 of the data for training and tests the models performance by predicting costs and calculating the RMSE on the remaining 1/3. The RMSE from all three models are averaged to determine the cross-validated error for that set of hyperparameters. I choose hyperparameters to minimize the cross-validated error on the training set. Finally, given a choice of hyperparameters I train the boosted regression tree using all of the training data.

2.7.3 Robsutness of the Initial Uncertainty Results

In this appendix I replicate the analysis in Section 2.5.1 controlling for different sets of fixed effects. In Section 2.5.1 I ran a regression to understand the relationship between the size of the pricing error and the novelty of a new product when controlling for fixed effects for stores, the year a product was introduced to a store, and the pasta's type. In Figure 2.8 I show the results of repeating this analysis using different fixed effects. Each panel of Figure 2.8 presents results when controlling for different sets of fixed effects. The bottom-middle panel are the results presented in Section 2.5.1. As seen in

[21] Note, a small fraction of costs recovered from the demand model appear to be outliers, similar to how some of the predicted own-price elasticity values are outliers. To prevent the outlier costs from skewing the boosted regression tree's predictions I exclude costs that are greater than \$10 in absolute value (the 98.5 percentile cost I recover is \$9.98). The 0.518 training RMSE and 0.614 testing RMSE values I calculate exclude the outlier costs. If I include the outliers in the testing set the resulting RMSE is 213.52.

Figure 2.8, the sign and magnitude of the effects are very consistent regardless of which fixed effects are used.

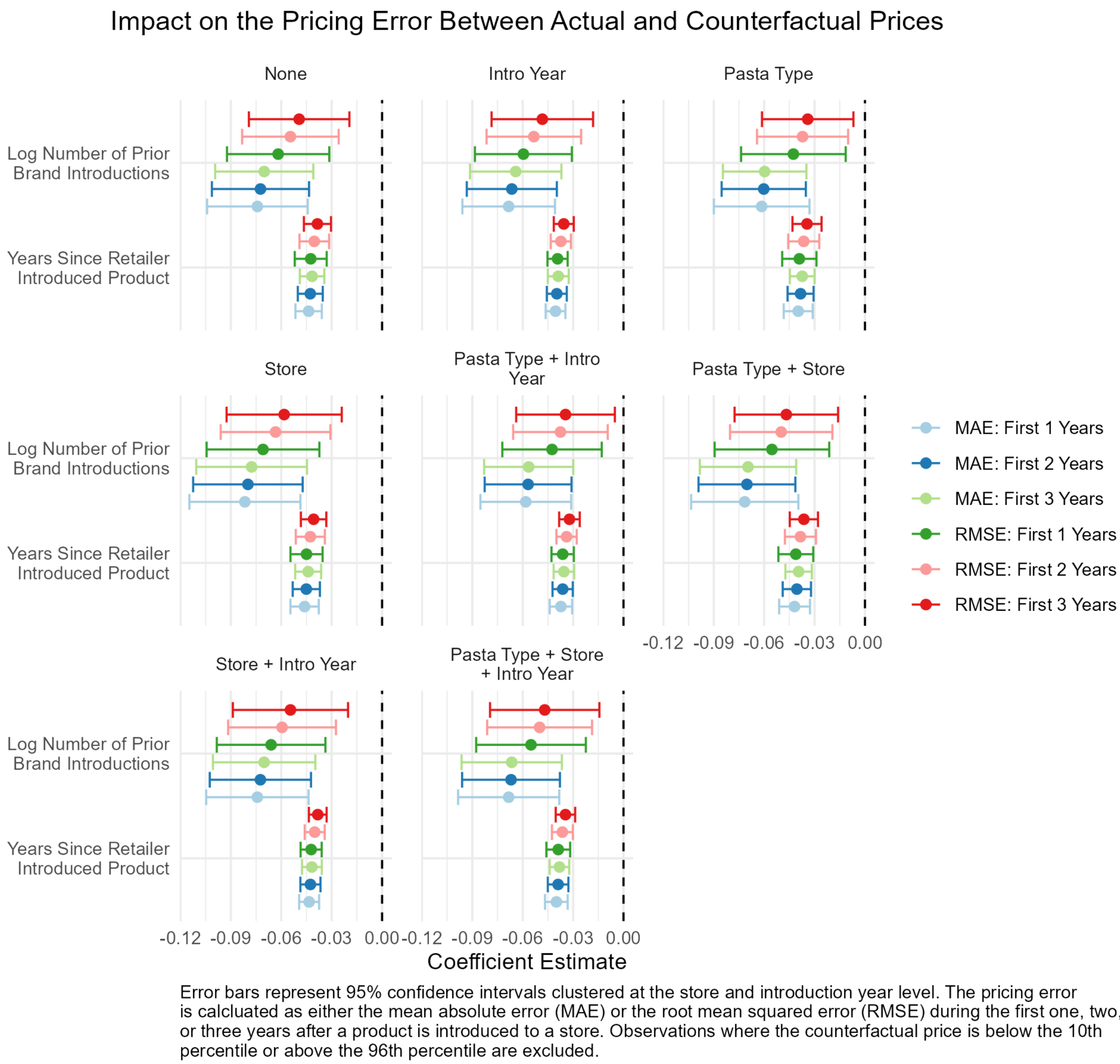


Figure 2.8: Impact of Product Novelty on Initial Mispricing Using Varying Fixed Effects

2.7.4 Robustness of the Learning Over Time Results

In this appendix I demonstrate the robustness of the results in Section 2.5.2, specifically the regression results presented in Table 2.5. In Table 2.5 I showed the impact of the time since a product was introduced to a store or by the retailer on the difference between observed and counterfactual prices. In this appendix I conduct a similar analysis however I use the log of the time since a product was introduced to a store or by the retailer. These results are shown in Table 2.6.

Consistent with the results in Table 2.5, Table 2.6 shows that the pricing error is largest during the learning period and declines the longer the product has been available. The results in both tables are similar in magnitude and sign indicating that regardless of if the length of a product's availability is measured in levels or logs the impact on the pricing error is similar.

Table 2.6: Impacts on the Log Pricing Error Over Time

	(1)	(2)	(3)
Informed Period (Testing Set)	−0.009*	−0.009*	−0.009
	[−0.017, −0.0008]	[−0.017, −0.0008]	[−0.022, 0.005]
Learning Period	0.361***	0.285***	0.371***
	[0.334, 0.387]	[0.259, 0.312]	[0.328, 0.415]
Log Years Since Store Introduction x Informed Period (Training Set)	0.002	0.024***	0.010**
	[−0.005, 0.009]	[0.017, 0.031]	[0.004, 0.016]
Log Years Since Store Introduction x Informed Period (Testing Set)	0.003	0.026***	0.011***
	[−0.003, 0.010]	[0.019, 0.032]	[0.005, 0.018]
Log Years Since Store Introduction x Learning Period	−0.100***	−0.047***	−0.027**
	[−0.117, −0.084]	[−0.064, −0.031]	[−0.045, −0.010]
Log Years Since Retailer Introduction		−0.070***	
		[−0.079, −0.062]	
Log Years Since Retailer Introduction x Informed Period (Training Set)			−0.024**
			[−0.041, −0.008]
Log Years Since Retailer Introduction x Informed Period (Testing Set)			−0.024**
			[−0.040, −0.009]
Log Years Since Retailer Introduction x Learning Period			−0.097***
			[−0.110, −0.083]
FE: Stores	X	X	X
FE: Month	X	X	X
FE: Pasta Type	X	X	X
N	7 129 997	7 129 997	7 129 997

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note:
95% confidence intervals clustered at the store and month level are shown. Observations are categorized as either belonging to 1) the set of observations in the informed period used to train the cost model, 2) the set of observations in the informed peroid used to test the cost model, or 3) the set of observations in the learning period. Indicators for these three sets are included in the regressions and the set of observations in the informed period used to train the cost model is the excluded set. Observations where the counterfactual price is below the 10th percentile or above the 96th percentile are excluded.

2.8 References

Berry, Steven T, "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, (1994), 242–262 (JSTOR).

Berry, Steven, Levinsohn, James, and Pakes, Ariel, "Differentiated products demand systems from a combination of micro and macro data: The new car market," *Journal of Political Economy*, 112 (2004), 68–105 (The University of Chicago Press).

Cardell, N Scott, "Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity," *Econometric Theory*, 13 (1997), 185–213 (Cambridge University Press).

Ching, Andrew T, "A dynamic oligopoly structural model for the prescription drug market after patent expiration," *International Economic Review*, 51 (2010), 1175–1207 (Wiley Online Library).

Conlon, Christopher, and Gortmaker, Jeff, "Best practices for differentiated products demand estimation with PyBLP," *The RAND Journal of Economics*, 51 (2020), 1108–1161.

Conlon, Christopher, and Mortimer, Julie Holland, "Empirical properties of diversion ratios," *The RAND Journal of Economics*, 52 (2021), 693–726 (Wiley Online Library).

Dellavigna, Stefano, and Gentzkow, Matthew, "Uniform pricing in U.S. retail chains," *Quarterly Journal of Economics*, 134 (2019), 2011–2084.

Doraszelski, Ulrich, Lewis, Gregory, and Pakes, Ariel, "Just starting out: Learning and equilibrium in a new market," *American Economic Review*, 108 (2018), 565–615.

Gandhi, Amit, and Houde, Jean-François, "Measuring substitution patterns in differentiated-products industries," (2020) (NBER).

Grigolon, Laura, and Verboven, Frank, "Nested logit or random coefficients logit? A comparison of alternative discrete choice models of product differentiation," *Review of Economics and Statistics*, 96 (2014), 916–935 (The MIT Press).

Hausman, Jerry A, "Valuation of new goods under perfect and imperfect competition," *The Economics of New Goods*, (T. F. Bresnahan and R. J. Gordon, eds.) (University of Chicago Press, 1996).

Huang, Yufeng, Ellickson, Paul B., and Lovett, Mitchell J., "Learning to set prices," *Journal of Marketing Research*, 59 (2022), 411–434.

Jeon, Jihye, "Learning and investment under demand uncertainty in container shipping," *The RAND Journal of Economics*, 53 (2022), 226–259.

Morrow, W Ross, and Skerlos, Steven J, "Fixed-point approaches to computing bertrand-nash equilibrium prices under mixed-logit demand," *Operations research*, 59 (2011), 328–345 (INFORMS).

Osborne, Matthew, "Consumer learning, switching costs, and heterogeneity: A structural examination," *Quantitative Marketing and Economics*, 9 (2011), 25–70.

CHAPTER 3

High Frequency Traders Slow Information Revelation

In modern financial markets, investors seeking to profit from new information compete with high frequency traders (HFTs) who invest in the ability to trade quickly. This paper examines their competition and its impact on market outcomes. Using a dynamic model, I find that while HFTs' speed advantage improves market liquidity (as indicated by bid-ask spreads), it also slows down the speed at which new information is incorporated into prices. HFTs decrease the profitability of new information by trading ahead of slower investors, prompting investors to trade less to avoid revealing valuable information. As a result, new information takes longer to be incorporated into prices. However, HFTs also improve market liquidity by ensuring that prices better reflect widely known information.

3.1 Introduction

In this chapter, I study how newly discovered information becomes incorporated into market prices. New information is profitable to investors who can transact in financial markets. If prices do not reflect an investor's new information, the investor can profit by buying the asset if the asset's true value is higher than the current price (or conversely by selling if the true value is lower than the current price). As long as prices don't reflect the new information, investors can increase their profits by trading more shares. But trading large amounts of the asset comes at a cost. It has long been recognized in the literature (see for example, Easley and O'Hara (1987)) and by financial market practitioners that large trades are indicative of an asset's value and can move prices. So, investors with new information face a trade-off: trade aggressively and earn large profits now or trade more slowly to preserve the new information's value and the ability to profit in the future.

The investor's trade-off between trading aggressively now and trading slowly over time does not only matter for the investor but is also important for the general public. Prices aggregate new information and reflect the public's common knowledge. So, it is important how quickly new information is incorporated into prices. If investors decide to trade less aggressively, they preserve their ability to profit in the future by delaying the time until the new information becomes common knowledge. Thus, for prices to be maximally informative, the public should want investors with new information to trade aggressively, revealing their new information quickly.

This chapter studies investors' decisions about how aggressively to trade and, as a result, how quickly new information becomes public, in the context of high frequency traders (HFTs). Unlike investors with new information, HFTs don't participate in the market because of an informational advantage. Instead, HFTs have invested in a speed advantage that allows them to trade faster than other market participants. Modern financial markets are spread across a variety of platforms and locations. As a result of this diffuse market structure, HFTs can use their speed advantage to quickly trade across markets profiting from any differences in prices.

At first glance, HFTs would appear to benefit the public who rely on prices as a signal of information. HFTs help ensure that prices are uniform across markets so that regardless of which market is checked prices provide the same information. However, this chapter shows that the impact of HFTs is mixed. HFTs provide a benefit by ensuring that prices better reflect all common knowledge information. However, HFTs also slow the spread of new information.

Investors with new information want to trade as much as possible before prices update to reflect that information. Due to the fractured nature of markets, large trades often require transacting across multiple exchanges. But HFTs' speed advantage makes informed investors' ability to profitably trade across multiple exchanges difficult. Upon observing an informed investor's trade on one market, HFTs can use their speed advantage to preempt the informed investor elsewhere by trading ahead of the informed investor and updating prices (this is referred to as active order anticipation). HFTs' ability to quickly update prices across exchanges limits investors' opportunities to profit from new information.[1] By reducing the profitability of new information, HFTs cause investors to both do

[1] Active order anticipation is significant enough that institutional investors have invested in technology such as smart order routers like Royal Bank of Canada's THOR that attempt to account for different latencies across exchanges to synchronize order arrival times.

less research into discovering new information (as shown in Baldauf and Mollner (2020)) and trade less aggressively, slowing the speed new information becomes common knowledge (as shown in this chapter).

In addition to impacting the incentives of investors with new information, HFTs also impact the incentives of the market participants who set prices. Most prices in modern markets are set by liquidity providers who—rather than taking a position on the value of assets—offer to buy and sell assets to profit from trading volume. Liquidity providers earn a profit by offering to buy assets for less than they sell them, or—using the language of financial markets—by charging a positive bid-ask spread. As shown in Glosten and Milgrom (1985), bid-ask spreads allow liquidity providers to profit when trading with uninformed traders and compensate the liquidity provider for the adverse selection risk of potentially trading with an investor who has new information.

