The 99¢ Dissertation: Essays on Psychological Pricing and Search

By

Jonah Jung Hao Yuen

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

June 30, 2018

Nashville, Tennessee

Approved:

Benjamin Eden, Ph.D.
Kathryn Anderson, Ph.D.
Robert Driskill, Ph.D.
Hyunseung Oh, Ph.D.
Paige Skiba, Ph.D.

ACKNOWLEDGMENTS

I would like to thank my advisor, Benjamin Eden, and the other members of my dissertation committee, Kathryn Anderson, Robert Driskill, Hyunseung Oh, and Paige Skiba, for their input and advice. I would also like to thank Andrew Dustan and numerous professors, graduate students, friends, and family for helpful comments and suggestions. I am especially thankful for my parents, Bernard and Chicky, for their steadfast support over all of these years.

I would like to thank Information Resources, Inc. (IRI) for making the data used in this dissertation available. All estimates and analysis in this dissertation based on data provided by IRI is by the author and not by IRI. This research was partially funded by a grant from the Vanderbilt University Council of Economics Graduate Students (CEGS), for which I am grateful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Prices play a critical role in economics by guiding the efficient allocation of resources in market-based economies and this function of prices makes understanding their behavior important for both macroeconomic dynamics and microeconomic decision-making. In this dissertation I utilize data on store prices and household purchases to address two puzzles regarding prices. First, the psychological pricing puzzle of why so many prices end in 99 cents, and second, the price dispersion puzzle of why the same product may have different prices in the same area. Chapter 1 tests whether a model of inattentive consumers can explain why many, but not all, prices end in 99 cents. Chapters 2 and 3 then address the puzzle of price dispersion in the context of search models, first by testing the empirical validity of a critical assumption in search models of price dispersion and then by proposing a new model to explain observed price behavior.

In the first chapter, I present a model of consumers who are sometimes inattentive to cent endings in prices, which gives firms an incentive to charge higher cent endings as inattention increases. A macroeconomic analysis testing if aggregated proxies for consumers' ability to focus on shopping, such as the unemployment rate or average number of hours worked, are related to the proportion of prices ending in 99 cents yields mixed results that generally do not support the inattention model. However, a microeconomic analysis relating households' purchases of goods with 99-cent price endings with their weekly shopping intensity yields results consistent with the model. Households who go on more shopping trips in a week are weakly associated with buying fewer goods with prices in ending 99 cents. In addition to these results, I also characterize new observations regarding the difference in 99-cent price endings between different types of products.

If the shopping behavior in the first chapter is only weakly related to purchasing goods with 99-cent price endings, do consumers pay attention to prices in general when shopping? Many models of price dispersion depend on the assumption that some consumers make multiple shopping trips in order to find lower prices for goods and in the second chapter I test if this assumption is valid empirically. The results suggest increased shopping is related to only small amounts of price savings on average, at most a penny per dollar. The small impact on price savings from shopping persists even when only considering goods households are known to buy in multiple stores or when dividing shopping trips into regular and irregular trips. These results suggest that seeking lower prices cannot be the primary motivation for the observed amount of shopping. This finding undermines the key building block of search models of price dispersion and I suggest a few reasons why large variation in shopping intensity may be related to only small changes in price savings.

The final chapter, written with Benjamin Eden and Maya Eden, constructs a model of uncertain and sequential trade that incorporates consumer search and storage of goods by firms to explain empirical observations regarding price dispersion and other aspects of prices. The key contribution of this chapter's empirical analysis is to split prices into different bins according to magnitude and document that the variation in price and quantity is lower for higher priced bins and that temporary sales contribute substantially to variation in the average price of a good. Furthermore, we show that the quantities sold by stores with high prices do not respond one-for-one to changes in the quantities sold by stores with low prices, but are positively related to the prices charged by stores in the low price bin.

Taken as a whole, this dissertation documents that existing theories explaining 99-cent price endings and price dispersion models do not adequately explain the relationships observed in the data. For the psychological pricing puzzle, a model of consumers who are inattentive to price endings can only explain a small amount of the variation in the proportion of goods with prices ending in 99 cents at the microeconomic level and generally fails at the macroeconomic level. In terms of the price dispersion puzzle, consumers who undertake additional shopping trips do not save much compared to their non-shopping compatriots, which is at odds with the assumption of price dispersion models. However, in a broader context, these results do not imply an outright rejection of the economic theories. Instead, they guide how we should consider refining models that attempt to explain the puzzles of psychological pricing and price dispersion since current theory can only explain a fraction of the behavior observed.

CHAPTER 1

THE 99¢ PUZZLE: EVALUATING THE INATTENTION MODEL OF 99¢ PRICE ENDINGS

1.1 Introduction

A brief stroll through almost any store in the United States will reveal a peculiar pricing practice - a tremendous amount of prices end in 99 cents and most prices end in a 9. This practice, called "customary", "odd", "just-below", or "psychological" pricing in the literature, has puzzled economists as far back as the early 20th century [41]. The retailing and marketing literature has scrutinized the effects of the practice but has focused on determining whether or not 99-cent endings increase revenue rather than asking why consumers prefer 99-cent endings, e.g. Schindler and Kibarian [72]. Behavioral economists have suggested a few theories, suggesting that 0 and 5 serve as reference points and thus 9 and 4 are seen as discounts from those reference points [72], but the key idea to emerge from that literature is that consumers have limited cognition and economize on price information. Why remember the cents digits when doing so may only save you less than a dollar and you have a thousand other things to remember?

In a model of inattentive consumers addressing cent endings, some agents do not bother looking at the cents digits in prices, instead comparing prices solely using the dollar digit. Firms recognize this behavior and capitalize on inattentive consumers by charging the maximum number of cents conditional on the dollar digit, i.e. 99. However, not all prices end in 99 so there must be some flexibility in the proportion of prices ending in 99. I construct a simple model where attention and 99-cent price endings have a negative relationship and test this relationship empirically at both the macroeconomic and microeconomic scale using two sets of data, both provided by Information Resources Inc. (IRI). For the macroeconomic analysis I uase a large set of scanner data of grocery store items covering many markets across the United States from 2001 to 2011 while the microeconomic analysis uses a set of data on household purchases in two American cities from 2008 to 2011.

I test the inattention model of 99-cent price endings at the macroeconomic level using unemployment as a measure of consumer inattention at an aggregated "market" level along with other measures related to consumer attention while shopping, such as average earnings and poverty levels. Research suggests that unemployment is associated with an increase in shopping behavior at the household level [52, 53], which justifies treating higher unemployment in a market as an increase in attention or decrease in market-wide inattention as long as shoppers are more attentive than non-shoppers or less frequent shoppers.

This empirical approach leads to the surprising result that the proportion of prices ending in 99 cents is positively related to the unemployment rate, which contradicts the inattention model of cent endings. These results are fairly robust and I estimate that a 1 percentage point increase in the unemployment rate is related to a 0.17 to 0.29 percentage point increase in the proportion of 99-cent endings. In addition, most aggregate proxies for inattention tested in this paper do not support the model except for labor force participation, where a 1 percentage point increase in the labor force to population ratio yields a 0.03 to 0.15 percentage point increase in 99-cent endings.

Analysis at the microeconomic level utilizes direct observation of how much each household shops in a week. The household-level analysis provides stronger support for the inattention model as increasing the average number of weekly trips by 1 is associated with 0.3% less spent on goods with prices ending in 99 cents and having 2 retired individuals as the heads of a household is related to 2% less spent on goods with 99-cent price ending. However, there are also results that do not support the inattention model at the household-level. The average number of stores a household visits in a week is not associated with any change in the 99-cent proportion and the number of unemployed members of a household is positively related to 99-cent price endings like the macroeconomic result, albeit not statistically significant. Moreover, the greatest determinant of the proportion of 99-cent price endings is a characteristic of the market itself, even after controlling for household demographics and the characteristics of the basket of goods a household purchases.

In addition to addressing whether the inattention model can explain 99-cent price endings, I also characterize the proportion of 99-cent price endings across product categories and time at the national level. The proportion of prices ending in 99 cents varies drastically across different product categories, with a low of around 10% for yogurt and rising to over 60% for diapers. However, there is no clear trend over time across all categories. While some categories show distinct upward trends in the proportion of 99-cent price endings, such as razors, laundry detergent, and beer, many others show no distinct trend over time.

The conclusion reached is that the prevalence of prices ending in 99 cents is still a puzzle: 99 is empirically the most popular cent ending by over twice as much as the next most popular ending, but not for the reasons suggested by a model of inattention. While this paper does not propose a definitive answer to why 99-cent price endings are so popular in light of the inattention model's failings, a few future avenues of research are proposed in the concluding remarks. The following section provides a brief summary of the literature on psychological prices and continues with a simple illustrative model to demonstrate the expected relationship between attention and 99-cent endings in section 1.3. Section 1.4 describes the data utilized in the empirical analysis in section 1.5 before concluding in section 1.6.

1.2 Previous Research

The overabundance of "odd', "just-below", "customary", or "psychological" prices, typically characterized as ending in a 9, has piqued economists' curiosity since the 1930s [41], but much of the research on the topic comes from the retail and marketing literature, asking whether or not 99-cent endings or odd-pricing in general increase sales. An early attempt at testing this phenomenon experimentally involved test subjects choosing between bundles of goods of similar value that were priced entirely in odd or even prices [60]. Other researchers constructed randomized controlled trials by partnering with mail-order catalogs to offer the same product at both odd prices and even prices, meaning a price ending in a 0, usually finding that 99-cent and 9-endings generally increased revenues relative to prices that end in 0 [41, 72].

As for why odd prices were associated with higher sales, the retail literature suggests consumers may suffer from "price illusion" [60], erroneously believing that prices are lower than they truly are when a price ends in a 9 or treating 99 cents as a signal of a low or sale price. However, the idea that a 99-cent ending acts as a signal has a few issues. First, such a signal should be meaningless since not only are many prices that end in 99 not temporary price reductions, but it is to mimic a 99-cent ending so consumers should have little reason to believe that 99 implies a sale price[1]. Second, Schindler and Kibarian [72] coincidentally test price endings acting as a signal when the company in their experiment insisted they include another experimental group that was shown prices ending in 88 cents since the company claimed their customers knew 88 cent endings were clearance prices. They find that prices with 88-cent endings do not lead to noticeably different revenues from prices with 00-cent endings, suggesting that cent endings are not a good signal for sales.

In the field of economics, interest in just-below prices has not only asked why they exist in such high frequencies, but also whether such prices can impact the frequency of price changes and lead to price stickiness. Blinder et al.'s [12] survey of price setters at American firms put a dent in the importance of psychological prices since 59.5% of surveyed firms said psychological pricing points, or "threshold prices" in the survey, were "totally unimportant" in deterring price increases. While they find that firms who sell to consumers rather than larger entities, such as businesses or the government, claim psychological prices are more important, the average importance across all firms in the survey is low, a finding which Hoeberichts and Stokman [49] corroborate using a survey of Dutch businesses. Despite those conclusions, the continued overabundance of prices ending in 99 or other just-below prices seems contrary to the observed statistics. How can 67.4% of prices end in a 9 for the cents digit and not be important?[2] Supporting this notion is research using German scanner data, finding evidence of psychological prices contributing to sticky prices when defining psychological prices as the two or five most common prices a firm uses, and other work in e-commerce [47, 48, 44].

[1]By this I mean a firm that thinks it should charge \$2.73 for a good could easily decide to charge \$2.99 and increase its revenue if consumers blindly assume \$2.99 is a temporary price reduction and believe a 99-cent price ending indicates the lowest price during their search.

[2]This percentage comes from the scanner data set used in this paper. A detailed description of the data is in Section 1.4.

From a theoretical perspective, a few economic models have been proposed to explain the abundance of psychological prices. Basu [7, 8] proposes a rational expectations model for cent endings. If consumers economize on price information by not looking at cents in prices, they form expectations for the price ending based on the distribution of cent endings and in equilibrium this expectation is correct, which happens to occur at 99 cents. Shy [75] suggests a different model, implementing a transportation cost that generates 99-cent price endings when the transportation cost is sufficiently high. Neither Basu nor Shy offer empirical evidence to support their models, but a third type of model relying on inattentive consumers has received both theoretical and empirical attention in the economic literature.

1.2.1 Rational Inattention, Salience, and Psychological Prices

The final theory used to explain psychological prices, and the theory this paper aims to test, is the idea that consumers are rationally inattentive. In contrast to the standard macroeconomic rational inattention literature that relies on the work by Sims or Mankiw and Reis [62, 77, 69, 70], explaining 99-cent endings using the concept of rational inattention appeals more to the behavioral economics pioneered by Kahneman and Tversky [50]. Both rational inattention and behavioral economics rely on the idea that decision makers have limited cognitive ability, leading to behavior that seems irrational under full information.

Questioning the empirical rationality of consumers is not a new topic in economics, e.g. Simon [76], and the modern literature has developed numerous models focusing on the limits of processing or remembering information. Dow [30] proposes a model in which consumers make decisions based on the price history for goods but are limited in their ability to remember price histories. In the two-good case, Dow concludes that a consumer is better off focusing attention on one good. While not addressed in the paper, such behavior opens the door to firms overcharging consumers who are focused on the price of another good.

Mullainathan et al. [63] instead proposes a model of "coarse thinking" in which consumers apply information from previous events to new situations erroneously. Although they implement their model in the context of advertisements and branding, coarse thinking could also be used to support a model where 99-cent price endings are erroneously believed to be a signal of a good price, as proposed in Lambert [60]. Another possible model is the existence of "comparison frictions". Kling et al. [56] show that consumers do not use information available at a low cost in an experiment comparing Medicare prescription plans. They find that directly providing personalized prescription pricing information is associated with a higher frequency of changing prescription plans and a price savings of about $100 compared to individuals who were only informed that personalized information comparing plans was available on a website. If consumers are not willing to make comparisons when costs are trivial, e.g. visiting a personalized website, and the benefits are potentially large, one might expect consumers to be even less likely to pay attention to prices when the costs are higher and the benefits are less than $1.

While the previous models lend support to the idea of inattentive consumers, a few models utilizing consumer inattention suggest 99-cent endings should be related to more attention, rather than less attention. de Clippel et al. [26] design a model with a distribution of consumers who vary in their level of inattention and the existence of a firm in a price leadership position. When the fraction of consumers who are completely inattentive is held fixed, a distribution of consumers with lower average attention, or more inattention, is associated with lower prices as the price leader lowers its price to deter consumers from searching by making sure its price does not "stick out" relative to the cheaper firms. If a 99-cent ending is associated with a high price, this model would predict fewer such price endings as average inattention increases.

Similarly, Gabaix [39] examines the theoretical implications of consumers who only value a subset of factors in the classical consumer maximization problem. Among numerous theoretical results, Gabaix finds that consumers will be more attentive to goods with higher price volatility, a higher share of expenditures, and goods that are more "important", meaning they have a high price elasticity. While paying more attention to goods that are a large share of expenditures is intuitive, higher attention paid to such goods proves problematic if those goods are also high in price and consumers pay attention to the digits of a price to varying degrees. If consumers do not pay equal attention to all digits in a price, one might expect the cent ending to be less important as price increases since it forms a smaller percent of the expenditure on the good.[3]

Another relevant field of research is the idea of "salience". Bordalo et al. [13] defines salience as the characteristic that stands out most when making a decision. In their model, they define goods by two characteristics - price and quality - and describe how consumers pay attention to the characteristic furthest away from its average value and overweight their preferences towards that characteristic. Further refinements of the concept include modeling how markets can switch between quality-salient and price-salient equilibria based on the cost of quality or producing the good and how consumers form norms based on memory of past events that influence the most salient factor [14, 15].

In their current form, models of salient characteristics cannot directly address 99-cent price endings because price is a continuous, linear characteristic of a good. Instead, treating the dollar digit as a separate characteristic of a good would be necessary to generate a discontinuous response to specific, minor changes in price. As is, increasing the price of a good with an average price of \$4.99 to \$5.00 would not be a salient change since the increase is only a small deviation from the average price. The change in dollar digit would need to be overemphasized for such a change to become the salient factor. However, if memory and norms guide which characteristics are salient, then 99-cent price endings could be explained by firms' reluctance to change from a 99-cent ending due to fears of consumers overreacting to price increases relative to consumers' norms.[4]

In contrast to traditional economic models, Schindler and Kirby [73] exemplifies the behavioral economics perspective by arguing that the numbers 0 and 5 serve as reference points for shoppers. When consumers see a price such as \$0.99, \$1.00 acts as their reference point and they view \$0.99 as \$1.00 with a 1 cent discount, which might convince them that the price is low enough to justify purchasing. However, one weakness of this behavioral heuristic is that the predictions fail for

[3] For example, the difference between \$1,000.01 and \$1,000.99 is trivial, and firms would therefore have a greater incentive to set the cent ending to 99.

[4] As an anecdotal example, 20 ounce bottles of soda at most stores were priced at \$0.99 for many years before eventually increasing in price to \$1.25 but were soon followed by 1 liter (35.2 ounce) bottles available for \$1.00. Perhaps the introduction of the larger product at a price closer to the old norm was a reaction to consumers incensed by the price increase relative to their norm for the 20 ounce bottle.

prices below 5-endings, such as 24 versus 25 cents. In addition to 0, 5 is the next most accessible numbers from a cognitive perspective, so the "just-below" effect should also lead to more prices ending in a 4 since they are just below prices ending in a 5. However, Schindler and Kirby find no evidence to support such an effect when using a sample of prices from newspapers.[5]

The idea of limited cognition has also inspired economists to study the idea that consumers have a "left-digit bias" and only look at a certain number of digits due to reading numbers from left to right.[6] Lacetera et al. [57] document such a bias in the context of used cars. They find that the prices of used cars fall dramatically when crossing odometer mileage thresholds for the ten-thousands and thousands digit while equal-magnitude decreases in odometer readings that do not cross a threshold do not yield an equal increase in value. Although their results provide useful evidence of left-digit bias, their research uses price as a means of measuring inattention to quality. Limited attention and left-digit bias with respect to price is a little more difficult to identify.

Basu [7] used the concept of sequentially processing digits in a number to develop his rational expectations model of cent endings, but the idea also finds its way into strictly rational inattention models of how consumers perceive prices. Stiving and Winer [80] provide evidence of this by using scanner data for tuna and yogurt prices under $1.00, reporting that the hundredths digit in a price has more predictive power on sales when the tenths digit is the same, suggesting that inattention and left-to-right reading of prices have some empirical validity. However, the literature testing if inattention and 99-cent endings are related to each other has been somewhat sparse. Bergen et al. [11] represents the most relevant research, testing the theory by regressing the presence of a 9 ending on the length of a price but they find only mixed evidence that longer prices lead to a higher probability of a price ending in 9. This lack of satisfactory empirical support for the inattention model of 99-cent endings suggests that the topic is ripe for more research.

[5] A histogram of price endings for the data used in this paper supports this conclusion since the bins ending in a 9, such as 0.29 and 0.79, contain many more observations than the bins containing not only 4-endings, but also 5-endings, e.g. 0.20 to 0.28 cents and 0.70 to 0.78 cents (Figure 1.2).

[6] Poltrock and Schwartz [65] provides some experimental evidence that subjects sequentially compare digits from left to right when comparing 4 or 6 digit numbers

1.3 An Illustrative Model

To illustrate the expected relationship between cent endings and consumer inattention, consider the simple model of a firm that determines the price of its good in a two step process: First, it decides on the dollar digit, D, then, it determines the cent ending, C, conditional on the chosen dollar digit. The first stage is not of primary concern in this paper, but assume firms choose the dollar digit optimally such that D^* maximizes profit compared to any other choice for the dollar digit, $\pi(D^*) \geq \pi(D_i)$ for all positive integers D_i.[7] After choosing D^*, the firm must choose its cent ending to maximize profits given the proportion of attentive shoppers, ω, according to the following profit function net of the profit from choosing D^*:

$$\pi(C|D^*) = \omega\left(C^M - a(C^M - C)^2\right) + (1-\omega)C \tag{1.1}$$

subject to $C \in [0, 99]$, where C^M is the profit maximizing cent ending when all consumers are attentive, i.e. $\omega = 1$, and a captures the sensitivity of attentive shoppers to the cent ending.[8] The first term, $\omega\left(C^M - a(C^M - C)^2\right)$ represents the profit from attentive shoppers.[9] Any deviation from C^M, the optimal cent ending under perfect attention, would lead to a fall in profit as the firm either forgoes revenue by charging a lower cent ending without a commensurate gain in customers, $C < C^M$, or loses customers and profit by charging more than the perfect attention optimal ending, $C > C^M$. The second term, $(1-\omega)C$, is the key expression that captures the linear increase in profits from inattentive consumers as the cent ending rises. For an inattentive consumer, demand is perfectly inelastic within a given dollar digit so any increase in the cent ending will be additional

[7] Here I assume that D^* is a single integer that does not depend on the proportion of inattentive consumers because the existing literature assumes that consumers at least pay attention to some of the dollar digits. One can widen the definition of inattention to include inattention to the ones or tens digits in the case of larger prices, e.g. $5,699.99, but that is not relevant to this paper given the small range of prices for goods in the data.

In addition, this dollar digit is considered optimal even when compared to its just-below price. The firm's profits are always greater with a price of $D.00 compared to $(D-1).99 so the firm is only considering how high the cent ending should be.

[8] While cent endings are a clear case of discrete numbers, I treat them as continuous to avoid the unnecessary complications of discrete choice when intuition is sufficient for understanding the basic arguments of the model.

[9] The specific functional form of the profit from attentive consumers is not critical and merely captures the idea that the profits from attentive shoppers falls when deviating from C^M as one would expect from a demand curve with some monopoly power.

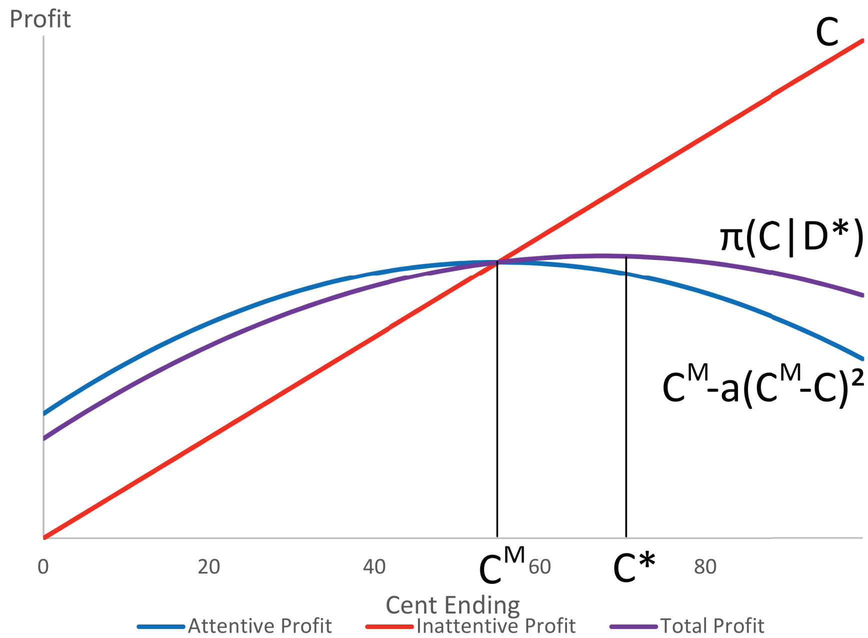


Figure 1.1: Diagram of profit per inattentive consumer, C, and attentive consumer, $C^M - a(C^M - C)^2$, and total profit, $\pi(C|D^*)$.

profit for the firm. If consumers were completely inattentive to the cent ending, $\omega = 0$, the firm would need only solve the simple equation:

$$\max_{C \in [0,99]} \pi(C|D^*) = C \tag{1.2}$$

which is trivially solved with $C = 99$. In other words, in a world consisting of only inattentive consumers, all prices should end in 99 cents.

Figure 1.1 illustrates the tension when choosing a particular cent ending in the model. In the figure, the increase in profit from raising the cent ending that comes from inattentive consumers is diminished by the fall in profit from attentive shoppers as the cent ending deviates from C^M. With a population of only attentive consumers, $\omega = 1$, the profit function conditional on the dollar digit is the hump-shaped function $C^M - a(C^M - C)^2$. However, as the proportion of attentive shoppers falls and ω approaches 0 the profit function begins to look more like the function $\pi(C|D^*) = C$.

One can solve for the firm's optimal cent ending using the first order condition of the profit function with respect to C:

$$C^* = C^M + \frac{1-\omega}{2a\omega}. \tag{1.3}$$

When $\omega = 1$, meaning all consumers are attentive, the optimal cent ending is C^M. Moreover, the derivative of C^* with respect to the level of attention is:

$$\frac{\partial C^*}{\partial \omega} = -\frac{1}{2a\omega^2} < 0. \tag{1.4}$$

meaning that the optimal cent ending is strictly decreasing in the level of attention and more inattention should lead to a cent ending closer to 99. Since the cent ending is bounded above by 99, the optimal ending for values above a critical value of ω is 99. Specifically, a 99-cent ending is optimal when

$$\omega < \frac{1}{2a(99 - C^M) + 1}, \tag{1.5}$$

This equation reflects the intuition that increases in attentive consumers' price sensitivity, a, raises the level of inattention necessary to justify a 99-cent ending as profit maximizing. As attentive shoppers become more sensitive to deviations from C^M, cent endings away from C^M have a larger negative effect on the firm's profits so a 99-cent ending is only ideal if the proportion of attentive consumers falls. In addition, a higher optimal attentive cent ending, C^M, decreases the amount of inattention required to make 99 an optimal ending as the loss in profits from deviating from C^M is smaller if 99 is fewer cents away from C^M.

While this model provides a framework for understanding how a firm chooses a cent ending for a price, it also implies a single cent ending for a firm so modeling the observed variation in cent ending across goods requires a richer theoretical framework. Since this study's analysis uses aggregated variables, the specific mechanism is not as important, but possible extensions of the model to explain variation in cent endings could allow for product-specific inattention parameters or firm-specific inattention levels. Product-specific inattention seems reasonable since consumers may not pay the same level of attention to the cent ending for every good, e.g. a necessity versus a luxury good. As for firm-specific inattention levels, one could think that either firms vary in their expectation of consumer inattention or that some firms receive more attentive shoppers, e.g. a discount store may attract bargain hunters who are more attentive to all digits in a price. Combining product-specific and firm-specific inattention measures could better model why not all products within a store have 99-cent price endings and why some stores have more 99-cent endings than others.

1.3.1 Potential Macroeconomic Correlates with Inattention

Measuring inattention, ω, directly is difficult and there is little literature addressing the topic, although some pioneering work by Khaw et al. [55] attempts to quantify inattention in a lab setting. The lack of a clear empirical measure of inattention forces any test of the inattention model of 99-cent price endings to rely on measures one would think are ex ante correlated with inattention to indirectly determine the role of inattention on 99-cent endings. Kaplan and Menzio's [52] empirical work finding a relationship between unemployment and shopping serves as the main inspiration for this paper's method, but other measures of consumers' opportunity cost of time are also possible correlates of interest.

One finding in the literature is that households with more unemployed people pay lower prices primarily by shopping at more stores [52]. While paying a lower price does not necessarily mean one is buying fewer prices ending in 99 cents, a reasonable assumption is that those who visit more stores are more attentive to price as evidenced by the lower price they pay relative to others. Presumably, as unemployment increases, there are more unemployed individuals who are able to spend time shopping in addition to time spent searching for a job relative to full-time employment. In addition, unemployed individuals may have a greater incentive to pay attention to prices due to a lack of income and tighter budget constraint. Both factors would suggest that as unemployment rises, the prevalence of 99-cent endings should fall.

However, the appropriateness of using the Bureau of Labor Statistic's unemployment measure is somewhat questionable since it is relative to the population of people employed or looking for work and does not capture discouraged workers or people not in the labor force who may devote their time to non-work activities, such as shopping. To address these concerns, one may want to instead consider the labor force participation rate or the level of employment relative to the total population as alternative measures of aggregate attentiveness since they may better capture how many people are too busy to pay attention to prices. For these measures, a higher labor force participation rate or a higher ratio of employment to population, meaning fewer people able to spend time shopping, would be expected to be correlated with more 99 cent endings.

Earnings could also be a factor influencing how much consumers pay attention while shopping. As earnings increase, the opportunity cost of spending time shopping increases. The gains from paying attention to prices on each individual may be small and a trivial fraction of income once a person earns a sufficiently high wage so higher earnings could potentially be related to purchasing more goods with a price ending in 99 cents. As an extreme measure of the lack of earnings, the poverty rate may also have predictive power for the fraction of prices ending in 99 cents. Similar to how unemployed people may be more sensitive to price due to less income, those living in poverty have a strong incentive to pay attention to prices due to a tight budget constraint. One would then expect that higher poverty rates would be associated with a lower proportion of 99-cent endings as more people pay attention to how they spend their income. However, Banerjee and Mullainathan's [6] research into limited attention and income models poorer households as households unable to afford "distraction saving goods and services" in regards to problems at home, e.g. unreliable heating, which leads to less of an ability to focus on other tasks. Under their model, one would instead expect a positive relationship between poverty and 99-cent endings as consumers are distracted by other issues.

1.3.2 Measuring Household-Level Inattention

In contrast to the macroeconomic perspective, the existence of microdata on household-level purchases and shopping behavior allows a more direct measure of how household attention is related to purchases. Although one cannot yet track precisely which goods consumers peruse while visiting a grocery store, a reasonable measure of how much attention a household pays to prices is the household's intensity of shopping behavior. Undertaking more trips or visiting more stores may indicate a household who pays careful attention to price and is willing to search for minor differences in price.

Data collected using scanner data methods typically contain information on when and where a purchase occurred, which allows one to calculate the number of shopping trips made, as measured by trips to the checkout counter, and the number of different stores a household has bought goods at. However, what is unclear is whether the number of trips or the number of different stores should be the appropriate measure of consumer attention.[10] A high number of shopping trips could be related to a household intensively searching and making use of temporary sales and daily deals. A high number of trips could instead be related to less attention if multiple trips reflect a household forgetting to purchase an item or if the household lives close to a store and stops by frequently as a matter of convenience. On the other hand, a high number of different stores is more likely to reflect a household purposefully seeking deals at multiple stores. Similar work studying search behavior using scanner data has failed to reach a consensus on the most appropriate measure to use [1, 52], so this chapter considers both measures as micro-level proxies for consumer attention.

In addition, one weakness of scanner data in general is that store visits that do not result in a purchase are not recorded in the data. A household could visit multiple stores or go on multiple trips, or even look at grocery advertisements ahead of time, but ultimately make purchases at only one or two stores in a single trip or two. This could become a source of error due to downwardly-biased mismeasurement of the level of attention paid by a household.[11] A pessimistic researcher should therefore treat the measured amount of shopping as a lower bound and underestimate of the true amount of search and attention paid by the household, which would then attenuate the effect on the proportion of products purchased ending in 99 cents as the negative effect from attentive shoppers making a low number of trips is counteracted by the positive effect from inattentive

[10] Chapter 2, which focuses on search behavior and price savings, considers even finer measures of shopping behavior, such as defining "regular" and "irregular" shopping trips and stores.

[11] This issue could be a more nefarious source of error if the savviest of shoppers not only pay the most attention to price, but are also wise enough to minimize how many stores they visit and how many shopping trips they make that result in a purchase.

individuals who are also observed making a low number of shopping trips.

1.4 Data Description

Ideally, the consumer inattention model of 99-cent price endings should hold at both the macroeconomic and microeconomic level. To verify this claim, I utilize data collected by a single data firm to conduct a macroeconomic, market-level analysis of prices in stores across the United States and a microeconomic, household-level analysis using household purchases in two American cities.

1.4.1 Market-Level Data

The price data for the market-level analysis come from a large set of scanner data from Information Resources, Inc., known as the IRI Academic Scanner Dataset, and covers 31 categories of items sold in grocery stores in 47 markets across the United States of America for the years 2001 through 2011. A market is usually associated with a city, such as Los Angeles, but sometimes covers a larger area, such as New England or the state of South Carolina. More specifically, the market-level analysis covers all 47 markets and utilizes all 31 categories for the years 2002 to 2011 but 30 categories for 2001.[12] A complete description of the data is available in Bronnenberg et al. [16].