In addition, liquidity providers can also reduce their adverse selection risk by investing in the ability to trade quickly and becoming HFTs. When liquidity providers are HFTs, they can mitigate the risk of trading with an investor with new information by using their speed to update prices once new information is revealed. Updating prices when new information is revealed elsewhere, but before investors can trade, is called passive order anticipation. Passive order anticipation reduces the liquidity provider's adverse selection risk, but it also reduces the profits to investors with new information.

By engaging in active and passive order anticipation, HFTs reduce the benefit of revealing new information and influence the decisions of investors with that information. Baldauf and Mollner (2020) showed in a static setting that by reducing the profitability of new information HFTs reduce the research efforts of investors to acquire new information. This chapter extends Baldauf and Mollner (2020) to a dynamic setting to understand how investors with new information best trade on that information and, as a result, how quickly new information becomes common knowledge. In a dynamic setting, investors with new information face a trade-off between trading aggressively on multiple exchanges (potentially encountering order anticipation) or trading slowly over time (reducing trading volumes in the short term to avoid order anticipation and preserve future profits). The dynamic setting in this chapter allows us to explore this strategic tension and show that HFTs cause investors to initially trade on fewer exchanges to avoid order anticipation. Trading on fewer exchanges allows investors to better disguise their trades among the normal market order flow and extends the time until investors' new information becomes common knowledge.

This chapter also contributes to our understanding of how HFTs impact price setting. As established in Glosten and Milgrom (1985), a positive bid-ask spread arises from the adverse selection risk faced by liquidity providers. When liquidity providers are HFTs that changes these risks. When liquidity providers have a speed advantage, they reduce their risk from investors with new information by engaging in passive order anticipation. In addition, investors with new information respond to passive order anticipation by trading less aggressively which further reduces adverse selection. Both forces act in the same direction to reduce the liquidity provider's risk, resulting in smaller bid-ask spreads.

The approach of this chapter is to add HFTs to a two-period Glosten and Milgrom (1985) model that is solved similarly to Back and Baruch (2004). Trade of an asset of unknown value occurs simultaneously across multiple exchanges, so trade on any given exchange cannot account for the simultaneous order flow elsewhere. An investor is randomly either informed about the asset's value or an uninformed noise trader. The investor incurs a random latency across exchanges, so orders arrive asynchronously. The market contains an infinite number of HFTs, who can act as either a liquidity provider, setting prices on all exchanges, or as snipers, who seek to exploit the investor's asynchronous trading by engaging in active order anticipation.

At the beginning of each period, the liquidity provider sets prices on all exchanges and the informed investor chooses how many exchanges to trade on. Trading on a greater number of exchanges is more profitable in the short term but reveals the value of the asset to HFTs and precludes future trade. If the informed investor trades on enough exchanges to reveal the value of the asset, HFT liquidity providers and HFT snipers compete to cancel stale prices (passive order anticipation) and exploit stale prices (active order anticipation) respectively.

In analyzing this model, I primarily focus on the informed investor's first-period, strategic decision about how many exchanges to send orders to. In the first period, the informed investor trades off the first period and second period profits when making this choice. Increasing the number of exchanges to send orders increases first-period profits, but reveals the value of the asset precluding future, profitable trade. Sending orders to fewer exchanges limits first-period profits but conceals the value of the asset to preserve second -period profits. The addition of HFTs to the model reduces the profitability of sending orders to multiple exchanges because order anticipation reduces the number of the investor's orders that are executed. Order anticipation increases the incentive for the investor to trade on fewer exchanges in the first period to preserve future profits.

By reducing the investor's trading intensity, HFTs cause it to take longer for new information to become common knowledge. The investor's new information becomes common knowledge when market participants learn the value of the asset from observing total order flows. By reducing the trading intensity of the informed investor, HFTs reduce the likelihood that order flows reveal the value of the asset and increase the time until new information becomes common knowledge.

In this model, the impact of HFTs on markets is mixed. The introduction of HFTs reduces the value of new information and increases the time until that new information becomes common knowledge. Simultaneously, HFTs reduce the risks to liquidity providers, allowing for tighter bid-ask spreads and prices that better reflect common knowledge about the asset's value.

3.2 Related Literature

This chapter is primarily related to two strands of literature, the first is the market microstructure and strategic trading literature (e.g. Kyle (1985), Glosten and Milgrom (1985), Easley and O'Hara (1987)) and the second is the more recent and related HFT literature (for a survey see Menkveld (2016)).

Within the market microstructure literature, this chapter is related to Kyle (1985) and Glosten and Milgrom (1985) who are among the first to model the strategic trading decision of investors given their impact on prices. In both this chapter and Glosten and Milgrom (1985) model orders arrive sequentially and market makers set bid and ask prices accordingly. This is in contrast to Kyle (1985) who models market makers as only observing aggregate order flows. Back and Baruch (2004) link the two approaches by solving a version of Glosten and Milgrom (1985) where the informed trader chooses their trading times optimally and shows convergence to the Kyle (1985) model.

This chapter adapts the framework of Back and Baruch (2004) to solve a two-period version of their model that incorporates HFTs. As a result, this chapter can be viewed as an application of the market microstructure literature to a setting with HFTs. In this chapter, as in prior market microstructure models, an informed trader strategically reduces trade in early periods to dampen their price impact and preserve the ability to earn future profits. Although the methods of trading aggressively are different across models — in Back and Baruch (2004) aggressive trade is more frequent while here aggressive trade occurs on more exchanges — in both settings aggressive trade reveals information about the value of the asset resulting in a large price impact that decreases the profitability of future trade. The relationship between aggressive or large trades and information revelation underlies our model and is foundational in the market microstructure literature (see for example Easley and O'Hara (1987)).

In addition to this chapter's similarity to the market microstructure literature as a whole, this chapter is part of the literature's more recent focus on HFTs. HFTs can engage in two patterns of behavior, active (trading ahead of an informed investor) or passive (canceling outstanding prices) order anticipation, and most papers focus on one or the other. For example, Yang and Zhu (2020) study a two-period version of the Kyle (1985) model where HFTs engage in only active order anticipation which induces informed traders to add noise to their order sizes. Foucault, Hombert, and Roşu (2016) and Li (2018) study only active order anticipation in a similar model and find that when HFTs' speeds increase market makers face greater adverse selection and reduce liquidity. Ait-Sahalia and Saglam (2017a) and Ait-Sahalia and Saglam (2017b) study only passive order anticipation in a model where the only HFTs are market makers. This speed advantage to market makers reduces the adverse selection they face and results in tighter spreads. While this prior literature is focused on HFTs' impact on market participants by focusing on either active or passive order anticipation this chapter includes both mechanisms to study the overall impact of HFTs.

However, this chapter is not alone in studying both active and passive order anticipation. Examples that do so include Foucault, Kozhan, and Tham (2017), Menkveld and Zoican (2017), Budish, Cramton, and Shim (2015), Baldauf and Mollner (2020). This chapter is most closely related to Baldauf and Mollner (2020) which builds on Budish, Cramton, and Shim (2015). Budish, Cramton, and Shim (2015) consider a model with exogenous public information while Baldauf and Mollner (2020) endogenize new information by considering an informed investor who undertakes costly research. Both of these models are static and model new information as immediately revealed, allowing the authors to examine how HFTs react to this new information. This chapter extends the static framework developed by Baldauf and Mollner (2020) to a dynamic, two-period setting to allow an informed investor to strategically trade over time. Considering HFTs in a dynamic setting allows us to consider the impacts of HFTs on the speed of trading and the speed that new information is revealed. In a dynamic setting, investors have an incentive to trade more slowly to maintain an informational advantage when trading with faster HFTs. By considering both active and passive order anticipation in a dynamic setting, this chapter can contribute to the literature by studying the full impact of HFTs on both liquidity and information revelation.

3.3 Model

An asset is traded across multiple exchanges and over time. There are two types of traders, an investor and HFTs. Investors can be either uninformed or informed about the asset's value while HFTs do not know the asset's value.

Asset The asset has value which is either 0 or 1 with prior probability $P(v = 1) = q$. The asset's value is determined by nature prior to $t = 1$ and is initially unknown to all traders.

Exchanges There are $X \geq 2$ exchanges that operate separate limit order books (LOB). A LBO operates by processing orders that arrive to the exchange sequentially in the order they are received. If multiple orders arrive at the same time then ties are broken uniformly at random (this is equivalent to assigning a small, random latency to orders). New orders are processed by either being matched with an existing order in the LOB and executing a trade or being added to the LOB. Two orders are matched and a trade is executed if the bid price of one order is greater than or equal to the ask price of the other. Orders are made public and are common knowledge to all traders after they are processed.

Time Time is primarily divided into two periods $\{1, 2\}$ and further subdivided into periods of length $\epsilon \ll \frac{1}{3}$. In this model we only need to consider the time periods $T = \{1, 1+\epsilon, 1+2\epsilon, 2, 2+\epsilon, 2+2\epsilon\}$.

Investor A single investor is always present in the market and only acts at $t \in \{1, 2\}$ by choosing the number of exchanges to send orders to in that period. The investor is either informed or uninformed about the value of the asset. They are informed with probability $\mu \in (0, 1)$. Informed investors learn the value of the asset at $t = 1$. Informed investors who learn $v = 1$ are denoted high type while informed investors who learn $v = 0$ are denoted as low type. Uninformed investors never learn the value of the asset, but in each period $t \in \{1, 2\}$ have either a buying or selling liquidity motive (each with equal probability) for a single share of the asset. The uninformed investor's buying or selling motives are independent across periods 1 and 2. As a result of only demanding one share per period, uninformed investors only ever send an order to a single exchange every period $t \in \{1, 2\}$.[2]

HFTs There are an infinite number of HFTs who can act at any time $t \in T$ by sending orders to exchanges. HFTs are uninformed about the asset's value. HFTs endogenously sort themselves into two roles, liquidity providers and snipers. Liquidity providers act to set bid and ask prices at times $t \in \{1, 2\}$ and cancel any outstanding prices in intermediate periods $t \in \{1 + \epsilon, 2 + \epsilon\}$. Snipers may send orders at $t \in \{1 + \epsilon, 2 + \epsilon\}$ in an attempt to trade with the liquidity provider's outstanding orders.

Latency Latency is the time between when an order is sent by a trader and the time at which it is processed by an exchange. Latency has the support $\{\epsilon, 2\epsilon\}$, so an order sent at time t is processed at either $t + \epsilon$ or $t + 2\epsilon$. Orders sent by HFTs always have latency ϵ while orders sent by investors may have latency of either ϵ or 2ϵ. When an investor sends orders one order is guaranteed to have latency ϵ while the remaining orders have latency ϵ with independent probabilities $1 - \rho \geq 0.5$.[3] A decrease in ρ decreases both the expected value and variance of latency and can be seen as in improvement in the investor's trading technology.[4]

Actions: Uninformed investors, informed investors, and snipers act by sending orders to buy or sell the asset (these can be thought of as market orders or as limit orders to buy at any price $\leq$

[2] This assumption that uninformed investors only send orders to a single exchange is restrictive, but does not change the model's predictions. The assumption identifies the investor as informed if orders are sent to more than one exchange. This simplifies the strategies of the informed investor and makes solving for equilibrium easier. Relaxing this assumption makes the strategic decision of the informed investor more complex, but the same strategic trade-offs hold.

[3] The investor's latency can be viewed as an inability to coordinate the timing of orders sent to multiple exchanges. One exchange always receives an order immedietly and is guarneteed to have latency ϵ. However, the difficulty of coordinating orders across exchanges is modeled as additional orders randomly receiving different latencies.

[4] This framing is useful because we often think of HFTs as the actors who improve their trading technology (represented as an increase in ρ).

1 and sell at any price $\geq$ 0). Further, high type informed investors are restricted to sending buy orders and low type informed investors are restricted to sending sell orders. Informed investors and HFTs can send orders to any number of exchanges and choose the number of exchanges to target. Uninformed investors only send orders to a single exchange, which is randomly selected with uniform probability among all exchanges each period. Uninformed and informed investors are restricted to sending orders at times $t \in \{1, 2\}$ while HFTs can send orders at any time $t \in T$. Orders sent by all players are for a single share. This restriction on order sizes is because it is only ex-post profitable for liquidity providers to trade with uninformed investors who only submit orders for one share.

Formally, denote z_t^+ as the cumulative number of buy orders sent by snipers and investors that have arrived to all exchanges through time t, z_t^- as the cumulative number of sell orders sent by snipers and investors that have arrived to all exchanges through time t, and $z_t = z_t^+ - z_t^-$ as the net cumulative orders that have arrived. The process **z** contains the complete history of anonymous trade. Denote the σ-field generated by $\{z_s | s \leq t\}$ by $\mathcal{F}_t$. Informed investors choose the number of exchanges to send orders $x^I(z_t) : \mathcal{F}_t \to [1, X] \subset \mathbf{N}$ and snipers choose the number of exchanges to send orders along with their sign (buy or sell) $x^S(z_t) : \mathcal{F}_t \to \{\text{buy, sell}\} \times [1, X] \subset \{\text{buy, sell}\} \times \mathbf{N}$. Uninformed investors are not strategic and their orders are deterministic based on their liquidity motive.

Liquidity providers act by sending or canceling orders with specified prices (prices for buy orders are bids and prices for sell orders are asks). These prices are conditional on all public information which consists of the past order flow on all exchanges $z_t \in \mathcal{F}_t$. Exchanges are ex-ante equivalent, so prices are equivalent across exchanges. Formally, bid and ask prices set at time t are a function of the cumulative order flow history across all exchanges at that time $a(z_t), b(z_t) : \mathcal{F}_t \to [0, 1]$.

Payoffs From acquiring a portfolio of y dollars and s shares informed investors and HFTs receive utility $y + sv$. I have said that uninformed investors have an illiquid demand/supply for one share every period depending on if they have a buying or selling motive that period. In order to rationalize this we can trivially define the per-period payoffs to a uninformed investor as follows: an uninformed investor with a buying motive gets utility $y + \mathbf{1}\{s = 1\}$ and an uninformed investor with a selling motive gets utility $y - \mathbf{1}\{s \neq -1\}$.