The IRI scanner dataset contains a wealth of information down to the store level at a weekly frequency for thousands of Consumer Packaged Goods (CPG) at the Universal Product Code (UPC), or barcode, level. However, matching markets to unemployment data requires aggregating to the market level at a monthly frequency. To aggregate, I calculate the proportion of 99-cent endings using two approaches. The first approach, which I refer to as the "posted price" or simply "price"

[12]The omitted category is carbonated beverages. IRI states that there are anomalies in the 2001 carbonated beverage data and suggests not using it. The other 30 categories covered are: beer, razor blades, carbonated beverages, cigarettes, coffee, cold cereal, deodorant, diapers, facial tissue, frozen dinner entrees, frozen pizza, household cleaners, hotdogs, laundry detergent, margarine and butter, mayonnaise, milk, mustard and ketchup, paper towels, peanut butter, photographic film, razors, salty snacks, shampoo, soup, spaghetti sauce, sugar substitutes, toilet tissue, toothbrushes, toothpaste, and yogurt.

measure, treats each UPC-store-week observation as a single observation when calculating the proportion of 99-cent endings, regardless of the number of units sold by each store. This measure can be thought of as reflecting how often firms choose 99-cent price endings. The second measure calculates the fraction of all units bought by consumers that end in 99 cents.

For the inattention measure, directly measuring aggregate consumer inattention directly is currently not possible, but Kaplan and Menzio [52, 53] suggest there is a close relationship between unemployment and shopping. Presumably, more frequent shoppers pay more attention to prices so higher unemployment should be a proxy for higher levels of attention or lower levels of inattention. To measure unemployment, I collect seasonally adjusted unemployment data compiled by the U.S. Bureau of Labor Statistics (BLS) at the Metropolitan Statistical Area (MSA) level over the relevant time period similar to the method used by Coibion et al. [20]. However, due to the differing composition of markets, I construct an unemployment rate for each market using a population-weighted average of the MSAs in each market based on the number of each MSA's counties present in the market.[13]

I also construct measures of monthly labor force relative to population and employment relative to population ratios to address concerns about unemployment's appropriateness as a proxy for the level of inattention. Specifically, the alternative measures I collect related to the labor market are the size of the labor force and number of nonfarm employees at the MSA level for 2001-2011 from the BLS, which are then aggregated and weighted as described in the preceding paragraph. From these, I construct the labor force to population ratio and number of employees to population ratio. Finally, I collect average weekly hours at the MSA level from the BLS for 2007-2011 and aggregate

[13] For instance, the San Francisco/Oakland market contains 5 of the 5 counties in the San Francisco-Oakland-Fremont MSA and 1 of the 2 counties from the San Jose-Sunnyvale-Santa Clara MSA. The market's unemployment rate is calculated using a weighted average of the unemployment rates of both MSAs using weights constructed by dividing 100% of the San Francisco-Oakland-Fremont MSA population or 50% of the San Jose-Sunnyvale-Santa Clara MSA population by the sum of 100% of the San Francisco-Oakland-Fremont MSA population and 50% of the San Jose-Sunnyvale-Santa Clara population. However, some counties are present in multiple markets and some multi-county MSAs were represented by a single county in some markets. These MSAs and counties are treated on a case-by-case basis with a rule of thumb that "small" counties relative to the overall market population are omitted since they have a negligible effect on the calculated unemployment rate. A more accurate measure would be to gather all of the county unemployment rates and construct an unemployment rate based on each market's labor force and population for each month but the gains in accuracy would likely be small while requiring gathering data on over 1,000 counties.

and weight to the market level as a measure of the time employed individuals have to shop. This measure is conditional on employment and cannot capture the behavior and role non-employed individuals' shopping behavior has on a market's cent endings, but offers a direct measure of the average time employed individuals have to spend on non-work activities.

Earnings data come from the BLS and are collected at the MSA level at a monthly frequency for 2007-2011. The specific measures collected are seasonally adjusted average hourly wages and seasonally adjusted weekly earnings in dollars, which are then aggregated to the IRI market level. One final measure of economic status for a market is the poverty rate from the U.S. Census' Small Area Income and Poverty Estimates (SAIPE), which is available for 2001 to 2011 at an annual frequency. More poverty may be related to more miserly consumers who pay attention to price out of necessity, which would result in a lower proportion of prices ending in 99 cents. However, it could also put a strain in consumers' ability to process additional price information of little marginal value, leading to a higher proportion of purchases ending in 99 cents.

1.4.1.1 Market-Level Descriptive Statistics

The dataset utilized with the labor market measures is a fully balanced panel of 6,204 observations - 132 monthly observations for each of the 47 markets present - describing the proportion of all prices or units sold ending in 99 cents for each market-month along with the labor market measures described in the preceding section. Table 1.1 provides summary statistics of key variables at the market-month or market-year level. Notably, 23.2% of prices ended in 99 cents while only 17.4% of units sold had a 99-cent price ending.

Figure 1.2 provides a visual depiction of the distribution of prices over the relevant time period, broken up into the recessionary and expansionary periods between 2001 and 2011.[14] On the surface, these diagrams do not show any clear pattern between the fraction of goods ending in 99 cents and recessionary periods. If anything, there seems to be a downward time trend that ends

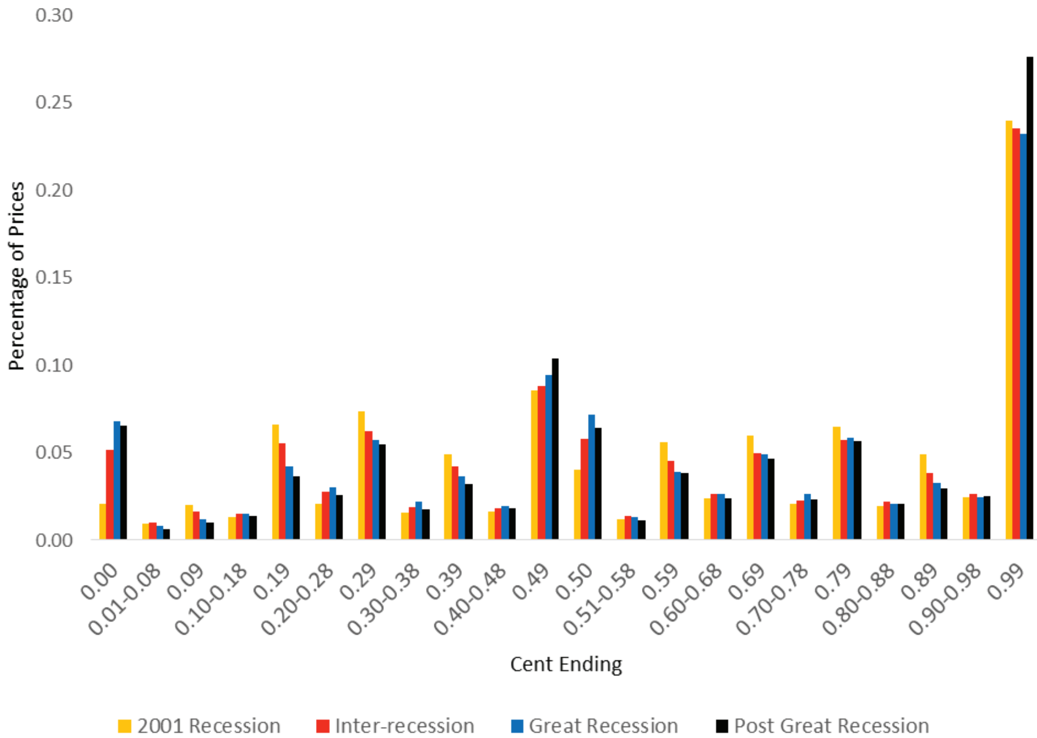
[14] The NBER dates the 2001 recession to March 2001 until November 2001 and the Great Recession from December 2007 to June 2009.

with a distinct increase after the Great Recession.

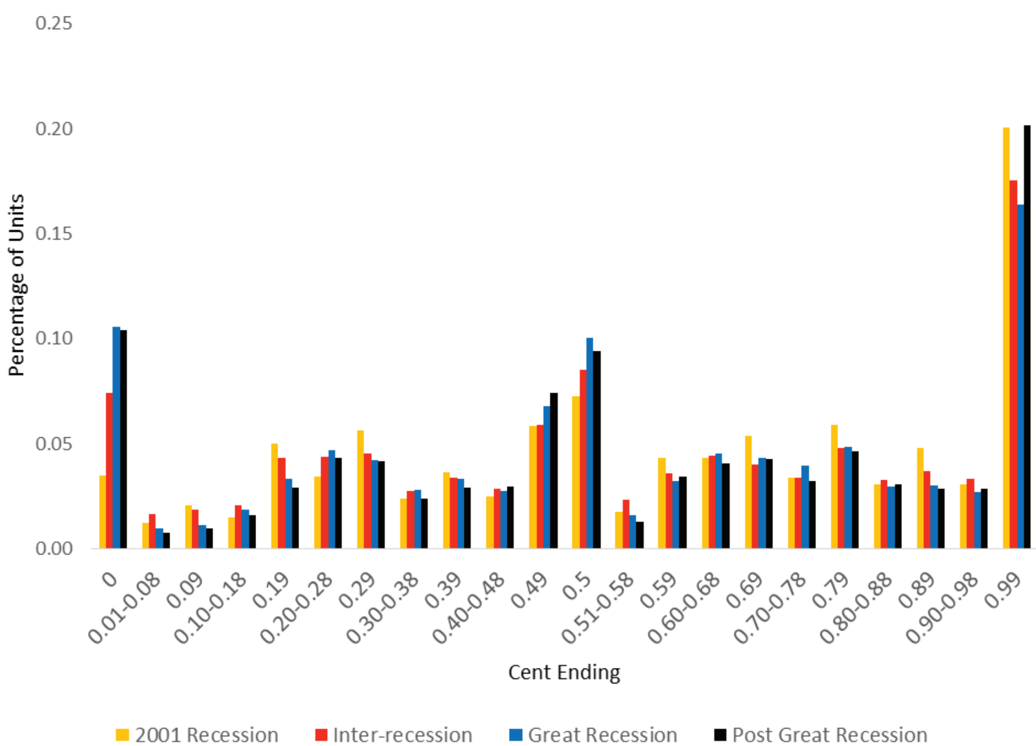
A simple stylized fact, the correlation between 99-cent endings and inattention measures, can demonstrate if the inattention theory has some merit worth further investigation. Table 1.2 presents correlations when pooling markets into one sample. While the correlation with the coarsest measure of attention, a recession dummy, is negative and supports the inattention hypothesis, it is relatively small in magnitude. Furthermore, correlation is a rough measure of the relationship since it cannot be separated from a time effect and also ignores any variation at the market level. Adding to the puzzle are the various labor market variables. First, the unemployment rate is positively correlated with 99-cent endings, contradicting the theory that more unemployment allows for more attentive shopping and reduces the amount of prices ending in 99 cents. Both the labor force and employment measures exhibit the same puzzle with even larger correlations. Also perplexing is the different signs for the correlation between weekly hours worked and the proportion of prices versus units ending in 99 cents.

For units sold, hours worked is positively related to 99-cent endings, which supports the idea that markets with people who work more have less time to shop and are more susceptible to 99-cent endings. However, the correlation with the percent of prices is negative and larger, contradicting the inattention theory. The only potential support for the inattention theory comes from hourly wages or weekly earnings. As income rises, one would expect less of an incentive to pay attention to cent endings since the savings would be minimal relative to the opportunity cost of paying attention. As such, hourly wages and weekly earnings are positively correlated with 99-cent endings. Based on these stylized facts, the inattention model of 99-cent endings appears to be on rocky ground but requires further analysis using the regression analysis in section 1.5.

The frequency of the poverty data necessitates aggregation to the annual frequency. In this case, the measure of 99-cent endings is the proportion of all prices or units purchased in a market each year. Of note, the measure of poverty provides support for the model of attention and income in Banerjee and Mullainathan [6] rather than a model having poverty associated with more attention. The positive relationship between poverty and 99-cent endings matches a model in which



(a) Posted prices



(b) Units sold

Figure 1.2: Distribution of price endings

2001-2011 (n = 6,204)	Mean	S.D.	Min	Max
Proportion of Prices Ending in 99	0.232	0.050	0.083	0.369
Proportion of Units Sold Ending in 99	0.174	0.047	0.048	0.350
Unemployment Rate	6.15	2.16	2.30	15.80
Labor Force to Population %	51.40	4.50	36.73	68.79
Employment to Population %	47.85	5.33	33.56	66.76
2007-2011 (n = 2,820)				
Proportion of Prices Ending in 99	0.238	0.056	0.108	0.369
Proportion of Units Sold Ending in 99	0.170	0.043	0.0478	0.288
Hourly Wages ($)	22.85	3.5215	17.11	35.30
Weekly Earnings ($)	791.3	122.3	578.2	1204.0
Weekly Hours Worked	34.64	1.22	31.25	39.10

(a) Monthly Frequency Variables

2001-2011 (n = 517)	Mean	S.D.	Min	Max
Proportion of Prices Posted Ending in 99	0.232	0.049	0.096	0.362
Proportion of Units Sold Ending in 99	0.174	0.045	0.061	0.281
Poverty Rate	0.127	0.031	0.061	0.242

(b) Annual Frequency Variables

Table 1.1: Summary statistics across market-month or market-year pairs using seasonally adjusted labor variables.

disadvantaged households cannot spare resources to process minor price differences and end up paying 99-cent endings more often.

1.4.2 Category-Level Descriptive Statistics

In addition to market-level statistics, the proportion of goods ending in 99 cents across categories may be of interest, especially if the proportion has changed over time. Table 1.3 presents the proportion of UPC-store-week prices, units, and revenue from goods with prices ending in 99 cents across the entire 2001 to 2011 sample period when pooling all markets together.

The proportion ranges from a high of 61% for the proportion of revenue for diapers down to 6% for the fraction of yogurt units sold. Across the three different measures, the proportion of units is typically lower than the other measures while the proportion of revenue is almost always the highest of the three measures. The observation that 99-cent endings are most abundant when looking at

Variable	% of Prices Ending in 99	% of Units Ending in 99
Recession dummy	-0.115	-0.075
Unemployment Rate	0.311	0.187
Labor Force to Population Ratio	-0.305	-0.218
Employment to Population Ratio	-0.492	-0.312
Hourly Wages	0.343	0.188
Weekly Earnings	0.236	0.090
Weekly Hours Worked	-0.104	0.033
Poverty Rate	0.200	0.236

Table 1.2: Correlation between 99-cent endings and seasonally adjusted proxies for inattention.

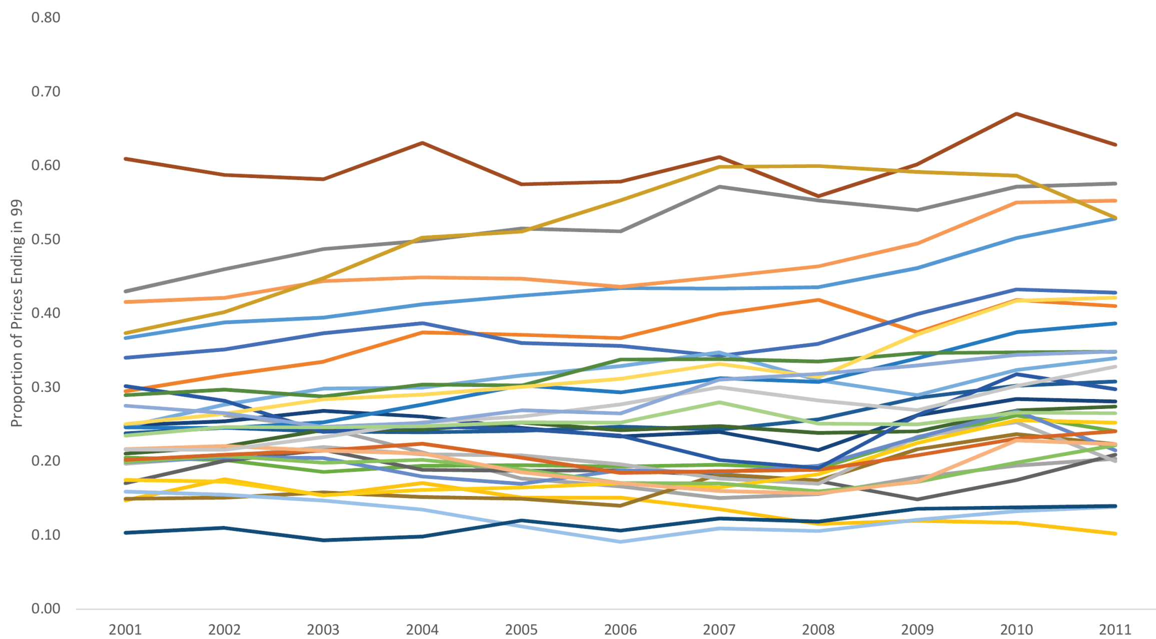


Figure 1.3: Annual proportion of prices ending in 99 cents over time by category, 2001-2011.

revenue corroborates the business literature's conclusion that 99-cent endings "just work" in terms of raising revenue [41, 72].

Figure 1.3 provides a visualization of the wide disparity in the proportion of prices ending in 99 cents and how the proportions have changed over time by charting the annual average proportion of prices ending in 99 cents for each category. Figures 1.4 to 1.6 offer more detail by splitting the 31 categories into the top ten (Figure 1.4), middle ten (Figure 1.5), and bottom eleven categories ranked according to the average proportion of prices ending in 99 cents in Table 1.3 (Figure 1.6).

One trend among the top 10 and middle 10 categories is a general increase in the proportion

Category	Proportion of Prices	Proportion of Units	Proportion of Revenue
Diapers	0.60	0.59	0.61
Razors	0.52	0.52	0.54
Photographic Film	0.50	0.47	0.54
Laundry Detergent	0.47	0.46	0.50
Beer	0.44	0.42	0.50
Coffee	0.38	0.37	0.43
Blades	0.37	0.36	0.40
Toilet Tissue	0.32	0.25	0.35
Shampoo	0.32	0.31	0.35
Hot Dogs	0.31	0.26	0.32
Paper Towels	0.30	0.24	0.36
Toothbrushes	0.30	0.27	0.34
Sugar Substitutes	0.27	0.27	0.31
Deodorant	0.26	0.24	0.27
Salty Snacks	0.26	0.22	0.24
Toothpaste	0.25	0.24	0.27
Frozen Pizza	0.25	0.19	0.25
Household Cleaners	0.25	0.24	0.26
Peanut Butter	0.21	0.19	0.22
Cold Cereal	0.21	0.18	0.19
Margarine & Butter	0.21	0.19	0.21
Milk	0.20	0.17	0.18
Carbonated Beverages	0.20	0.17	0.21
Spaghetti Sauce	0.20	0.18	0.20
Mustard & Ketchup	0.19	0.19	0.19
Mayonnaise	0.19	0.20	0.21
Facial Tissues	0.19	0.19	0.20
Frozen Dinners	0.18	0.13	0.19
Cigarettes	0.14	0.11	0.21
Soup	0.13	0.10	0.12
Yogurt	0.12	0.06	0.13

Table 1.3: Proportion of posted prices, units sold, and revenue from goods with prices ending in 99 cents by category, 2001-2011.

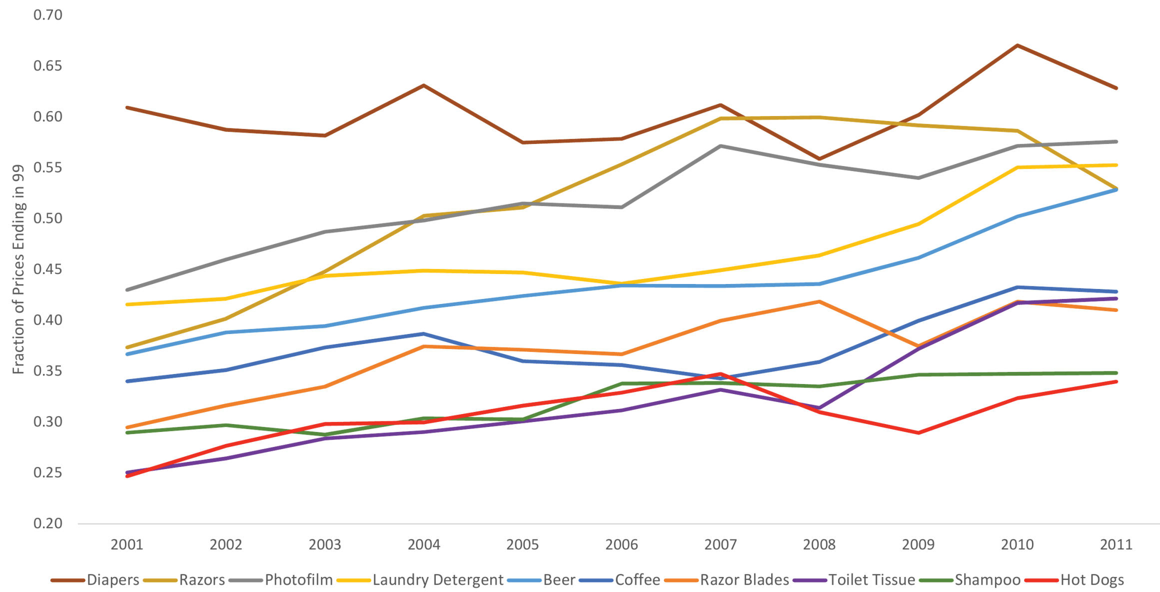


Figure 1.4: Annual proportion of prices ending in 99 cents over time for 10 categories with highest average proportions, 2001-2011.

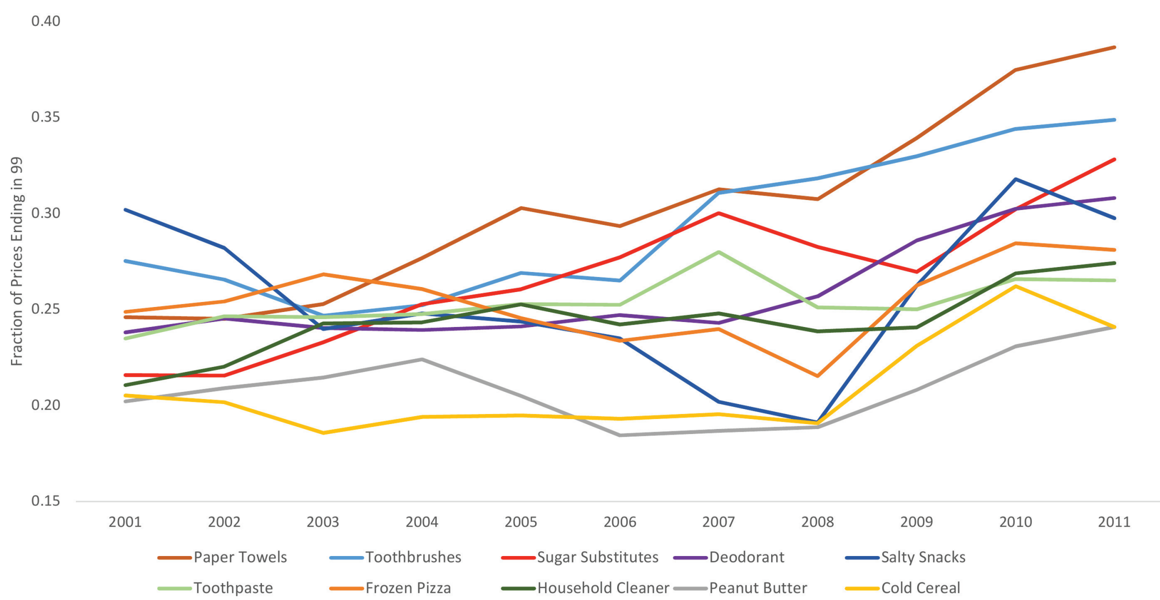


Figure 1.5: Annual proportion of prices ending in 99 cents over time for 10 categories with middle-level average proportions, 2001-2011.

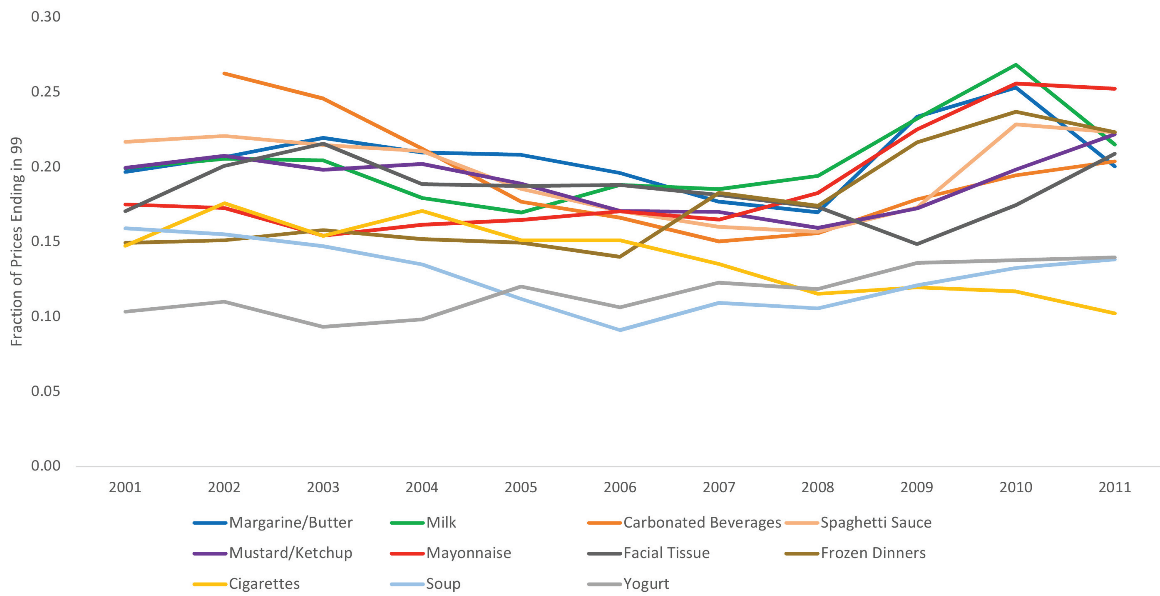


Figure 1.6: Annual proportion of prices ending in 99 cents over time for 10 categories with lowest average proportions, 2001-2011.

of prices ending in 99 cents over time not seen clearly in Figure 1.2. Categories like razors, photographic film, laundry detergent, and beer in Figure 1.4 and paper towels, toothbrushes, and sugar substitutes in Figure 1.5 are prime examples of this steady upward trend. In contrast, few, if any, categories have a persistent downward trend in the 99-cent proportion as initially suggested by Figure 1.2. However, some categories exhibit a U-shaped time trend, especially categories in the bottom 11, such as carbonated beverages and soup (Figure 1.6).[15]

Taking Table 1.3 and Figures 1.4 to 1.6 together, prices ending in 99 cents seem more likely for goods purchased in bulk or having high average prices. Goods like diapers and laundry detergent can be bought in large packages with a correspondingly high price. Similarly, razors and razor blades can be single-bladed and bought in large quantities or as expensive multi-blade razors. On the other end of the size and price spectrum are soup and yogurt, which are often bought in small, single-servings that are less than $1 as noticed in Stiving and Winer [80].

[15]The downward portion of the U-shape precedes the Great Recession, so why these categories had a noticeable decline in the proportion of prices ending in 99 is another puzzle left for future research.

1.4.3 Household-Level Data

The household-level data utilized in this paper comes from the Behaviorscan panel data component of the IRI Academic Scanner Dataset and consists of households in two American markets, Pittsfield, Massachusetts and Eau Claire, Wisconsin for the years 2008 through 2011. The Behaviorscan panel tracks households' purchases for the same 31 categories in the market-level data across all types of stores, unlike the market-level data that only contains transactions at grocery and drug stores.[16] Tracking initially relied on panelists using a wand to report purchases but later switched to a card swiped at the register to record purchases.

In addition to recording purchases of goods from the 31 categories, households also logged each trip that resulted in a purchase of a good in a tracked category and recorded the total pre-tax amount spent. Including only households that satisfy IRI's minimum reporting requirement each year for 2008 through 2011 yields a sample of 1,021 households from Pittsfield and 1,013 from Eau Claire for a total of 2,034.[17] The dataset also includes demographic data derived from a survey administered in the summer of 2012. The survey includes self-reported pre-tax income for the entire household, number of individuals in the household, and a wide range of characteristics for the head of household: race, age group, education level, occupation, and employment status.[18]

Purchase data are available at the Household-Store-UPC-Minute level and consists of a household identifier, the week of purchase using IRI's calendar system, the minute of the week when the trip occurred, the number of units purchased of the good, the type of store where the purchase occurred, the total pre-tax amount spent, a store identifier, and the product's UPC. In addition, the Behaviorscan dataset includes summarized trip data covering any trip that included a purchase of any good in the tracked categories, reporting the week and minute of the trip, an identifier for the

[16] In addition to grocery and drug stores, the Behaviorscan panel data includes a "Mass" category for other types of stores, such as Wal-Mart

[17] IRI requires households satisfy a minimum reporting requirement to be included in the data for each individual year. The requirement is that a household made at least one purchase in any of the 31 categories every 4 weeks throughout that year.

[18] Demographic information is included for the co-head of household if applicable and also contains a wide range of household descriptor variables not used in this analysis, such as home ownership, marital status, a variable describing children in the household, the number of cats owned, the number of dogs owned, the number of televisions owned, and the number of televisions connected to cable service.

store, and the total pre-tax amount spent by the household during that trip.

While a limited dependent variable estimation would seem appropriate when using a dummy variable for a good's price ending in 99 cents, relating this variable to attention through variation in shopping behavior requires aggregation since multiple goods may be part of a single trip's shopping basket and variation in shopping behavior only occurs over time. In other words, one cannot see variation in shopping or attention when using a probit or logit model when an observation is a UPC purchased by a household in a single trip. To address this issue, I aggregate to the weekly level, constructing variables for the weekly number of shopping trips, the weekly number of stores with a purchase, and a weekly measure of goods purchased ending in 99 cents for each household.

Similar to the market-level data, one measure I use for 99-cent endings at the household level is the proportion of units with a price ending in 99 cents for a household-week's purchases. In contrast to the market-level data, the proportion of posted prices ending in 99 cents may not be as relevant for households since that is more a function of how many goods the household is buying rather than how much attention it is paying to prices. Instead, I consider a different measure - the proportion of a household's weekly expenditures on goods ending in 99 cents - which captures how important 99 cent endings are to a household's spending.

Finally, the infrequency of the demographic variables complicates controlling for household characteristics over time, so the relevant price-ending and shopping variables used in the analysis in section 1.5.2 and the descriptive statistics in the following section are the weekly values averaged across the entire four-year period. This treatment of the data balances the desire to leverage the micro-level data with the need to use infrequent demographic variables and is similar to the yearly average of monthly aggregates used in Aguiar and Hurst [1].[19]

[19] While a lower level of averaging could be used by sacrificing the demographic control variables, even monthly averages of weekly behavior for some households contain no goods ending in 99 cents, which is problematic for estimating linear probability models due to observations at the lower bound of the dependent variable.

1.4.3.1 Household-Level Descriptive Statistics

Table 1.4 provides summary statistics for the 2,034 households in the data set. Statistics describing the proportion of purchases ending in 99 cents and shopping behavior of households are the four-year average of the weekly variables.[20] For example, the "Mean Weekly Proportion of Units Ending in 99" variable is a household's proportion of units bought with prices ending in 99 cents in a week averaged over the number of weeks with a purchase for that household. The value in the Mean column is then average of this variable across the 2,034 households.

The table shows that the average household spends 18.6% of their total expenditures on goods with prices ending in 99 cents while 14.8% of the units purchased had prices ending in 99 cents. However, there is a large amount of variation in these averages with a low of 1.4% of units and 2.0% for expenditures, but ranging to 48.9% of units and 62.7% for expenditures. In terms of the basket of goods households buy weekly, the average expenditure is \$25.31 across 6.32 different goods. Supporting some concerns about the reliability of households having to initially record purchases with a wand is the low minimum for average weekly expenditure, \$5.07, which means a household in the sample reported or was recorded spending only \$5.07 across all 31 categories on average, conditional on making a purchase in a week.