The game proceeds as follows starting at $t = 1$ and an example of events is depicted below in

Figure 3.1.

t = 1: The investor is informed about the value of the asset with probability μ. If the investor is uninformed they randomly either send a buy or sell order to a single exchange. If the investor is informed they choose the number of exchanges to send orders $x^I(0)$ (if $v = 1$ these are buy orders and if $v = 0$ these are sell orders). One HFT acts as a liquidity provider by sending orders to all exchanges to set bid $b(0)$ and ask $a(0)$ prices.[5] The remaining HFTs act as snipers. Snipers have the same information as liquidity providers, so they cannot profit off of prices $b(0)$, $a(0)$ and take no action.

t = 1+ϵ: The orders from the liquidity provider to set prices $b(0)$, $a(0)$ arrive to all exchanges. The orders from the investor that receive latency ϵ arrive to a subset of exchanges and are executed against the prices set by liquidity providers. The cumulative order flow $z_{1+\epsilon}$ is the number of orders sent by the investor that receive latency ϵ. Players update their beliefs about the type of the investor and the value of the asset based on $z_{1+\epsilon}$. If $|z_{1+\epsilon}| = 1$ the investor could be uninformed or informed, if $z_{1+\epsilon} > 1$ the investor must be informed and $v = 1$, and if $z_{1+\epsilon} < -1$ the investor must be informed and $v = 0$.

Exchanges that did not receive orders from the investor have outstanding bid $b(0)$ and ask $a(0)$ prices. If $|z_{1+\epsilon}| > 1$ snipers can profit off of the outstanding prices $b(0)$, $a(0)$ and send orders to exchanges with outstanding prices to buy if $z_{1+\epsilon} > 1$ i.e. $x^S(z_{1+\epsilon}) = buy \times (X - z_{1+\epsilon})$ or sell if $z_{1+\epsilon} < -1$ i.e. $x^S(z_{1+\epsilon}) = sell \times (X + z_{1+\epsilon})$. If $|z_{1+\epsilon}| \geq 1$ liquidity providers send orders to cancel all outstanding bid $b(0)$ and ask $a(0)$ prices because any orders arriving at $t + 2\epsilon$ are from either an informed investor or a sniper.

t = 1 + 2ϵ: On exchanges that did not receive orders from the investor at $t = 1 + \epsilon$ any cancellations sent by liquidity providers, orders sent by snipers, and orders sent by the investor with latency 2ϵ arrive and are processed in a random order. There is at most a single outstanding bid $b(0)$ and ask $a(0)$ price per exchange, so only the order processed first is executed. There are an infinite number of snipers, so if snipers sent orders at $t = 1 + \epsilon$ these orders are processed first almost surely. After orders are processed the cumulative order flow is $|z_{1+2\epsilon}| = 1$ if the investor is uninformed or if the

[5] It is equivalent if there is one liquidity provider sending orders to all exchanges or one liquidity provder per exchange. For simplicity I will assume only one liquidity provider sets bid and ask prices on all exchanges.

informed investor sent a single order and is $|z_{1+2\epsilon}| = X$ if the investor is informed and sent orders to multiple exchanges. No exchanges have outstanding bid or ask prices and liquidity providers do not send orders to set new prices.

t = 2: If $|z_2| = X$ the investor must be informed and the value of the asset is common knowledge. In this case there is no more trade and the game ends. If $|z_2| = 1$ the investor may be informed or uninformed and the asset value remains unknown to HFTs. Liquidity providers send orders to all exchanges to set bid $b(z_2)$ and ask $a(z_2)$ prices. If the investor is uninformed they randomly either send a buy or sell order to a single exchange. If the investor is informed they choose the number of exchanges to send orders $x^I(1)$ (if $v = 1$ these are buy orders and if $v = 0$ these are sell orders). Snipers have the same information as liquidity providers, so they cannot profit off of prices $b(1), a(1)$ and take no action.

If $|z_2| = 1$ events at $t = 2 + \epsilon$ and $t = 2 + 2\epsilon$ occur equivalently to $t = 1 + \epsilon$ and $t = 1 + 2\epsilon$ with the exception that the cumulative order flow is augmented by one.

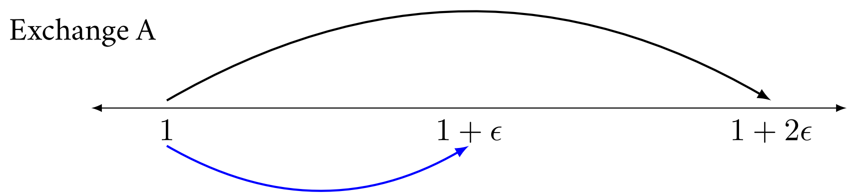
In this model the strategies to be solved for are the prices $\{a(0), b(0), a(1), b(-1)\}$ set by liquidity providers and the number of exchanges targeted by the informed investor $\{x^I(0), x^I(1), x^I - 1)\}$.[6]

3.4 Equilibrium Description

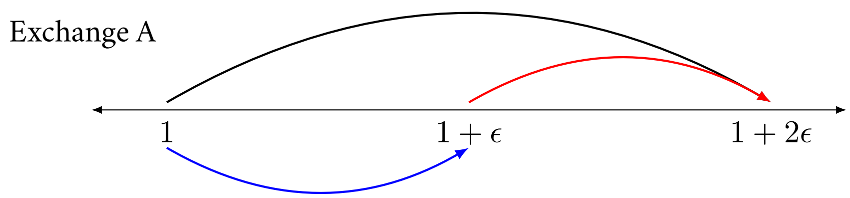
In this game, and in Glosten and Milgrom (1985), the liquidity provider sets a positive bid-ask spread to compensate for adverse selection when the investor's type is unknown. By charging a positive bid-ask spread the liquidity provider is able to earn a profit from trading with the uninformed investor that compensates for the expected loss when trading with informed investors and snipers. This ability to profit from the uninformed investor provides an incentive for the liquidity provider to participate in trade, but because HFTs endogenously compete to perform the role of the liquidity provider expected profits are zero.

Informed investors face a strategic tension at $t = 1$ between sending orders to multiple exchanges and sending orders to a single exchange. Uninformed investors send orders to a single exchange,

[6] For completness it is also important to specify $a(-1), b(1)$. However, these prices are only relevant if the investor is uninformed so they are trivially equal to q.



(a) At time $t = 1$ an informed investor sends a buy order to exchange A and this order is randomly given latency 2ϵ (black arrow). The liquidity provider also sends a sell order to exchange A at time $t = 1$ and this orders has latency ϵ (blue arrow). The liquidity provider's order is the only order to arrive to exchange A at time $t = 1 + \epsilon$, so it is not executed and remains in the LOB. At time $t = 1 + 2\epsilon$ the informed investor's order arrives and is executed against the liquidity provider's order.



(b) If in addition to the actions described in (a), a sniper was to send an order to buy to exchange A at time $t = 1 + \epsilon$ (red arrow) either the sniper's order or the informed investor's order would be matched with the liquidity provider's order at $t = 1 + 2\epsilon$ (either the sniper's order or informed investor's order is randomly processed first according to the tie breaking mechanism) and the remaining order is automatically canceled because it is not executed. As a result, either the informed investor or the sniper trades with the liquidity provider and the LOB is empty at the end of the $t = 1 + 2\epsilon$ period.

Figure 3.1: Example of Events on a Single Exchange

so by sending orders to multiple exchanges the informed investor reveals their type and the value of the asset, precluding future trade. By sending a single order, the informed investor can disguise themselves as an uninformed investor and continue profitably trading at $t = 2$. In choosing the number of orders to send at $t = 1$ the informed investor must balance short term profits with their price impact and ability to profit in the future. The difference in order flow from uninformed and informed investors allows HFTs to learn information about the value of the asset and the type of the investor.

Snipers act by trading in the same direction as the informed traders in order to profit from any information that is revealed but not yet incorporated into prices. Snipers reduce the profits of informed investors because their speed advantage allows them to trade in advance of any long latency orders.[7] Snipers also create losses for liquidity providers by exploiting stale prices before they can be

[7] Technically snipers' orders and the investor's long latency orders arrive simultaneously, but because there are infinite

cancelled. Snipers earn positive profits by reacting quickly when the order flow reveals information, but because they are infinite in number their individual profits are zero almost surely.

I characterize the equilibrium in terms of bid and ask prices given past order flow $\{a(0), b(0), a(1), b(-1)\}$ and the probability an informed investor sends X orders at time $t = 1$ (the reason this probability is sufficient to capture the informed investor's strategy is explained in the following section). The equilibrium concept is weak perfect Bayesian equilibrium where the relevant beliefs are about if the investor is informed. I begin deriving the equilibrium by considering the actions of the informed investor, deriving beliefs and actions of HFTs, and finally solving for the optimal actions of the informed investor.

3.4.1 Informed Investor

Lemma 3.4.1. $x^I(z_t) \in \{1, X\}$ *is strictly dominant.*

Although the informed investor can send orders to any number of exchanges $x^I(z_t) \in [1, X] \subset \mathbf{N}$, sending orders to a number other than one or all of the exchanges is strictly dominated. Sending orders to more than one exchange is information revealing because uninformed investors only send orders to a single exchange. Snipers don't send orders at $t = 1, 2$ so if $|z_{1+\epsilon}| > 1$ or $|z_{2+\epsilon}| > 1$ these orders must have come from an informed investor. This is perfectly informative about the asset's value because only a high type informed investor sends buy orders to more than one exchange and only a low type informed investor sends sell orders to more than one exchange. All $x^I > 1$ reveal equivalent information, but payoffs are increasing in the number of orders filled and thus the number of orders sent. As a result, if multiple orders are sent then sending X orders is dominant.[8]

At time $t = 2$ the informed investor sends orders to all X exchanges. Informed investors cannot send orders after $t = 2$, so there is no cost to revealing information by sending multiple orders. Orders at all exchanges are filled at the same price, so informed investor's payoffs are strictly increasing in the number of orders sent.

---

snipers a sniper's order is processed first almost surely.

[8] In a model where uninformed inventors can send orders to more than one exchange it is still dominant among fully revealing strategies to send orders to all exchanges. However, when the informed investor does not wish to fully reveal the value of the asset the investor's strategy may be more complex.

The only remaining choice made by informed investors is to send either one or X orders at $t = 1$. Sending X orders at $t = 1$ reveals the value of the asset and precludes trade at $t = 2$, so by sending X orders at $t = 1$ the investor is trading off higher profits at $t = 1$ for future profits at $t = 2$. To formalize this let X_I be the expected number of the informed investor's orders that are executed when X orders are sent. Let $a(z)$ be the ask price, and $b(z)$ the bid price set by liquidity providers when z total orders have previously arrived across all exchanges.[9] The payoffs to an informed investor sending orders at $t = 1$ can be expressed as:

$$\text{High type sending 1 order}: \quad 1 - a(0) + X_I(1 - a(1)) \tag{3.1}$$

$$\text{High type sending X orders}: \quad X_I(1 - a(0)) \tag{3.2}$$

$$\text{Low type sending 1 order}: \quad b(0) + X_I b(-1) \tag{3.3}$$

$$\text{Low type sending X orders}: \quad X_I b(0) \tag{3.4}$$

where the $1 - a(0)$ and $b(0)$ terms are payoffs from orders sent at $t = 1$ and the $1 - a(1)$ and $b(-1)$ terms are payoffs from orders sent at $t = 2$. Equating the expressions for sending 1 or X orders, the high type investor is indifferent between sending orders to one and X exchanges at $t = 1$ if

$$\frac{1 - a(0)}{a(1) - a(0)} = X_I \tag{3.5}$$

and the low type investor is indifferent at $t = 1$ if

$$\frac{b(0)}{b(0) - b(-1)} = X_I \tag{3.6}$$

Denote the probability that the high type sends X orders as α and the probability that the low type sends X orders at β. We can summarize the actions of a high type informed investor as follows:

- $t = 1$: Send orders to all X exchanges with probability α (or β for a low type investor) and to only one exchange with probability $1 - \alpha$ ($1 - \beta$ for a low type investor).

[9] Prices are set with at ϵ delay, so z represents the information available to the liquidity provider when $a(z), b(z)$ are set.

- $t = 2$: Send orders to all X exchanges.

3.4.2 HFTs' Beliefs

3.4.2.1 $\mathbf{t} \in [\mathbf{1}, \mathbf{2})$

At $t = 1$ the probability that the asset value is 1 is $P(v = 1) = q$ and the probability that the investor is informed is μ. Orders arrive with of a delay of at least ϵ so it is important at $t = 1$ to understand how an order arriving at $t = 1 + \epsilon$ impacts the HFTs' beliefs. If an order is received at a given exchange i at $t = 1 + \epsilon$ (denoted $r_{i,1+\epsilon} = 1$ for a buy and $r_{i,1+\epsilon} = -1$ for a sell) and no exchanges have previously received orders ($z_1 = 0$) then the probability that order came from an uninformed investor can be calculated using Bayes rule. It is important to note that although multiple exchanges may receive orders at $t = 1 + \epsilon$, prices are set at $t = 1$ so the only information available for setting prices on exchange i is the number of orders arriving at exchange i ($r_{i,1+\epsilon}$) and the history of orders at all exchanges ($z_1 = 0$).

$$P(\text{uninformed}|r_{i,1+\epsilon} = 1, z_1 = 0) = \frac{\frac{1}{2X}(1-\mu)}{\frac{1}{2X}(1-\mu) + \frac{1}{X}q\mu(1-\alpha) + q\mu\alpha(1-\rho\frac{X-1}{X})} \tag{3.7}$$

$$P(\text{uninformed}|r_{i,1+\epsilon} = -1, z_1 = 0) = \frac{\frac{1}{2X}(1-\mu)}{\frac{1}{2X}(1-\mu) + \frac{1}{X}(1-q)\mu(1-\beta) + (1-q)\mu\beta(1-\rho\frac{X-1}{X})} \tag{3.8}$$

where $\frac{1}{X}$ is the probability that when an investor sends a single order at $t = 1$ that order arrives to exchange i. Additionally, $\frac{1}{2}$ is the probability that the uninformed investor has a buying (or selling) motive, and $1 - \mu$ is the probability that the investor is uninformed.