For the measures of shopping behavior, households typically undertake 2.81 trips per week, conditional on making at least one purchase. The minimum of 1 trip per week on average is not entirely surprising, but the maximum of 15.62 sounds rather high. However, the maximum of the number of average weekly stores, 9.57 stores, suggests that averaging 15.62 trips per week may not be too unreasonable. Despite having such a high maximum value, the average number of stores visited weekly by households is significantly lower, with households only going to 1.98 stores per week on average conditional on shopping that week.

Demographic variables, which are most often category variables, have median values reported in Table 1.5. When combining the two markets, the typical household consists of 2 people with a

[20] Since households do not make purchases every week, the four-year average is over a sample ranging from 92 to 208 weeks, depending on the household. The average number of weekly observations across households is 171 weeks.

Household-Level Statistic	Mean	S.D.	Min	Max
Mean Weekly Proportion of Units Ending in 99	0.148	0.085	0.014	0.489
Mean Weekly Proportion of Expenditures Ending in 99	0.186	0.102	0.020	0.627
Mean Weekly Expenditure	$25.31	$12.13	$5.07	$173.24
Mean Weekly Units	10.25	4.67	2.15	57.59
Mean Number of Weekly Products	6.32	2.62	1.73	24.86
Mean Number of Weekly Trips	2.81	1.49	1	15.62
Mean Number of Weekly Stores	1.98	0.96	1	9.57
Number of Retired Heads of Household	0.118	0.396	0	2
Number of Not Working Heads of Household	0.393	0.588	0	2
Proportion With Retired Head of Household	0.091	0.288	0	1
Proportion With a Not Working Head of Household	0.339	0.474	0	1

Table 1.4: Summary statistics across households for household shopping behavior and labor-related demographics (n = 2,034).

Variable	Median Category
Family Size	2
Age of Head of Household	55 to 64
Total Household Income	$35,000-$44,999
Household Education*	Graduated High School

Table 1.5: Summary statistics for household demographic category variables (n = 2,034).
* There are only 2,031 observations for household education.

head of household between the ages of 55 and 64 with only a high school diploma or equivalent. The typical household's total pre-tax income is between $35,000 and $44,999, which is below the national median income of $53,889 during this time period.[21] In terms of labor market variables, despite a large proportion of heads of households are over age 65 as indicated by the median head of household age group being 55 to 64, only 9.1% of households report a retired head of household and the average number of retired heads of households per household is 0.118 (Table 1.4). Not as surprising is the fact that 33.9% of households have at least one head of household who is not working, which could be an unemployed individual or someone not in the labor force.

A natural question to ask is which demographic characteristics are related to purchases of with 99-cent price endings. In addition to unemployment status, one might think the number of retired individuals, age, family size, income, or education may be related to such purchases. Retired

[21] 2015 dollars, BLS American Community Survey 2011-2015

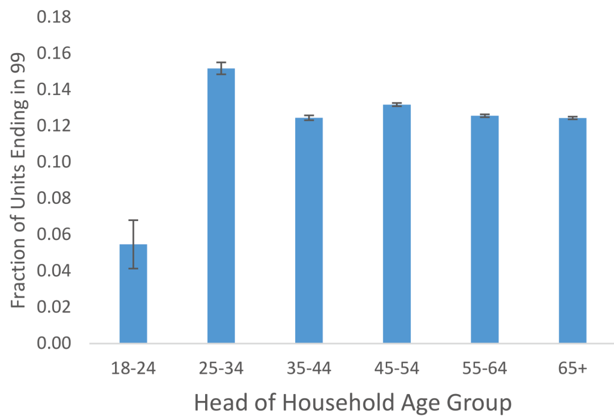
individuals may be a clear example of an attentive shopper since they spend almost no time on labor market activities and have a lower opportunity cost of time so one would expect more retirees in a household to be related to more attentive shopping and paying prices with 99-cent endings less often. Similarly, individuals at different ages may have different opportunity costs of attention and vary in their attentiveness while shopping. Larger families could possibly take advantage of the additional members and divide shopping responsibilities across family members, allowing for more attention. Alternatively, larger families may pay less attention if the heads of household have to devote more energy to taking care of other members of the household. Income and education level also have ambiguous effects as households with more income or education may have a higher opportunity cost of attention while also having characteristics that may be associated with financial acuity.

Figures 1.7 to 1.9 offer a glimpse at the variation in 99-cent endings across these demographic categories by reporting the proportion of units purchased with prices ending in 99 cents by households in each category over the sample period. In terms of age, the proportions of 99-cent endings are statistically different from each other for most age groups as indicated by the 95% confidence interval bars, but the magnitudes are relatively similar, ranging from around 12% of purchases for age groups 35 and above to a high of 15% for the 25 to 34-year-old age group (Figure 1.7b).[22]

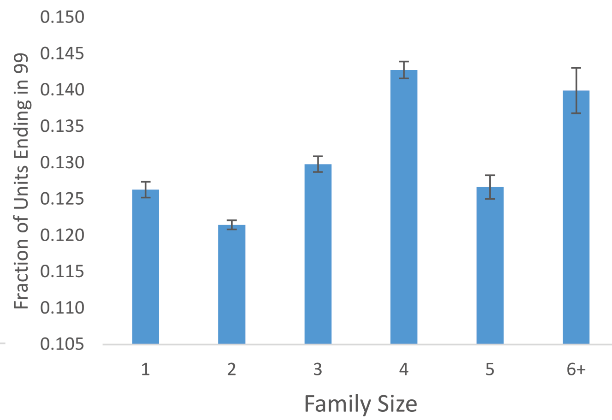
Family size also exhibits statistical differences, but the values only range from 12% to 14.3% (Figure 1.7b). Two-person households exhibit the lowest proportion of purchases with 99-cent endings while households with 4 or 6 or more members have the highest proportion, around 13% to 14%. Oddly, households with 5 members are similar to households with 1 or 3 individuals, buying goods with prices ending in 99 cents around 12.7% of the time.

The income and education categories exhibit statistically significant differences between specific levels, but no linear trend emerges (Figure 1.8). Magnitudes are again small, ranging from 12% to under 14% for income groups and 12% to 16% for education. One noticeable characteristic is that the highest and second lowest categories for each demographic variable have the highest pro-

[22]The proportion for the head of household in the 18 to 24 age group is anomalously low at 5%, but it consists entirely of observations from a single household, hence the large confidence interval.



(a) Age of primary head of household

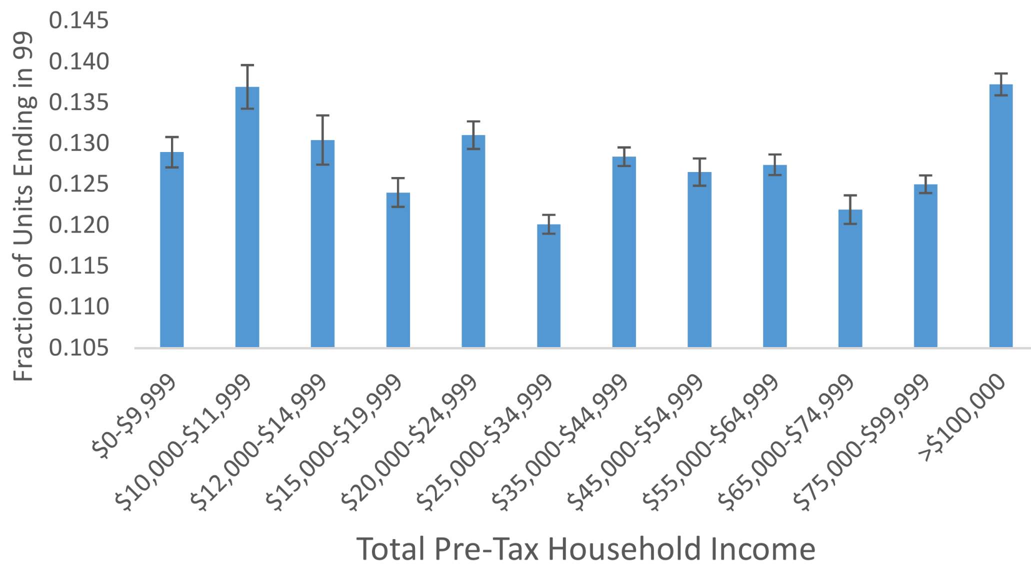


(b) Size of family

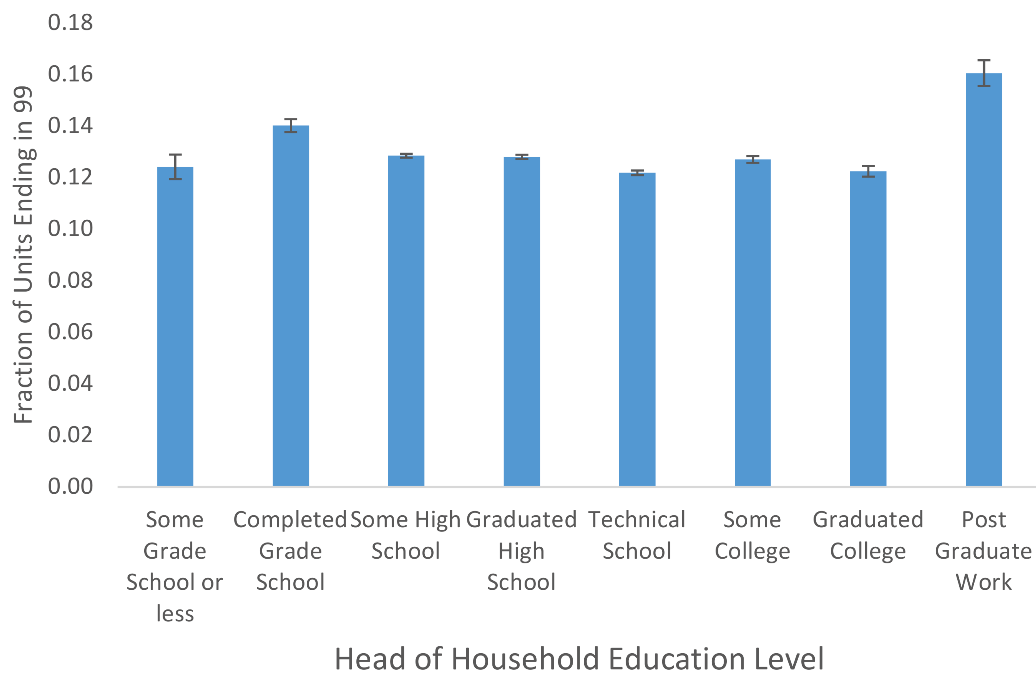
Figure 1.7: Proportion of units with prices ending in 99 cents by head of household age group or family size with 95% confidence interval bars.

portion of units with prices ending in 99 cents; incomes greater than \$100,000 or between \$10,000 and \$11,999 and education levels of Post-Graduate Work or Completed Grade School are the highest proportions for their respective demographic variables. Why the second lowest category has a higher proportion compared to the lowest category is somewhat puzzling, but could be part of a story in which less educated and less wealthy households don't have the resources available to be attentive as suggested in Banerjee and Mullainathan [6]. On the other hand, the high proportion of 99-cent endings for the highest-level categories for income and education is in line with the idea that the opportunity cost of paying attention is too high for such individuals to take note of cent endings.

In terms of the number of retired or not employed heads of household, the graphs show a difference between categories but are somewhat misleading (Figure 1.9). While it is true that the proportion of 99-cent endings is statistically different across the three possible values for not employed heads of households and the number of retired heads of households to a lesser degree, the actual magnitudes differ by tenths of a percent. The 99-cent proportion is not statistically different between having one and two retired heads of household, but households with no retired heads of households have a statistically significant lower proportion. However, this proportion is 12.7% compared to the 13.0% and 13.1% of households with one and two retired heads of

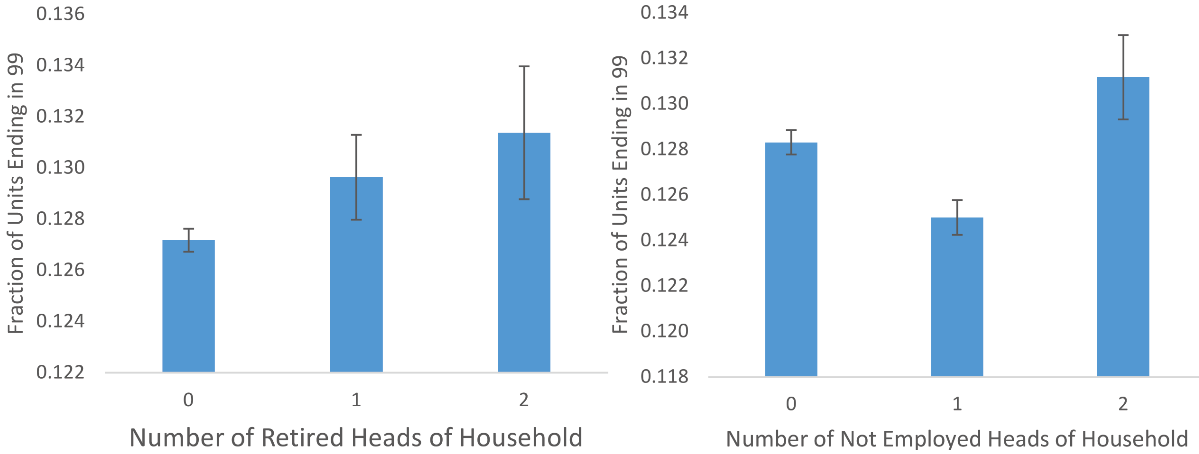


(a) Total household pre-tax income



(b) Education level of primary head of household

Figure 1.8: Proportion of units with prices ending in 99 cents by income or education level with 95% confidence interval bars.



(a) Retired heads of household (b) Not employed heads of household

Figure 1.9: Proportion of units with prices ending in 99 cents by number of retired heads of household or not employed heads of household with 95% confidence interval bars.

household, respectively. In a similar vein, households with varying numbers of not employed heads of households are statistically different from each other in terms of 99-cent proportion. Households with one not employed head of household exhibit the lowest proportion while households with two not employed heads of household have the highest proportion, which supports the notion that too much unemployment may have negative effects on the ability pay attention to prices due to more salient resource constraints. However, the magnitudes of the proportions are similar, ranging from 12.5% to 13.1% for households with one and two not employed heads of household, respectively.

Despite the statistically significant difference in the proportion of 99-cent endings for the number of retirees or not employed heads of households, the small difference in actual magnitudes does not bode well for the market-level analysis that relies on unemployment as a measure of market-wide consumer inattention. Also problematic is the non-linearity in the proportion as the number of not employed heads of household increases (Figure 1.9b). This suggests that while having one not employed family member may allow that person to split his or her time between attentive shopping and other tasks, having both heads of household not employed puts too much of a financial strain on the household to put resources towards paying attention to cent endings.[23] Nevertheless,

[23]The number of retired heads of households also seems to support this part of the story since the proportion increases with the number of retired individuals.

regression analysis may yield some surprising results and section 1.5.1 presents regression results relating the proportion of purchases involving goods with 99-cent price endings and market-level proxies for consumer inattention and section 1.5.2 reports the results at the household level using shopping behavior as the measure of attention.

1.5 Analysis & Discussion

Regression analysis using the IRI Academic Scanner Data yields mixed results. At the macroeconomic, market-level level, regressions contradict theory by indicating a positive relationship between unemployment and the proportion of purchases ending in 99 cents. On the other hand, the household-level analysis supports the inattention theory, suggesting that increased shopping intensity is related to a small but statistically significant decrease in the fraction of expenditures on goods ending in 99 cents.

1.5.1 Market-Level Analysis

A basic test of the hypothesis that 99-cent endings are related to inattention using a proxy for attention is to perform the following pooled OLS regression:

$$P_{i,t} = \beta A_{i,t} + \gamma_1 t + \mathbf{X}_i \cdot \Gamma_2 + \mathbf{X}_T \cdot \Gamma_3 + \varepsilon_{i,t}, \tag{1.6}$$

where $P_{i,t}$ is the proportion of posted prices ending in 99 cents for market i in month t, $A_{i,t}$ is the proxy for attention, e.g. the unemployment rate, for market i in month t, t is a linear time trend at the monthly frequency, X_i is a vector of dummy variables for the markets, X_T is a month of the year dummy variable, e.g. February, March, April, etc., to account for seasonality, and $\varepsilon_{i,t}$ is an error term.

Table 1.6 presents the results from the regression described by equation (1.6). The primary result, initially suggested by the correlation stylized fact, is that 99-cent endings are in fact positively related to the unemployment rate after controlling for time and market effects with a magnitude

	$P_{i,t}$	$Q_{i,t}$	$P_{i,t}$	$Q_{i,t}$	$P_{i,t}$	$Q_{i,t}$
	(1)	(2)	(3)	(4)	(5)	(6)
Unemployment Rate	0.0079***	0.0094***				
	(0.0003)	(0.0003)				
Labor Force to Pop. %			0.0015***	0.0003		
			(0.0004)	(0.0005)		
Employment to Pop. %					-0.0062***	-0.0074***
					(0.0003)	(0.0003)
Time Trend	Linear	Linear	Linear	Linear	Linear	Linear
Market Dummies	Y	Y	Y	Y	Y	Y
Month Dummies	Y	Y	Y	Y	Y	Y
R^2	0.7315	0.6093	0.6808	0.5273	0.7036	0.5656
N	6,204	6,204	6,204	6,204	6,204	6,204

Table 1.6: Results for OLS regressions relating proportion of 99-cent endings to labor market measures, 2001-2011. $*-$ 10 % significance level, $**-$ 5% significance level,$***-<1\%$ significance level.

of 0.0079 (Table 1.6, Column 1). Since the dependent variable is measured as a proportion, this coefficient can be interpreted by considering a 1 percentage point increase in the unemployment rate being related to a 0.79 percentage point increase in the fraction of prices that end in 99 cents.[24]

An alternative approach to the problem is to measure how consumers respond to 99-cent endings by changing the dependent variable to the proportion of units purchased that ends in 99 cents rather than the proportion of posted prices ending in 99 cents. Equation (1.7) summarizes the regression with the only difference from equation 1.6 being the dependent variable changing to the proportion of units purchased ending in 99 cents, $Q_{i,t}$:

$$Q_{i,t} = \beta A_{i,t} + \gamma_1 t + \mathbf{X}_i \cdot \Gamma_2 + \mathbf{X}_T \cdot \Gamma_3 + \varepsilon_{i,t}, \tag{1.7}$$

The results reported in Column 2 of Table 1.6 shows a higher impact on units sold compared to the

[24]Robustness checks accounting for sale prices, lowest-priced goods, and omitting data from 2001 are reported in Appendix A.1. In addition, a regression using heteroskedastic robust standard errors clustered at the market level does not change the main result of this regression but is included in the appendix, section A.2, for interested readers.

proportion of prices with a coefficient suggesting a 1 percentage point increase in unemployment is related to a 0.94 percentage point increase in the number of units sold ending in 99 cents.

To put both coefficients into perspective, one can calculate an elasticity of the percentage of 99-cent endings with respect to the unemployment rate using the average values for the relevant measures:

$$\varepsilon \quad = \frac{\hat{\beta}/\text{Mean percentage of 99-cent endings}}{\text{1\% Change in Unemployment Rate}/\text{Mean Unemployment Rate}} \tag{1.8}$$

Given the mean values for the percentage of 99-cent endings and mean unemployment rate (Table 1.1), the calculated elasticity for the percentage of goods with a posted price ending in 99 is 0.21 and the elasticity for the percentage of goods purchased that end in 99 cents is 0.33. While these elasticities are somewhat small, one should keep in mind that the percentage change in unemployment rates during recessions can exceed 100%, which would translate into a not insignificant shift in the percentage of prices and units sold ending in 99 cents.

As for the other measures related to the labor market, the employment to population ratio follows a similar story as the unemployment rate. More employed individuals relative to the population is related to fewer 99-cent endings, with a 1 percentage point increase in the employment to population ratio related to a 0.62 to 0.74 percentage point drop in the frequency of 99-cent endings (Table 1.6, Columns 5 and 6). However, the labor force to population ratio does in fact support the inattention model after adding sufficient control variables. Once time and market controls are added, the labor force to population ratio has a positive relationship with the proportion of prices ending in 99 cents, which matches the story of more people working or searching for jobs leading to less attentive shoppers and more 99-cent endings. In this case, a 1 percentage point increase in the labor force to population ratio is related to a 0.15 percentage point increase in the fraction of prices ending in 99 cents and a 0.03 percentage point increase in the fraction of units purchased that end in 99 cents, although the unit measure is not statistically significant. However, one must note that this result is not statistically significant when using standard errors clustered at the market level, unlike the unemployment and employment results (see Table A.11 in section A.2).

1.5.1.1 Other Macroeconomic Proxies for Consumer Inattention

When looking at wages, earnings, and hours, the results are mixed (Table 1.7). Unlike the correlations in Table 1.2, wages and earnings no longer have a strong positive relationship with 99-cent endings after adding time and market controls. While hourly wages are still positively related to the number of units sold that end in 99 cents, the coefficient is not statistically significant and is challenged by the negative and significant coefficient when looking at the fraction of prices ending in 99 cents. Furthermore, weekly earnings are now negatively correlated with both prices and units ending in 99 cents, although not always statistically significant, with a $100 increase in weekly earnings related to a 0.09 to 0.40 decrease in the frequency of 99-cent endings. The regressions involving weekly hours worked are puzzling on two levels: first, the signs are mixed between the price and unit measure of 99-cent ending, and second, the signs are now the reverse of the correlations reported in Table 1.2. In the regression, 1 more weekly hour worked is related to a 0.15 percentage point increase in the prevalence of prices ending in 99 cents and a 0.28 percentage point fall in the fraction of units purchased ending in 99 cents.[25]

The relationship between poverty and 99-cent endings is the opposite of the preliminary correlations in Table 1.2, being negative and statistically significant for both price and unit dependent variables (Table 1.8). The results suggest that a 1 percentage point increase in the poverty rate is related to approximately a 0.32 percentage point decrease in the frequency of 99 cent endings. In contrast to the correlation, these coefficients support a story in favor of the inattention model as increased poverty reflects a greater proportion of individuals with an incentive to be sensitive and attentive to price down to the penny rather than not having the spare mental resources to pay attention proposed by Banerjee and Mullainathan [6].[26]

[25]Of wages, earnings, and hours, only hours' coefficient for the price measure of 99-cent endings is statistically significant in a fixed effect regression (Table A.12).

[26]However, the coefficients are not significant when clustered at the market level but still negative in sign (Table A.13 in Section A.2).

	$P_{i,t}$	$Q_{i,t}$	$P_{i,t}$	$Q_{i,t}$	$P_{i,t}$	$Q_{i,t}$
	(1)	(2)	(3)	(4)	(5)	(6)
Hourly Wages	-0.0013*	0.0004				
	(0.0008)	(0.0007)				
Weekly Earnings			-9.13e-06	-4.03e-05**		
			(1.74e-05)	(1.59e-05)		
Weekly Hours Worked					0.0015**	-0.0028***
					(0.0007)	(0.0006)
Time Trend	Linear	Linear	Linear	Linear	Linear	Linear
Market Dummies	Y	Y	Y	Y	Y	Y
Month Dummies	Y	Y	Y	Y	Y	Y
R^2	0.8556	0.7863	0.8554	0.7867	0.8557	0.7877
N	2,820	2,820	2,820	2,820	2,820	2,820

Table 1.7: Results for OLS regressions relating proportion of 99-cent endings to wage and earning measures, 2007-2011. $*-$ 10 % significance level, $**-$ 5% significance level,$***-<$ 1% significance level.

	$P_{i,t}$	$Q_{i,t}$
	(1)	(2)
Poverty Rate	-0.3280**	-0.3164**
	(0.1581)	(0.1585)
Year Dummies	Y	Y
Market Dummies	Y	Y
R^2	0.8024	0.7548
N	517	517

Table 1.8: Results for OLS regressions relating proportion of 99-cent endings to poverty rate, 2001-2011. $*-$ 10 % significance level, $**-$ 5% significance level,$***-<$ 1% significance level.

1.5.2 Household-Level Analysis

Estimating the relationship between consumer attention, as measured by average shopping intensity, and the average proportion of units purchased with prices ending in 99 cents, $\overline{q}_i$, is done using the following equation:

$$\overline{q}_i = \beta \overline{a}_i + \mathbf{G}_i \cdot \Gamma_1 + \mathbf{C}_i \cdot \Gamma_2 + \mathbf{X}_i \cdot \Gamma_3 + \gamma_4 Eau_i + \varepsilon_i, \tag{1.9}$$

or:

$$\overline{m}_i = \beta \overline{a}_i + \mathbf{G}_i \cdot \Gamma_1 + \mathbf{C}_i \cdot \Gamma_2 + \mathbf{X}_i \cdot \Gamma_3 + \gamma_4 Eau_i + \varepsilon_i, \tag{1.10}$$

when using the proportion of household expenditure on goods with 99-cent price endings as the dependent variable, $\overline{m}_i$.[27]

The subscript i denotes denotes the household while $\overline{a}_i$ reflects the level of consumer attention measured by average weekly shopping trips or weekly number of stores visited. The variable $\mathbf{G}_i$ is a vector of controls for a household's average weekly basket of goods, consisting of the average amount spent and the average number of different products purchased. To control for possible variation due to the type of goods purchased, the household's average expenditure share for each category is represented by the vector $\mathbf{C}_i$.[28]

Other controls include $\mathbf{X}_i$, a vector of household demographic control variables covering the number of retired heads of household, number of not working heads of household, and dummy variables for family size, head of household age, household income, and head of household education. Eau_i is a dummy variable equal to 1 if the household resides in the Eau Claire market and ε_i is an error term. Of particular interest in addition to the estimate for β are the Γ_1 coefficients

[27] Recall that, unlike the market-level measures of 99-cent endings, the proportion of posted prices ending in 99 cents is not used since the decision to have prices end in 99 cents is a firm decision and does not adequately vary with a consumer's amount of attention. Also unlike the market-level analysis, there is no time element to the regression since averages are taken over the sample period in order to match the availability of the demographic data.

[28] Photographic film is the omitted category when running the regression because it has the smallest maximum expenditure across households at 7.0%.

on household basket controls, $\mathbf{G}_i$, and the Γ_3 coefficients on retirement and not employed head of household variables in $\mathbf{X}_i$.

Table 1.9 presents the regression results for equations (1.9) and (1.10) when measuring consumer attention using a household's average number of weekly trips using OLS (Tables 1.9a and 1.9b, respectively). Across both specifications, the average number of weekly trips is negative and statistically sginficant at less than the 1% level even with the full complement of control variables. Moreover, the coefficient is generally robust to household and category control variables once the Eau Claire dummy and basket controls are included. The coefficient $\hat{\beta}$ when relating to the average proportion of units with prices ending in 99 cents to average weekly shopping trips ranges from -0.0027 to -0.0031, meaning -0.27 to -0.31 percentage points when controlling for at least geography and basket. Similarly, the estimated effect when using the average proportion of expenditures spent on goods with 99-cent price endings as the dependent variable ranges from -0.0032 to -0.0036, or -0.32 to -0.36 percentage points.

Supposing that a coefficient of -0.0030, or -0.3 percentage points, generally captures the effect of increasing average weekly shopping trips on the proportion of 99-cent price endings, one would conclude that attention measured in this way does support the inattention model explaining 99-cent price endings, unlike the market-level results. However, these results suggest that the inattention model cannot be the definitive explanation because the shopping trip variable explains little of the observed variation in the 99-cent price ending proportion. Given the standard deviation in average weekly trips of 1.49, a two standard deviation change in the average number of trips, meaning increasing or decreasing average weekly trips by 2.98, would only be associated with a 0.9 percentage point change in the proportion of units or expenditures on goods with prices ending in 99 cents. Compared to the average proportions of 14.8% for units and 18.6% for expenditures, very little can be explained by attention as measured by weekly trips.[29]

In terms of the household demographic variables that are comparable to the market-level analysis, households with two retired heads of households are less likely to pay 99 cents relative to

[29]This conclusion is further supported by the low R^2, 0.0038 to 0.0039, when not including any other variables other than the average number of weekly trips (Table 1.9, Columns 1 and 6).

	$\overline{q}_i$	$\overline{q}_i$	$\overline{q}_i$	$\overline{q}_i$	$\overline{q}_i$
	(1)	(2)	(3)	(4)	(5)
Mean Weekly Trips	-0.0036***	-0.0021***	-0.0031***	-0.0028***	-0.0027***
	(0.0011)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Mean Weekly Expenditure			0.0016***	0.0015***	0.0012***
			(0.0002)	(0.0002)	(0.0002)
Mean Weekly # of Products			-0.0070***	-0.0074***	-0.0063***
			(0.0008)	(0.0008)	(0.0008)
One Retired HH				0.0032	0.0040
				(0.0047)	(0.0045)
Two Retired HH				-0.0196**	-0.0172**
				(0.0095)	(0.0088)
Eau Claire Dummy		-0.1376***	-0.1401***	-0.1409***	-0.1384***
		(0.0022)	(0.0022)	(0.0022)	(0.0031)
Household Controls	N	N	N	Y	Y
Category Controls	N	N	N	N	Y
R^2	0.0039	0.6611	0.6806	0.6922	0.7354
n	2,034	2,034	2,034	2,012	2,012

(a) Dependent Variable: Average proportion of units with price ending in 99 cents

	$\overline{m}_i$	$\overline{m}_i$	$\overline{m}_i$	$\overline{m}_i$	$\overline{m}_i$
	(6)	(7)	(8)	(9)	(10)
Mean Weekly Trips	-0.0042***	-0.0025***	-0.0036***	-0.0033***	-0.0032***
	(0.0014)	(0.0009)	(0.0009)	(0.0009)	(0.0008)
Mean Weekly Expenditure			0.0020***	0.0018***	0.0012***
			(0.0003)	(0.0003)	(0.0002)
Mean Weekly # of Products			-0.0051***	-0.0057***	-0.0036***
			(0.0011)	(0.0011)	(0.0009)
One Retired HH				0.0052	0.0072
				(0.0017)	(0.0049)
Two Retired HH				-0.0233**	-0.0188*
				(0.0113)	(0.0102)
Eau Claire Dummy		-0.1668***	-0.1691***	-0.1702***	-0.1687***
		(0.0026)	(0.0026)	(0.0026)	(0.0036)
Household Controls	N	N	N	Y	Y
Category Controls	N	N	N	N	Y
R^2	0.0038	0.6703	0.6926	0.7047	0.7669
n	2,034	2,034	2,034	2,012	2,012

(b) Dependent Variable: Average proportion of expenditures on 99-cent price ending goods.

Table 1.9: Results for OLS regressions relating proportion of 99-cent endings to average number of weekly trips.

Heteroskedastic robust standard errors in parentheses. * = 10% significance, ** = 5% significance, *** $<$ 1% significance.

households with no retired heads of household, unlike what was indicated in Figure 1.9a. The effect is 1.96% points when regressed against the proportion of units and 2.33% points when regressed against the proportion of expenditures, which is over twice the impact of what a two standard deviation increase in weekly shopping trips could achieve. However, households with only one retired head of household are not statistically different from households with no such retired persons. Dummy variables for one or two not working heads of household do not have a statistically different effect from no not-working heads of households and the coefficients are in fact positive.[30]

Controls for a household's average weekly expenditures and number of different products are both statistically significant in all cases but have opposite effects (Table 1.9, Columns 3-5 and 8-10). A one dollar increase in average weekly household expenditure is related to a 0.12 and 0.16 percentage points increase in the proportion of units with 99-cent price endings or a 0.12 and 0.20 percentage points increase in the proportion of expenditures. As a lower bound estimate, a one standard deviation increase in a household's weekly average expenditures of $12.13 would be associated with a 1.46% increase in the proportion of purchases with a 99-cent price ending, which is fairly sizable and roughly a 10% increase relative to the average values for the proportions.

Increasing the number of different products purchased on average is associated with fewer prices ending in 99 cents. Adding another good to a household's weekly shopping basket is related to a fall in the proportion of 99-cent price endings of 0.36 percentage points for the proportion of expenditure up to 0.74 percentage points for the proportion of units. However, the negative coefficient could just reflect the fact that not all goods have prices ending in 99 cents rather than being causal in a meaningful way. If not all goods have 99-cent price endings, buying a wider variety of goods should result in buying goods with other cent endings at least by chance, which would lower the proportion of purchases with 99-cent price endings.

[30]Not shown in tables due to lack of significance. As for the other household demographic controls not reported in the table, omitting the lowest categories for each variable indicates that all age groups have statistically significant differences in proportion relative to the 18 to 24 age group, which isn't surprising given Figure 1.7a. Education and Income categories are not statistically different from the lowest levels of education and income while families of 4 have a higher proportion of 99-cent endings relative to households of size 1, but all other household sizes do not have statistically significant differences in proportion relative to one-person households.

Combining the negative relationship for the average number of goods with the positive effect related to increased expenditures suggests that 99-cent endings are related with higher priced goods. If spending more raises the proportion of 99-cent endings but buying more goods decreases the proportion, it must be that higher priced goods are more likely to have 99-cent endings, which matches a model of consumers who are inattentive and ignore cent endings when the dollar value is high enough. The smaller effect of the number of products when using the proportion of expenditures as the dependent variable compared to the proportion of units supports this conclusion as higher priced goods that are more likely to end in 99 cents would have a larger positive effect on expenditures and mitigate the average negative effect on the proportion of expenditure.

The wide dispersion in the proportion of prices ending in 99 cents across categories in Figure 1.3 would suggest that controlling for the categories purchased by a household is necessary. However, including variables for the proportion of expenditures spent on each category by a household barely affects the regression results (Table 1.9, Columns 5 and 10). Only the razor blade category is statistically significant at the 10% or lower level, increasing the proportion of units with 99-cent endings by about 0.6 percentage points or proportion of revenue spent on 99-cent endings by 0.8 percentage points for every 1 percentage point increase in expenditure share spent on razor blades.[31] One possible reason for the effects from expenditure share by category not being significant is the relatively low shares for each household. Although certain households are observed spending large shares of their income on a single category, the average expenditure share for each category varied from 0.08% for razor blades to 11% for carbonated beverages.[32] Spread across all of the categories, the expenditure share for each category likely has too little of an effect on the proportion of 99-cent price endings to stand out from the other categories.[33]

While there are many insights gained from the average weekly trip variable and household

[31] Not shown for brevity.

[32] One household spent 88% of their observed expenditures on cigarettes and another spent 81% on beer. Other categories with households spending most of their expenditures in one category include households spending 77% on carbonated beverages, 71% on frozen dinners, and 54% on milk. See Table A.14 in appendix section A.3 for summary statistics for share of expenditure by category.

[33] Restricting the share of categories to only the top few categories with the highest proportion of 99-cent price endings does not affect this since the average expenditure for the top categories, such as diapers, razors, and photographic film, are each less than 0.4%.

basket controls, the most problematic coefficient is the large impact of the Eau Claire dummy variable. Adding the geography dummy variable increases R^2 values by over 0.65, accounting for the majority of variation observed in the proportion of 99-cent price endings. Moreover, the coefficient is roughly -0.14, or -14%, when related to the proportion of units and -0.17, or -17% when related to the proportion of expenditures. These effects are tremendous compared to the average values of 14.8% and 18.6% for units and expenditures, respectively, reported in Table 1.4.[34]

Why Eau Claire has such a lower proportion of prices ending in 99 cents is somewhat of a mystery because Eau Claire and Pittsfield are very similar in terms of age, income, education, and race demographics. One possible explanation could be a difference in tax rates that may affect how consumers perceive prices. Initially, both Eau Claire and Pittfield had similar sales taxes, 5.5% in Eau Claire and 5% in Pittsfield with similar tax-exemptions for unprepared food, i.e. groceries[35]. During this sample period, one major tax change occurred in the Pittsfield market in 2009, when the sales tax was raised to 6.25%. Conlon and Rao [21] provide evidence that firms use changes in taxation to time increase prices in excess of the change in tax by having prices jump to psychological price points in the alcohol industry. This sort of practice could explain why Pittsfield has a relatively higher proportion of purchases with prices ending in 99 cents and also the jump in 99-cent price endings after the Great Recession ended in 2009. One caveat for this explanation is that many goods in this sample fall under the category of groceries and are therefore tax-exempt in both markets, so the effect should depend on category. However, including category expenditure share variables in a regression do not diminish the effect of the Eau Claire dummy variable. A more in-depth analysis of the issue could pursue a difference-in-difference method to test if these two markets differ due to the sales tax increase.[36]

[34] Since Eau Claire households make up approximately half of the sample and are therefore a key contributor to the average proportions in Table 1.4, one can imagine a household that is average in all aspects switching from being the average value for the Eau Claire dummy, about 0.5, to being equal to the full value for the Eau Claire dummy and having the proportion of 99-cent price endings fall by 7% to 8.5%, or almost half of the average proportion.

[35] http://www.mass.gov/dor/businesses/help-and-resources/legal-library/documents-by-tax-type/sales-and-use-tax.html, https://www.revenue.wi.gov/WisconsinTaxBulletin/127tr.pdf, and https://www.revenue.wi.gov/Pages/FAQS/pcs-taxrates.aspx. Accessed May 17, 2018.

[36] As a first step towards understanding this difference, Table A.15 in appendix section A.3 reports the proportion of

As an alternative measure of attention, Table 1.10 displays the OLS estimation results for equations (1.9) and (1.10) when consumer attention is measured by the average weekly number of unique stores a household made a purchase in (Tables 1.10a and 1.10b, respectively). At odds with the notion that the number of stores is a better measure of attentive consumers is the fact that the average number of stores is not statistically significant for the proportion of 99-cent price endings once the dummy variable for the Eau Claire market is included. Ignoring the lack of significance, the coefficients are always negative but roughly one-third the magnitude of the coefficients for the effect when using average weekly trips. The control variables for basket and household characteristics also exhibit remarkably similar magnitudes to their values under the average weekly trip specification.[37]

Taken together, the persistent, statistically significant negative effect of additional shopping trips across all specifications and the surprising lack of a significant effect for households who visit more stores paint a picture of consumers who go on many trips on average paying the most attention to the smallest details of price compared to those who shop around at multiple stores. Those shopping at multiple stores may be motivated by other factors, such as product variety, that may not be reflected by prices ending in 99 cents. Households who go on numerous trips, possibly to the same store, may instead be able to build up a better sense of price changes over time rather within stores and know when small price fluctuations occur and are less likely to pay 99-cent ending prices.

Compared to the macroeconomic analysis using market-level labor-related variables, the microceconomic analysis of household purchases and shopping frequency offers consistent but weak support for the the model of consumer inattention explaining why prices end in 99 cents. Al-

---

prices ending in 99 cents for each category by market. While a few categories are fairly similar across markets, such as the 1% difference for carbonated beverages, most categories differ by at least 5% between the markets and some differences are fairly large, such as the 26% difference for beer. A much finer analysis at the good-market-household level could possibly tease out the nature of this difference, but is beyond the scope of this paper.

[37]To address the possibility of the shopping intensity variables being endogenous, the number of retired heads of household or not employed heads of households were considered as instrumental variables for average weekly shopping intensity but both were weak instruments from a statistical perspective. Furthermore, diagnostic tests cannot reject the null hypothesis that the shopping trip and store variables are not endogenous relative to the 99-cent price ending proportion.

	$\overline{q}_i$	$\overline{q}_i$	$\overline{q}_i$	$\overline{q}_i$	$\overline{q}_i$
	(1)	(2)	(3)	(4)	(5)
Mean Weekly Stores	-0.0073***	-0.0013	-0.0016*	-0.0011	-0.0013
	(0.0017)	(0.0010)	(0.0010)	(0.0010)	(0.0009)
Mean Weekly Expenditure			0.0016***	0.0014***	0.0011***
			(0.0002)	(0.0002)	(0.0002)
Mean Weekly # of Products			-0.0067***	-0.0072***	-0.0062***
			(0.0009)	(0.0008)	(0.0008)
One Retired HH				0.0031	0.0039
				(0.0047)	(0.0045)
Two Retired HH				-0.0194**	-0.0172*
				(0.0096)	(0.0089)
Eau Claire Dummy		-0.1376***	-0.1400***	-0.1409***	-0.1387***
		(0.0022)	(0.0022)	(0.0022)	(0.0031)
Household Controls	N	N	N	Y	Y
Category Controls	N	N	N	N	Y
R^2	0.0069	0.6599	0.6780	0.6900	0.7337
n	2,034	2,034	2,034	2,012	2,012

(a) Dependent Variable: Average proportion of units with price ending in 99 cents

	$\overline{m}_i$	$\overline{m}_i$	$\overline{m}_i$	$\overline{m}_i$	$\overline{m}_i$
	(6)	(7)	(8)	(9)	(10)
Mean Weekly Stores	-0.0088***	-0.0015	-0.0012	-0.0007	-0.0011
	(0.0021)	(0.0012)	(0.0012)	(0.0012)	(0.0010)
Mean Weekly Expenditure			0.0019***	0.0018***	0.0012***
			(0.0003)	(0.0003)	(0.0002)
Mean Weekly # of Products			-0.0048***	-0.0054***	-0.0035***
			(0.0012)	(0.0011)	(0.0009)
One Retired HH				0.0053	0.0072
				(0.0055)	(0.0050)
Two Retired HH				-0.230**	-0.0188*
				(0.0114)	(0.0104)
Eau Claire Dummy		-0.1668***	-0.1691***	-0.1703***	-0.1692***
		(0.0026)	(0.0026)	(0.0026)	(0.0036)
Household Controls	N	N	N	Y	Y
Category Controls	N	N	N	N	Y
R^2	0.0069	0.6692	0.6900	0.7026	0.7651
n	2,034	2,034	2,034	2,012	2,012

(b) Dependent Variable: Average proportion of expenditures on 99-cent price ending goods.

Table 1.10: Results for OLS regressions relating proportion of 99-cent endings to average number of weekly stores.
Heteroskedastic robust standard errors in parentheses. * = 10% significance, ** = 5% significance, *** $<$ 1% significance.

though households who make more weekly shopping trips purchase a lower proportion of goods with prices ending in 99 cents, the effect is small relative to other factors, such as the number of different goods purchased, and the fraction of variation explained by the average number of weekly trips is small as indicated by R^2 values. The negative and statistically significant effect associated with a household having two retired heads of households is another piece of evidence supporting the inattention model even after controlling for the type of goods purchased. However, the overwhelming importance of the dummy variable for the Eau Claire market suggests that the true explanation is neither consumer inattention, the category of goods purchased, or any observable household characteristics.

1.6 Conclusion

In an attempt to test whether consumer inattention can explain the prevalence of 99-cent price endings, this paper uses both macroeconomic and microeconomic data to proxy for consumer inattention and determine the relationship with the proportion of goods with prices ending in 99 cents. Using the unemployment rate as an aggregate measure of a market's general level of consumer attention leads to a result that contradicts the inattention model since the relationship between 99-cent endings and levels of attention is positive. Back of the envelope calculations suggests that the elasticity of 99-cent endings with respect to unemployment ranges from 0.17 to 0.29 for prices posted by stores and 0.23 to 0.34 for units purchased by consumers. In absolute terms, a one standard deviation increase in the unemployment rate for the relevant time period, 2001-2011, is predicted to relate to a 0.8 to 1.7 percentage point increase in the percentage of prices ending in 99-cents and an increase between 0.9 to 2.1 percentage points for units sold ending in 99 cents.

Further undermining the inattention model is the fact that the employment to population ratio and weekly earnings are negatively related to 99-cent endings while hourly wages and weekly hours worked exhibit mixed results depending on whether one looks at the percentage of prices or percentage of units sold ending in 99 cents. However, two market-level variables weakly support the inattention model: labor force participation and the poverty rate. In terms of elasticities with the proportion of 99 cent endings, the labor force measure has an elasticity between 0.09 and 0.33 while poverty's elasticity is roughly 0.18 to 0.23. These elasticities are on par with unemployment's elasticity to 99-cent endings so perhaps poverty and labor force participation are as important for understanding the story behind 99-cent endings. One caveat for the results in favor of the inattention model is that they are not significant after clustering standard errors at the market level, unlike the results contradicting the model.

In terms of the household-level analysis, there is some support for the inattention model of 99-cent endings when measuring consumer inattention using the average number of weekly shopping trips. A household taking one additional trip on average is expected to have a 0.3% lower share of units or expenditures on goods with prices ending in 99 cents. Further support for the inattention argument is that having two retired heads of households is associated with a 2% decrease in the proportion of units or expenditures on goods with 99-cent price endings.

While those effects are consistent across specifications and in the direction predicted by the inattention model, the average number of trips explains little of the variation in 99-cent endings and its effects are outweighed by control variables for the number of different products purchased and market. Also problematic is the lack of significance for a household's average number of stores with a purchase in a week and the lack of significance but positive effect from having more not employed members of a household, similar to the market-level results. These results force one to question the validity of the model, but the coefficients with respect to shopping trips and retired household members prevent the outright rejection of inattention as an explanation of the popularity of 99-cent price endings. Also problematic is the large effect explained by the dummy variable for the Eau Claire market that persists even after controlling for household demographics and basket characteristics. Households in Eau Claire on average buy 14% fewer units with prices ending in 99 cents and spend 17% less on goods with prices ending in 99 cents. This disparity varies by category so the difference could be driven by variation in the types of goods purchased by each households in each market, which could be determined by a finer analysis at the market-good-household level. Another possible explanation could be firms in Pittsfield using a 2009 sales tax increase to mask large changes in price to higher 99-cent ending prices, similar to behavior seen in Connecticut [21].

Another finding of this paper is the wide disparity in the proportion of prices ending in 99 cents across categories with an upward trend over time in 99-cent proportion within the top categories (Figure 1.3). Proportions range from as high as 60% of prices for diapers to a low around 10% for goods like yogurt and soup when considering the macro-level dataset covering 47 markets across the United States. However, when taken to the microdata, the share of expenditures spent on a category by a household does not have a statistically significant relationship with the proportion of goods purchased with a 99-cent ending price.

If inattention cannot fully explain the high percentage of prices ending in 99 cents, what other models are plausible? Could Basu's [7] concept of rational expectations of cent endings or Shy's [75] transportation costs generate a positive relationship between attentive consumers and 99-cent endings? This author's intuition suggests the answer is no, but further research and tweaks to their models could lead to interesting results. Or perhaps the current tendency to end prices in 99-cents is a result of historical prices and inertia, a sort of long-term stickiness in price endings. In earlier times, when goods were much cheaper, the difference between a dinner entree priced at 0.79 rather than 0.99 was relatively large and might make the difference between buying something extra for dessert or going home with only the main dish, rewarding attentive shoppers. However, as inflation takes its toll on prices, that 20 cent difference becomes trivial and only the most attentive, extreme shoppers pay attention to such a difference, weakening the incentive to pay any attention at all. Such shoppers may be so few, even during economic downturns, that firms see no reason not to charge the maximum number of cents without going to the next dollar. One could test this hypothesis by using data from much earlier, when prices were vastly lower in nominal terms. If the relationship between 99-cent endings and unemployment is reversed in the historical data, then this explanation may hold some validity. However, finding data spanning a sufficient time period may be somewhat difficult.

Another anecdote that could address the question of 99-cent price endings is the case of Canada, which eliminated the 1 cent coin from circulation and now rounds prices paid in cash to the nearest

5 cents (“Phasing out the Penny”, Canada Mint 2012). In such a case, why should prices ending in 99 cents exist if the consumer ultimately pays a price ending in 00, e.g. posted prices of $1.99 and $2.00 yield the same cash price at the register, $2.00? Canadian scanner data from this particular period that could be used to see how price endings change and such an analysis will be an intriguing study in the future to address the mystery behind 99-cent endings.

CHAPTER 2

A PENNY SAVED: QUANTIFYING THE VALUE OF SEARCH

2.1 Introduction

How valuable is search? The literature on search models assumes there is a benefit from searching but this assumption has not often been the focus of empirical tests. Suppose a household naïvely believes that the Law of One Price holds for typical consumer goods despite evidence to the contrary and buys all of its goods from a single store without considering other sellers.[1] Search costs aside, how costly is such behavior relative to what the household could achieve by searching? Studies attempting to quantify the value from search in the context of grocery shopping have used the number of shopping trips a household makes in a quarter or month as a measure of search intensity, finding price savings between 0.04% and 10% [1, 52]. However, empirical data offers evidence that the marginal decision to go shopping happens more than once a month or quarter, suggesting that previous studies could be improved by quantifying the value of increased shopping frequency within a week.[2]

This study estimates the value of search by taking advantage of weekly data and using a measure of search more appropriate for the weekly frequency at which new information is provided in the context of grocery shopping, finding small but statistically significant price savings as measured by the price paid for a basket of goods relative to the basket's average price that week. Using the number of shopping trips taken in a week or number of stores in a week in which a household makes a purchase as a measure of search intensity, OLS regressions suggest that doubling search intensity, most often meaning increasing stores shopped at or shopping trips from 1 to 2, lowers

[1] See any empirical paper on price dispersion for evidence that identical goods are not always the same price. Eden [35] and Eden et al. [36] describe price dispersion in data from the same source as the data in this paper.

[2] For the data used in this paper, households on average made at least 1 shopping trip in 90% of weeks over a 4 year period and went on 131.4 shopping trips annually, or 2.53 per week, which indicates that households shop for grocery store products fairly frequently. Appendix section A.4 has further evidence that suggests that households do not seem to exhibit a fixed time between shopping trips one would expect if households planned the number of shopping trips a month or quarter in advance. The existence of grocery store ads updated every week further supports the idea that households face a decision to shop each week.

the price a household pays for goods by 0.20% to 1.00% relative to the average, or about a penny per dollar.

The empirical analysis of this paper relies on data from Information Resources Inc. covering the weekly shopping trips and purchases of consumer packaged goods across a range of categories for a sample of households in two small American cities, Pittsfield, Massachusetts and Eau Claire, Wisconsin, from 2008 to 2011. The data includes information on purchase quantity and price at the transaction level and the number of shopping trips at the weekly level, allowing an analysis of the relationship between shopping frequency and price savings relative to others purchasing the same good. Price savings are measured by following the existing literatures method of comparing the price a household paid versus the average price paid for a good.

One descriptive finding is that the distribution of the number of shopping trips or stores visited in a single week is highly skewed, with over 50% of households going on one or two trips and over 75% visiting only one or two stores in a week, conditional on going shopping that week. This measure of shopping behavior exhibits much less variation compared to other studies that measured monthly or quarterly shopping behavior and suggests an analysis of the benefits from additional weekly shopping behavior is nontrivial. Furthermore, this observation somewhat supports the Burdett and Judd [17] search model's prediction that consumers do not search more than twice for a good, which lends further support to the appropriateness of using weekly shopping as the measure of search.

The small price savings effect of additional weekly shopping trips is robust to alternative estimation procedures. Controlling for possible selection bias by analyzing changes in shopping behavior within a household using a household fixed effect yields lower estimates for the price savings effect that are statistically significant. In addition, restricting a household's basket of goods to products purchased in at least two different stores, which may better capture which goods a household pays attention to when shopping, yields mildly higher estimates of the savings effect with estimates as high as 1.00% when doubling shopping frequency. Alternatively, dividing shopping trips into "regular" and "irregular" trips results in similarly tiny estimates for the price savings effect.

In light of non-trivial search costs, such as travel costs when going to multiple stores, these results suggest that price savings cannot be the primary motivation for search as assumed in many price dispersion models since saving of 1% at most can rarely overcome said costs.[3] Although identifying the exact alternative model to explain the observed search behavior is beyond the purview of the data, alternative theories include treating shopping trips as a consumption good that generates positive utility or allowing consumers to perceive value based on the discount from some reference point for price leading to a greater willingness to incur search costs, as described in the behavioral economics and marketing literature, e.g. Kalyanaram and Little [51]. In addition to the implication for price dispersion models, this paper's findings have implications for tax policy, suggesting that minor, less than 1%, increases in taxes may not be perceived by consumers since price savings of a similar magnitude cannot reasonably explain the range of consumer search for grocery store products.

2.2 Previous Research

The assumption that there is a benefit from search is so fundamental that one could cite every paper on the topic, but here I will only present a few highlights from the literature. Stigler [79] was one of the first to propose a model of individuals searching in order to purchase a product. He proposed a model characterized by fixed sample size search, where an individual chooses ex ante how many price quotations to seek and then decides where to make a purchase by choosing the store with lowest price observed among the sample. However, Stigler assumed that firms post prices from an exogenous non-degenerate price distribution and his focus was on how search generates a dispersed distribution of transaction prices as some consumers are able to find lower prices than others.

The next major step in the literature was to create model that loosened the assumption that the

[3] The average weekly basket of goods in this data costs approximately $25, which implies an average price savings of $0.25 from doubling shopping frequency in a week.

distribution of posted price was exogenously non-degenerate, which Reinganum [67] achieves by implementing heterogeneous marginal costs for firms. In addition, Reinganum also changed the type of search to sequential search, in which consumers visit one firm after another, receiving a price quote from each. A major contribution of Reinganum's model was its focus on balancing the marginal benefit gained from an additional search, which is something this paper wishes to quantify, versus the marginal cost of an additional search, which is something this paper unfortunately cannot address. Reinganum's key idea then led to the groundbreaking model by Burdett and Judd [17]. Burdett and Judd did away with heterogeneous firms and crafted a model with homogeneous firms and consumers that generated price dispersion by using fixed sample size search. In their model, each consumer receives one free search and then must decide whether to make an additional search. Given the homogeneous nature of consumers, equilibrium requires that they be indifferent between their one free price quotation and seeking a second quotation. In their model, the benefit of an additional search in terms of finding a lower price is critical in maintaining equilibrium non-degenerate price dispersion.

One final type of search model that is particularly relevant to the grocery store industry is the information clearinghouse model [82]. In a clearinghouse model, there is a centralized repository of price information that consumers must pay to access. The search decision is binary, with consumers choosing whether or not to pay the price to access the clearinghouse, which can be thought of as purchasing a newspaper that is filled with grocery store advertisements. Estimating the benefits of search in the clearinghouse model is somewhat more difficult since consumers who search have access to all prices, which is unrealistic from an empirical standpoint, but the model still utilizes the idea that those who search and obtain more price information must expect to benefit from doing so otherwise they would not have an incentive to access the clearinghouse.

In terms of empirically investigating how people shop, prior research on consumer shopping behavior typically focused on how consumers were able to achieve lower prices rather than their frequency of search. For instance, Griffith et al. [43] used a nationally representative sample of households from the United Kingdom in 2006 to study how buying on sale, buying in bulk, buying different brands, and buying at different outlets affected the price paid for food, finding that price savings from bulk purchases were largest, followed by buying on sale, purchasing generics, and finally shopping at a supermarket chain with the largest market share, Tesco. Despite describing the frequency of shopping by transportation method at the weekly level, Griffith et al. never attempted to connect price savings to the frequency of trips.

Of particular interest for earlier research on consumer shopping behavior was the intersection of income, race, culture, and shopping behavior [29, 61]. Studying 39 households in Philadelphia, Pennsylvania, Dixon and McLaughlin found that Puerto Rican households mostly frequented grocery stores owned by Puerto Ricans, paying higher prices than African-American households who typically shopped at large supermarkets. However, Puerto Rican and African-American households differed in terms of not only the set of goods typically purchased, but also in preferred brands. In another study, Lloyd and Jennings investigated the relationship between income and distance to most frequented grocery store for households in Columbia, South Carolina. They found that lower income households were more likely to primarily shop at stores closest to their residence while higher income households more often shopped for groceries at a store close to other stores they frequented, i.e. shopping centers. However, higher income households were shown to avoid nearby stores possibly due to pressure to not shop at stores associated with African-Americans. While these insights are useful in understanding how those demographics may affect the price households pay, one unanswered question is how race affects the frequency of shopping. Unfortunately, directly addressing that question is beyond the scope of this paper due to the available data consisting of households that are 98% White (Figure A.18 in Appendix section A.5). For a more representative dataset, not accounting for racial diversity in shopping behavior could bias estimates in unanticipated ways. This issue is mitigated for this paper by presenting results with a household fixed effect and by the fact that the sample of households used is overwhelmingly White but estimates may not be applicable for households in the racial minority.

In terms of studying how the number of shopping trips affects the price paid for goods, Aguiar and Hurst [1] offered one of the first studies using novel scanner data that included household shopping behavior. Using the Nielsen Homescan panel of households in Denver from January 1993 to March 1995, Aguiar and Hurst estimated the relationship between the number of shopping trips a household made in a month and a price index constructed by calculating a household's expenditure on its monthly basket of goods relative to what the basket would have cost if the basket was purchased at the average monthly price. In their preferred specifications using a household's income, age, or family size as instrumental variables for the number of shopping trips in a month, they estimated a decrease in the price index of 7% to 10% when doubling the number of monthly trips. However, one must keep in mind that these estimates came from data collected in an outdated search environment before the widespread adoption of the Internet and e-commerce. While access to price information on the Internet has not completely eliminated price dispersion, increased use of online price comparison sites has been shown to decrease price dispersion [81], which implies that a retail environment without such an online resource, e.g. Denver in the mid-1990s, could have a high degree if price dispersion that allows for shoppers to find lower prices compared to their non-shopping counterparts.

Kaplan and Menzio [52] also estimated the price savings from search using the much richer Kilts-Nielsen Consumer Panel dataset covering the purchases of approximately 50,000 households across 54 markets in the United States from 2004 to 2009. When defining a good by its brand, product characteristics, and size rather than UPC or barcode as in Aguiar and Hurst [1], they estimated that shopping at an additional store in a quarter lowered price paid by 0.6% relative to the average while an additional shopping trip in a quarter reduced price paid by only 0.04% after including household coupon usage. They also reported price savings estimates of 1.06% from visiting one more store and 0.14% from making one more shopping trip when not controlling for coupons. In addition to these estimates, the primary finding of their work was that search behavior and search frictions could only explain 35% to 55% of observed dispersion while intertemporal substitution to take advantage of temporary price reductions accounted for the remaining dispersion. Their conclusion offered one answer to the question of whether price dispersion models should rely on search as the primary mechanism that generates the observed dispersion, suggesting that intertemporal models are more important However, their use of quarterly aggregation when households may decide to search weekly necessitates further research into the question.

Although Kaplan and Menzio [52] use a much richer dataset than Aguiar and Hurst [1] and even this paper, one drawback from their estimation is that they used a larger level of aggregation, aggregating the basket of goods and measures of search to the quarterly level. Aguiar and Hurst are also guilty of this to a lesser degree, aggregating to the month and then performing an annual-level regression after averaging over months. Given the weekly frequency of the raw scanner data in both papers, aggregation may throw away valuable information about the value of shopping. Furthermore, such aggregation implies that households decide on how much to shop only once a quarter or once a month. This seems unrealistic intuitively and the number of days between trips does not exhibit a periodicity one would expect from planning shopping trips far in advance (See Appendix section A.4). Perhaps greater evidence in support of analyzing shopping at the weekly frequency is the schedule of grocery store ads. Visiting most U.S.-based grocery stores or grocery store websites will tell you that they release a weekly ad, featuring the prices of particular products. This suggests that there is new price information every week for grocery products, which means households have an incentive to consider shopping every week. Given this omission in the literature, it is only natural to study the relationship between shopping behavior at the weekly level and price savings.

2.3 Data Description

The data utilized in this paper comes from the Behaviorscan panel data that is part of Information Resources Inc.'s Academic Scanner Dataset.[4] The Behaviorscan panel tracks households' purchases of consumer packaged goods measured at the Universal Product Code (UPC), or barcode, level for 31 categories,[5] relying on either a card similar to a grocery store loyalty card to track

[4] A complete description of the data is available in Bronnenberg et al. [16].

[5] The 31 categories covered are: beer, razor blades, carbonated beverages, cigarettes, coffee, cold cereal, deodorant, diapers, facial tissue, frozen dinner entrees, frozen pizza, household cleaners, hotdogs, laundry detergent, margarine and butter, mayonnaise, milk, mustard and ketchup, paper towels, peanut butter, photographic film, razors, salty snacks, shampoo, soup, spaghetti sauce, sugar substitutes, toilet tissue, toothbrushes, toothpaste, yogurt

transactions or a wand that households use to manually record their purchases from any store, although all panelists were eventually converted to using a card to track purchases.[6] In addition to recording purchases of goods from the 31 categories, households also logged each trip that resulted in a purchase of a good in any of the tracked categories and recorded the total pre-tax amount spent for the entire trip on all goods. The full panel covers households in two markets, Pittsfield, Massachusetts and Eau Claire, Wisconsin, from 2001 to 2011, but this study focuses on the years 2008 through 2011 in order to maintain a sizable number of panelists who fulfill the minimum reporting requirement for the entire period of analysis.[7] Including only households that satisfy the minimum reporting requirement each year for 2008 through 2011 yields a sample of 2,034 households, 1,021 from Pittsfield and 1,013 from Eau Claire.

Purchase data consists of a household identifier, the week of purchase using IRI's calendar system, the minute of the week when the trip occurred, the number of units purchased of the good, the type of store where the purchase occurred, the total pre-tax amount spent, a store identifier, and the product's UPC. Figure 2.1 illustrates a sample purchase observation that describes household 1100180 purchasing 1 unit of product 0013410057341 on January 26, 2008 at 5:45 PM from a grocery store identified as 9999879 for $19.99 before tax.[8] The trip data in the Behaviorscan dataset captures each trip that included a purchase from any good in the tracked categories, reporting the household, week and minute of the trip, an identifier for the store, and the total pre-tax amount spent during that trip. Figure 2.2 provides two sample observations from the trip data, showing a household, identified by 11000016, shopping at the same store, store number 650679, approximately 35 weeks apart, spending $8.77 during the first trip and $19.43 on the latter trip.

[6]The set of stores covered by this data is unlike the national scanner data portion of the IRI Academic Dataset that only covers transactions at grocery and drug stores. The Behaviorscan panel data includes a "Mass" category for other types of stores, such as Wal-Mart.

[7]IRI requires households to satisfy a minimum reporting requirement to be included in the data for each individual year. The requirement is that a household make at least one purchase in any of the 31 categories every 4 weeks within a year. Like other panel data on consumer purchases, there is some concern with the accuracy of households reporting their purchases. While some households, 760 of 2,034, started by using a wand to scan the items purchased, making households responsible for the accuracy of the data, all households either started with or transitioned to a system similar to a shopper loyalty card that accurately recorded purchases at checkout.

[8]The product happens to be 360 ounces of Miller Lite beer packaged as cans in a cardboard box.

PANID	WEEK	MINUTE	UNITS	OUTLET	DOLLARS	IRI_KEY	COLUPC
1100180	1482	8265	1	GK	19.99	9999879	0013410057341

Figure 2.1: Example purchase observation.
PANID = household identifier, WEEK = week the purchase occurred, MINUTE = minute of the week the purchase occurred, UNITS = number of units purchased, OUTLET = the type of store where the item was bought (GK = grocery store), DOLLARS = total pre-tax amount spent, IRI_KEY = store identifier, COLUPC = Universal Product Code of good.

PANID	WEEK	MINUTE	IRI_KEY	KRYSCENTS
1100016	1636	8408	650679	877
1100016	1672	5228	650679	1943

Figure 2.2: Examples of trip observations.
PANID = household identifier, WEEK = week the trip occurred, MINUTE = minute of the week the trip occurred, IRI_KEY = store identifier, KRYSCENTS = total cents spent on that trip.