In the denominator $q\mu$ is the probability the investor is informed and sends a buy order because the asset value is 1. The term $(1 - \rho\frac{X-1}{X})$ corresponds to the probability that exchange i receives an order at $1 + \epsilon$ when an investor sends orders to all exchanges. When an investor sends orders to all exchanges one order is guaranteed to arrive at $1 + \epsilon$. Of the remaining $X - 1$ orders each has probability $1 - \rho$ of arriving at $1 + \epsilon$. As a result, the expected number of orders that arrive at $1 + \epsilon$ is $1 + (X - 1)(1 - \rho)$. There is a $\frac{1}{X}$ probability that each order arrives to exchange i so the probability of exchange i receiving an order at $1 + \epsilon$ is $\frac{1+(X-1)(1-\rho)}{X} = 1 - \rho\frac{X-1}{X}$.

The probability that an order came from an informed investor sending orders to only one exchange is:

$$P(\text{informed sending }1|r_{i,1+\epsilon}=1,z_1=0)=\frac{\frac{1}{X}q\mu(1-\alpha)}{\frac{1}{2X}(1-\mu)+\frac{1}{X}q\mu(1-\alpha)+q\mu\alpha(1-\rho\frac{X-1}{X})} \tag{3.9}$$

$$P(\text{informed sending }1|r_{i,1+\epsilon}=-1,z_1=0)=\frac{\frac{1}{X}(1-q)\mu(1-\beta)}{\frac{1}{2X}(1-\mu)+\frac{1}{X}(1-q)\mu(1-\beta)+(1-q)\mu\beta(1-\rho\frac{X-1}{X})} \tag{3.10}$$

The probability that an order came from an informed investor sending orders to all X exchanges is:

$$P(\text{informed sending }X|r_{i,1+\epsilon}=1,z_1=0)=\frac{q\mu\alpha(1-\rho\frac{X-1}{X})}{\frac{1}{2X}(1-\mu)+\frac{1}{X}q\mu(1-\alpha)+q\mu\alpha(1-\rho\frac{X-1}{X})} \tag{3.11}$$

$$P(\text{informed sending }X|r_{i,1+\epsilon}=-1,z_1=0)=\frac{(1-q)\mu\beta(1-\rho\frac{X-1}{X})}{\frac{1}{2X}(1-\mu)+\frac{1}{X}(1-q)\mu(1-\beta)+(1-q)\mu\beta(1-\rho\frac{X-1}{X})} \tag{3.12}$$

Any orders arriving at $t = 1 + 2\epsilon$ are guaranteed to be from an informed investor. All investors have at least one order arriving at $t = 1 + \epsilon$, so the uninformed investor, who only sends one order, will never have an order arrive at $t = 1 + 2\epsilon$. However, if one order arrives at $t = 1 + \epsilon$ and none arrive at $t = 1 + 2\epsilon$ this does not guarantee the investor is informed. If an informed investor only sends one order this would produce the same order arrival pattern as an uninformed investor. As a result, orders arriving at $t = 1+2\epsilon$ guarantee the investor is informed, but only one order at $t = 1+\epsilon$ does not guarantee the investor is uninformed.

3.4.2.2 t $\geq$ 2

If multiple orders were received before $t = 2$ both the investor's type and the asset's value are revealed. At this point all information is common knowledge and no further trade occurs. If a single order was received before $t = 2$ (denoted $z_2 = 1$ for a buy and $z_2 = -1$ for a sell) and at $t = 2 + \epsilon$ an exchange

i receives an order (denoted $r_{i,2+\epsilon} = 1$ or $r_{i,2+\epsilon} = -1$) the probability the investor is uninformed is:

$$P(\text{uninformed}|r_{i,2+\epsilon} = 1, z_2 = 1) = \frac{\frac{1}{4X}(1-\mu)}{\frac{1}{4X}(1-\mu) + q(1-\alpha)\mu(1-\rho\frac{X-1}{X})} \tag{3.13}$$

$$P(\text{uninformed}|r_{i,2+\epsilon} = -1, z_2 = -1) = \frac{\frac{1}{4X}(1-\mu)}{\frac{1}{4X}(1-\mu) + (1-q)(1-\beta)\mu(1-\rho\frac{X-1}{X})} \tag{3.14}$$

Here $\frac{1}{4}$ is the probability the uninformed investor has a buying (or selling) liquidity motive at both $t = 1$ and $t = 2$.

The probability the order came from an informed investor is:

$$P(\text{informed sending } X|r_{i,2+\epsilon} = 1, z_2 = 1) = \frac{q(1-\alpha)\mu(1-\rho\frac{X-1}{X})}{\frac{1}{4X}(1-\mu) + q(1-\alpha)\mu(1-\rho\frac{X-1}{X})} \tag{3.15}$$

$$P(\text{informed sending } X|r_{i,2+\epsilon} = -1, z_2 = -1) = \frac{(1-q)(1-\beta)\mu(1-\rho\frac{X-1}{X})}{\frac{1}{4X}(1-\mu) + (1-q)(1-\beta)\mu(1-\rho\frac{X-1}{X})} \tag{3.16}$$

These probabilities can be used to calculate the probability the asset value is 1:

$$\begin{aligned} P(v = 1|r_{i,2+\epsilon} = 1, z_2 = 1) =& qP(\text{uninformed}|r_{i,2+\epsilon} = 1, z_2 = 1) \\ &+ P(\text{informed sending } X|r_{i,2+\epsilon} = 1, z_2 = 1) \end{aligned} \tag{3.17}$$

$$\begin{aligned} P(v = 0|r_{i,2+\epsilon} = -1, z_2 = -1) =& (1-q)P(\text{uninformed}|r_{i,2+\epsilon} = -1, z_2 = -1) \\ &+ P(\text{informed sending } X|r_{i,2+\epsilon} = -1, z_2 = -1) \end{aligned} \tag{3.18}$$

Similar to $t = 1 + 2\epsilon$, any orders received at $t = 2 + 2\epsilon$ must have come from an informed investor. The uninformed investor's order sent at $t = 2$ is guaranteed to receive latency ϵ and arrive at $t = 2 + \epsilon$. As a result, any orders arriving at $t = 2 + 2\epsilon$ are orders sent by an informed investor at $t = 2$ that received latency 2ϵ.

3.4.3 HFTs' Actions

3.4.3.1 $\mathbf{t} \in [\mathbf{1}, \mathbf{2})$

There are an infinite number of HFTs who compete in Bertrand competition to act as the liquidity provider and set prices. This ensures that expected profit made by liquidity providers is zero. At $t = 1$ the liquidity provider sends orders to all exchanges to post bid and ask prices $a(0), b(0)$. These prices persist until they are traded against or cancelled. At $t = 1 + \epsilon$ there are three potential traders of the asset: an uninformed investor, an informed investor sending orders to a single exchange, and an informed investor sending orders to all X exchanges. If a price on exchange i is not traded against at $t = 1 + \epsilon$ that price can either be traded against or cancelled at $t = 1 + 2\epsilon$. At $t = 1 + 2\epsilon$ there are two potential traders of the asset: an informed investor sending orders to all X exchanges or a sniper who has learned the value of the asset.

At $t = 1 + \epsilon$ the value of the asset is revealed if $|z_{1+\epsilon}| > 1$. Knowing the value of the asset, HFTs acting as snipers would like to trade on the outstanding prices at the $X - |z_{1+\epsilon}|$ exchanges that did not receive orders from the investor. Snipers respond by sending limit orders to these exchanges to buy at $p \leq 1$ if $z_{1+\epsilon}$ is positive and sell at $p \geq 0$ if $z_{1+\epsilon}$ is negative. These orders arrive at $t = 1 + 2\epsilon$ and—because there are infinite snipers—are processed first almost surely. In expectation when the investor sends orders to all exchanges snipers learn the value of the asset and are able to trade on X_S exchanges.

$$X_S = (X - 1)\rho - (X - 1)\rho^{X-1} \tag{3.19}$$

The first term is the expected number of exchanges that do not receive orders from the informed investor at $t = 1 + \epsilon$ and have prices available for trade by snipers. The second term represents the case when all of the informed investor's random latency orders arrive at $t = 1 + 2\epsilon$ and the asset's value is not perfectly revealed ($|z_{1+\epsilon}| = 1$).[10] As a result, X_S represents the expected number of outstanding prices at $t = 1 + 2\epsilon$ when the asset's value is revealed ($|z_{1+\epsilon}| > 1$). When the asset's value

[10] $(X - 1)$ is the number of exchanges with outstanding prices and ρ^{X-1} is the probability that all of the investor's random latency orders arrive at $t = 1 + 2\epsilon$.

is revealed an infinite number of snipers respond and are able to trade against all outstanding prices almost surely.

When an informed investor sends orders to all X exchanges, in expectation X_I of the informed investor's orders will execute.

$$X_I = 1 + (X-1)(1-\rho) + \frac{1}{2}(X-1)\rho^{X-1} \tag{3.20}$$

The first term, 1, is the informed investor's order that is guaranteed to arrive at $t = 1+\epsilon$. The second term, $(X-1)(1-\rho)$, is the expected number of orders that receive a random latency and also arrive at $t = 1+\epsilon$. If $|z_{1+\epsilon}| > 1$ then an infinite number of snipers respond and none of the informed investor's orders arriving at $t = 1+2\epsilon$ execute. However if $|z_{1+\epsilon}| = 1$ and snipers do not act, then the informed investor can expect an additional $\frac{1}{2}(X-1)$ orders to execute at $t = 1+2\epsilon$.[11] When $|z_{1+\epsilon}| = 1$ the investor is only able to trade against $\frac{1}{2}$ of the $X-1$ outstanding prices because the liquidity provider sends orders to cancel all prices at $t = 1+\epsilon$. The liquidity provider's cancellations and the investor's long latency orders arrive simultaneously and are processed in a random order that is independent across exchanges.

Proposition 3.4.1. *If $|z_{1+\epsilon}| = 1$ it is strictly dominant for snipers not to act at $t = 1+\epsilon$.*

Proof. I present the on-path case when $z_{1+\epsilon} = 1$ here and leave the off-path case as well as the case when $z_{1+\epsilon} = -1$ for Appendix 3.6. At $t = 1+\epsilon$ snipers do not send buy orders to trade on the liquidity provider's outstanding ask prices if $a(0) > E[v|z_{1+\epsilon} = 1]$. The expected asset value does not depend on the sniper's actions and can be calculated using Bayes rule as

$$E[v|z_{1+\epsilon} = 1] = P(v = 1|z_{1+\epsilon} = 1) = \frac{\frac{1}{2}q(1-\mu) + q\mu(1-\alpha) + q\mu\alpha\rho^{X-1}}{\frac{1}{2}(1-\mu) + q\mu(1-\alpha) + q\mu\alpha\rho^{X-1}} \tag{3.21}$$

where $\frac{1}{2}q(1-\mu)$ is the probability that $v = 1$ and $z_{1+\epsilon} = 1$ because the investor is uninformed. $q\mu(1-\alpha)$ is the probability $v = 1$ and $z_{1+\epsilon} = 1$ because the investor is informed and sends a single order. Finally, $q\mu\alpha\rho^{X-1}$ is the probability $v = 1$ and $z_{1+\epsilon} = 1$ because the investor is informed and sends orders to all X exchanges.

[11] This occurs with probability ρ^{X-1}.

There are infinite HFTs competing in Bertrand competition to set the ask price at $t = 1$ so the ask price $a(0)$ is determined by a zero profit condition. If snipers do not send orders when $z_{1+\epsilon} = 1$ then the zero profit condition is:

$$
\begin{aligned}
0 =&(a(0) - 1)P(v = 1)P(\text{uninformed and } z_{1+\epsilon} = 1) \\
&+ (a(0) - 0)P(v = 0)P(\text{uninformed and } z_{1+\epsilon} = 1) \\
&+ (a(0) - 1)P(v = 1)P(\text{informed sending } 1) \\
&+ (a(0) - 1)P(v = 1)P(\text{informed sending } X)(X_I + X_S) \\
=&(a(0) - 1)\frac{1}{2}q(1 - \mu) + a(0)\frac{1}{2}(1 - q)(1 - \mu) \\
&+ (a(0) - 1)q\mu(1 - \alpha) + (a(0) - 1)q\mu\alpha(X - \frac{1}{2}(X - 1)\rho^{X-1})
\end{aligned} \tag{3.22}
$$

This implies that the ask price $a(0)$ is

$$
a(0) = \frac{\frac{1}{2}q(1-\mu) + q\mu(1-\alpha) + q\mu\alpha(X - \frac{1}{2}(X-1)\rho^{X-1})}{\frac{1}{2}(1-\mu) + q\mu(1-\alpha) + q\mu\alpha(X - \frac{1}{2}(X-1)\rho^{X-1})} \tag{3.23}
$$

The sniper prefers not send orders when $z_{1+\epsilon} = 1$ if $a(0) > E[v|z_{1+\epsilon} = 1]$. Using Equations 3.21 and 3.23 we can show that the sniper does not send orders when:

$$
X - \frac{1}{2}(X - 1)\rho^{X-1} > \rho^{X-1} \tag{3.24}
$$

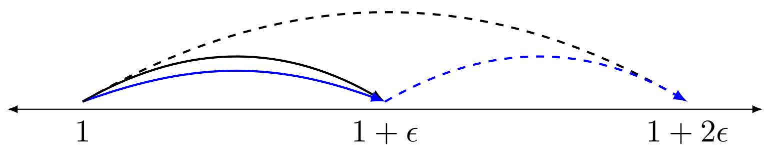
This condition holds true because $X \geq 2$ and $\rho \in [0, \frac{1}{2}]$. Thus snipers do not send buy orders at $t = 1 + \epsilon$ when $z_{1+\epsilon} = 1$. □

We can summarize the HFTs' actions that occur in the first period as follows (an example of these actions is shown in Figure 3.2):

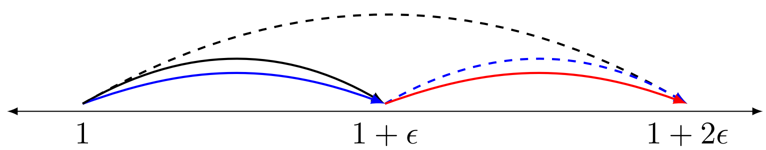
- $t = 1$: One HFT acts as a liquidity provider by sending orders to all exchanges to set bid and ask prices $b(0), a(0)$.
- $t = 1 + \epsilon$: The investor's orders arrive and execute against the bid or ask prices.
  - If only one exchange receives an order ($|z_{1+\epsilon}| = 1$) snipers do nothing and the liquidity

provider sends orders to cancel all outstanding bid and ask prices (Figure 3.2a).

- If more than one exchange receives an order ($|z_{1+\epsilon}| > 1$) snipers respond by sending buy or sell orders to all exchanges with outstanding prices. Liquidity provider also respond by sending orders to cancel all outstanding bid and ask prices (Figure 3.2b).



(a) At $t = 1$ the liquidity provider sends orders to all exchanges (solid blue arrow) to set bid and ask prices $b(0), a(0)$. The investor sends orders to either 1 or all X exchanges. If at $1+\epsilon$ only a single exchange receives an order (solid black arrow), any additional orders arriving at $1+2\epsilon$ must come from an informed investor (dashed black arrow). Upon observing an order at $1+\epsilon$, the liquidity provider sends orders to cancel all outstanding prices (dashed blue arrow). The liquidity provider's cancellation orders and the informed investor's long latency orders both arrive at $1+2\epsilon$ and the order in which they are processed is determined at random and independently across exchanges.



(b) At $t = 1$ the liquidity provider sends orders to all exchanges (solid blue arrow) to set bid and ask prices $b(0), a(0)$. The investor sends orders to either 1 or all X exchanges. If more than one order arrives at $1+\epsilon$ (solid black arrow), these orders must have come from an informed investor and the value of the asset is revealed. Liquidity providers respond by sending orders to cancel all outstanding prices (dashed blue arrow) and snipers respond by sending orders to trade on any outstanding prices (solid red arrow). The liquidity provider's cancellation orders, the informed investor's long latency orders, and the snipers' orders all arrive at $1+2\epsilon$. There are an infinite number of snipers, so a sniper's order is processed first almost surely.

Figure 3.2: HFTs' First Period Actions

3.4.3.2 $\mathbf{t \geq 2}$

At $t = 2$ if more than one order has previously been executed the investor must be informed. In this case no further actions or trade occur because the investor's type and the asset's value are common knowledge.

However, if $|z_2| = 1$ the asset's value is uncertain to HFTs and trade will occur. At $t = 2$ the liquidity provider sends orders to all exchanges to post bid and ask prices. These prices persist until they are traded against or cancelled. At $t = 2 + \epsilon$ there are two potential traders of the asset: an uninformed investor and an informed investor sending orders to all X exchanges. If a price on exchange i is not traded against at $t = 2 + \epsilon$ that price can either be traded against or cancelled at $t = 2 + 2\epsilon$. At $t = 2 + 2\epsilon$ there are two potential traders of the asset: an informed investor sending orders to all X exchanges or a sniper who has learned the value of the asset.

The actions that occur in the second period are similar to the first period with one exception. Informed investors always send orders to all X exchanges in the second period. The second period is the final round of trading so there is no incentive for the informed investor to reduce their trading to conceal the value of the asset.

Similar to the first-period, the value of the asset is revealed if $|z_{2+\epsilon}| > 2$. Knowing the value of the asset, HFTs acting as snipers respond by sending buy or sell orders to all exchanges with outstanding prices. As in the first period, snipers are able to execute X_S orders in expectation while the informed investor is able to execute X_I orders. An equivalent version of Proposition 3.4.1—describing sniper's actions when $|z_{2+\epsilon}| = 2$—exists for the second period as well.

Proposition 3.4.2. *If $|z_{2+\epsilon}| = 2$ it is strictly dominant for snipers not to act at $t = 2 + \epsilon$.*

Proof. I present the on-path case when $z_{2+\epsilon} = 2$ here and leave the off-path case as well as the case when $z_{2+\epsilon} = -2$ for Appendix 3.6. At $t = 2 + \epsilon$ snipers do not send buy orders to trade on the liquidity provider's outstanding ask prices if $a(1) > E[v|z_{2+\epsilon} = 2]$. The expected asset value does not depend on the sniper's actions and can be calculated using Bayes rule as

$$E[v|z_{2+\epsilon} = 2] = P(v = 1|z_{2+\epsilon} = 2) = \frac{\frac{1}{4}q(1-\mu) + q\mu(1-\alpha)\rho^{X-1}}{\frac{1}{4}(1-\mu) + q\mu(1-\alpha)\rho^{X-1}} \tag{3.25}$$

where $\frac{1}{4}q(1-\mu)$ is the probability that $v = 1$ and $z_{2+\epsilon} = 2$ because the investor is uninformed. $q\mu(1-\alpha)\rho^{X-1}$ is the probability $v = 1$ and $z_{2+\epsilon} = 2$ because the investor is informed and sent a single order in the first period.

There are infinite HFTs competing in Bertrand competition to set the ask price at $t = 2$ so the ask price $a(1)$ is determined by a zero profit condition. If snipers do not send orders when $z_{2+\epsilon} = 2$ then the zero profit condition is:

$$\begin{aligned}0 =&(a(1)-1)P(v=1 \text{ and uninformed and } z_{2+\epsilon}=2|z_2=1)\\&+(a(1)-0)P(v=0 \text{ and uninformed and } z_{2+\epsilon}=2|z_2=1)\\&+(a(1)-1)P(v=1 \text{ and informed sent 1 } |z_2=1)(X_I+X_S)\\=&(a(1)-1)q\frac{\frac{1}{4}(1-\mu)}{\frac{1}{4}(1-\mu)+q\mu(1-\alpha)}+a(1)(1-q)\frac{\frac{1}{4}(1-\mu)}{\frac{1}{4}(1-\mu)+q\mu(1-\alpha)}\\&+(a(1)-1)\frac{q\mu(1-\alpha)}{\frac{1}{4}(1-\mu)+q\mu(1-\alpha)}(X-\frac{1}{2}(X-1)\rho^{X-1})\end{aligned} \tag{3.26}$$

This implies that the ask price $a(1)$ is

$$a(1)=\frac{\frac{1}{4}q(1-\mu)+q\mu(1-\alpha)(X-\frac{1}{2}(X-1)\rho^{X-1})}{\frac{1}{4}(1-\mu)+q\mu(1-\alpha)(X-\frac{1}{2}(X-1)\rho^{X-1})} \tag{3.27}$$

The sniper prefers not send orders when $z_{2+\epsilon}=2$ if $a(1)>E[v|z_{2+\epsilon}=2]$. Using Equations 3.25 and 3.27 we can show that the sniper does not send orders when:

$$X-\frac{1}{2}(X-1)\rho^{X-1} > \rho^{X-1} \tag{3.28}$$

This condition holds true because $X \geq 2$ and $\rho \in [0, \frac{1}{2}]$. Thus snipers do not send buy orders at $t=2+\epsilon$ when $z_{2+\epsilon}=2$. □

Similar to the first period we can summarize the HFTs' actions that occur in the second period when $|z_2|=1$ as follows:

- $t=2$: One HFT acts as a liquidity provider by sending orders to all exchanges to set bid and ask prices.
- $t=2+\epsilon$: The investor's orders arrive and execute against the bid or ask prices.
  - If only one exchange receives an order ($|z_{2+\epsilon}|=2$) snipers do nothing and the liquidity provider sends orders to cancel all outstanding bid and ask prices.
  - If more than one exchange receives an order ($|z_{2+\epsilon}|>2$) snipers respond by sending buy or sell orders to all exchanges with outstanding prices. Liquidity provider also respond

by sending orders to cancel all outstanding bid and ask prices.

3.4.4 Equilibrium Characterization

Returning to the problem of the informed investor, we can now use equations for the bid and ask prices at $t = 1$ (Equations 3.23, 3.36) and at $t = 2$ (Equations 3.27, 3.46) to solve for the indifference condition at $t = 1$ between sending one and X orders that we previously represented in Equations (3.5, 3.6).

Proposition 3.4.3. *The probability of sending X buy orders α^* and X sell orders β^* that solve the informed investor's indifference condition (Equations 3.5 and 3.6) given ask prices $a(0), a(1)$ and bid prices $b(0), b(-1)$ (Equations 3.23, 3.27, 3.36, and 3.46) are*

$$\alpha^* = \frac{q\mu(X_I + X_S)(X_I - 1) - \frac{1}{2}q\mu X_I - \frac{1}{4}(1-\mu)}{q\mu(X_I + X_S)(\frac{3}{2}X_I - 1) - \frac{1}{2}q\mu X_I} \quad \text{when } (X_I + X_S)(X_I - 1) - \frac{1}{2}X_I > \frac{1-\mu}{4q\mu} \tag{3.29}$$

$$\beta^* = \frac{(1-q)\mu(X_I + X_S)(X_I - 1) - \frac{1}{2}(1-q)\mu X_I - \frac{1}{4}(1-\mu)}{(1-q)\mu(X_I + X_S)(\frac{3}{2}X_I - 1) - \frac{1}{2}(1-q)\mu X_I} \quad \text{when } (X_I + X_S)(X_I - 1) - \frac{1}{2}X_I > \frac{1-\mu}{4(1-q)\mu} \tag{3.30}$$

Corollary 3.4.1. $\lim_{X\to\infty} \alpha^*, \beta^* = \frac{2}{3}$

Proofs of all Propositions and Corollaries are provided in Appendix 3.6.

A mixed strategy equilibrium arises from the strategic tension between increasing profits in period 1 by sending orders to all exchanges and preserving the ability to profit in period 2 by sending an order to a single exchange. The relative benefit to both strategies depends primarily on how prices change between periods. Sending X orders in the first period results in X_I total trades, all of which occur in the first period. Sending a single order in the first period increases the total number of trades to $X_I + 1$, but X_I of these trades occur in the second period. As a result, if prices were the same across periods the investor would prefer to send a single order at $t = 1$ because the investor can execute more total trades. However if prices worsen dramatically between periods (prices rise after a buy order or fall after a sell order), the benefit of executing an additional trade—by sending a single order at $t = 1$—is not worth the cost that most trades execute at worse prices.

The investor's probability of sending X orders (α^*, β^*) in Proposition 3.4.3 balances a reduction in the expected total trades ($X_I + 1 - \alpha^*$) with "better" prices when $| z_2 |= 1$. Increasing α, β makes prices at $t = 2$ ($a(1), b(-1)$) more attractive to the informed investor. As α, β increase, the less likely it is that the investor is informed at $t = 2$ when $| z_2 |= 1$. This reduces the liquidity provider's adverse selection risk, and results in prices $a(1), b(-1)$ that are closer to q (which increases the informed investor's profits).

Corollary 3.4.2. *There is no pure strategy equilibrium where the informed investor sends orders to all exchanges in the first period ($\alpha^* = 1, \beta^* = 1$).*

Taking this logic to one extreme gives Corollary 3.4.2. If $\alpha = 1$ the informed investor always sends X orders at $t = 1$ and in expectation earns profits $X_I(1 - a(0))$. By always sending X orders at $t = 1$ the liquidity provider is certain at $t = 2$ that the investor is uninformed if $z_2 = 1$. The liquidity provider faces no adverse selection risk and sets $a(1) = q$. But, the informed investor's would rather trade at $t = 2$ than at $t = 1$ because $a(1) = q < a(0)$.[12] This causes the informed investor to deviate from their pure strategy by sending one order at $t = 1$ and X orders at $t = 2$. When deviating the informed investor can expect to earn profits $1 - a(0) + X_I(1 - a(1))$ which is an improvement over expected profit when $\alpha = 1$.

Corollary 3.4.3. *If $(X_I + X_S)(X_I - 1) - \frac{1}{2}X_I \leq \frac{1-\mu}{4q\mu}$ a high-type informed investor plays a pure strategy equilibrium of always sending one buy order at $t = 1$ ($\alpha^* = 0$).*

If $(X_I + X_S)(X_I - 1) - \frac{1}{2}X_I \leq \frac{1-\mu}{4(1-q)\mu}$ a low-type informed investor plays a pure strategy equilibrium of always sending one sell order at $t = 1$ ($\beta^* = 0$).

If either the probability the investor is informed (μ) or the probability the asset value is 1 (q) are sufficiently low then a high-type investor always sends an order to only one exchange at $t = 1$. When μ and/or q are low it is unlikely that the investor is high-type, so the liquidity provider's adverse selection risk when setting ask prices is low. As a result, the liquidity provider will set ask prices close

[12] $q < a(0)$ becasue at $t = 1$ the liquidity provider charges a positive bid-ask aspread to compensate for the risk that the investor is informed.

to q as long as the investor's type is concealed. Prices closer to q are more attractive to an informed investor, so the investor will act to conceal their type. The result is Corollary 3.4.3, that the informed investor will conceal their identity for as long as possible by only trading on a single exchange.

An implication of Corollary 3.4.3 is that when an investor has more valuable information (because the prior expected value, q, is further from the actual value) or an investor is less likely to have information (μ is low), the more the investor wants to conceal their information by trading on fewer exchanges. This implies that more surprising information is revealed more slowly.[13] More surprising information is more valuable and investors with valuable information want to maximize their ability to profit by trading on fewer exchanges and delaying the time until that information becomes common knowledge.

A related finding can be seen by varying the trading speed of HFTs relative to investors, ρ.[14] As the relative speed of HFTs increases it becomes harder for investors to profit from new information. Investors respond, as seen in Proposition 3.4.4, by reducing their trading intensity to preserve their ability to profit in the future.

Proposition 3.4.4. *In the first period, the probability of sending orders to all exchanges α^*, β^* and the bid-ask spread are weakly decreasing in the speed of the HFTs, ρ.*

When HFTs become faster relative to investors, then investors trade less aggressively (by sending orders to fewer exchanges) and bid-ask spreads become tighter. Faster HFTs can better engage in order anticipation which reduces the number of orders executed by investors trading on multiple exchanges. When facing faster HFTs, investor's profits from sending orders to all exchanges are lower, incentivizing them to spread their orders out over time to conceal the true value of the asset and avoid order anticipation. In the model this can be seen by investors reducing the probability of sending orders to all exchanges in the first period (α^*, β^*).