2.3.1 Shopping Behavior

One critical question with no definitive answer currently is how to define a search using shopping behavior. Although researchers have distinguished different methods of searching in terms of fixed sample, sequential, or information clearinghouse search, what exactly characterizes a search is not discussed. Should one consider a trip to a store, possibly the same store multiple times, a search or should one only count the number of different stores a person visits as the number of searches? Empirically and intuitively, the number of different stores seems to be the most important measure [52]. However, going on multiple trips, perhaps even to the same store, may be indicative of search intensity and therefore an important aspect of shopping frequency.[9] As such, I describe two measures of shopping frequency in this section and include both in the analysis in Section 2.4. In particular, I aggregate to the week level, summing the number of trips a house-

[9]Another concern may be the unmeasured search consumers perform online. However, there are numerous factors that mitigate this concern for the IRI data. First, the households are located in small American cities and are much older than the average American household (see Figure 2.5), which may be associated with a lower likelihood of using online shopping due to the learning costs associated with the technology. Second, the time frame of the data, 2008 to 2011, was before the ubiquity of smartphones so it is likely that only a small proportion of households had access to online prices while in stores. Finally, the type of grocery store products in the data may not have been profitable to sell online with the current shipping infrastructure given how long grocery stores have taken to implement online ordering systems, e.g. a large chain, Kroger, only began offering the option to a limited number of cities in 2014 (http://www.cincinnati.com/story/money/2015/06/10/kroger-rolling-online-shopping/71009794/ Accessed October 18, 2017.).

hold made in a week and counting the number of unique stores a household made a purchase at that week, which results in a sample of 348,283 Household-Week observations across the 2,034 households.[10]

Table 2.1 summarizes shopping frequency at the household-week level, meaning the number of trips or the number of unique stores a household made a purchase at in a particular week. On average, households made 2.886 observed shopping trips per week and buying items in 2.007 different stores each week. However, there is a wide range in behavior for specific household-week observations given that at least one household went on 26 trips in one week and a household bought goods at 18 different stores in a single week. Figures 2.3 and 2.4 provide a better picture of typical shopping behavior by plotting the distribution of household-week observations. Figure 2.3 shows that most households made 1 or 2 trips per week, possibly to the same store. However, over a quarter of household-weeks consisted of 3 or 4 trips and over 15% consisted of 5 to 10 trips in a week.

Figure 2.4 offers an even starker result for weekly shopping frequency. Nearly half of all household-weeks consisted of trips to a single store and the proportion of observations rapidly declines as the number of stores increases. In fact, approximately 75% of observations made a purchase in at most 2 stores and fewer than 10% bought goods at 4 or more stores. What is somewhat intriguing is that the empirical data supports the Burdett and Judd [17] model's assumption that no consumers seek more than two price quotations in equilibrium. While not strictly true, over 50% of household-week observations in terms of trips and over 75% of observations in terms of unique stores with a purchase exhibit this behavior, suggesting that the model's equilibrium condition is not entirely unrealistic.

Similar to Aguiar and Hurst [1] and Kaplan and Menzio [52], I measure price savings by constructing a price index for each household that calculates the price the household paid for its

[10] A key weakness of this data, and all scanner-based data addressing this topic, is that one cannot observe searches that did not result in a purchase of a good. For instance, if a consumer walks into a store and looks around but does not buy anything, that should most certainly be counted as a search but it is impossible to account for that trip since there is no record of the visit due to the lack of a purchase. The empirical analysis in section 2.4.2 proposes one way to address this problem using goods known to have been purchased at more than one store.

Shopping Measure	Mean	s.d.	Min	Max
Number of Trips in Week	2.886	2.103	1	26
Number of Stores in Week	2.007	1.355	1	18

Table 2.1: Summary statistics for shopping behavior at Household-Week level (n = 348,283).

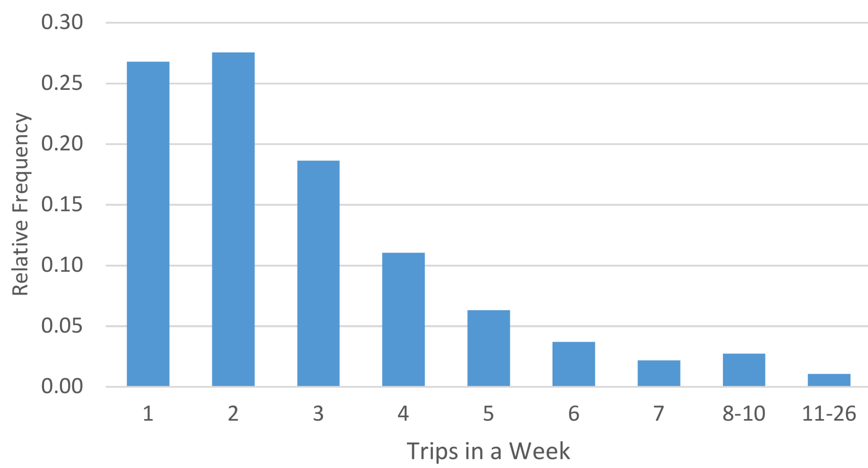


Figure 2.3: Trips per week

Figure 2.4: Unique stores with a purchase per week

basket of goods relative to the average price paid for the same basket across all households who purchased the basket's goods, but instead I use weekly average price for the index rather than monthly or quarterly average price. Mathematically, I calculate the price index, $P_{i,t}$:

$$P_{i,t} = \frac{\sum\limits_{j \in J_{i,t}} p_{i,j,k,t} \cdot q_{i,j,k,t}}{\sum\limits_{j \in J_{i,t}} \overline{p_{j,k,t}} \cdot q_{i,j,k,t}} \tag{2.1}$$

where i indexes the household, j indexes goods at the UPC level, $J_{i,t}$ is the set of goods purchased by household i in week t, k is the market (Eau Claire or Pittsfield), t indexes the week, $p_{i,j,k,t}$ is the price household i living in market k paid for good j in week t, $q_{i,j,k,t}$ is the quantity of good j purchased in week t by the household, and $\overline{p_{j,k,t}}$ is the average price paid for good j in market k in week t across all consumers who purchased good j that week. Intuitively, the price index compares what a household paid to the price it would have paid if the household were able to purchase its basket of goods at the average prices paid that week.[11]

2.3.2 Demographic Data

Demographic data describing the households is available in two snapshots, one from surveys administered to households throughout 2007 and the other administered in the summer of 2012. Since some households entered the panel in 2008 and therefore do not have demographic data from 2007, I use only the demographic data from the 2012 survey. The survey includes self-reported combined pre-tax income for the household, number of individuals in the household, home ownership, marital status, a variable describing children in the household, the number of cats owned, the number of dogs owned, the number of televisions owned, the number of televisions connected to cable service, and a wide range of characteristics for the head of household: race, age group, education level, occupation, and employment status, with the same information available for the head of household's partner when applicable.

[11] See Appendix section A.6 for a numerical illustration of how the price index is constructed.

Variable	Median Category
Family Size	2
Head of HH Age	55 to 64
Head of HH Education	Graduated High School
Total HH Annual Pre-tax Income	$35,000-$44,999

Table 2.2: Median demographic values for households, n = 2,034.

Variable	Fraction of Households
Any Retired Head of HH	0.0910
Any Not Working Head of HH	0.3392
White Head of Household	0.9851
Own Residence	0.1300
Have at least 1 child	0.1794

Table 2.3: Fraction of households exhibiting other demographic characteristics, n = 2,034.

Since the demographic variable are usually delineated by category, the most meaningful summary statistic of the demographic data is the median, which is reported for a few characteristics in Table 2.2. However, a few categories are better described as fractions of households exhibiting the characteristic, which are reported in Table 2.3. If one were to summarize the typical household in the sample, it would be a white, two-person household headed by someone between the ages of 55 and 64 who graduated from high school but did not attend college and earns between $35,000 and $44,999 annually.

The sample of households from Pittsfield and Eau Claire differs significantly from the general American population in numerous ways.[12] Of primary concern is the skewness in age of primary head of households towards 65 years old and older (Figure 2.5). Aguiar and Hurst [1] find that older households adjust their shopping behavior, spending more time shopping, which suggests

[12] Section A.5 has additional information on how this sample differs from the general American population. In particular, the racial composition of the sample's households is vastly different from the American population, with over 95% of households reporting as white (Figure A.18), but it is not clear how race should affect the frequency of shopping outside of shopping at specific ethnic markets and the effects correlated with income, education, and employment status. As mentioned in section 2.2, Dixon and McLaughlin [29] studied potential differences in shopping behavior due to race using a set of 39 households in Philadelphia, finding that African-American households were more likely to frequent a large chain supermarket while Puerto Rican households preferred small grocery stores owned by Puerto Ricans and paid more relative to African-American households. Lloyd and Jennings [61] also found that there was some pressure to patronize stores based on race in South Carolina, but income was more important for predicting where households shopped. Nevertheless, both papers do not address the frequency of shopping.

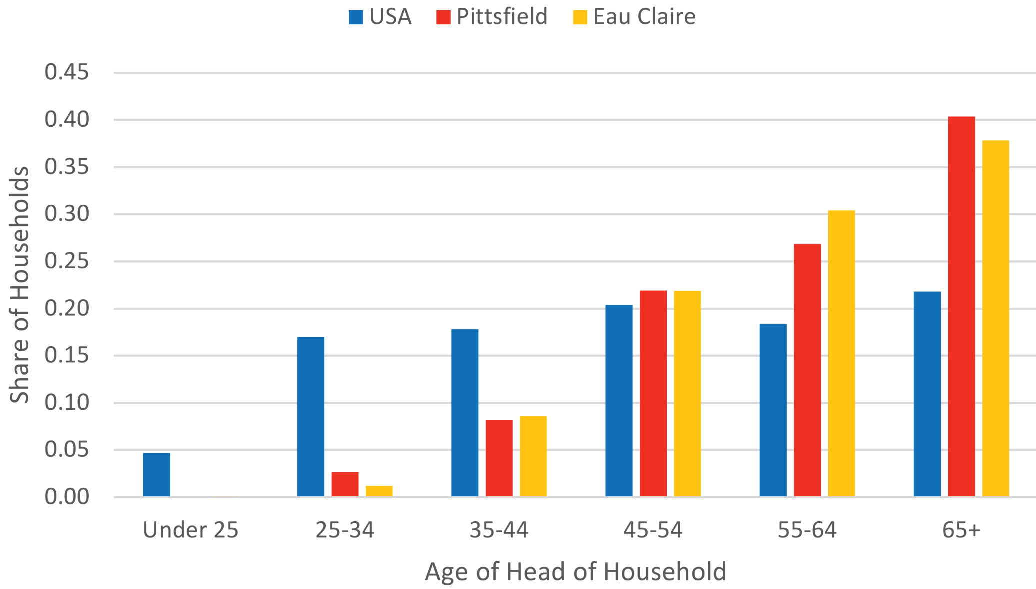


Figure 2.5: Distribution of age of head of household for the United States versus Pittsfield and Eau Claire Markets, $n_{Pittsfield}$ = 1,013, $n_{EauClaire}$ = 1,010*.
*8 households in Pittsfield and 3 households in Eau Claire did not report age.

that one would expect the IRI sample would yield estimates of the savings effect of search that are biased downwards towards 0 as the sample consists of households who are most prone to shop frequently. This biased sample would not only be capturing shopping frequency occurring at the extreme margin of search behavior, where the marginal benefits of additional search would be minimal, but also result in a smaller spread in the price index as the majority of households are shoppers who are able to find the lowest price and lower the average price paid. Mitigating this concern is the fact that less than 10% of households have at least one retired head of household, suggesting that heads of households continue to work past age 65. However, over one-third of households have at least one head of the household not currently employed so the opportunity cost of time may still be lower than average for these households and lead to above average shopping frequency. Despite these concerns, Figures 2.3 and 2.4 still suggest that 25% to 50% of households in this sample do not perform more than two searches in a week.

To test the hypothesis that the IRI sample may yield estimates that are biased towards 0, I replicate select regressions from Aguiar and Hurst [1] and Kaplan and Menzio [52].[13] Table 2.4 summarizes the results of the replicated regressions. The estimated coefficients are almost always lower when using the IRI data, especially when using the instrumental variable approach of Aguiar and Hurst, suggesting that a downward bias exists. However, diagnostic tests for the instrumental variables suggests that the instruments are weak when used with the IRI data and the OLS estimates when using their method are in fact higher when using the IRI data so the bias in the IRI data is not entirely conclusive.[14] In addition, as mentioned in section 2.2, the data used by Aguiar and Hurst comes from the mid 1990s, when current information technology was still in its infancy, so the search environment could have allowed for more price dispersion and a greater benefit from search. As for the replication of Kaplan and Menzio's regressions, the estimated effect for increasing the number of quarterly trips is significantly smaller when using the IRI data, 75% smaller than their estimate, but the estimate for the effect of increasing the number of stores is only about 30% smaller. Despite the possibility of a downward bias in the estimate of the savings effect from search as suggested by the replicated regressions, the estimates from the IRI data can still be meaningful when treated as a lower bound for how much one can benefit from additional search.

2.4 Analysis & Discussion

Search models are typically static models where search only provides information on prices in the current period. However, if prices are serially correlated across time for stores, there could be additional price information gleaned from shopping in a particular week since a store with a relatively low price in a prior week is likely to have a relatively low price in the current period.[15]

[13] To the author's knowledge, the replication of Aguiar and Hurst [1] are nearly exact replications of the regressions aside from some minor differences in how age and income groups are defined. For the replication of Kaplan and Menzio [52], dummies for different age groups are used instead of a cubic polynomial in age due to the data available in the IRI dataset. This change may explain the much lower estimates for the price savings effect as age dummies may control for more of the observed variation compared to a polynomial.

[14] In addition to being weak instruments, a Hausman test suggests that the number of trips is not an endogenous variable when comparing OLS to the IV estimates, but measurement error is still a valid concern.

[15] Note that this effect is distinct from identifying low versus high priced stores or accounting for store quality. A store could consistently price goods lower than the average price, but if temporary price reductions from the store's regular price last over weeks, visiting that store in one week would provide information on price over time not captured by a store fixed effect.

Aguiar and Hurst [1]			
	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$
ln(Number of Trips)	-0.010	-0.189***	-0.074***
	(0.006)	(0.035)	(0.032)
Type	OLS	IV: Age	IV: Income
Frequency	Monthly	Monthly	Monthly
Dataset	ACNielsen '93-'95	ACNielsen '93-'95	ACNielsen '93-'95
Replication	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$
ln(Number of Trips)	-0.014***	-0.043***	-0.040***
	(0.001)	(0.008)	(0.010)
Type	OLS	IV: Age	IV: Income
Frequency	Monthly	Monthly	Monthly
Dataset	IRI '08-'11	IRI '08-'11	IRI '08-'11

Kaplan and Menzio [52]		
	$P_{i,t}$	$P_{i,t}$
No. of Trips $\times 100$	-0.144***	
	(0.005)	
No. of Stores $\times 100$		-1.063***
		(0.027)
Frequency	Quarterly	Quarterly
Dataset	Kilts-Nielsen '04-'09	Kilts-Nielsen '04-'09
Replication	$P_{i,t}$	$P_{i,t}$
No. of Trips $\times 100$	-0.036***	
	(0.001)	
No. of Stores $\times 100$		-0.706***
		(0.017)
Frequency	Quarterly	Quarterly
Dataset	IRI '08-'11	IRI '08-'11

Table 2.4: Replication of prior research using the IRI dataset.
$P_{i,t}$ is the ratio of price for the basket of goods a household, i, paid relative to the price of the basket at the average price paid in time t.

Under this condition, a buyer who searches in one week will accumulate a stock of useful information about prices that will have a benefit in future periods. The weekly data from IRI is particularly suited for testing this hypothesis.

To test the hypothesis that search in a particular week has benefits over time, I run a regression in the following form:

$$\begin{aligned} P_{i,t} = \alpha_0 + \alpha_1 (Shopping\ Measure)_{i,t} + \alpha_2 (Shopping\ Measure)_{i,t-1} \\ + \alpha_3 (Shopping\ Measure)_{i,t-2} + \alpha_4 (Shopping\ Measure)_{i,t-3} + \mathbf{X}_{i,t} \cdot B + \varepsilon_{i,t} \end{aligned} \tag{2.2}$$

where $P_{i,t}$ is the price index described in equation (2.1), i indexes the household and t indexes the week, $\mathbf{X}_{i,t}$ is a vector of controls for the household's basket, and $\varepsilon_{i,t}$ is an error term.[16] The shopping measures for t through $t-3$ capture shopping behavior of the current week and prior 3 week period, as suggested by monthly aggregation, in terms of trips or stores shopped at for that week, and includes zeroes for the weeks a household is not observed making a shopping trip or buying a good at a store. A reasonable expectation is that the value of accumulated information decays over time as prices from many weeks ago are less predictive of current price, which would suggest:

$$\alpha_1 > \alpha_2 > \alpha_3 > \alpha_4 \tag{2.3}$$

In contrast, aggregating above the week level, to the month level as in Aguiar and Hurst [1] or the quarterly level in Kaplan and Menzio [52], assumes that information from searches in the aggregated time period provide equally valuable information with no decay. In the case of monthly aggregation over 4 weeks, this would imply:

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 \tag{2.4}$$

[16]Following Aguiar and Hurst [1], the basket controls are the log of the number of UPCs purchased, the log of the number of different categories purchased, and the log of the total cost of the basket of goods purchased with the addition of a dummy variable for households in the Eau Claire market.

In order to simplify the measurement of shopping behavior in the empirical analysis, I use dummies variables for whether a household went on two or more trips (or stores) in a particular week rather than using the raw number of trips (or stores).[17] This results in the equation that I estimate:

$$P_{i,t} = \alpha_0 + \alpha_1 I(High\ Shopping)_{i,t} + \alpha_2 I(High\ Shopping)_{i,t-1} + \alpha_3 I(High\ Shopping)_{i,t-2} + \alpha_4 I(High\ Shopping)_{i,t-3} + \mathbf{B} \cdot X_{i,t} + \varepsilon_{i,t} \tag{2.5}$$

where $I(High\ Shopping)_{i,j}$ is equal to 1 if household i went on two or more trips (or stores) in week j.

Table 2.5 provides estimates for the α_i coefficients using OLS.[18] Whether measuring additional search by trips or stores, the assumption that $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ fails. With both measures, shopping in the current week is two to three times more valuable in terms of lowering the price index compared to other weeks, 0.44% for making two or more trips in the current week and 0.56% for buying goods at two or more stores in the current week. Including both store and trip shopping dummies seems to fulfill the condition that $\alpha_2 = \alpha_3 = \alpha_4$ in terms of stores, but shopping at two or more stores in the current week is still twice as valuable as any of the prior 3 weeks. Furthermore, the estimated effect from going on multiple trips either becomes statistically insignificant or puzzlingly positive, albeit small in magnitude. The general results also suggest a violation of equation (2.3) that assumes that the lagged coefficients are strictly decreasing. Regardless of the measure of shopping, the benefit of searching three weeks ago, captured by α_4, is higher than searching one or two weeks ago. This may be capturing a pricing strategy by stores that changes prices every three weeks and suggests a more sophisticated model of how information from search decays is necessary.

One observation useful for comparison to earlier research and the results later in this paper

[17] This specification mitigates the effect of outliers and is inspired by search models of price dispersion that predict consumers do not search more than twice, e.g. Burdett and Judd [17].

[18] Note that the sample size is smaller than later household-week samples in this paper due to losing the first three weeks when constructing the lagged shopping variables.

	$P_{i,t}$	$P_{i,t}$	$P_{i,t}$
	(1)	(2)	(3)
$I(High\ Trips)_{i,t}$	-0.0044***		-3.20×10^{-07}
	(0.0003)		(0.0004)
$I(High\ Trips)_{i,t-1}$	-0.0012***		0.0008**
	(0.0002)		(0.0003)
$I(High\ Trips)_{i,t-2}$	-0.0016***		0.0007**
	(0.0002)		(0.0003)
$I(High\ Trips)_{i,t-3}$	-0.0022***		0.0001
	(0.0002)		(0.0003)
$I(High\ Stores)_{i,t}$		-0.0056***	-0.0055***
		(0.0003)	(0.0003)
$I(High\ Stores)_{i,t-1}$		-0.0017***	-0.0023***
		(0.0002)	(0.0003)
$I(High\ Stores)_{i,t-2}$		-0.0021***	-0.0026***
		(0.0002)	(0.0003)
$I(High\ Stores)_{i,t-3}$		-0.0026***	-0.0026***
		(0.0002)	(0.0003)
Controls	Market & Basket	Market & Basket	Market & Basket
R^2	0.0070	0.0096	0.0096
n	341,557	341,557	341,557

Table 2.5: OLS results when relating price index to lagged shopping behavior at household-week level.

Standard errors in parentheses, * = 10% significance, ** = 5% significance, *** $<$ 1% significance.

is that summing the four coefficients gives an idea of the total savings expected from increased shopping when treating the information gained from shopping as a stock of information that decays entirely after 4 weeks. In the case of trips, the total value of going on two or more trips in a given week is a 0.94% reduction in the price index while the total value from buying goods at two or more stores is a 1.20% reduction. These values are vastly different from the IV estimates from Aguiar and Hurst [1] but are on par with their OLS estimates. However, the more important result is the coefficients on lagged shopping variables are not identical, suggesting that monthly aggregation, and likely quarterly aggregation, is not appropriate from an empirical standpoint and therefore the next step is to quantify the value of shopping at the more relevant weekly frequency.

2.4.1 The Weekly Value of Search

To measure the value of increased shopping frequency at the weekly level, I perform a regression similar to Aguiar and Hurst [1] of the form:

$$\log P_{i,t} = \beta_0 + \beta_1 \log(Shopping\ Measure)_{i,t} + \mathbf{B} \cdot X_{i,t} + \varepsilon_{i,t} \tag{2.6}$$

where variables are indexed in the same way as equation (2.1), Shopping Measure is in terms of number of trips or unique stores shopped at in a week, $X_{i,t}$ is the same vector of controls for the market and basket of goods purchased used in (2.5), and ε is an error term.[19]

Results from the OLS regression described in equation (2.6) are reported in Table 2.6. Although the regression coefficients can be understood as the elasticity between the price index and shopping behavior, a reasonable way to interpret the coefficients is to consider the effect of doubling the number of trips or stores visited, which would be a 100% change in shopping behavior. On the surface, this may seem like a large change, but Figures 2.3 and 2.4 suggest that this is not unreasonable given that most people go on 1 or 2 trips and buy goods at 1 or 2 stores in a week. This interpretation seems especially apt for the number of stores with a purchase given the sharp

[19] i indexes the household and t indexes time.

	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$
	(1)	(2)	(3)	(4)
ln(Number of Trips)	-0.0042***		-0.0055***	
	(0.0002)		(0.0002)	
ln(Number of Stores)		-0.0076***		-0.0082***
		(0.0002)		(0.0002)
Controls	Market	Market	Market & Basket	Market & Basket
R^2	0.0020	0.0046	0.0146	0.0168
n	348,283	348,283	348,283	348,283

Table 2.6: Results for OLS regressions relating price index to shopping behavior at household-week level.
Standard errors in parentheses, * = 10% significance, ** = 5% significance, *** $<$ 1% significance.

drop off and seemingly exponential decline in the distribution of that measure of search.

For the number of trips, the coefficient is negative, as predicted, statistically significant, and robust to the addition of the basket control variables. In terms of magnitude, a doubling in the number of shopping trips in a week, most commonly going from 1 trip to 2 trips, is associated with a 0.42% lower price paid for the entire basket. The number of stores with a purchase also exhibits the expected sign and maintains statistical significance after adding basket controls. In this case, doubling the number of stores in a week is related to a 0.82% decline in the price index, almost double the savings from a similar increase in the number of trips and in line with Kaplan and Menzio's [52] estimate of a 0.6% to 1.1% discount for visiting an additional store in a quarter.

In addition to the regression using OLS, I run equation (2.6) with a household fixed effect instead of separate demographic variables and cluster standard errors at the household level. In this case, identification comes from variation in shopping behavior across weeks for individual households, which might be considered a more valid measure of the marginal benefit of additional searches. Table 2.7 reports the results from the fixed effect regressions. The general results are robust after adding the fixed effect, with both the log number of trips and log number of stores having a negative and statistically significant effect on the log of the price index. However, the coefficients are much smaller in magnitude, suggesting that doubling the number of trips reduces the

	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$
	(5)	(6)	(7)	(8)
ln(Number of Trips)	-0.0007***		-0.0020***	
	(0.0002)		(0.0003)	
ln(Number of Stores)		-0.0022***		-0.0030***
		(0.0003)		(0.0003)
Controls	None	None	Basket	Basket
Fixed Effect	Household	Household	Household	Household
R^2 Within	0.0000	0.0002	0.0114	0.0115
R^2 Between	0.0742	0.1407	0.0501	0.0710
R^2 Overall	0.0019	0.0046	0.0130	0.0143
n	348,283	348,283	348,283	348,283

Table 2.7: Household fixed effect regressions results relating price index to shopping behavior. Standard errors clustered at the household level, * = 10% significance, ** = 5% significance, *** $<$ 1% significance.

price index by 0.20% while doubling the number of stores with a purchase reduces the price index by 0.30%. These coefficients are much lower than the OLS results in Table 2.6, but demonstrate that there are observable benefits to additional shopping even when focusing on within household variation rather than cross-household behavior, albeit of questionable economic significance.[20]

2.4.2 Goods Purchased at Two or More Stores

One concern with measuring search behavior using the number of trips or stores is how relevant each trip or visit to a store is for a particular good. If a person only considers buying apples from store A, visits to store B should not be considered a search for apples. Furthermore, such goods may be what truly drive the decision to search more as the household is aware of the good's availability at potentially competing stores. To address this possibility, I construct a sample of household-

[20]Aguiar and Hurst [1] suggest that there may be endogeneity between the price index and shopping behavior and address the issue using an instrumental variable approach. After averaging across the sample time period as Aguiar and Hurst do, similar IV regressions were considered along with using education level of the head of household, the number of retired heads of household, or the number of not working heads of households as potential instruments. All but age were weak instruments for shopping behavior. Moreover, Hausman tests for endogeneity indicate that one cannot reject the null hypothesis that the number of trips or stores shopped at are exogenous to the price index. The one strong instrument, age instrumenting for the number of trips, yields an estimated coefficent of -0.0135, or an elasticity of 1.35% (see section A.7).

week observations where each household's basket of goods only consists of goods purchased by the household in at least two different stores. For convenience, I will refer to these goods that have been purchased by a household in two or more different stores as a "multi-store good".

For this restricted sample, the measures of shopping for each household-week are the number of trips to the set of stores that the household has purchased a multi-store good from and the number of stores in that set that the household made a purchase in that week. Defining shopping in this manner has the added benefit of providing a better measure of search in terms of shopping trips. In the full sample, one weakness was the inability to account for visits to a store and possible observation of a price that did not result in a purchase. However, with the multi-store good sample, one can observe a trip to a store that previously resulted in the purchase of a multi-store good that that does not result in a current purchase of the multi-store good as long as the household purchased other items from the IRI set of goods at that store. For example, if a basket of multi-store goods consists of bananas that have been observed to be purchased at store A, a visit to store A that is observed to result in a purchase of goods besides bananas could be interpreted as a search for bananas that did not result in a purchase. This interpretation allows for the number of trips to a store selling the multi-store goods to be more accurately considered a search compared to the variable used in Tables 2.6 and 2.7 as one could plausibly believe the shopper observed the price of multi-store goods they did not purchase while shopping.

The resulting sample consists of 1,491 of the original 2,034 households and 111,924 of the initial 348,283 household-week observations. Table 2.8 presents the OLS results when estimating (2.6) using this sample while Table 2.9 shows the results when applying a household fixed effect. When considering only multi-store goods, the estimated savings effect of additional trips falls while the estimate of the effect from buying goods at additional stores increases. The smaller estimated effect for increasing the number of trips is somewhat puzzling under the assumption that the number of trips to stores selling multi-store goods is more meaningful than the number of trips for the full sample, but the estimated effect of shopping at additional stores reaches a peak elasticity of 1.00%, suggesting that multi-store goods do indeed capture the goods that benefit most

	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$
	(1)	(2)	(3)	(4)
ln(Num. of Trips to 2+ Stores)	-0.0028***		-0.0037***	
	(0.0005)		(0.0005)	
ln(Num. of 2+ Stores Shopped)		-0.0095***		-0.0100***
		(0.0007)		(0.0007)
Controls	Market	Market	Market & Basket	Market & Basket
R^2	0.0020	0.0035	0.0105	0.0119
n	111,924	111,924	111,924	111,924

Table 2.8: Results for OLS regressions relating price index to shopping behavior at household-week level using sample of multi-store goods.
"2+ Stores" is the set of stores in which the household buys goods that have been purchased in at least two stores. Standard errors in parentheses, * = 10% significance, ** = 5% significance, *** $<$ 1% significance.

from additional search. In addition, the estimates of the savings effect after including a household fixed effect are unambiguously higher. Although the estimates are still small, with elasticities of 0.23% for trips and 0.36% for stores, the increased coefficients again suggest that search has a greater impact on goods that households are known to have purchased from different stores.

2.4.3 Regular vs. Irregular Shopping Trips

Consumers may have a regular shopping day or a particular set of stores they regularly shop at and purchase goods without regards to price. Such shopping trips should not count as a marginal search since consumers do not weigh the costs and benefits of these trips, instead they may be thought of as the costless initial search in Burdett and Judd [17]. In addition, consumer attention to prices may differ between regular shopping trips and what may be considered irregular trips. Given the habitual nature of a regular shopping trip, a consumer may be insensitive to price, purchasing some fixed basket of goods. In contrast, irregular trips may fall into two types: one, they may be trips in which consumers are willing to pay the search cost with the expectation of finding lower prices, as prescribed in price dispersion models, or two, they may be trips performed because the consumer forgot to purchase a good or needs a good immediately and makes the irregular trip as a matter of convenience. To address the potential difference in the price savings effect of regular

	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$
	(5)	(6)	(7)	(8)
ln(Num. of Trips to 2+ Stores)	-0.0013**		-0.0023***	
	(0.0006)		(0.0006)	
ln(Num. of 2+ Stores Shopped)		-0.0029***		-0.0036***
		(0.0009)		(0.0009)
Controls	None	None	Basket	Basket
Fixed Effect	Household	Household	Household	Household
R^2 Within	0.0000	0.0001	0.0101	0.0101
R^2 Between	0.0001	0.0045	0.0054	0.0067
R^2 Overall	0.0001	0.0014	0.0081	0.0086
n	111,924	111,924	111,924	111,924

Table 2.9: Household fixed effect regressions results relating price index to shopping behavior at household-week level using sample of multi-store goods.
"2+ Stores" is the set of stores in which the household buys goods that have been purchased in at least two stores. Standard errors clustered at the household level, * = 10% significance, ** = 5% significance, *** $<$ 1% significance.

and irregular shopping behavior, I estimate the equation:

$$P_{i,t} = \beta_0 + \beta_1 \, Regular \ Shopping_{i,t} + \beta_2 \, Irregular \ Shopping_{i,t} + \mathbf{B} \cdot X_{i,t} + \varepsilon_{i,t} \tag{2.7}$$

which is similar to equation (2.6) except the shopping behavior variable has been split into regular and irregular shopping behavior and the equation must be estimated in levels rather than logs due to the possibility of having zero regular trips or irregular trips in a week.

How one defines a "regular" shopping trip may be in terms of time of the week, e.g. every Monday, or in terms of a particular set of stores. I therefore consider three possible definitions for regular shopping behavior:

1. The number of trips on regular shopping days in a week
2. The number of trips to regular stores in a week
3. The number of regular stores visited in a week

Irregular trips are then defined as:

	Mean	s.d.	Min	Max
Number of Regular Trips in a Week	1.248	1.565	0	22
Number of Irregular Trips in a Week	1.638	1.291	0	22
Number of Trips to Regular Stores in a Week	1.875	1.639	0	23
Number of Trips to Irregular Stores in a Week	1.011	1.244	0	19
Number of Regular Stores Shopped at in a Week	1.125	0.719	0	8
Number of Irregular Stores Shopped at in a Week	0.881	1.041	0	14

Table 2.10: Summary statistics for regular and irregular shopping behavior at the Household-Week level (n = 348,283).
Number of Regular Trips in a Week and Number Irregular Trips in a Week uses the a trip on a regular shopping day as the definition for a regular shopping trip.

1. The number of trips on non-regular shopping days in a week
2. The number of trips to non-regular stores in a week
3. The number of non-regular stores visited in a week

To determine the number of regular shopping days for each household, I first calculate the floor of the average number of shopping days in a week over the entire four year period for a household and then rank the days of the week by how often the household shops on those days over the four year period. Let this integer be denoted by D_i. The household's regular shopping days are then the top D_i most frequent days of the week the household goes shopping. For example, if a household went shopping on 3.14 days in a week on average, the household would be classified as having 3 regular shopping days. The household's regular shopping days of the week would then be the 3 most frequent days of the week it went shopping, say Monday, Friday, and Saturday. Any shopping trips on Monday, Friday, or Saturday would be considered a regular shopping trip for this household while trips on other days are considered irregular trips. Similarly, a household's list of regular stores is calculated using the floor of the average number of different stores it visits in a week over the entire period, denoted by S_i, and then classifying the S_i most frequently visited stores over the four year period as regular stores while trips to stores not in this list are classified as irregular stores. Table 2.10 presents summary statistics for the six different measures of regular and irregular shopping.

Ex ante, one would expect regular shopping trips to have no effect on the price index if consumers are inattentive to price during such trips, $\beta_1 = 0$. On the other hand, irregular shopping trips might have a negative impact on the price index as lower prices during other times or at other stores motivate households to pay the cost of the additional shopping trip but could possibly have no effect or even a positive effect on the price index if irregular trips are instead last minute trips in which consumers are insensitive to price.

Tables 2.11, 2.12, and 2.13 present the OLS and household fixed effect results when using the number of trips on regular shopping days, the number of trips to regular stores, and the number regular stores visited as the definitions for the number of regular shopping trips, respectively. As one might expect if consumers are not price sensitive during regular shopping trips after including household fixed effects, $\hat{\beta}_1$ is not statistically different from zero when regular shopping is defined using regular stores, either the number of trips to regular stores or the number of different regular stores visited in a week (Tables 2.12 and 2.13, columns (4) and (6)). These results corroborate the idea that stores are the appropriate measure of increased search since consumers are not price sensitive when shopping at the stores they regularly visit.

In contrast, the OLS results suggest that regular shopping behavior may be just as important as irregular shopping behavior since the coefficients on both variables are somewhat similar to each other when measuring regular shopping by the number of trips on regular shopping days or the number of regular stores visited in a week. Only measuring regular shopping by the number of trips to regular stores yields a meaningful difference between regular and irregular shopping behavior, with trips to irregular stores having three times as much of an impact on the price index. In addition, the coefficients on regular shopping and irregular shopping are each generally robust to the inclusion of the other shopping variable, suggesting that regular and irregular shopping decisions may be independent of each other.[21]

In terms of magnitude, the effects of additional irregular shopping are still relatively small and

[21] This is further supported by the 0.08 correlation between regular trips and irregular trips and 0.05 correlation between the number of trips to regular and irregular stores. The correlation between the number of regular stores and irregular stores is 0.16, however.

	$P_{i,t}$	$P_{i,t}$	$P_{i,t}$	$P_{i,t}$	$P_{i,t}$	$P_{i,t}$
	(1)	(2)	(3)	(4)	(5)	(6)
Regular Trips	-0.0017**		-0.0016***	-0.0003***		-0.0004***
	(0.0001)		(0.0001)	(0.0001)		(0.0001)
Irregular Trips		-0.0016***	-0.0015***		-0.0004***	-0.0005***
		(0.0001)	(0.0001)		(0.0001)	(0.0001)
Controls	B, M	B, M	B, M	B	B	B
Fixed Effect	None	None	None	Household	Household	Household
R^2 Within	-	-	-	0.000	0.0001	0.0001
R^2 Between	-	-	-	0.0435	0.0522	0.0611
R^2 Overall	0.0107	0.0101	0.0117	0.0013	0.0008	0.0019
n	348,283	348,283	348,283	348,283	348,283	348,283

Table 2.11: OLS and household fixed effect regressions results relating price index to regular and irregular shopping behavior when measuring regular shopping by trips on regular shopping days.
B = Basket Controls, M = Market Controls. Standard errors clustered at the household level for fixed effect regressions.
* = 10% significance, ** = 5% significance, *** $<$ 1% significance.

similar to the other estimates in this paper. A one standard deviation increase in the various shopping measures in Table 2.10 using the OLS estimates in column (3) from Tables 2.11 to 2.13 would only change the price index by 0.15 percentage points for the number of trips to regular stores to 0.34 percentage points for the number of trips to irregular stores. Estimates after accounting for a household fixed effect suggest an effect ranging from 0.002 percentage points for the number of trips to regular stores to 0.15 percentage points for the number of irregular stores visited when shopping behavior increases by one standard deviation using the coefficients in columns (6) of Tables 2.11 to 2.13. These values again suggest that price savings cannot be the primary motivation for increased search or shopping behavior.

2.5 Conclusion

The regression results in section 2.4 suggest that not searching does not have a large negative impact on the price consumers pay since doubling weekly shopping intensity yields prices 0.37% to 1.00% lower than the average weekly price or 0.20% to 0.36% when analyzing the within household change in search from week to week by using a household fixed effect. Although the results

	$P_{i,t}$	$P_{i,t}$	$P_{i,t}$	$P_{i,t}$	$P_{i,t}$	$P_{i,t}$
	(1)	(2)	(3)	(4)	(5)	(6)
Trips to Regular Stores	-0.0010**		-0.0009***	-0.0001		-0.00001
	(0.0001)		(0.0001)	(0.0001)		(0.00011)
Trips to Irregular Stores		-0.0027***	-0.0027***		-0.0009***	-0.0009***
		(0.0001)	(0.0001)		(0.0001)	(0.0001)
Controls	B, M	B, M	B, M	B	B	B
Fixed Effect	None	None	None	Household	Household	Household
R^2 Within	-	-	-	0.0000	0.0002	0.0002
R^2 Between	-	-	-	0.0156	0.1001	0.1011
R^2 Overall	0.0097	0.0118	0.0124	0.0004	0.0024	0.0025
n	348,283	348,283	348,283	348,283	348,283	348,283

Table 2.12: OLS and household fixed effect regressions results relating price index to regular and irregular shopping behavior when measuring regular shopping by trips to regular stores.
B = Basket Controls, M = Market Controls. Standard errors clustered at the household level for fixed effect regressions.
* = 10% significance, ** = 5% significance, *** $<$ 1% significance.

	$P_{i,t}$	$P_{i,t}$	$P_{i,t}$	$P_{i,t}$	$P_{i,t}$	$P_{i,t}$
	(1)	(2)	(3)	(4)	(5)	(6)
Number of Regular Stores	-0.0036***		-0.0029***	0.0003		-0.00001
	(0.0002)		(0.0002)	(0.0003)		(0.00030)
Number of Irregular Stores		-0.0035***	-0.0032***		-0.0014***	-0.0014***
		(0.0001)	(0.0001)		(0.0002)	(0.0002)
Controls	B, M	B, M	B, M	B	B	B
Fixed Effect	None	None	None	Household	Household	Household
R^2 Within	-	-	-	0.0000	0.0004	0.0004
R^2 Between	-	-	-	0.0577	0.1035	0.1040
R^2 Overall	0.0107	0.0122	0.0133	0.0014	0.0030	0.0030
n	348,283	348,283	348,283	348,283	348,283	348,283

Table 2.13: OLS and household fixed effect regressions results relating price index to regular and irregular shopping behavior when measuring regular shopping by the set of regular stores with a purchase.
B = Basket Controls, M = Market Controls. Standard errors clustered at the household level for fixed effect regressions,
* = 10% significance. ** = 5% significance, *** $<$ 1% significance.

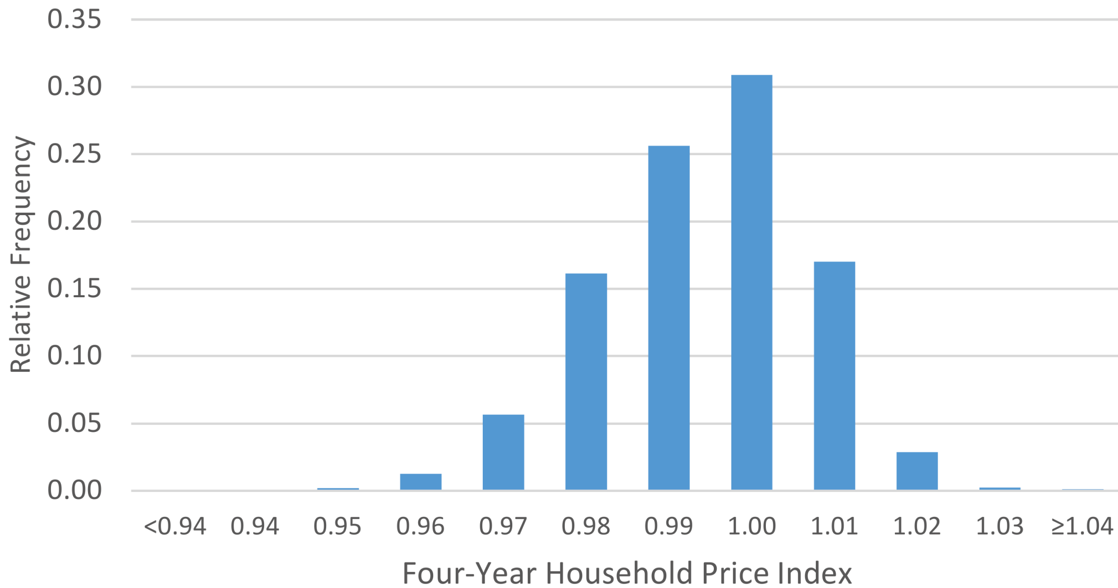
are statistically significant and robust, is this a reasonable estimate?

To shed some light on this question, Figure 2.6a plots the total amount paid relative to weekly average prices for each household from 2008 to 2011 to capture how much each household saved relative to average prices over the entire sample period.[22] When considering the total amount paid relative to weekly average prices shown in the figure, over 90% of households pay within 3 percentage points of the average weekly price for their basket of goods. Only 5 of the 2,034 households are able to persistently pay a price 5% or more lower than average weekly price. Likewise, only 2 households pay more than 4% of the average weekly price. These observations suggest that households cannot be significantly separated by how much they pay relative to the average.[23] However, that is not to say that there is no benefit to search at all. Figure 2.6b instead plots the individual price indexes, $P_{i,t}$, for all household-week observations in the full sample. While approximately 36% of household-week price indexes are within 1% of the average, there is much wider dispersion with 14% of observations paying at least 4% less than average and 16% paying 4% more than the average weekly price.

If the average effect over long periods of time is only moderate savings, what can explain why shopping intensity has such a long tail in contrast to models that predict consumers search no more than twice? Why regularly shop more than twice a week or visit more than two stores if savings are small in terms of long term averages? The typical assumption that households measure the costs and benefits does not hold up if there are even trivial search costs. Given the average weekly basket's price of \$25.81, the highest estimate of savings from doubling search intensity, 1%, suggests that the average household saves less than \$0.26 by doubling the number of stores shopped at in a week. In order for consumers to be rationally choosing the amount of search, the search cost of doubling the number of searches must be less than \$0.26, which seems unreasonable given the transportation and opportunity cost of time involved in shopping at an additional store.

[22] Mathematically, this is $P_i = \dfrac{\sum_t \sum_{j \in J_{i,t}} p_{i,j,k,t} \cdot q_{i,j,k,t}}{\sum_t \sum_{j \in J_{i,t}} \overline{p_{j,k,t}} \cdot q_{i,j,k,t}}$

[23] The caveat to this conclusion is that this sample consists of older households so they may all shop enough that they all pay the same low price so deviation from the average price is not a useful measure.



(a) Four Year Household Price Index, 2008-2011 (n = 2,034). Minimum value: 0.94, Maximum value: 1.05.

(b) Household-Week Price Index, 2008-2011 (n = 348,283). Minimum value: 0.25, Maximum value: 3.41.

Figure 2.6: Distribution of price indexes

While this bound on the search cost increases depending on the size of the basket purchased, 75% of household-weeks observed spent less than \$36 in a week, implying a maximum upper bound on search costs of \$0.36 for the vast majority of the sample.

One should note that the findings of this paper do not invalidate models of price dispersion that rely on search. Instead of invalidating search as an explanation for price dispersion, at least for grocery store products, the results of this paper suggest a refinement is necessary for price dispersion models utilizing search. The fact is that consumers do search by going on multiple shopping trips and making purchases at more than one store in a week (Figures 2.3 and 2.4). However, the motivation for searching needs to be reconsidered.

Alternative explanations for why people search may be models utilizing behavioral economics or models of consumers who treat shopping as a consumption good that provides utility regardless of the price savings outcome. If consumers are anchored to a specific price, they may value the perceived discount rather than comparing the marginal benefits and marginal costs of the search and purchase decisions.[24] In addition, empirical research has shown that consumers who purchase a good more frequently have a smaller range of price acceptance, meaning the range of prices considered low versus what they consider a high price [51]. This suggests that consumers who shop more often may do so because they are only willing to purchase their basket of goods at a lower range of prices, e.g. "I'm only willing to pay \$0.99 and not a penny more", which requires a more extensive search to find the desired price.

Shopping may also be considered a consumption good or even a recreation activity. Some people may enjoy the experience of shopping and not actually care about how much less they pay for their goods.[25] In either the behavioral or shopping as consumption good model, the price savings is not the sole contribution to utility and consumers are more willing to incur search costs due to the increased benefit from appreciating the size of their discount or enjoyment from being in a grocery store. While testing these hypotheses is beyond the capabilities and scope of the

[24] For instance, someone who drives across town to pay 5 cents per gallon less for gasoline despite spending more than the saved amount by driving to the gas station.

[25] This type of consumer may explain the observed long tail in the distribution of shopping trips and stores. A household finding savings in excess of search costs when going on 26 shopping trips in a week seems unlikely.

data utilized in this paper, merging behavioral valuations of price savings or treating shopping as a consumption good with search models of price dispersion may yield a coherent model that can explain this paper's observations.

Finally, one other possible explanation for the small observed savings effect is that the products in the dataset are somewhat small in price and may only be a small part of a household's expenditures.[26] For larger items, such as automobiles or houses, the benefits of a marginal search may be rather large. However, dismissing the importance of search for items that are either low in price or small in terms of expenditure share would ignore the fact that price dispersion exists in these prices [35, 36]. If nothing else, the small estimate for the price savings from search suggests that price dispersion models should reconsider the motivation for search.

[26] The median price of goods in the dataset is $2.29 and the maximum observed price is $99.99.

CHAPTER 3

INSIDE THE PRICE DISPERSION BOX: EVIDENCE FROM U.S. SCANNER DATA

With Benjamin Eden and Maya Eden

3.1 Introduction

Models of price dispersion use various assumptions. Among these are monopolistic competition, menu costs, search frictions and uncertainty about aggregate demand. Different assumptions may lead to drastically different policy implications and therefore attempting to differentiate between various models is important. Here we study some implications of monopolistic competition models that assume a CES utility function and versions of the Prescott [66] model.

The paper complements Eden [35] who focuses on variations in the cross sectional price dispersion across goods. Here we focus on the within good behavior of prices and quantities. We divide stores that sell a given product in a given week into bins according to their posted price. For example, we look at stores with price above the median and below the median. Stores may be above the median in one week and below the median in another week. In the same week a store may have goods that are priced above the median and below the median. We therefore define the bins for each good-week combination (UPC-week cell).

We argue that alternative theories can be distinguished by the coefficients in the following regression:

$$\ln(x_1) = a + b_1 \ln(p_1) + b_2 \ln(x_j) + b_3 \ln(p_j) \tag{3.1}$$

where $\ln(x_1)$ is the average (log) quantity sold during the week by stores in the highest price bin, $\ln(p_1)$ is the average (log) price posted at the beginning of the week by stores in the highest price bin, $\ln(p_j)$ is the average price posted by stores in bin j and $\ln(x_j)$ is the average quantity sold by stores in bin j, where j index a lower price bin.

We find that: (a) the own price elasticity is negative ($b_1 < 0$), (b) the quantity elasticity is between zero and unity ($0 < b_2 < 1$), (c) the cross price elasticity is positive but less than the absolute value of the own price elasticity ($0 < b_3 < -b_1$), (d) b_3 and b_2 are lower for lower price bins. We also look at variations over weeks and find that: (e) the variations over weeks in the average price and quantity sold is lower for higher price bins and (f) Temporary sales contribute substantially to variations over weeks in the (cross sectional) average price of the typical good.[1]

We attempt to explain the above findings with a model in which price dispersion arises as a result of uncertainty about aggregate demand. We start with a simple monopolistic competition model in which there are many buyers who belong to a single household. Each buyer visits one store only and is instructed by the head of the household to buy a quantity that depends on the price. We derive the prediction of this model about the elasticities within UPC-week cell under the assumption that the head of the household maximizes a CES utility function. We then explore a flexible price version of the Prescott model, the Uncertain and Sequential Trade (UST) model, where buyers can costlessly move across stores and buy at the cheapest available price. The "simple versions" of the two models do not explain the main empirical findings. Observation (e) is a challenge to Prescott type models that assume a tradeoff between the probability of making a sale and the price. Observations (b) and (c) are a challenge to versions of monopolistic competition models that imply a quantity elasticity of unity and a cross price elasticity that is equal to the absolute value of the own price elasticity. We attempt to explain the main findings with a model in which some buyers shop around (as in the UST model) and some buyers do not shop around (as in our monopolistic competition model).

Previous empirical studies of price dispersion focused on the implications of search models. See for example, Sorensen [78], Lach [58] and Kaplan and Menzio [52]. There is also an empirical

[1] The focus here is on the behavior of posted prices. Glandon [42] study the behavior of transaction prices, obtained by dividing aggregate revenues across stores in each UPC-week cell by the aggregate quantity sold in the cell. He finds that temporary sales have a large impact on the price actually paid by consumers. Coibion et al. [20] find that effective (transaction) prices are procyclical while posted prices are acyclical. They explain this difference by changes in shopping activities: In recessions consumers spend more time shopping and tend to reallocate expenditures towards lower price retailers. This is consistent with Kaplan and Menzio [52] who found a significant effect of the employment status of the head of the household on the average price paid.

literature that focus on sticky price models. See for example, Reinsdorf [68], Eden [33], Baharad and Eden [5], and Ahlin and Shintani [3]. And there is a literature that studies price dispersion in the airline industry. See for example, Escobari [38], Gerardi and Shapiro [40], and Cornia et al. [22]. Here we use the UST model to discuss the empirical findings. One reason for our focus on the UST model is the finding that demand uncertainty is important in explaining differences in price dispersion across goods Eden [35]. Another rather obvious reason is that we are more familiar with this model. We hope however that the findings in this paper will be of interest also to readers who are not familiar with the UST model and who wish to regard the UST framework as an organizational device.

Section 3.2 is about the monopolistic competition model. Section 3.3 is about a flexible price version of the Prescott model: the Uncertain and Sequential Trade (UST) model. Section 3.4 describes the data and variations in the (cross sectional) average price over weeks. Section 3.5 is about elasticities. Section 3.6 repeats the calculations after controlling for store effects and section 3.7 repeats the calculations after controlling for UPC specific store effects. Section 3.8 assesses the importance of temporary sales in the variation of the (cross sectional) average price over weeks. Section 3.9 computes the probabilities of attracting shoppers by lower price stores. Section 3.10 provides concluding remarks.

3.2 Monopolistic Competition

We start with a model in which the household uses a CES utility function to allocate expenditure over stores as in Dixit and Stiglitz [28].[2]

There is a single household with $2N$ members: N workers (sellers) and N buyers. Members live in N neighborhoods. There are two members per neighborhood: A buyer and a seller. Each seller produces a good and offers it for sale in his neighborhood. There are thus N goods that are differentiated by location but have identical physical characteristics. In week t, the head of the

[2]This model is an important component of the New Keynesian literature. See for example, Christiano et al. [19] and Coibion et al. [20].

household decides how much to buy from each of the N goods and instructs the buyer who lives in neighborhood i to buy x units from the seller in his neighborhood. The head of the household chooses to spend a total of I_t dollars on the N goods. He faces the prices $(p_{1t},\ldots,p_{Nt})$ and chooses the quantities $(x_{1t},\ldots,x_{Nt})$ to maximize a CES utility function. In week t, the head of the household solves the following problem:

$$\max_{x_{jt}} \left(\sum_{j=1} N(x_{jt})^{\gamma} \right)^{1/\gamma} \quad \text{s.t.} \quad \sum_{j=1}^{N} p_{jt} x_{jt} = I_t \tag{3.2}$$

where $0 < \gamma < 1$.

The first order conditions to this problem requires:

$$x_{1t} = x_{jt} \left(\frac{p_{1t}}{p_{jt}} \right)^{\theta} \quad \text{for all} \quad j \tag{3.3}$$

where $\theta = \dfrac{1}{\gamma - 1} < 0$. We take logs and add a classical measurement error to obtain:

$$\ln(x_{1t}) = \ln(x_{jt}) + \theta \ln(p_{1t}) - \theta \ln(p_{jt}) + e_t \tag{3.4}$$

This regression imposes strong restrictions on the coefficients. It says that the quantity elasticity (the coefficient of $\ln(x_{jt})$) is unity and the absolute value of the own price elasticity is equal to the cross price elasticity. In terms of equation (3.1) it imposes: $b_2 = 1$ and $b_1 = -b_3$.

In the above model each buyer visits one store. We now turn to the other extreme in which buyers can costlessly move between stores and buy at the cheapest available price.

3.3 Sequential Trade

The original Prescott [66] model assumed that prices are set in advance. Eden [31] relaxed the price rigidity assumption and described a sequential trade process in which cheaper goods are sold first.[3] In his Uncertain and Sequential Trade (UST) model, buyers arrive at the market place sequentially. Each buyer sees all available offers, buys at the cheapest available price and disappears. Sellers must make irreversible selling decisions before they know the aggregate state of demand. In equilibrium they are indifferent between prices that are in the equilibrium range because the selling probability is lower for higher prices. Sellers in the model make time consistent plans and do not have an incentive to change prices during the trading process. Prices are thus completely flexible.

We start with a simple version and then augment it to account for various features of the data.

3.3.1 A simple version

There are many goods and many sellers who can produce the goods at a constant unit cost. Here we focus on one good with a unit cost of λ. Production occurs at the beginning of the period before the arrival of buyers. Storage is not possible. The number of buyers $\tilde{N}$ is an i.i.d. random variable that can take two possible realizations: N with probability $1-q$ and $N+\Delta$ with probability q.

Sellers take prices and the probability of making a sale as given. They know that they can sell at the price P_1 for sure. They may also be able to sell at a higher price, P_2, if demand is high, with probability q. In equilibrium sellers are indifferent between the two price tags: The expected profits is the same for both tags.

It is useful to think of two hypothetical markets. The price in the first market is P_1 and the probability that this market opens is 1. The price in the second market is P_2 and the probability that it opens is q. From the seller's point of view, he can sell any quantity at the price announced in the market, if the market opens and cannot sell anything at that market if the market does not open. A unit with a price tag of P_1 will be sold in the first market. A unit with a price tag of P_2 will be sold in the second market, if this market opens.

Buyers arrive sequentially in batches. The first batch of N buyers buys in the first market at

[3] For other extensions of the Prescott model, see Dana [23, 24, 25] and Deneckere and Peck [27].

the price P_1. The second market opens only if the second batch of Δ buyers arrives. If this second batch arrives the second market opens at the price P_2.

The demand of each of the active buyer at the price P is: $D(P)$. In equilibrium sellers supply x_1 units to the first market and x_2 units to the second market.

Equilibrium is thus a vector (P_1, P_2, x_1, x_2) such that the expected profits for each unit is zero:

$$P_1 = qP_2 = \lambda \tag{3.5}$$

And markets that open are cleared:

$$x_1 = ND(P_1) \quad \text{and} \quad x_2 = \Delta D(P_2) \tag{3.6}$$

Figure 3.1 illustrates the equilibrium solution. The demand in market 1 at the price λ, $ND(\lambda)$ is equal to the supply to the first market (x_1). When market 2 opens at the price λ/q, the demand in this market, $\Delta D(\lambda/q)$, is equal to the supply (x_2).

Note that in this simple version posted prices do not change over time. The quantity sold at the low price does not change over time but the quantity sold at the high price fluctuates over time.

3.3.2 Storage

Bental and Eden [9] studied a UST model that allows for storage. In their model both quantities and prices fluctuate with the beginning of period level of inventories. The amount available for sale fluctuates as a result of both i.i.d. demand and supply shocks. An increase in the amount of inventories carried from the previous period reduces all prices. A temporary reduction in the cost of production has a similar effect. Prices will in general depend on the amount available for sale (inventories + current production) and when this amount is high, prices will remain low until inventories are back to "normal". Thus, "temporary sales" may be the result of both demand and supply shocks. Eden [34] uses aggregate NIPA data and VAR analysis to test the implications of the model. The model is also consistent with the findings of Aguirregabiria [2] who used a unique

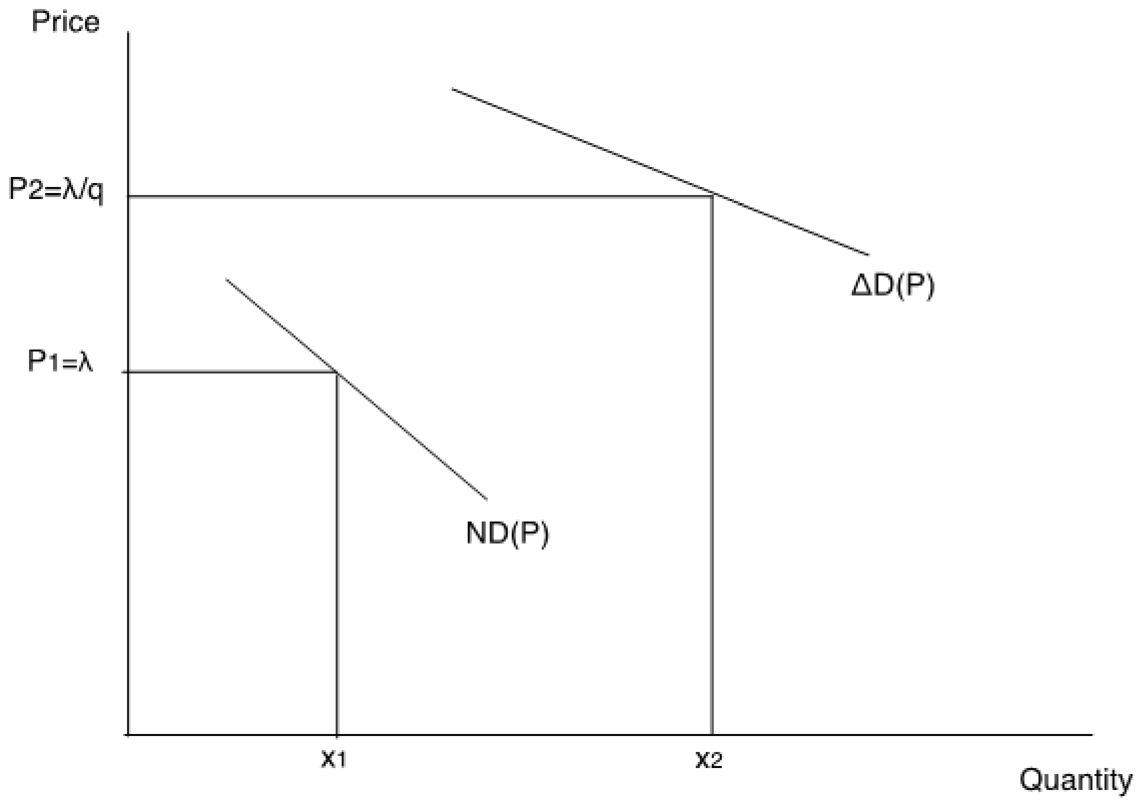


Figure 3.1: Prices and quantities in the simple version of the UST model

data set from a chain of supermarket stores in Spain and found a very significant and robust effect of inventories at the beginning of the month on current price.[4]

The BE model assumes a convex cost function and exponential depreciation. Here we assume a constant per unit cost and one-hoss-shay depreciation. The constant per unit cost simplifies the analysis. The one-hoss-shay depreciation is realistic because most supermarket items have an

[4] Aguirregabiria [2] provides a description of the negotiation between the firm (chain's headquarter) and its suppliers. The toughest part of the negotiation with suppliers is about the number of weeks during the year that the brand will be under promotion, and about the percentage of the cost of sales promotions that will be paid by the wholesaler (e.g. cost of posters, mailing, price labels). A similar description is in Anderson and Simester [4] who present institutional evidence that sales (accompanied by advertising and other demand generating activities) are complex contingent contracts that are determined substantially in advance. There is also some flexibility. For many promotions manufacturers allow for a "trade deal window" of several weeks where the seller can execute the promotion. These descriptions are consistent with the hypothesis that temporary sales are used to respond to high inventories. Sometimes the delivery schedule allows the firm to predict the level of inventories and as a result temporary sales are set in advance. The flexibility in the timing of sales reflects the need to respond to inventories that were accumulated as a result of demand shocks. The observation that temporary sales do not respond to cost shocks [4] and are largely acyclical [20] is also consistent with the hypothesis that temporary sales are used to respond to transitory changes in the level of inventories that last for a relatively short time and not to changes that last for relatively long time (like changes in cost and the level of unemployment). This is different from the view that temporary sales are not used to respond to changes in fundamentals and are merely a discrimination device.

	$\tilde{N}_t = N + \Delta$	$\tilde{N}_t = N$
$\tilde{N}_{t-1} = N + \Delta$	$x(1,NI) + x_2$	$x(1,NI)$
$\tilde{N}_{t-1} = N$	$x(1,I) + x_2$	$x(1,I)$

Table 3.1: Total amount sold in period t

expiration date. It also serves as a tiebreaker and yields predictions about temporary sales that are an important feature of the data.

To simplify, we assume that the good can be stored for one period only. Thus, if a good is not sold in the first period of its life, it can still be sold in the second period but it has no value if it is not sold within the two periods.

As before, the number of buyers $\tilde{N}_t$ is i.i.d. and can take two possible realizations: $\tilde{N}_t = N$ with probability $1 - q$ and $\tilde{N}_t = N + \Delta$ with probability q.

At the beginning of period t the economy can be at two states. In state I (I for inventories) the demand in the previous period was low $\left(\tilde{N}_{t-1} = N\right)$ and the second market did not open. As a result, inventories were carried from the previous period. In state NI (NI for no inventories) demand was high $\left(\tilde{N}_{t-1} = N + \Delta\right)$ and there are no inventories. The price in the first market is $P(1,I)$ in state I (with inventories) and $P(1,NI)$ in state NI (with no inventories). The quantity offered for sales in market 1 is $x(1,I)$ in state I and $x(1,NI)$ in state NI. The price in the second market (P_2) and the supply (x_2) do not depend on the level of inventories. The quantity sold in the first market is equal to the quantity offered for sale. The quantity sold in the second market is zero if demand is low and x_2 if demand is high. Table 3.1 describes the total amount sold (over the two markets) as a function of last period's demand and this period's demand.

A formal analysis and the equilibrium definition is in Eden [35, Appendix C]. To make this paper self-contained we repeat here the description of the model. In allocating the available amount of goods (from new production and inventories) across the two markets, the older units get a "priority" in the first market (and the younger units get a "priority" in the second market). Given prices the allocation rule is as follows. If the amount of old units (from inventories) is less than the demand in the first market then all old units are supplied to the first market. If the amount of old units is greater than the demand in the first market then only old units are supplied to the first market. To motivate this allocation rule, we consider the following example. There are two stores: Store O with old units and store Y with young units. Suppose further that store Y posts the first market low price and store O posts the second market high price. In this case if aggregate demand is low and store O does not sell and the units supplied by store O expire. Alternatively, if store O posts the first market price and store Y posts the second market price, the unsold units supplied by store Y do not expire and can be sold next period. It follows that the joint profits of both stores can be increased if they do not follow our allocation rule. This cannot occur in equilibrium.

A young unit that is not sold in the current period will be sold in the next period at the price $P(1,I)$. The value of a young unit that is not sold in the current period (the value of inventories) is $\beta P(1,I)$, where $0 < \beta < 1$ is a constant that captures discounting, storage costs and depreciation. The value of an old unit that is not sold is zero. Given the above allocation rule and given that production is strictly positive in each period, newly produced units are supplied to the second market and in equilibrium the following arbitrage condition must hold.[5]

$$qP_2 + (1-q)\beta P(1,I) = \lambda \tag{3.7}$$

The left hand side of (3.7) is the expected present value of revenues from a newly produced unit allocated to the second market. If the second market opens (with probability q) the seller gets P_2. Otherwise he will get the unit value of inventories, $\beta P(1,I)$. The right hand side of (3.7) is the unit cost of production. Thus, (3.7) says that the marginal cost is equal to expected revenues.