[13] I use the word surprising to capture both that the investor's information is less likely to exist (because the probability the investor is informed, μ, is low) and that the investor's information is less likely to be true (because in the case of a high-type investor the actual asset value, 1, is further from the prior expected value, q).

[14] Recall that we defined the probability an investor's order receives long latency (2ϵ) as ρ. HFTs orders, meanwhile, always receive short latency (ϵ). This allows us to interpret an increase in ρ as an increase in HFTs' speed relative to informed investors.

An increase in HFTs' relative speed reduces bid-ask spreads through the combination of two forces. Increasing HFTs' speed increases their ability to engage in passive order anticipation—the ability of liquidity providers to cancel outstanding prices before investors can trade—which reduces the risk of setting bid and ask prices. Passive order anticipation reduces the number of exchanges on which liquidity providers incur losses, reducing risk, and decreasing the spread liquidity providers need to charge. In addition to a reduction in liquidity provider's risk from increased passive order anticipation, less aggressive trading by informed investors also reduces liquidity provider's adverse selection risk in the first period. Liquidity providers are less likely to trade with and incur losses from informed investors who trade on fewer exchanges. A decrease in liquidity provider's risk from both more passive order anticipation and from less aggressive trading by informed investors allows liquidity providers to charge smaller bid-ask spreads, increasing market liquidity.

While smaller bid-ask spreads improve market liquidity, less aggressive trading by informed investors has negative implications for the spread of new information.

Corollary 3.4.4. *Increasing the relative speed of HFTs, ρ, delays the expected time when the asset's value becomes common knowledge.*

An increase in HFTs' speed relative to informed investors reduces the informed investor's ability to profit once the investor's information becomes common knowledge. In the model, the investor's information becomes common knowledge when the investor sends orders to multiple exchanges. Investors respond to HFTs' increased speed by reducing their probability of sending orders to all exchanges (Proposition 3.4.4). By trading on fewer exchanges, investors preserve their ability to profit by delaying the time until their new information becomes common knowledge (Corollary 3.4.4). The consequence of this is that as HFTs grow faster relative to investors it takes longer for investor's information to become public.

As a result of Proposition 3.4.4 and Corollary 3.4.4 we can view the impact of HFTs on the spread of information as mixed.[15] When information is already common knowledge HFTs' speed advantage

[15] The results that HFTs decrease bid-ask spreads and increase the time until new information is common knowledge can be seen in greater contrast by considering a version of the model without HFTs. This can be accomplished by setting $\rho = 0$ so that investors and HFTs have the same latency. In this case, when an informed investor sends orders to all exchanges the orders arrive simultaneously with latency ϵ. There is no opportunity for order anticipation because no

ensures that prices better reflect this information through tighter bid-ask spreads. However when information is not widely known, HFTs cause that information to spread more slowly and delay the time until new information becomes common knowledge. The result is that any policies that might impact the speed difference between market participants need to balance the impacts on how prices reflect both new and existing information.

3.5 Conclusion

This chapter uses a simple, two-period model to find that HFTs improve market liquidity through smaller bid-ask spreads but delay the time until new information becomes common knowledge. One limitation of the model's two-period structure is that the informed investor always trades on all exchanges in the second period, revealing the value of the asset. This is due to the effect of the game ending rather than any strategic tensions. If the model were extended to a multi-period structure this would allow us to make claims about how informed investors trade more or less aggressively over time as information is slowly revealed. In a multi-period model the informed investor would continue to face a trade-off between trading on more exchanges to increase profits in the current period and trading on fewer exchanges to preserve their private information and future profits. In a multi-period model the informed investor would have an even greater incentive in the first period to trade on fewer exchanges because future profits—when the asset's value remains private information—can be earned over a longer period of time. Extending this chapter's model to multiple periods is a valuable extension in order to more fully examine these and other effects on the spread of new information.

This chapter's finding that HFTs decrease bid-ask spreads is similar to the findings in Baldauf and Mollner (2020) who consider a similar, static model. However unlike this chapter, Baldauf and Moll-

---

orders are delayed and there are no stale prices to either be exploited by snipers or cancelled by market makers.

In a model without HFTs the informed investor may still play a mixed equilibrium because there is a strategic trade-off between trading more in the first period and trading at better prices in the second period. Compared to a setting with HFTs ($\rho > 0$) the informed investor trades on more exchanges in the first period. Removing HFTs results in greater profits from sending orders to all exchanges because there is no longer order anticipation. Without HFTs first-period bid-ask spreads are wider, because if an informed investor chooses to trade on all exchanges the liquidity provider will incur losses on every exchange. Finally, Corollary 4.4 tells us that in a model without HFTs new information becomes common knowledge faster. Overall, markets without HFTs are characterized by less liquidity (wider bid-ask spreads) and the faster spread of new information.

ner (2020) consider a static model and are unable to study how quickly new information spreads. Instead, Baldauf and Mollner (2020) use their model to show that faster HFTs discourage costly research by investors. Both the results in this chapter, that HFTs increase in the time for new information to become common knowledge, and Baldauf and Mollner (2020)'s findings, that HFTs reduce costly research, result in less public dissemination of new, valuable information. The intuition behind both findings is similar too, in both this chapter and in Baldauf and Mollner (2020) faster HFTs engage in more order anticipation reducing investor's ability to profit from new information. The impact of HFTs is that in Baldauf and Mollner (2020) investors engage in less costly research to learn new information and in this chapter that investors trade less aggressively to preserve their private information.

Overall, this chapter finds a mixed effect of HFTs on the spread of new information. HFTs benefit the marketplace by reducing the adverse selection risk of providing liquidity. When the liquidity provider's speed increases bid-ask spreads are smaller resulting in more liquid markets and prices that better reflect common knowledge information. However, HFTs are also detrimental to the spread of new information. When HFTs speed increases this reduces the ability for investors to profit from new information. Investors respond by trading in a manner that conceals their new information and delays the time until new information becomes common knowledge. Without placing weights on the value of prices to reflect both new and existing information it is not possible to conclude if HFTs are categorically good or bad for financial markets. Rather this chapter's main contribution is to study the mechanisms through which information is incorporated into prices. New information is only discovered and made public when it is profitable to do so. HFTs reduce the profitability of new information and, as a result, slow the revelation of new information.

3.6 Appendix: Proof of Propositions and Corollaries

3.6.1 Propsition 3.4.1

In the main text I showed that when the liquidity provider believes that snipers do not act when $z_{1+\epsilon} = 1$ it is optimal for snipers not to act. Here I show the case that when the liquidity provider believes that snipers will send buy orders to all exchanges when $z_{1+\epsilon} = 1$ it is also optimal for snipers not to act.

If snipers send buy orders when $z_{1+\epsilon} = 1$ then at $t = 1+\epsilon$ snipers send buy orders to all exchanges whenever $z_{1+\epsilon} > 0$. This ensures that regardless of the investor's information if the investor sends buy orders at $t = 1$ then snipers will send buy orders to all exchanges at $t = 1 + \epsilon$. The result is that all X of the liquidity provider's ask price are executed against by the end of $t = 1 + 2\epsilon$. The liquidity provider's zero profit condition when setting ask prices $a(0)$ under this set of beliefs is

$$\begin{aligned}
0 =&[(a(0)-1)P(v=1)P(\text{uninformed and } z_{1+\epsilon}=1) \\
&+ (a(0)-0)P(v=0)P(\text{uninformed and } z_{1+\epsilon}=1) \\
&+ (a(0)-1)P(v=1)P(\text{informed})]X \\
=&[(a(0)-1)\frac{1}{2}q(1-\mu) + (a(0)-0)\frac{1}{2}(1-q)(1-\mu) + (a(0)-1)q\mu]X
\end{aligned} \tag{3.31}$$

This implies that the ask price $a(0)$ is

$$a(0) = \frac{\frac{1}{2}q(1-\mu) + q\mu}{\frac{1}{2}(1-\mu) + q\mu} \tag{3.32}$$

Snipers do not act if $a(0) > E[v|z_{1+\epsilon} = 1]$. Using Equations 3.21 and 3.32 the sniper does not send orders when:

$$1 > \rho^{X-1} \tag{3.33}$$

which is true because $X \geq 2$ and $\rho \in [0, \frac{1}{2}]$.

As a result, regardless of the liquidity providers beliefs about snipers' actions when $z_{1+\epsilon} = 1$ it is always strictly dominant for the snipers not to act.

We can show that similar inequalities hold when $z_{1+\epsilon} = -1$. In this case the sniper is deciding between not acting and sending sell orders to all exchanges. Snipers prefer not to send sell orders when $E[v|z_{1+\epsilon} = -1] > b(0)$. We can calculate $E[v|z_{1+\epsilon} = -1]$ as

$$E[v|z_{1+\epsilon} = -1] = P(v = 1|z_{1+\epsilon} = -1) = \frac{\frac{1}{2}q(1-\mu)}{\frac{1}{2}(1-\mu) + (1-q)\mu(1-\beta) + (1-q)\mu\beta\rho^{X-1}} \tag{3.34}$$

If the liquidity provider believes that snipers will not send sell orders when $z_{1+\epsilon} = -1$ then the liquidity provider's zero profit condition when setting bid prices $b(0)$ is

$$\begin{aligned} 0 =&(1-b(0))P(v=1)P(\text{uninformed and } z_{1+\epsilon} = -1) \\ &+ (0-b(0))P(v=0)P(\text{uninformed and } z_{1+\epsilon} = -1) \\ &+ (0-b(0))P(v=0)P(\text{informed sending } 1) \\ &+ (0-b(0))P(v=0)P(\text{informed sending } X)(X_I + X_S) \\ =&(1-b(0))\frac{1}{2}q(1-\mu) - b(0)\frac{1}{2}(1-q)(1-\mu) \\ &- b(0)(1-q)\mu(1-\alpha) - b(0)(1-q)\mu\alpha(X - \frac{1}{2}(X-1)\rho^{X-1}) \end{aligned} \tag{3.35}$$

This implies that the bid price $b(0)$ is

$$b(0) = \frac{\frac{1}{2}q(1-\mu)}{\frac{1}{2}(1-\mu) + (1-q)\mu(1-\beta) + (1-q)\mu\beta(X - \frac{1}{2}(X-1)\rho^{X-1})} \tag{3.36}$$

The sniper prefers not send sell orders when $z_{1+\epsilon} = -1$ if $E[v|z_{1+\epsilon} = -1] > b(0)$. Using Equations 3.34 and 3.36 the sniper does not send orders when:

$$X - \frac{1}{2}(X-1)\rho^{X-1} > \rho^{X-1} \tag{3.37}$$

which is true because $X \geq 2$ and $\rho \in [0, \frac{1}{2}]$.

If instead the liquidity provider believed that snipers send sell orders when $z_{1+\epsilon} = -1$ then at

$t = 1+\epsilon$ snipers send sell orders to all exchanges whenever $z_{1+\epsilon} < 0$. This ensures that regardless of the investor's information if the investor sends sell orders at $t = 1$ then snipers will send sell orders to all exchanges at $t = 1 + \epsilon$. The result is that all X of the liquidity provider's bid price are executed against by the end of $t = 1+2\epsilon$. The liquidity provider's zero profit condition when setting bid prices $b(0)$ under this set of beliefs is

$$\begin{aligned}
0 =& [(1 - b(0))P(v = 1)P(\text{uninformed and } z_{1+\epsilon} = -1) \\
&+ (0 - b(0))P(v = 0)P(\text{uninformed and } z_{1+\epsilon} = -1) \\
&+ (1 - b(0))P(v = 0)P(\text{informed})]X \\
=& [(1 - b(0))\frac{1}{2}q(1 - \mu) - b(0)\frac{1}{2}(1 - q)(1 - \mu) - b(0)(1 - q)\mu]X
\end{aligned} \tag{3.38}$$

This implies that the bid price $b(0)$ is

$$b(0) = \frac{\frac{1}{2}q(1 - \mu)}{\frac{1}{2}(1 - \mu) + (1 - q)\mu} \tag{3.39}$$

Snipers do not act if $E[v|z_{1+\epsilon} = -1] > b(0)$. Using Equations 3.34 and 3.39 the sniper does not send orders when:

$$1 > \rho^{X-1} \tag{3.40}$$

which is true because $X \geq 2$ and $\rho \in [0, \frac{1}{2}]$.

3.6.2 Propsition 3.4.2

In the main text I showed that when the liquidity provider believes that snipers do not act when $z_{2+\epsilon} = 2$ it is optimal for snipers not to act. Here I show the case that when the liquidity provider believes that snipers will send buy orders to all exchanges when $z_{2+\epsilon} = 2$ it is also optimal for snipers not to act.

If snipers send buy orders when $z_{2+\epsilon} = 2$ then at $t = 2+\epsilon$ snipers send buy orders to all exchanges whenever $z_{2+\epsilon} > 1$. This ensures that regardless of the investor's information if the investor sends buy orders at $t = 2$ then snipers will send buy orders to all exchanges at $t = 2 + \epsilon$. The result is that all X of the liquidity provider's ask price are executed against by the end of $t = 2 + 2\epsilon$. The liquidity provider's zero profit condition when setting ask prices $a(1)$ under this set of beliefs is

$$\begin{aligned}
0 =&[(a(1) - 1)P(v = 1 \text{ and uninformed and } z_{2+\epsilon} = 2|z_2 = 1) \\
&+ (a(1) - 0)P(v = 0 \text{ and uninformed and } z_{2+\epsilon} = 2|z_2 = 1) \\
&+ (a(1) - 1)P(v = 1 \text{ and informed sent } 1\ |z_2 = 1)]X \\
=&[(a(1) - 1)q\frac{\frac{1}{4}(1-\mu)}{\frac{1}{4}(1-\mu) + q\mu(1-\alpha)} + a(1)(1-q)\frac{\frac{1}{4}(1-\mu)}{\frac{1}{4}(1-\mu) + q\mu(1-\alpha)} \\
&+ (a(1) - 1)\frac{q\mu(1-\alpha)}{\frac{1}{4}(1-\mu) + q\mu(1-\alpha)}]X
\end{aligned} \tag{3.41}$$

This implies that the ask price $a(1)$ is

$$a(1) = \frac{\frac{1}{4}q(1-\mu) + q\mu(1-\alpha)}{\frac{1}{4}(1-\mu) + q\mu(1-\alpha)} \tag{3.42}$$

Snipers do not act if $a(1) > E[v|z_{2+\epsilon} = 2]$. Using Equations 3.25 and 3.42 the sniper does not send orders when:

$$1 > \rho^{X-1} \tag{3.43}$$

which is true because $X \geq 2$ and $\rho \in [0, \frac{1}{2}]$.

As a result, regardless of the liquidity providers beliefs about snipers' actions when $z_{2+\epsilon} = 2$ it is always strictly dominant for the snipers not to act.

We can show that similar inequalities hold when $z_{2+\epsilon} = -2$. In this case the sniper is deciding between not acting and sending sell orders to all exchanges. Snipers prefer not to send sell orders when $E[v|z_{2+\epsilon} = -2] > b(-1)$. We can calculate $E[v|z_{1+\epsilon} = -2]$ as

$$E[v|z_{2+\epsilon} = -2] = P(v = 1|z_{2+\epsilon} = -2) = \frac{\frac{1}{4}q(1-\mu)}{\frac{1}{4}(1-\mu) + (1-q)\mu(1-\beta)\rho^{X-1}} \tag{3.44}$$

If the liquidity provider believes that snipers will not send sell orders when $z_{2+\epsilon} = -2$ then the liquidity provider's zero profit condition when setting bid prices $b(-1)$ is

$$\begin{aligned}
0 =&(1-b(-1))P(v=1 \text{ and uninformed seller}|z_2=-1)\\
&+(0-b(-1))P(v=0 \text{ and uninformed seller}|z_2=-1)\\
&+(0-b(-1))P(v=0 \text{ and informed sent } 1|z_2=-1)(X_I+X_S)\\
=&(1-b(-1))\frac{\frac{1}{4}q(1-\mu)}{\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta)}-b(-1)\frac{\frac{1}{4}(1-q)(1-\mu)}{\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta)}\\
&-b(-1)\frac{(1-q)\mu(1-\beta)}{\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta)}(X-\frac{1}{2}(X-1)\rho^{X-1})
\end{aligned} \tag{3.45}$$

This implies that the bid price $b(-1)$ is

$$b(-1)=\frac{\frac{1}{4}q(1-\mu)}{\frac{1}{4}(1-\mu)+(1-q)\mu(1-\beta)(X-\frac{1}{2}(X-1)\rho^{X-1})} \tag{3.46}$$

The sniper prefers not send sell orders when $z_{2+\epsilon}=-2$ if $E[v|z_{2+\epsilon}=-2]>b(-1)$. Using Equations 3.44 and 3.46 the sniper does not send orders when:

$$X-\frac{1}{2}(X-1)\rho^{X-1}>\rho^{X-1} \tag{3.47}$$

which is true because $X\geq 2$ and $\rho\in[0,\frac{1}{2}]$.

If instead the liquidity provider believed that snipers send sell orders when $z_{2+\epsilon}=-2$ then at $t=2+\epsilon$ snipers send sell orders to all exchanges whenever $z_{2+\epsilon}<-1$. This ensures that regardless of the investor's information if the investor sends sell orders at $t=2$ then snipers will send sell orders to all exchanges at $t=2+\epsilon$. The result is that all X of the liquidity provider's bid price are executed against by the end of $t=2+2\epsilon$. The liquidity provider's zero profit condition when setting bid prices

$b(-1)$ under this set of beliefs is

$$\begin{aligned}
0 =& [(1-b(-1))P(v=1 \text{ and uninformed seller}|z_2=-1) \\
&+ (0-b(-1))P(v=0 \text{ and uninformed seller}|z_2=-1) \\
&+ (0-b(-1))P(v=0 \text{ and informed sent } 1|z_2=-1)]X \\
=& \left[(1-b(-1))\frac{\frac{1}{4}q(1-\mu)}{\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta)} - b(-1)\frac{\frac{1}{4}(1-q)(1-\mu)}{\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta)}\right. \\
&\left. - b(-1)\frac{(1-q)\mu(1-\beta)}{\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta)}\right]X
\end{aligned} \tag{3.48}$$

This implies that the bid price $b(-1)$ is

$$b(-1) = \frac{\frac{1}{4}q(1-\mu)}{\frac{1}{4}(1-\mu)+(1-q)\mu(1-\beta)} \tag{3.49}$$

Snipers do not act if $E[v|z_{2+\epsilon}=-2] > b(-1)$. Using Equations 3.44 and 3.49 the sniper does not send orders when:

$$\begin{aligned}
\frac{\frac{1}{4}q(1-\mu)}{\frac{1}{4}(1-\mu)+(1-q)\mu(1-\beta)\rho^{X-1}} >& \frac{\frac{1}{4}q(1-\mu)}{\frac{1}{4}(1-\mu)+(1-q)\mu(1-\beta)} \\
\implies 1 >& \rho^{X-1}
\end{aligned} \tag{3.50}$$

which is true because $X \geq 2$ and $\rho \in [0, \frac{1}{2}]$.

3.6.3 Proposition 3.4.3

Recall the indifference conditions to be solved are:

$$\frac{1-a(0)}{a(1)-a(0)} = X_I$$

defines α^* and

$$\frac{b(0)}{b(0)-b(-1)} = X_I$$

defines β^*.

Starting with the condition for α we can use Equations 3.23 and 3.27 that define the ask prices.

$$\begin{aligned}
1-a(0) =& \frac{\frac{1}{2}(1-q)(1-\mu)}{\frac{1}{2}(1-\mu)+q\mu(1-\alpha)+q\mu\alpha(X_I+X_S)} \\
a(1)-a(0) =& \frac{\frac{1}{4}q(1-\mu)+q\mu(1-\alpha)(X_I+X_S)}{\frac{1}{4}(1-\mu)+q\mu(1-\alpha)(X_I+X_S)} - \frac{\frac{1}{2}q(1-\mu)+q\mu(1-\alpha)+q\mu\alpha(X_I+X_S)}{\frac{1}{2}(1-\mu)+q\mu(1-\alpha)+q\mu\alpha(X_I+X_S)} \\
=& \frac{\frac{1}{2}(1-\mu)\mu(1-q)q\left[(1-\alpha)(X_I+X_S)-\frac{1}{2}(1-\alpha)-\frac{1}{2}\alpha(X_I+X_S)\right]}{\left[\frac{1}{4}(1-\mu)+q\mu(1-\alpha)(X_I+X_S)\right]*\left[\frac{1}{2}(1-\mu)+q\mu(1-\alpha)+q\mu\alpha(X_I+X_S)\right]}
\end{aligned}$$

We can then solve for α^* using the indifference condition in Equation 3.5:

$$\frac{1}{2}(1-q)(1-\mu) = \frac{\frac{1}{2}(1-\mu)\mu(1-q)q\left[(1-\alpha)(X_I+X_S)-\frac{1}{2}(1-\alpha)-\frac{1}{2}\alpha(X_I+X_S)\right]}{\left[\frac{1}{4}(1-\mu)+q\mu(1-\alpha)(X_I+X_S)\right]}X_I$$

Canceling terms gives the expression:

$$\frac{1}{4}(1-\mu) = q\mu(1-\alpha)(X_I+X_S)(X_I-1) - \frac{1}{2}q\mu(1-\alpha)X_I - \frac{1}{2}q\mu\alpha(X_I+X_S)X_I$$

Which when solved for α gives the expression:

$$\alpha^* = \frac{q\mu(X_I+X_S)(X_I-1)-\frac{1}{2}q\mu X_I-\frac{1}{4}(1-\mu)}{q\mu(X_I+X_S)(\frac{3}{2}X_I-1)-\frac{1}{2}q\mu X_I}$$

Now turning to the indifference condition for β we can use Equations 3.36 and 3.46 that define the bid prices.

$$\begin{aligned} b(0) =& \frac{\frac{1}{2}q(1-\mu)}{\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta)+(1-q)\mu\beta(X_I+X_S)} \\ b(0)-b(-1) =& \frac{\frac{1}{2}q(1-\mu)}{\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta)+(1-q)\mu\beta(X_I+X_S)} \\ & - \frac{\frac{1}{4}q(1-\mu)}{\frac{1}{4}(1-\mu)+(1-q)\mu(1-\beta)(X_I+X_S)} \end{aligned}$$

We can then solve for β^* using the indifference condition in Equation 3.6:

$$\begin{aligned} & \frac{\frac{1}{2}q(1-\mu)}{\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta)+(1-q)\mu\beta(X_I+X_S)} \\ =& \frac{\left[\frac{1}{2}q(1-\mu)\right] * \left[\frac{1}{4}(1-\mu)+(1-q)\mu(1-\beta)(X_I+X_S)\right]X_I}{\left[\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta)+(1-q)\mu\beta(X_I+X_S)\right] * \left[\frac{1}{4}(1-\mu)+(1-q)\mu(1-\beta)(X_I+X_S)\right]} \\ & - \frac{\left[\frac{1}{4}q(1-\mu)\right] * \left[\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta)+(1-q)\mu\beta(X_I+X_S)\right]X_I}{\left[\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta)+(1-q)\mu\beta(X_I+X_S)\right] * \left[\frac{1}{4}(1-\mu)+(1-q)\mu(1-\beta)(X_I+X_S)\right]} \end{aligned}$$

Canceling terms gives the expression:

$$X_I - 1 = \frac{\frac{1}{2}\left[\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta)+(1-q)\mu\beta(X_I+X_S)\right]X_I}{\left[\frac{1}{4}(1-\mu)+(1-q)\mu(1-\beta)(X_I+X_S)\right]}$$

Which when solved for β gives the expression:

$$\beta^* = \frac{(1-q)\mu(X_I+X_S)(X_I-1)-\frac{1}{2}(1-q)\mu X_I-\frac{1}{4}(1-\mu)}{(1-q)\mu(X_I+X_S)(\frac{3}{2}X_I-1)-\frac{1}{2}(1-q)\mu X_I}$$

3.6.4 Corollary 3.4.1

Repeated application of L'Hôpital's rule gives the desired result.

3.6.5 Corollary 3.4.2

A high-type informed investor sends X buy orders when $\alpha \geq 1$:

$$\frac{q\mu(X_I + X_S)(X_I - 1) - \frac{1}{2}q\mu X_I - \frac{1}{4}(1-\mu)}{q\mu(X_I + X_S)(\frac{3}{2}X_I - 1) - \frac{1}{2}q\mu X_I} \geq 1$$
$$-\frac{1}{4}(1-\mu) \geq q\mu(X_I + X_S)\frac{1}{2}X_I$$

Which never holds true.

A low-type informed investor sends X buy orders when $\beta \geq 1$:

$$\frac{(1-q)\mu(X_I + X_S)(X_I - 1) - \frac{1}{2}(1-q)\mu X_I - \frac{1}{4}(1-\mu)}{(1-q)\mu(X_I + X_S)(\frac{3}{2}X_I - 1) - \frac{1}{2}(1-q)\mu X_I} \geq 1$$
$$-\frac{1}{4}(1-\mu) \geq (1-q)\mu(X_I + X_S)\frac{1}{2}X_I$$

Which never holds true.

3.6.6 Corollary 3.4.3

A high-type informed investor will send one buy order when $\alpha \leq 0$:

$$\frac{q\mu(X_I + X_S)(X_I - 1) - \frac{1}{2}q\mu X_I - \frac{1}{4}(1-\mu)}{q\mu(X_I + X_S)(\frac{3}{2}X_I - 1) - \frac{1}{2}q\mu X_I} \leq 0$$
$$q\mu(X_I + X_S)(X_I - 1) - \frac{1}{2}q\mu X_I - \frac{1}{4}(1-\mu) \leq 0$$
$$4q\mu(X_I + X_S)(X_I - 1) \leq 2q\mu X_I + (1-\mu)$$
$$(X_I + X_S)(X_I - 1) \leq \frac{1}{2}X_I + \frac{(1-\mu)}{4q\mu}$$
$$(X_I + X_S)(X_I - 1) - \frac{1}{2}X_I \leq \frac{(1-\mu)}{4q\mu}$$

A low-type informed investor will send one buy order when $\beta \leq 0$:

$$\begin{aligned}
\frac{(1-q)\mu(X_I+X_S)(X_I-1)-\frac{1}{2}(1-q)\mu X_I-\frac{1}{4}(1-\mu)}{(1-q)\mu(X_I+X_S)(\frac{3}{2}X_I-1)-\frac{1}{2}(1-q)\mu X_I} \leq& 0\\
(1-q)\mu(X_I+X_S)(X_I-1)-\frac{1}{2}(1-q)\mu X_I-\frac{1}{4}(1-\mu) \leq& 0\\
4(1-q)\mu(X_I+X_S)(X_I-1) \leq& 2(1-q)\mu X_I+(1-\mu)\\
(X_I+X_S)(X_I-1) \leq& \frac{1}{2}X_I+\frac{(1-\mu)}{4(1-q)\mu}\\
(X_I+X_S)(X_I-1)-\frac{1}{2}X_I \leq& \frac{(1-\mu)}{4(1-q)\mu}
\end{aligned}$$

3.6.7 Proposition 3.4.4

I will fist show that the probability that a high-type investor sends orders to all exchanges (α^*) is decreasing in ρ. Recall from Equation 3.29 that α^* can be expressed as:

$$\begin{aligned}
\alpha^* =& \frac{q\mu(X_I+X_S)(X_I-1)-\frac{1}{2}q\mu X_I-\frac{1}{4}(1-\mu)}{q\mu(X_I+X_S)(\frac{3}{2}X_I-1)-\frac{1}{2}q\mu X_I}\\
X_I+X_S =& X-\frac{1}{2}(X-1)\rho^{X-1}\\
X_I =& 1+(X-1)(1-\rho)+\frac{1}{2}(X-1)\rho^{X-1}
\end{aligned}$$

We can show that the derivative $\frac{\partial\alpha^*}{\partial\rho} < 0$.