We now distinguish between two cases. In the first case, illustrated by Figure 3.2a, inventories in state I are relatively low and newly produced goods are supplied in state I to both markets. Since newly produced goods are supplied to the first market, the price in the first market is the marginal

[5]Production must be positive because the entire supply in the first market is always sold.

cost: $P(q,I) = P(1,NI) = \lambda$. Substituting this into (3.7) yields:

$$P_2 = \frac{\lambda(1-(1-q)\beta)}{q} \tag{3.8}$$

In the second case, illustrated by Figure 3.2b, newly produced goods are supplied to the first market only in state NI. In state I the entire supply to the first market is out of inventories and the supply to the second market is of both newly produced units and old units. Since old units are supplied to both markets, we must have:

$$aP_2 = P(1,I) \tag{3.9}$$

This says that the expected revenue of supplying an old unit to the second market is the same as the revenue from supplying it to the first market. The solution to (3.7) and (3.9) is:

$$P(1,I) = \frac{\lambda}{1+(1-q)\beta} < \lambda \quad \text{and} \quad P_2 = \frac{\lambda}{q(1+(1-q)\beta)} \tag{3.10}$$

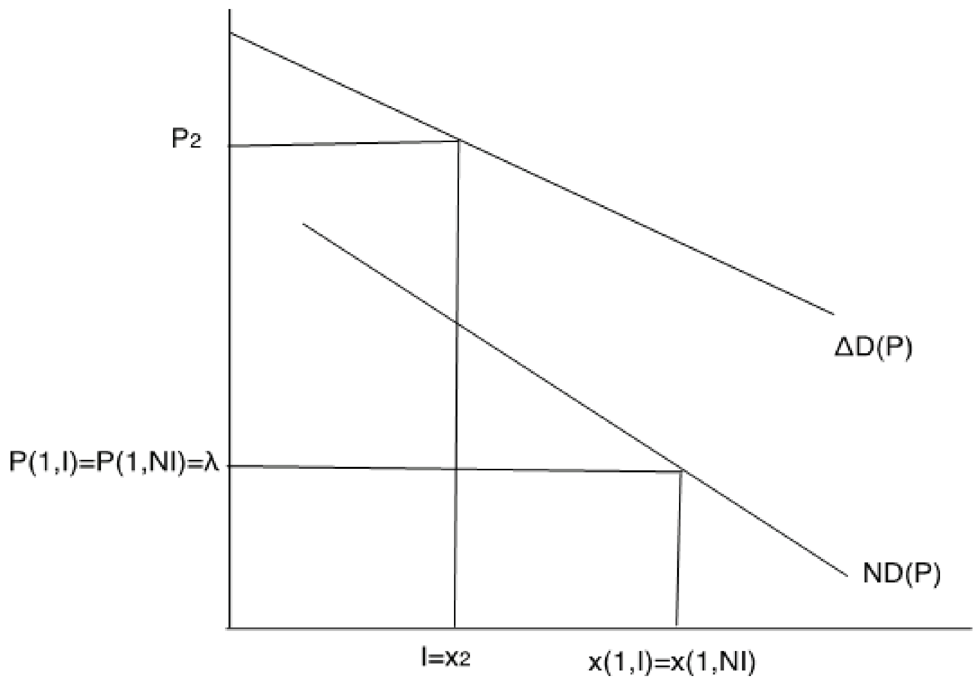
Note that the first market price in state I is below cost as in the loss-leaders model of Lal and Matutes [59].[6]

The model described by Figure 3.2 may account for temporary sales. Some stores offer their newly produced good at the high ("regular") price of market 2. Then if demand is low they accumulate inventories and offer the good for sale at the low price of market 1. We also note that the price and quantity in the first market may change over time.

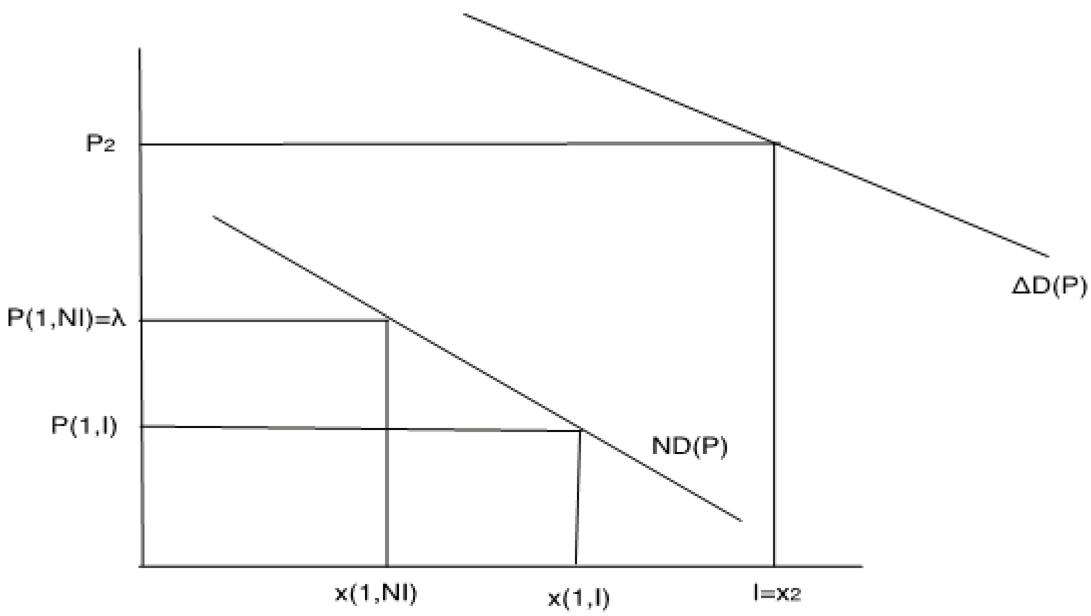
3.3.3 Using sales by low price stores to predict sales by high price stores

In section 3.3.1 and 3.3.2, the amount sold in the second market does not depend on the amount sold in the first market. This result is special to the assumption that $\tilde{N}$ can take only two possible realizations.

[6]It is also possible that all the old units are allocated to the first market and all the new units are allocated to the second market. Also in this case the first market price can be below cost: $P(1,I) \leq \lambda$. See, Eden [35].



(a) In state I, $I = x_2$ "old units" and $x(1,i) - I$ newly produced units are supplied to the first market.



(b) In state I, $x(1,I)$ "old units" are supplied to the first market and $I - x(q,I)$ "old units" are supplied to the second market. No new units are supplied to the first market.

Figure 3.2: Possible equilibria

In the more general case in which $\tilde{N}$ may take many possible realizations the quantity sold in market $k < i$ is correlated with the quantity sold in market i, because strictly positive sales in market k imply $\tilde{N} \geq N_k$ and leads to an upward revision in the probability that $\tilde{N} \geq N_i$ and market i will open.

We now establish this correlation under the assumption that the amount supplied to market s (x_s) does not change over time. This assumption holds if inventories are not too large as in Figure 3.2a.

We start with the case in which the number of buyers $\tilde{N}$ can take 3 possible realizations: (N_0, N_1, N_2) with probabilities (π_1, π_2, π_3), where $N_0 = 0 < N_1 < N_2$. There are two hypothetical markets. The first market opens with probability $1 - \pi_0$ and if it opens it serves N_1 buyers. The second market opens with probability π_2 and if it opens it serves $\Delta = N_2 - N_1$ buyers. The unconditional expected quantity sold in market 2 is: $E(x_2) = \pi_2 \Delta D(P_2)$. The expected quantity sold in market 2 conditional on sale in market 1 is[7]:

$$E(x_2|x_1 > 0) = \frac{\pi_2 \Delta D(P_2)}{1 - \pi_0}, \tag{3.11}$$

and the expected quantity sold in market 2 conditional on no sales in market 1 is:

$$E(x_2|x_1 = 0) = 0 \tag{3.12}$$

There is thus a positive relationship between the expected quantity sold in market 1 and the quantity sold in market 2.

We now consider the more general case. We assume that the number of buyers $\tilde{N}$ can take $m+1$ possible realizations: $(N_0, N_1, \ldots, N_m)$, where $(0 - N_0 < N_1 < N_2 < \cdots < N_m)$. The probability that the number of buyers is N_s is denoted by: $\pi_s = Prob\left(\tilde{N}\right) = N_s$. The probability that the number of

[7]Applying Bayes' rule leads to: $\mathbb{P}\left(\tilde{N} = N_w | \tilde{N} \geq N_1\right) = \dfrac{Prob\left(\tilde{N} = N_2 \cap \tilde{N} \geq N_1\right)}{Prob\left(\tilde{N} \geq N_1\right)} = \dfrac{\pi_2}{1 - \pi_0}$.

buyers is greater than N_s is denoted by:

$$q_s = Prob\left(\tilde{N} \geq N_s\right) = \sum_{i=s}^{m} \pi_i \tag{3.13}$$

The demand in market m if it opens is: $\Delta_m D(P_m)$, where $\Delta_m = N_m - N_{m-1}$. The expected quantity sold in market m given x_i is:

$$E(x_m|x_i > 0) = \frac{\pi_m \Delta_m D(P_m)}{q_i}, \quad E(x_m|x_i = 0) = 0 \tag{3.14}$$

There is thus a positive relationship between the quantity sold in market i and the expected quantity sold in market m. We now make the following additional observation.

Claim 1: The expected sales in market m conditional on $x_i > 0$ is increasing in the index i.

This Claim follows from (3.14) and from: $q_1 > q_2 > \cdots > q_m = \pi_m$. The intuition is in the observation that $Prob\left(\tilde{N} = n_m|\tilde{N} > N_i\right)$ is increasing in i. Therefore, from the point of view of the highest price stores, positive sales by medium price stores are more encouraging news than positive sales by low price stores.

Equations (3.14) may be tested by running the quantity sold by the highest price store (store m) on a dummy that is equal to 1 if sales in a lower price store was positive and zero otherwise. We did not pursue this route because we cannot distinguish in the data between the case in which the good was on the shelf and was not sold to the case in which the good was not on the shelf. More importantly, we think that a more realistic model should include non-shoppers as in Salop and Stiglitz [71], Shilony [74], and Varian [82].

3.3.4 Non-shoppers

We abstract from storage and extend the model in section 3.3.1 to include two types of buyers: shoppers and non-shoppers. The monopolistic competition model in section 3.2 and the UST model in section 3.3.1 may be obtained as special cases. The monopolistic competition model may be obtained if we eliminate shoppers. The UST model may be obtained if we eliminate non-shoppers.

We focus here on predicting the quantity sold by the highest price stores on the basis of the quantity sold by medium and low price stores. To simplify, we assume that shoppers' activity is not important for the highest price stores and focus on the quantity elasticity: The elasticity of the quantity sold by the highest price stores with respect to the quantity sold by lower price stores. (The coefficient b_2 in equation (3.1)).

We assume shocks to the demand of non-shoppers and shocks to the number of active shoppers. For predicting the quantity sold by the highest price stores only the demand of non-shoppers is relevant and therefore shoppers' activity introduces noise that reduces the quantity elasticity. When the shock to the number of shoppers is eliminated, we get the monopolistic competition result of unit elasticity. With shoppers' activity we get an elasticity that is less than unity.

Sales by medium price stores provide relatively more information about the demand of non-shoppers because they are less influenced by shoppers' activity. And therefore as in section 3.3.3, we find that for predicting sales by the highest price stores, sales by medium price stores are more relevant than sales by low price stores.

To model this idea, we assume n sellers and $nk+m$ buyers, where $n > m$. Some sellers advertise their price and some do not. There are $m > 0$ advertisers and $n-m > 0$ non-advertisers. Similarly there are some buyers who shop around and some who do not. There are $m > 0$ (potential) shoppers and $nk > 0$ non-shoppers.

At the beginning of week t, sellers (advertisers and non-advertisers) produce the good at the cost of λ_t, where λ_t is the realization of an i.i.d. random variable $\tilde{\lambda}$. The demand of an active buyer (active shopper and non-shopper alike) at the price P is $a_t P^{\theta}$ where $\theta < 0$ and a_t is the realization of an i.i.d. random variable $\tilde{a}$. The demand of the individual buyer is similar to the demand of the individual buyer in the monopolistic competition model (3.4) where the price of the numeraire good is unity and the level of consumption from the numeraire good is the realization of $\tilde{a}$.

Sellers and non-shoppers are distributed over n locations. In each location there is one store (seller) and k non-shoppers that always buy in the local store. The number of active shoppers is an i.i.d. random variable $\tilde{s}$ that can take $m+1$ possible realizations: $s = 0, \dots, m$, where realization s occurs with probability π_s. Thus, in state s there are s active shoppers and $m-s$ shoppers who are not active.

Advertisers post on the internet information about price and availability. Thus, when an advertiser is stocked out this information is immediately on the web site.[8] Active shoppers use the Internet.

Each of the $n-m > 0$ non-advertisers sells only to the k non-shoppers in his location. Each of the m advertisers may attract some shoppers in addition to the non-shoppers in his location. To simplify, we assume that each advertiser chooses capacity (production) to satisfy the demand of $1+k$ buyers. We thus assume that a store satisfies the demand of its k regular clients and has an additional capacity to serve one shopper if he arrives. Unlike New Keynesian models, here stores may be stocked out. Unlike some search models, here capacity depends on the price.

We also simplify by assuming that non-shoppers buy first. After non-shoppers have completed their trade there is still available capacity in m stores. At this point the shoppers form a hypothetical line. There may be no shoppers (if $\tilde{s} = 0$) and in this case there is no more trade. Otherwise, the first shopper buys at the cheapest advertised price. The second in line has less choice because one store is already stocked out. In general, the active shoppers who are "last" in line have less choice than those who are at the head of the line. Figure 3.3, describes the sequence of events within the week.

The price posted by an advertiser in location i in week t is P_{it}. We choose indices such that the price posted by advertisers increases with the location index: $P_{1t} < \cdots < P_{mt}$. The price posted by non-advertisers (indexed $i > m$) is the monopoly price P_t^m.

As was said above, shoppers buy at the cheapest available price. The first shopper chooses store 1 with the price P_1. The second shopper chooses store 2 with the price P_2 and so on. The

[8]This is different from Burdett et al. [18] who assume that buyers can see all prices but cannot see availability.

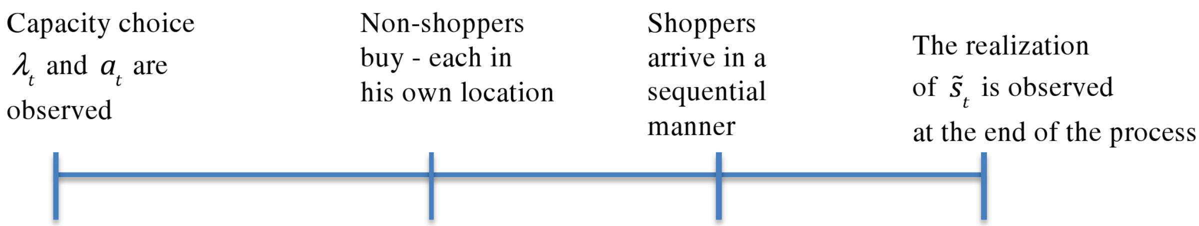


Figure 3.3: Sequence of events within the week

probability of attracting a shopper by a store that advertise the price P_i is:

$$q_i = Prob(\tilde{s} > i) = \sum_{s=i}^{m} \pi_s \tag{3.15}$$

In equilibrium all advertisers make the same expected profits, Π:

$$(k+q_i)P_{it}a_tP_{it}^{\theta} - (k+1)\lambda_t a_t P_{it}^{\theta} = \Pi_t \tag{3.16}$$

This leads to:

$$P_{it} = \frac{(k+1)\lambda_t}{k+q_i} + \frac{\Pi_t}{(k+q_i)a_tP_{it}^{\theta}} \tag{3.17}$$

In the simple version of the UST model in section 3.3.1, $k = \Pi = 0$ and the expected revenue per unit $q_iP_{it} = \lambda_t$ is the same across prices. Here the first term is the unit cost divided by the average capacity utilization $(k+q_i)/(k+1)$. The second term is the expected profit per unit sold. For the definition and existence of equilibrium, see Appendix A.9.

We now turn to study the relationship between the quantity sold by a non-advertiser and the quantity sold by an advertiser. The quantity sold by the non-advertisers is:

$$\ln(x_t^m) = \ln(k) + \ln(a_t) + \theta \ln(P_t^m) \tag{3.18}$$

The quantity sold by an advertiser is:

$$\ln(x_{it}) = \ln(\tilde{\omega}_i) + \ln(a_t) + \theta \ln(P_{it}) \tag{3.19}$$

where $\ln(\tilde{\omega}_i)$, equal to $\ln(1+k)$ with probability q and $\ln(k)$ otherwise, is the number of buyers that shop in the advertiser's store.

Subtracting (3.19) from (3.18) leads to:

$$\begin{aligned} \ln(x_t^m) &= \ln(k) + \theta \ln(P_t^m) + \ln(x_{it}) - \ln(\tilde{\omega}_i) - \theta \ln(P_{it}) \\ &= \ln(x_{it}) + \theta \ln(P_t^m) - \theta \ln(P_{it}) + D_{it} \end{aligned} \tag{3.20}$$

where D_{it} is the difference in the number of buyers between the non-advertiser and the advertiser:

$$D_i = \ln(k) - \ln(\tilde{\omega}_i) = \begin{cases} \ln(k) - \ln(1+k) & \text{if} \quad s \geq i \\ 0 & \text{otherwise} \end{cases} \tag{3.21}$$

This difference is negative if a shopper arrives at the advertiser's store and zero if he does not arrive. Since $\dfrac{x_{it}}{AP_{it}^{\theta}} = 1+k$ if $s \geq i$, we can write (3.21) as:

$$D_i = \begin{cases} \ln(k) - \ln\left(\frac{x_{it}}{a_t P_{it}^{\theta}}\right) & \text{if} \quad s \geq i \\ 0 & \text{otherwise} \end{cases} \tag{3.22}$$

Since $Prob(s \geq i) = q_i$, (3.22) implies: $E(D_{ij}|\tilde{a} = a) = q_i \left(\ln(k) - \ln\left(\frac{x_i}{aP_i^{\theta}}\right)\right) = q_i(\ln(k) - \ln(x_i) - \theta \ln(P_i) - \ln(a))$. The unconditional expectations are:

$$E(D_{it}) = q_i \ln(k) - q_i \ln(x_{it}) + q_i \theta \ln(P_{it}) + q_i E(\ln(\tilde{a})) \tag{3.23}$$

We write:

$$D_{it} = E(D_{it}) + \varepsilon_{it} \tag{3.24}$$

By construction ε_{it} has zero mean and is i.i.d.

Substituting (3.23) and (3.24) in (3.20) leads to:

$$\ln(x_t^m) = \psi_i + (1-q_i)\ln(x_{it}) + \theta\ln(P_t^m) - (1-q_i)\theta\ln(P_{it}) + \varepsilon_{it} \tag{3.25}$$

We can use (3.25) to interpret regression (3.1) of the average quantity sold by stores in the high price bin on the average price in the high price bin and the average price and quantity in the low price bin. Equation (3.25) has the following strong predictions.

Claim 2: (a) the quantity elasticity (the coefficient of $\ln x_{it}$) is between zero and unity and the own price elasticity (the coefficient of $\ln P_t^m$) is greater in absolute value than the cross price elasticity (the coefficient of $\ln P_{it}$); (b) the quantity elasticity and the cross price elasticity are decreasing in the index of the bin.

The quantity elasticity is less than unity because an increase in the quantity sold by the advertiser may be due to the arrival of a shopper rather than an increase in a. It is due to the arrival of a shopper with probability q and therefore the elasticity is only $1-q$. Since the quantity elasticity is decreasing in q it decreases with the index of the bin. In Appendix A.10 we generalize the results in Claim 2 to the case in which the dependent variable is the quantity sold by an advertiser.

3.4 Data Description

We use a rich set of scanner data from Information Resources Inc. (IRI).[9] The complete data set covers 48 markets across the United States, where a market is sometimes a city, e.g. Chicago,

[9] A complete description of the entire data set can be found in Bronnenberg et al. [16].

Los Angeles, New York, and sometimes states, e.g. Mississippi. There are 31 diverse categories of products found in grocery and drug stores, such as carbonated beverages, paper towels, and hot dogs. We define goods by the Universal Product Code (UPC). The data provide information about the total number of units and total revenue for each UPC-store-week cell. We obtain the price for each cell by dividing revenue by the number of units sold. We use data from grocery stores in Chicago during the years 2004 and 2005. We use 3 samples. The 52 weeks in the year 2004, the 52 weeks in the year 2005 and the 104 weeks in the combined sample of 2004-2005.[10]

We apply the following filtering (in a sequential manner):

(a) We drop all UPC-Store cells that do not have positive revenues in all of the samples weeks.[11]

(b) We drop all UPCs that were sold by less than 11 stores.

(c) We drop all categories with less than 10 UPCs.

(d) We drop UPC-Week observations with no price dispersion.

The first exclusion is applied because we cannot distinguish between zero-revenue observations that occur when the item is not on the shelf and zero-revenue observations that occur when the item is on the shelf but was not sold. It is also required for identifying "temporary sale" prices. The second exclusion is aimed at reliable measures of cross sectional price dispersion. The third economizes on the number of category dummies. After applying (a)-(c) we have "semi-balanced" samples in which the number of stores varies across UPCs but stores that are in the sample sold their products in all of the samples weeks.

The requirement that the product be sold continuously by more than 11 stores leads to a sample of fairly popular brands.[12] The focus on fairly popular items is likely to reduce the problem of close substitutes that have different UPCs. In addition, the exclusion of items sold by less than 11 stores

[10] We also replicated the results for other cities (New York, Los Angeles, Philadelphia, Raleigh/Durham and Washington, D.C.). We find strong agreement with the Chicago data presented here.

[11] We also dropped observations in which the quantity sold was zero but revenues were positive.

[12] This is not unique to this paper. Sorensen [78] has collected data on 152 top selling drugs. Lach [58] excluded products that were sold by a small number of stores. Kaplan and Menzio [52] exclude UPCs with less than 25 reported transactions during a quarter in a given market.

significantly reduce the number of items with very high price dispersion that may arise as a result of measurement errors.[13]

Temporary Sales

We assume that a temporary sale occurs when a drop in the price of at least 10% is followed by a price equal to or above the pre-sale price within four weeks. To study the effects of temporary sales we use samples of regular prices obtained from the original samples after deleting all observations in which the price was a "sale" price. After eliminating "sale prices" we used an additional filter that dropped all UPC-week cells that had less than 11 stores or had no price dispersion. Note that the original filter required that each UPC- store cell have strictly positive revenues in all weeks. This allowed for the implementation of our definitions of sales. We drop this requirement in the second round of filtering and as a result the number of stores that sell a given UPC (at a "regular" price) may vary across weeks.

Bins

We split the stores in each UPC-Week cell into bins of approximately equal size. For example, the 2 bins division split the stores in each UPC-Week cell into two categories: High and low price stores, where the price of the stores in the high price bin (bin 1) is greater than or equal to the median.[14]

The price of a UPC in a given store can be above the median in one week and below the median in another week. Indeed most UPC-Store combinations are sometimes above the median and sometimes below the median. Only about 4% of the UPC-Store combinations are above the median in more than 95% of the weeks.

[13] To get a sense of the effect of the sample exclusion on the result, Eden [35] studies one week in detail. Indeed there is a difference between the sample of 8,602 UPCs that were sold by more than 1 store during that week and the sample of 4,537 UPCs that were sold by more than 10 stores. Relative to the larger sample, price dispersion in the smaller sample is lower. The highest price dispersion was found in an item that was sold by 2 stores and for this item the ratio of the highest to lowest price was 15.

[14] For example, if there are 3 stores and the prices are: 5 in store 1, 6 in store 2 and 7 in store 3 then stores 2 and 3 are in bin 1. If the prices are 6 by stores 1 and 2, and 7 by store 3, then only store 3 will be in bin 1.

Summary statistics are in Table 3.1. The first rows are the number of UPCs and the number of observations for individual categories based on the 2 bins samples. In the 2004 sample there were 32 UPCs in the beer category. The number of observations (UPC-Week cells) is $32 \cdot 52 = 1664$. In 2005 there were 56 UPCs in the beer category. The number of observations is not equal to $56 \cdot 52$ because in 3 cells there was no price dispersion. The total number of observations for each sample is in the bottom of the Table. The combined 04-05 sample has fewer UPCs because criterion (a) in our filtering procedure is harder to satisfy when there are 104 weeks. As a result the combined sample includes relatively more popular brands. The total number of observations varies with the number of bins because of insufficient price dispersion. For example if there are 20 stores in a UPC-week cell with 10 stores posting the price 1 and 10 posting the price 2, the stores can be easily divided into 2 bins but not into 3 or 5 bins. For the same reason, the number of observations in the samples of regular prices is lower than the number of observations in the samples of all prices. The number of observations reported here is for the original dollar prices. Later, when we use residuals instead of the original prices, almost all cells have price dispersion and as a result the number of observations is closer to the number of UPCs times weeks.

Table 3.2 is about bin size. As was said before, the bins are only approximately the same size because of the discrete nature of the data. In the 2 bins division, 60% of the stores are in bin 1 and 40% in bin 2. Later, when we control for store effects, the size of the bins are more similar.

Table 3.3 is about the frequency of temporary sales. The last column labeled as "frequency of sales" is the number of "sale prices" divided by the number of prices in the sample. Since our sample size varies with the number of bins, the frequency of sales varies slightly between samples. It is 0.20 when dividing the stores in the 2005 sample into 2 or 3 bins and 0.22 when dividing the stores into 5 bins.

The first 5 columns in Table 3.3 are the frequency of sales by bin. This is calculated by dividing the number of "sale prices" in the bin (aggregating over all UPCs and weeks) by the number of prices in the bin. When using the 2005 sample and the 2 bins division, 10% of the prices in bin 1 are "sale prices". The number for bin 2 is 34%. Using the 2005 sample and the 5 bins division,

	2004		2005		2004-2005	
Category	# UPCs	Obs.	# UPCs	Obs.	# UPCs	Obs.
Beer	32	1,664	56	2,909	20	2,080
Carbonated Beverages	86	4,472	144	7,471	58	6,032
Cold Cereal	93	4,836	133	6,900	53	5,512
Facial Tissues	12	624	18	893	-	-
Frozen Dinner Entrees	36	1,871	75	3,765	-	-
Frozen Pizza	25	1,300	53	2,744	12	1,248
Hot Dogs	14	728	21	1,091	-	-
Margarine & Butter	25	1,300	40	2,060	18	1,872
Mayonnaise	17	884	19	988	-	-
Milk	32	1,664	64	3,294	23	2,392
Mustard & Ketchup	14	728	21	1,092	-	-
Paper Towels	-	-	19	901	-	-
Peanut Butter	18	936	24	1,245	11	1,144
Salty Snacks	94	4,887	120	6,226	42	4,368
Soup	49	2,548	74	3,826	22	2,288
Spaghetti Sauce	13	676	32	1,660	-	-
Toilet Tissue	13	676	19	958	-	-
Yogurt	92	4,783	152	7,870	65	6,760
Totals						
1 Bin, All Prices	665	34,580	1084	56,368	324	33,696
2 Bins, All Prices	665	34,577	1084	55,893	324	33,696
3 Bins, All Prices	665	34,273	1084	54,650	324	33,624
5 Bins, All Prices	665	29,533	1084	45,923	324	29,234
1 Bin, Regular Prices	80	4,160	215	11,180	18	1,872
2 Bins, Regular Prices	80	4,158	215	10,860	18	1,872
3 Bins, Regular Prices	80	3,286	215	10,060	18	1,640
5 Bins, Regular Prices	72	2,424	212	7,770	18	1,160

Table 3.1: Summary statistics for the three samples.
An observation is a UPC - Week cell. The first column is the category name. The two columns that follow are about the 2004 sample. The first is the number of UPCs in each category and the second is the number of UPC-Weeks in that category. The next two columns are for the 2005 sample and the last two columns are for the combined 2004-05 sample. Totals are in the last rows.

42% of the prices in the lowest price bin (bin 5) are sale prices. The number for the highest price bin (bin 1) is 5%. This says that the fact that an item is on sale does not guarantee that it is cheap relative to the prices offered in the same week. The fraction of prices on sale is increasing with the index of the bin suggesting that the probability that an item is cheap relative to other stores given that it is on "sale" is higher than the unconditional probability.

Table 3.4 estimates the conditional probabilities: The probability that a price is in bin *i* given that it is a "sale price". For example, when using the 2005 sample and a 2 bins division, the probability that a "sale price" is in bin 1 is 0.3. This conditional probability is calculated as follows. Using Table 2, the unconditional probability that a price is in bin 1 is: $Prob(bin1) = 0.6$. Using the last column in Table 2a, the unconditional probability that a price is a "sale price" is: $Prob(Sale) = 0.2$. The probability that a price in bin 1 is a "sale price" is in the first column of Table 2a. It is: $Prob(Sale|bin = 1) = 0.1$. The probability that a price is in bin 1 and it is a "sale price" is: $Prob(bin1 \cap Sale) = Prob(bin1)Prob(Sale|bin1) = (0.6)(0.1) = 0.06$. The probability that a price is in bin 1 given that it is a sale price is: $Prob(bin1|price = \text{"}sale\text{"}) = \dfrac{Prob(bin1 \cap Sale)}{Prob(Sale)} = \dfrac{0.06}{0.2} = 0.3$. There is a remarkable agreement about the estimates of the conditional probabilities across samples.

The observation that a "sale price" can be in the highest price bin is surprising. It is possible that some stores are more expensive than others and they are in the highest price bin even when they have a "sale". It is also possible that the timing of "sales" is correlated across stores. We will try to distinguish between these two explanations later when we remove store effects.

Table 3.3 provides the averages of the main variables using the 2 bins division. The difference in average log price between the high price stores and the low price stores (P1-P2) is about 20%.[15] The difference in the average log quantity sold (X2-X1) is 58% for the 2004 sample, 37% for the 2005 sample and 49% for the 04-05 sample. These differences are smaller when using the sample of regular prices. For regular prices, the average price is about 15% higher in the high price bin and the average quantity is about 25% higher in the low price bin. Thus temporary sales contribute

[15] It is 21% for the 2004 sample, 18% for the 2005 sample and 21% for the combined 04-05 sample

2 Bins	Bin 1	Bin 2			
2004	0.60	0.40			
2005	0.60	0.40			
2004-2005	0.60	0.40			
3 Bins	**Bin 1**	**Bin 2**	**Bin 3**		
2004	0.47	0.25	0.28		
2005	0.47	0.24	0.29		
2004-2005	0.46	0.25	0.29		
5 Bins	**Bin 1**	**Bin 2**	**Bin 3**	**Bin 4**	**Bin 5**
2004	0.34	0.16	0.16	0.15	0.19
2005	0.34	0.16	0.15	0.15	0.20
2004-2005	0.33	0.16	0.16	0.15	0.20

Table 3.2: Bin size.
The average fraction of stores in each bin. Averages are over weeks and UPCs.

2 Bins	Bin 1	Bin 2				Freq. Sale
2004	0.09	0.035				0.19
2005	0.10	0.034				0.20
2004-2005	0.10	0.37				0.21
3 Bins	**Bin 1**	**Bin 2**	**Bin 3**			**Freq. Sale**
2004	0.06	0.22	0.38			0.19
2005	0.07	0.24	0.37			0.20
2004-2005	0.07	0.25	0.41			0.21
5 Bins	**Bin 1**	**Bin 2**	**Bin 3**	**Bin 4**	**Bin 5**	**Freq. Sale**
2004	0.04	0.16	0.24	0.33	0.43	0.21
2005	0.05	0.18	0.27	0.32	0.42	0.22
2004-2005	0.05	0.18	0.27	0.35	0.45	0.23

Table 3.3: Frequency of temporary sales by bins.
The first 5 columns are the frequency of "temporary sales" by bins. These frequencies are obtained by dividing the number of "temporary sale prices" in the bin (aggregating over UPCs and weeks) by the total number of prices in the bin. The last column is obtained by dividing the number of "temporary sale prices" in the sample (aggregating over bins, weeks and UPCs) by the total number of prices.