$$\frac{\partial \alpha^*}{\partial \rho} =$$

$$\frac{(2X-2)\rho^{X+2}\Big(\mu(2qX^2-1)+1\Big)-4\rho^4\Big(4qX^2\mu-3X(\mu-1)+\mu-1\Big)}{\frac{1}{2}q(X-1)\mu\Big(2\rho^{X+1}-6\rho^{X+2}+3\rho^{2X}-3X(-2\rho^{X+2}+\rho^{2X}+4\rho^3-4\rho^2)+4\rho^3\Big)^2}$$

$$+\frac{2(X-1)X\rho^{X+3}\Big(\mu(4q(X+1)-3)+3\Big)-4q(X-1)^3\mu\rho^{X+4}}{\frac{1}{2}q(X-1)\mu\Big(2\rho^{X+1}-6\rho^{X+2}+3\rho^{2X}-3X(-2\rho^{X+2}+\rho^{2X}+4\rho^3-4\rho^2)+4\rho^3\Big)^2}$$

$$+\frac{-6(X-1)^2\rho^{2X+1}\Big(\mu(2qX-1)+1\Big)+4q(X-2)(X-1)^2\mu\rho^{2X+2}+q(X-1)^3\mu\rho^{3X}}{\frac{1}{2}q(X-1)\mu\Big(2\rho^{X+1}-6\rho^{X+2}+3\rho^{2X}-3X(-2\rho^{X+2}+\rho^{2X}+4\rho^3-4\rho^2)+4\rho^3\Big)^2}$$

The denominator is clearly positive because the majority of terms are squared. So we are left so that that the numerator is negative:

$$\begin{aligned}&(2X-2)\rho^{X+2}\Big(\mu(2qX^2-1)+1\Big)-4\rho^4\Big(4qX^2\mu-3X(\mu-1)+\mu-1\Big)\\&+2(X-1)X\rho^{X+3}\Big(\mu(4q(X+1)-3)+3\Big)-4q(X-1)^3\mu\rho^{X+4}\\&-6(X-1)^2\rho^{2X+1}\Big(\mu(2qX-1)+1\Big)+4q(X-2)(X-1)^2\mu\rho^{2X+2}+q(X-1)^3\mu\rho^{3X}\leq 0\end{aligned}$$

We can rearrange the above expression to collect terms by X, q, and μ:

$$\begin{aligned}
0 \leq & X^3 q\mu \Bigg[-4\rho^{X+2} - 8\rho^{X+3} + 4\rho^{X+4} + 12\rho^{2X+1} - 4\rho^{2X+2} - \rho^{3X} \Bigg] \\
& + X^2 q\mu \Bigg[16\rho^4 + 4\rho^{X+2} - 12\rho^{X+4} - 24\rho^{2X+1} + 16\rho^{2X+2} + 3\rho^{3X} \Bigg] \\
& + Xq\mu \Bigg[8\rho^{X+3} + 12\rho^{X+4} + 12\rho^{2X+1} - 20\rho^{2X+2} \Bigg] \\
& + q\mu \Bigg[-\rho^{X+4} + 8\rho^{2X+2} - 2\rho^{3X} \Bigg] \\
& + X^2(1-\mu) \Bigg[-6\rho^{X+3} + 6\rho^{2X+1} \Bigg] \\
& + X(1-\mu) \Bigg[12\rho^4 - 2\rho^{X+2} + 6\rho^{X+3} - 12\rho^{2X+1} \Bigg] \\
& + (1-\mu) \Bigg[-4\rho^4 + 2\rho^{X+2} + 6\rho^{2X+1} \Bigg]
\end{aligned}$$

Finally, we can split the above expression into the following terms all of which are greater than or equal to 0 when $\rho \in [0, \frac{1}{2}]; X \geq 2$.

$$X^3 q\mu\left[3\rho^{2X+1}-4\rho^{2X+2}-\rho^{3X}\right]\left(=\rho^{2X+1}(3-4\rho-\rho^{X-1})>\rho^{2X+1}(3-2-\frac{1}{2^{X-1}})>0\right)$$

$$+X^2 q\mu\left[8\rho^4-4X\rho^{X+2}\right]\left(=4\rho^4(2-X\rho^{X-2})\geq 0\right)$$

$$+X^2 q\mu\left[4X\rho^{X+4}-8\rho^{X+4}\right]\left(\geq 0\right)$$

$$+X^2 q\mu\left[\rho^{X+2}-4\rho^{X+4}\right]\left(=\rho^{X+2}(1-4\rho^2)\geq 0\right)$$

$$+X^2 q\mu\left[3\rho^{X+2}-6\rho^{2X+1}\right]\left(=3\rho^{X+2}(1-2\rho^{X-1})\geq 0\right)$$

$$+X^2 q\mu\left[8\rho^4-8X\rho^{X+3}\right]\left(=8\rho^4(1-X\rho^{X-1})\geq 0\right)$$

$$+X^2 q\mu\left[8X\rho^{2X+1}-16\rho^{2X+1}\right]\left(=8\rho^{2X+1}(X-2)\geq 0\right)$$

$$+Xq\mu\left[4\rho^{X+3}-2X\rho^{2X+1}\right]\left(=2\rho^{X+3}(2-X\rho^{X-2})\geq 0\right)$$

$$+Xq\mu\left[12\rho^{2X+1}-20\rho^{2X+2}\right]\left(>(12\rho^{2X+1}-10\rho^{2X+1})>0\right)$$

$$+q\mu\left[2\rho^{2X+2}-2\rho^{3X}\right]\left(>2\rho^{2X+2}(1-\rho^{X-2})\geq 0\right)$$

$$+q\mu\left[X^3\rho^{2X+1}+16X^2\rho^{2X+2}+3X^2\rho^{3X}+4X\rho^{X+3}+11X\rho^{X+4}+(X-1)\rho^{X+4}+6\rho^{2X+2}\right]\left(>0\right)$$

$$+X(1-\mu)\left[6\rho^4-6X\rho^{X+3}\right]\left(=6\rho^4(1-X\rho^{X-1})\geq 0\right)$$

$$+X(1-\mu)\left[6\rho^{X+3}-6\rho^{2X+1}\right]\left(=6\rho^{X+3}(1-\rho^{X-2})\geq 0\right)$$

$$+X(1-\mu)\left[2\rho^4-2\rho^{X+2}\right]\left(=\rho^4(2-2\rho^{X-2})\geq 0\right)$$

$$+X(1-\mu)\left[3X\rho^{2X+1}-6\rho^{2X+1}\right]\left(=3\rho^{2X+1}(X-2)\geq 0\right)$$

$$+(1-\mu)\left[3X^2\rho^{2X+1}+2(X-2)\rho^4+2X\rho^4+2\rho^{X+2}+6\rho^{2X+1}\right]\left(>0\right)$$

The proof that β^* is decreasing in ρ is analogous because α^* and β^* are equivalent if q is replaced by $1-q$.

We can show that the first period bid-ask spread $(a(0)-b(0))$ is decreasing in ρ using the result that α^*, β^* are decreasing in ρ. Recall from Equation 3.23 that the first period ask price is:

$$a(0) = \frac{\frac{1}{2}q(1-\mu) + q\mu(1-\alpha^*) + q\mu\alpha^*(X - \frac{1}{2}(X-1)\rho^{X-1})}{\frac{1}{2}(1-\mu) + q\mu(1-\alpha^*) + q\mu\alpha^*(X - \frac{1}{2}(X-1)\rho^{X-1})}$$

The derivative $\frac{\partial a(0)}{\partial \alpha^*}$ is positive:

$$\frac{\partial a(0)}{\partial \alpha^*} = \frac{\frac{1}{2}(1-q)(1-\mu)q\mu\Big(X - 1 - \frac{1}{2}(X-1)\rho^{X-1}\Big)}{\Big(\frac{1}{2}(1-\mu) + q\mu(1-\alpha^*) + q\mu\alpha^*(X - \frac{1}{2}(X-1)\rho^{X-1})\Big)^2} > 0$$

So the derivative $\frac{\partial a(0)}{\partial \alpha^*}\frac{\partial \alpha^*}{\partial \rho}$ is negative.

Further the derivative $\frac{\partial a(0)}{\partial \rho}$ is negative:

$$\frac{\partial a(0)}{\partial \rho} = -\frac{\frac{1}{2}q\mu\alpha^*(X-1)^2\rho^{X-2})\Big(X - 1 - \frac{1}{2}(X-1)\rho^{X-1}\Big)}{\Big(\frac{1}{2}(1-\mu) + q\mu(1-\alpha^*) + q\mu\alpha^*(X - \frac{1}{2}(X-1)\rho^{X-1})\Big)^2} < 0$$

So the overall impact of ρ is to decrease the ask price $a(0)$.

We can show that ρ has the opposite impact on the bid price. Recall from Equation 3.36 that the first period bid price is:

$$b(0) = \frac{\frac{1}{2}q(1-\mu)}{\frac{1}{2}(1-\mu) + (1-q)\mu(1-\beta^*) + (1-q)\mu\beta^*(X - \frac{1}{2}(X-1)\rho^{X-1})}$$

The derivative $\frac{\partial b(0)}{\partial \beta^*}$ is negative:

$$\frac{\partial b(0)}{\partial \beta^*} = -\frac{\frac{1}{2}(1-q)q(1-\mu)\mu(X - 1 - \frac{1}{2}(X-1)\rho^{X-1})}{\Big(\frac{1}{2}(1-\mu) + (1-q)\mu(1-\beta^*) + (1-q)\mu\beta^*(X - \frac{1}{2}(X-1)\rho^{X-1})\Big)^2} < 0$$

So the derivative $\frac{\partial b(0)}{\partial \beta^*}\frac{\partial \beta^*}{\partial \rho}$ is positive.

Further the derivative $\frac{\partial b(0)}{\partial \rho}$ is positive:

$$\frac{\partial b(0)}{\partial \rho} = \frac{\frac{1}{4}(1-q)q(1-\mu)\mu\beta^*(X-1)^2\rho^{X-2}}{\left(\frac{1}{2}(1-\mu)+(1-q)\mu(1-\beta^*)+(1-q)\mu\beta^*(X-\frac{1}{2}(X-1)\rho^{X-1})\right)^2} > 0$$

So the overall impact of ρ is to increase the bid price $b(0)$.

3.6.8 Corollary 3.4.4

The investor's information about the asset's value becomes common knowledge once the informed investor's orders arrive to more than one exchange within a given period. If the investor is informed and sends orders to all exchanges this occurs at $t = 1+\epsilon$ with probability $1-\rho^{X-1}$ and occurs at $t = 1+2\epsilon$ with probability ρ^{X-1}. If the investor is informed and sends orders to only one exchange this occurs at $t = 2+\epsilon$ with probability $1-\rho^{X-1}$ and occurs at $t = 2+2\epsilon$ with probability ρ^{X-1}. We can calculate the expected time at which the informed investor's private information about the asset's value becomes common knowledge as:

$$q\Big(\alpha^*((1+\epsilon)(1-\rho^{X-1})+(1+2\epsilon)\rho^{X-1})+(1-\alpha^*)((2+\epsilon)(1-\rho^{X-1})+(2+2\epsilon)\rho^{X-1})\Big)+$$
$$(1-q)\Big(\beta^*((1+\epsilon)(1-\rho^{X-1})+(1+2\epsilon)\rho^{X-1})+(1-\beta^*)((2+\epsilon)(1-\rho^{X-1})+(2+2\epsilon)\rho^{X-1})\Big)$$

This expression can be simplified to:

$$2+\epsilon+\epsilon\rho^{X-1}-q\alpha^*-(1-q)\beta^* \tag{3.51}$$

From Proposition 3.4.4 α^*, β^* are decreasing in ρ. So an increase in ρ results in a increase in the expected time at which the informed investor's private information becomes common knowledge.

3.7 References

Aisen, Daniel, Katsuyama, Bradley, Park, Robert, Schwall, John, Steiner, Richard, Zhang, Allen, and Popejoy, Thomas L, "Synchronized processing of data by networked computing resources," (Google Patents, 2020).

Ait-Sahalia, Yacine, and Saglam, Mehmet, "High frequency market making: Implications for liquidity," *Working Paper*, (2017a) (Available at SSRN 2908438).

Ait-Sahalia, Yacine, and Saglam, Mehmet, "High frequency market making: Optimal quoting," *Working Paper*, (2017b) (Available at SSRN 2331613).

Back, Kerry, and Baruch, Shmuel, "Information in securities markets: Kyle meets glosten and milgrom," *Econometrica*, 72 (2004), 433–465 (Wiley Online Library).

Baldauf, Markus, and Mollner, Joshua, "High-frequency trading and market performance," *The Journal of Finance*, 75 (2020), 1495–1526 (Wiley Online Library).

Budish, Eric, Cramton, Peter, and Shim, John, "The high-frequency trading arms race: Frequent batch auctions as a market design response," *The Quarterly Journal of Economics*, 130 (2015), 1547–1621 (MIT Press).

Easley, David, and O'Hara, Maureen, "Price, trade size, and information in securities markets," *Journal of Financial Economics*, 19 (1987), 69–90 (Elsevier).

Foucault, Thierry, Hombert, Johan, and Roşu, Ioanid, "News trading and speed," *The Journal of Finance*, 71 (2016), 335–382 (Wiley Online Library).

Foucault, Thierry, Kozhan, Roman, and Tham, Wing Wah, "Toxic arbitrage," *The Review of Financial Studies*, 30 (2017), 1053–1094 (Oxford University Press).

Glosten, Lawrence R, and Milgrom, Paul R, "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders," *Journal of Financial Economics*, 14 (1985), 71–100 (Elsevier).

Kyle, Albert S, "Continuous auctions and insider trading," *Econometrica*, (1985), 1315–1335 (JSTOR).

Li, Wei, "High frequency trading with speed hierarchies," *Working Paper*, (2018) (Available at SSRN

2365121).

Menkveld, Albert J, “The economics of high-frequency trading: Taking stock,” *Annual Review of Financial Economics*, 8 (2016), 1–24 (Annual Reviews).

Menkveld, Albert J, and Zoican, Marius A, “Need for speed? Exchange latency and liquidity,” *The Review of Financial Studies*, 30 (2017), 1188–1228 (Oxford University Press).

Yang, Liyan, and Zhu, Haoxiang, “Back-running: Seeking and hiding fundamental information in order flows,” *The Review of Financial Studies*, 33 (2020), 1484–1533 (Oxford University Press).