2 Bins	Bin 1	Bin 2			
2004	0.28	0.72			
2005	0.30	0.70			
2004-2005	0.28	0.72			
3 Bins	Bin 1	Bin 2	Bin 3		
2004	0.15	0.29	0.56		
2005	0.16	0.30	0.54		
2004-2005	0.15	0.30	0.56		
5 Bins	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
2004	0.07	0.12	0.18	0.24	0.39
2005	0.07	0.13	0.18	0.23	0.39
2004-2005	0.07	0.12	0.18	0.23	0.39

Table 3.4: The probability that the price is in bin *i* given that it is a "sale price".

to both price dispersion and unit dispersion.

Table 3.4 computes the standard deviation of the average price and the average quantity over weeks. We first calculate the average (over stores) price and units for each UPC-week-bin cell. We then calculate the standard deviation of these averages for each UPC-bin across weeks. Table 3.4 reports the average of these standard deviations across UPCs. In the two bins case, the standard deviation of P2 (the average weekly price in the low price bin) is more than 30% larger than the standard deviation of P1. It is larger by 54% for the 2004 sample, by 30% for the 2005 sample and by 40% for the 04-05 sample. The standard deviations of the quantities are also larger for the low price bin. The quantity standard deviation is larger by 47% for the 2004 sample, by 35% for the 2005 sample and by 39% for the 04-05 sample.

The following 3 rows in Table 3.4 describe the standard deviations when dividing each UPC-Week cell into three bins: High, medium and low. Also here the standard deviation of the price in the low price bin is higher than the standard deviation of the price in the high price bin. The last rows in Table 3.4 are the standard deviations when dividing each UPC-Week cell into 5 bins. The standard deviations in bin 5 (the lowest price bin) are higher than the standard deviations in bin 1 (the highest price bin). The ratio of the standard deviations of the average price in bin 5 to

All Prices	P1	P2	X1	X2	# Stores
2004	0.81	0.59	2.76	3.35	15.43
2005	0.86	0.68	2.63	2.99	21.05
2004-2005	0.76	0.55	3.07	3.56	14.56
Regular Prices					
2004	0.90	0.76	2.97	3.27	15.89
2005	1.08	0.93	2.70	2.93	22.54
2004-2005	1.17	1.03	3.01	3.33	15.70

Table 3.3: Mean values.

The table uses the 2 bins division to provide the mean of the variables. P1 is the average log price for high price stores, P2 is the average log price for low price stores, X1 is the average log of the quantity sold for the high price stores and X2 is the average for the low price stores. The first rows use the sample of all prices and the last rows use the sample of regular prices obtained by deleting observations that are labeled as "sale prices". The last column is the average number of stores (average across UPCs).

the standard deviation in bin 1 is 1.8 on average (2 for 2004, 1.6 for 2005 and 1.76 for 2004-05). For quantities the average ratio is 1.6 (1.76 for 2004, 1.46 for 2005 and 1.61 for 2004-05). When using the samples of regular prices (Table 3.5) these ratios are smaller. For prices the average ratio is 1.62 (1.64, 1.46 and 1.77). And for quantities the average ratio is 1.36 (1.42, 1.23 and 1.44).

Figure 3.4a plots the standard deviations in the last rows of Table 3.4 (5 bins division) and the 2005 sample. The standard deviations are increasing with the index of the bin and there is a good fit between the quantity standard deviation and the price standard deviation. The quantity standard deviation is about 5.7 times the price standard deviation.[16] Figure 3.4b uses the sample of regular prices (Table 3.5). The standard deviation of the regular price (SD RP) in the highest price bin (bin 1) is lower than the standard deviations in lower price bins. The standard deviation of the "regular" quantity (SD RX) is relatively low both for the highest price bin and the lowest price bin. Figure 3.4c compares the standard deviations. The standard deviations are higher in the all prices sample. The standard deviations of quantities are increasing in the all price sample but are more like a "hump shape" in the sample of regular prices.

The observation that variations over weeks in posted price are lower for higher price bins is

[16]The coefficient varies across bins and samples from 5 to 6.5. For the 2004 sample the average coefficient is 6 and for the 2005 sample it is 5.3.

One bin	2004	2005	2004-2005
P	0.0765	0.0908	0.0837
X	0.3132	0.3166	0.3049

Two bins	2004	2005	2004-2005
P1	0.0690	0.0862	0.0784
P2	0.1066	0.1119	0.1106
X1	0.3276	0.3401	0.3289
X2	0.4824	0.4586	0.4562

Three bins	2004	2005	2004-2005
P1	0.0633	0.0758	0.0728
P2	0.0936	0.1065	0.1008
P3	0.1172	0.1129	0.1174
X1	0.3422	0.3543	0.3494
X2	0.4856	0.5001	0.4970
X3	0.5951	0.5138	0.5544

Five bins	2004	2005	2004-2005
P1	0.0612	0.0721	0.0712
P2	0.0822	0.0967	0.0941
P3	0.0953	0.1060	0.1029
P4	0.1140	0.1109	0.1148
P5	0.1248	0.1148	0.1253
X1	0.3865	0.3914	0.3972
X2	0.5317	0.5295	0.5533
X3	0.5751	0.5632	0.5902
X4	0.6520	0.5921	0.6359
X5	0.6806	0.5726	0.6394

Table 3.4: Standard deviations over weeks.

The table reports standard deviations over weeks. We first calculate the average price and units for each UPC-week-bin cell. We then calculate the standard deviation of these averages for each UPC-bin across weeks. The first rows report the standard deviation for the 2 bins case. The next rows report the standard deviation for the 3 bins case and the rows in the bottom report the standard deviation for the 5 bins case.

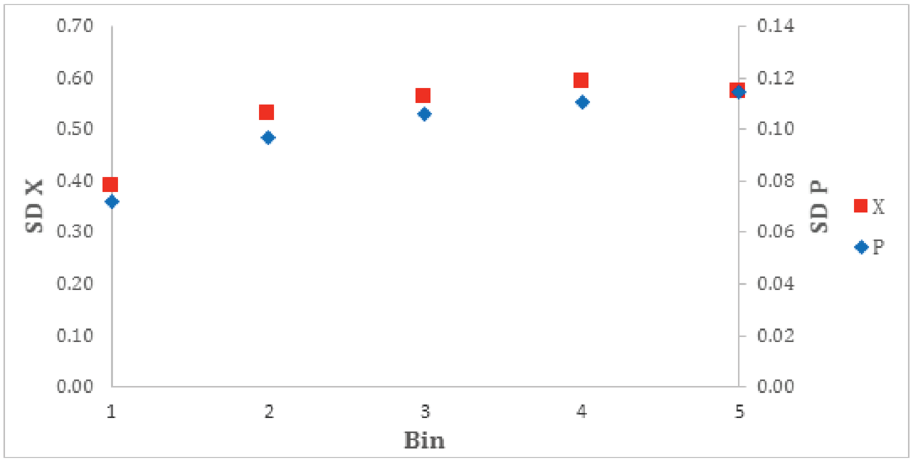
One bin	2004	2005	2004-2005
P	0.0219	0.0392	0.0318
X	0.1943	0.2068	0.2104

Two Bins	2004	2005	2004-2005
P1	0.0244	0.0392	0.0309
P2	0.0363	0.0544	0.0424
X1	0.2300	0.2595	0.2334
X2	0.3019	0.3288	0.3536

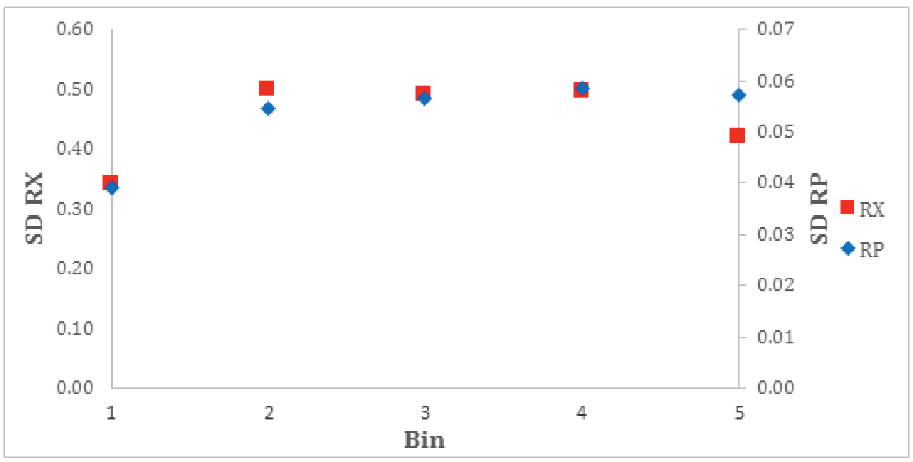
Three Bins	2004	2005	2004-2005
P1	0.0251	0.0386	0.0287
P2	0.0435	0.0548	0.0539
P3	0.0420	0.0554	0.0460
X1	0.2576	0.2963	0.2668
X2	0.4094	0.4206	0.4603
X3	0.3473	0.3586	0.3998

Five Bin	2004	2005	2004-2005
P1	0.0270	0.0393	0.0282
P2	0.0416	0.0547	0.0535
P3	0.0430	0.0566	0.0529
P4	0.0460	0.0585	0.0548
P5	0.0443	0.0573	0.0499
X1	0.2975	0.3412	0.2977
X2	0.4732	0.4980	0.5247
X3	0.4752	0.4916	0.5318
X4	0.4974	0.4961	0.5255
X5	0.4211	0.4202	0.4296

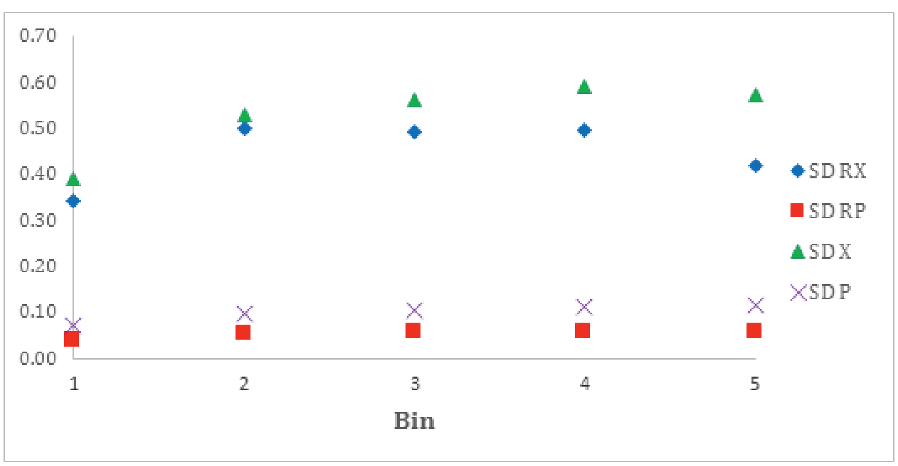
Table 3.5: The Samples of regular prices.
This table repeats the calculations in Table 3.4 after eliminating all “temporary sale” observations.



(a) The sample of all prices: SD X is the standard deviation of the quantity and SD P is the standard deviation of the price.



(b) The sample of regular prices: SD RX is the standard deviation of the quantity and SD RP is the standard deviation of the price.



(c) The sample of all prices (SD X and SD P) and the sample of regular prices (SD RX and SD RP)

Figure 3.4: The standard deviations across bins using the 2005 sample.

roughly consistent with the model in section 3.3.2. In this model there are no variations in the high price but the low price may depend on the amount of inventories. It seems that we need non-shoppers to account for the observation that variations over weeks in the quantity sold are lower for higher price bins. In section 3.3.4 high price stores specialize in servicing non-shoppers and the quantity sold by the high price stores do not fluctuate with the number of shoppers. The observation that the standard deviations are lower for the sample of regular prices and the difference in the standard deviations is especially large for the cheapest price bin is consistent with the model in section 3.3.2 because eliminating temporary sales observations reduces the variations over weeks in the first market price and quantity.

3.5 Elasticities

We now turn to examine the relationship (3.25) between the quantity sold by the highest price stores and the quantity sold by lower price stores. We start by splitting the stores in each UPC-week cell into two groups (below and above the median price) and run in Table 3.4 the average log quantity in the high price bin (X1) on the average log quantity and price in the low price bin (X2 and P2) and the average log price in the high price bin (P1). Averages are across the stores in the UPC-week-bin cell. The elasticity with respect to the average quantity sold per store in the low price group (X2) is about 0.6 in all the three samples. The own price elasticity is about -2 and the cross price elasticity is about 1.8. The rows in the bottom of the Table report the results when using the samples of regular prices. Part (a) of Claim 2 works. The quantity elasticity is less than unity and the cross price elasticity is less than the absolute value of the own price elasticity.

To examine Part (b) which says that the quantity elasticity and the cross price elasticity decline with the price distance, we need the 3 and 5 bins divisions. Table 3.5 describes the results when using the 3 bins division. The first 2 columns use the 2004 sample: The first uses the averages from the medium price stores as explanatory variables and the second uses the variables from the low price stores as explanatory variables. The quantity sold by the high price stores is more sensitive to the variables in the medium price stores. The elasticity with respect to the quantity sold in the medium price stores (the coefficient of X2) is 0.6 while the elasticity with respect to the quantity sold in the low price stores (the coefficient of X3) is 0.5. The elasticity with respect to the price in the medium price group is 1.7 while the elasticity with respect to the price in the low price group is 1.5. This pattern occurs also in the other two samples and is consistent with Claim 2. The rows that follow use samples of regular prices. The coefficients of X2 are slightly lower than the coefficients of X3. The coefficients of P2 are much higher than the coefficients of P3. Also here part (a) of Claim 2 works. Part (b) works for the sample of all prices but the results for the sample of regular prices are mixed.

Table 3.6 uses the 5 bins division and the largest 2005 sample. It describes the results when running the average quantity in the high price stores (bin 1) on the average quantity and price in the other 4 bins. In the first 4 columns the explanatory variables are from a single bin: From bin 2 in the first column, from bin 3 in the second and so on. In the last column we report the regression results when using all the explanatory variables. The first rows use the sample of all prices. The last rows use the sample of regular prices.

Part (a) of Claim 2 works. The quantity elasticity is less than unity and the cross price elasticity is less than the absolute value of the own price elasticity. Part (b) works for the sample of all prices. In the first four columns the elasticities decline with the distance from the high price group. The elasticity of the quantity sold with respect to the quantity sold by stores in bin 2 is 0.57 while the elasticity with respect to the quantity sold by stores in bin 5 (the lowest price stores) is 0.48. The elasticity with respect to the price posted by bin 2 stores is 1.75 while the elasticity with respect to the price posted by bin 5 stores is 1.3. Part (a) works also for the sample of regular prices but for these samples the results with respect to part (b) are mixed.

Figure 3.5 describes the results for the 3 samples of “all prices” when using three explanatory variables as in the first four columns of Table 3.6. Figure 3.5a is the elasticity with respect to the quantity sold (the quantity elasticity = the coefficient of Xj), Figure 3.5b is the cross price elasticities (the coefficient of Pj). As can be seen there is a strong agreement among the three samples. The quantity and cross price elasticities are both decreasing in the bin index. Figure

	2004	2005	2004-2005
	X1	X1	X1
Full sample	(1)	(2)	(3)
X2	0.6240***	0.6390***	0.6487***
	(0.0030)	(0.0024)	(0.0031)
P1	-1.9794***	-1.8803***	-2.0269***
	(0.0208)	(0.0157)	(0.0225)
P2	1.7941***	1.6484***	1.8618***
	(0.0209)	(0.0157)	(0.0218)
R^2	0.6846	0.6950	0.6977
N	34,070	54,273	33,279
	X1	X1	X1
Regular prices	(4)	(5)	(6)
X2	0.5919***	0.6759***	0.6593***
	(0.0085)	(0.0058)	(0.0121)
P1	-1.7665***	-1.2157***	-0.7332***
	(0.1138)	(0.0556)	(0.1938)
P2	1.5475***	1.0903***	0.6356***
	(0.1115)	(0.0537)	(0.1874)
R^2	0.8205	0.7063	0.7563
n	4,150	10,860	1.872

Table 3.4: Two bins regression.

Standard errors in parentheses. One star (*) denotes p-value of 10%, two stars (**) denote p-value of 5% and three stars (***) denote p-value of 1%. Category dummies are included in all the regressions. X1 is the average log of the quantity sold across stores in the high price bin, X2 is the average across stores in the low price bin, P1 is the average log price across stores in the high price bin and P2 is the average across stores in the low price bin.

	2004		2005		2004-2005	
	X1	X1	X1	X1	X1	X1
Full sample	(1)	(2)	(3)	(4)	(5)	(6)
X2	0.6125***		0.6019***		0.6062***	
	(0.0034)		(0.0028)		(0.0035)	
P2	1.7259***		1.7131***		1.8254***	
	(0.0271)		(0.0208)		(0.0280)	
X3		0.5026***		0.5466***		0.5418***
		(0.0032)		(0.0027)		(0.0033)
P3		1.5115***		1.4728***		1.6487***
		(0.0193)		(0.0152)		(0.0205)
P1	-1.8924***	-1.7559***	-1.9234***	-1.7478***	-1.9492***	-1.8375***
	(0.0274)	(0.0196)	(0.0208)	(0.0153)	(0.0289)	(0.0214)
R^2	0.6241	0.5838	0.6248	0.6057	0.6261	0.6081
n	34,273	34,273	54,650	54,650	33,624	33,624
	X1	X1	X1	X1	X1	X1
Regular prices	(7)	(8)	(9)	(10)	(11)	(12)
X2	0.5048***		0.6308***		0.6160***	
	(0.0118)		(0.0072)		(0.0151)	
P2	2.3644***		1.0252***		2.4977***	
	(0.1712)		(0.0811)		(0.2223)	
X3		0.5146***		0.6518***		0.6423***
		(0.0105)		(0.0066)		(0.0127)
P3		1.5361***		0.7428***		0.6728***
		(0.1098)		(0.0557)		(0.1797)
P1	-2.6050***	-1.7701***	-1.1542***	-0.8342***	-2.6847***	-0.7556***
	(0.1751)	(0.1138)	(0.0835)	(0.0583)	(0.2257)	(0.1856)
R^2	0.5532	0.5925	0.5943	0.6435	0.6607	0.7255
n	3,286	3,286	10,060	10,060	1,640	1,640

Table 3.5: Three bins regressions.
Each UPC-week cell is divided into three bins. Xj = the average log units in bin j. Pj = the average log price in bin j. Category dummies are included in all the regressions.

3.6 uses the samples of all prices to plot the coefficients of the regressions that use 9 explanatory variables as in the last column in Table 3.6. Also here there is a strong agreement among the three samples and the qualitative results do not change.

The results obtained when using the samples of all prices support the hypothesis that the quantity elasticities and the cross price elasticities decrease in the index of the bin. The results when using the samples of regular prices are mixed, possibly due to the fact that shoppers play a critical role in obtaining the results in Claim 2 and removing temporary sales prices may have reduced the role of shoppers who are looking for bargains.

3.6 Store Effect

Stores that are similar in price may be similar in other ways. For example, stores in rich neighborhoods may charge on average a price that is higher than the price charged by stores in poor neighborhoods. If shoppers shop with higher intensity in their own neighborhood, the quantity sold by a group of stores maybe more sensitive to the variables in a group of stores that is close in price because the two groups are also closer in locations.

In an attempt to address this problem we remove the store effect by running the following regressions.

$$\ln(P_{ijt}) = a_i + b_j(store - dummy) + e^P_{itj} \tag{3.26}$$

$$\ln(x_{ijt}) = a_i + b_j(store - dummy) + e^x_{ijt} \tag{3.27}$$

where P is price, x is quantity sold, i index the UPC, j index the store and t index the week. We then repeat the above tables after replacing $\ln(P)$ with the residuals e^P_{ijt} and $\ln(x)$ with the residuals e^x_{ijt}.

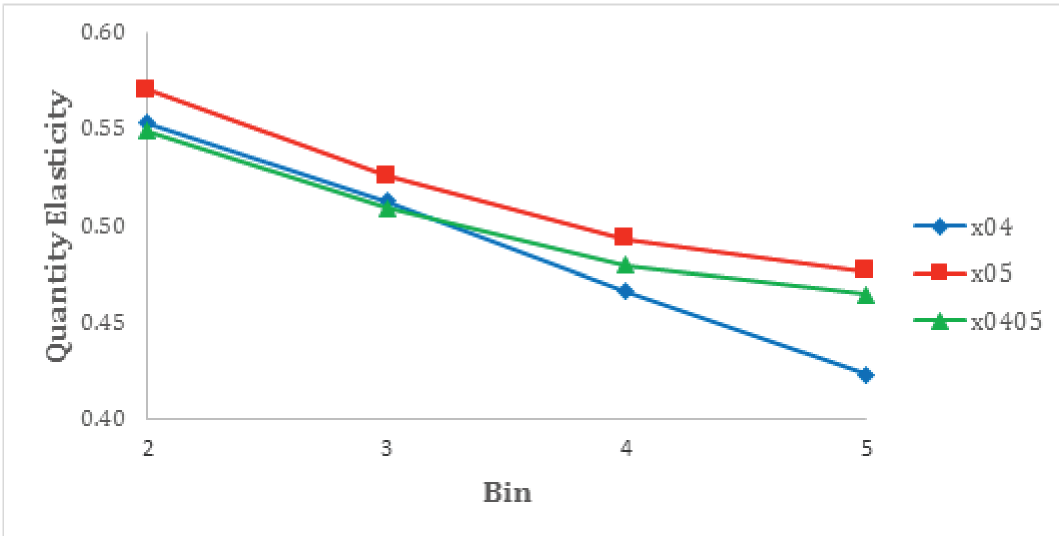
Tables 3.7 are comparable Tables 3.2. The bins are more equal in size because the residuals are different across stores and the problem of lack of price dispersion is less common. The conditional probability in Table 3.9 are not very different from the conditional probabilities in Table 3.4.

Full sample	X1 (1)	X1 (2)	X1 (3)	X1 (4)	X1 (5)
X2	0.5703***				0.2868***
X3		0.5259***			0.1718***
X4			0.4933***		0.1456***
X5				0.4764***	0.1468***
P1	-1.9572***	-1.8640***	-1.7809***	-1.6082***	-2.591***
P2	1.7537***				1.3498***
P3		1.6279***			0.6461***
P4			1.5063***		0.2972***
P5				1.3114***	0.1389***
R^2	0.582	0.5483	0.5399	0.5382	0.6684
n	45,923	45,923	45,923	45,923	45,923

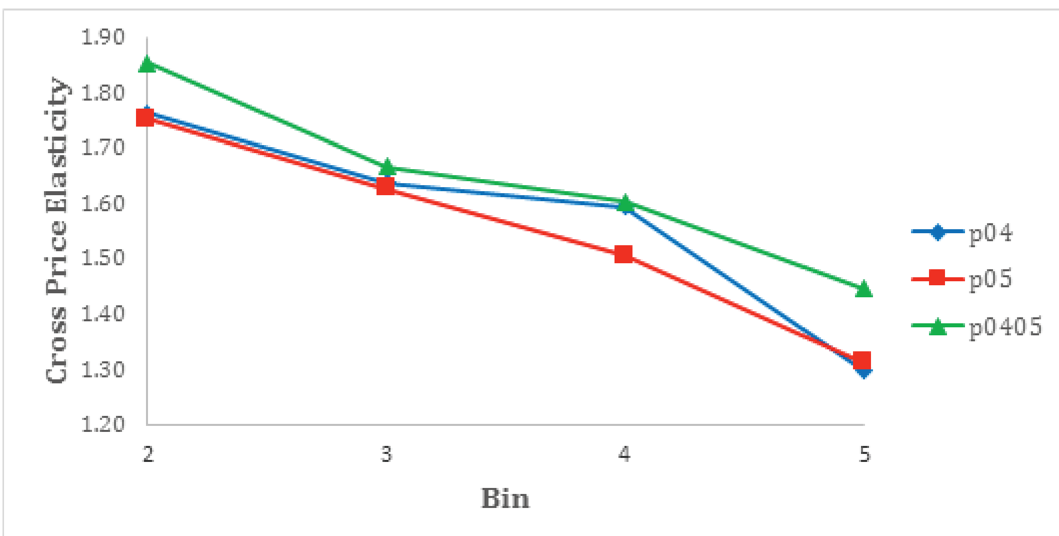
Regular prices	X1 (6)	X1 (7)	X1 (8)	X1 (9)	X1 (10)
X2	0.5534***				0.2384***
X3		0.5556***			0.1927***
X4			0.5424***		0.1784***
X5				0.5696***	0.1986***
P1	-1.3523***	-1.4064***	-1.2412***	1.0328***	-1.8160***
P2	1.2258***				1.0234***
P3		1.2685***			0.5513***
P4			1.0965***		0.1396
P5				0.9089***	0.0331
R^2	0.5692	0.5511	0.5534	0.5748	0.6799
n	7,770	7,770	7,770	7,770	7,770

Table 3.6: Five bins regression using the 2005 sample.

The first four columns report the results when using 3 explanatory variables. The last column is the results when using 9 explanatory variables. The first rows use the sample of all prices. The rows that follow use the sample of regular prices. Category dummies are included in all the regressions.

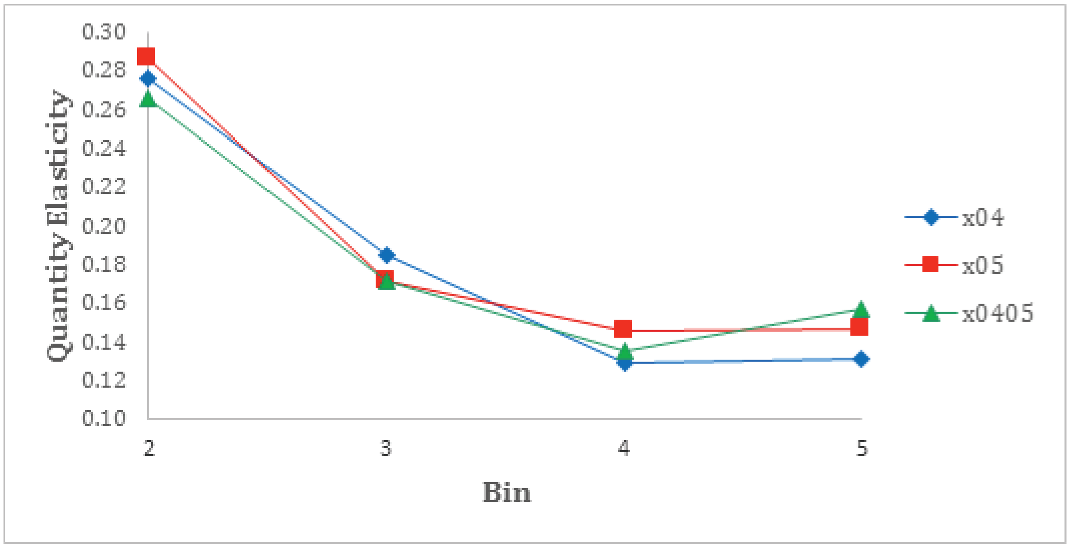


(a) Elasticity with respect to the quantity sold

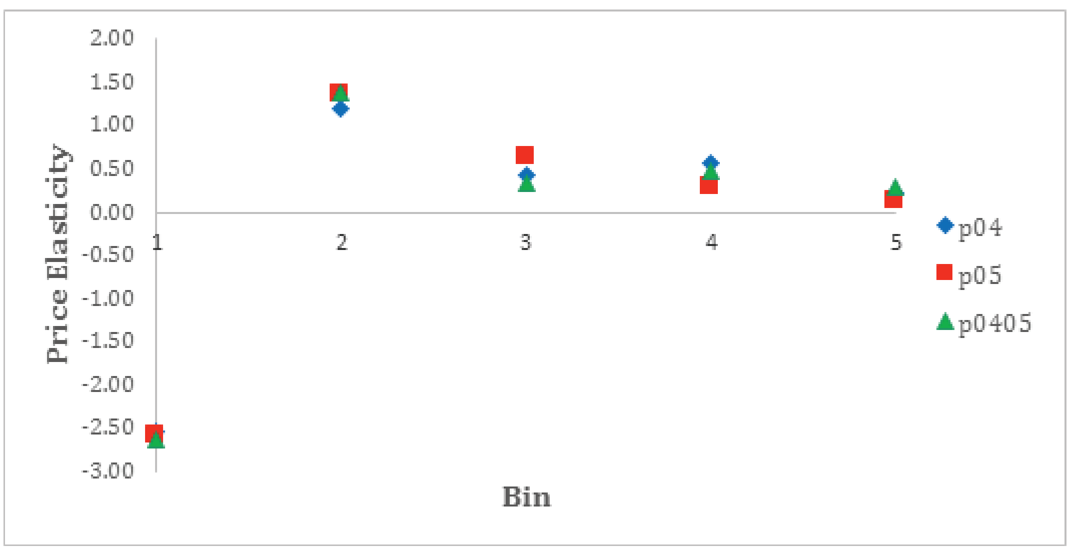


(b) Cross price elasticities

Figure 3.5: Elasticities based on a 3 explanatory variables regressions (First four columns in Table 3.6, sample of all prices).



(a) Quantity elasticities



(b) Price elasticities

Figure 3.6: Elasticities based on a 9 explanatory variables regression in the last column of Table 3.6.

2 bins	bin 1	bin 2			
2004	0.52	0.48			
2005	0.51	0.49			
2004-2005	0.52	0.48			
3 bins	bin 1	bin 2	bin 3		
2004	0.35	0.31	0.34		
2005	0.35	0.32	0.33		
2004-2005	0.36	0.31	0.34		
5 bins	bin 1	bin 2	bin 3	bin 4	bin 5
2004	0.23	0.19	0.19	0.19	0.21
2005	0.22	0.19	0.19	0.19	0.21
2004-2005	0.23	0.19	0.18	0.19	0.22

Table 3.7: Bin size accounting for store effect.

2 bins	bin 1	bin 2				Freq. Sale
2004	0.08	0.31				0.19
2005	0.11	0.29				0.19
2004-2005	0.10	0.34				0.21
3 bins	bin 1	bin 2	bin 3			Freq. Sale
2004	0.06	0.16	0.35			0.19
2005	0.08	0.19	0.32			0.19
2004-2005	0.07	0.19	0.38			0.21
5 bins	bin 1	bin 2	bin 3	bin 4	bin 5	Freq. Sale
2004	0.05	0.09	0.16	0.27	0.39	0.19
2005	0.07	0.12	0.18	0.26	0.35	0.19
2004-2005	0.06	0.11	0.18	0.30	0.42	0.21

Table 3.8: Frequency of temporary sales by bins accounting for store effect.

2 bins	bin 1	bin 2			
2004	0.23	0.77			
2005	0.28	0.72			
2004-2005	0.24	0.76			
3 bins	bin 1	bin 2	bin 3		
2004	0.11	0.27	0.62		
2005	0.15	0.30	0.55		
2004-2005	0.12	0.27	0.61		
5 bins	bin 1	bin 2	bin 3	bin 4	bin 5
2004	0.06	0.09	0.16	0.26	0.43
2005	0.08	0.11	0.18	0.25	0.37
2004-2005	0.06	0.10	0.16	0.26	0.43

Table 3.9: The probability that the price is in bin *i* given that it is a "sale price" accounting for store effect.

Table 3.10 and 3.11 are comparable to Tables 3.4 and 3.5. The results are qualitatively the same suggesting that store effects do not drive our findings about the standard deviations by bins.

Table 3.12 uses the residuals from (3.26) and (3.27) to estimate the two bins regression. The quantity elasticity is about 0.75 and is higher than the elasticity in Table 3.4. As in Table 3.4 and consistent with Claims 2 the absolute value of the own price elasticity is higher than the cross price elasticity.

Table 3.13 uses 3 bins division: high, medium and low price. The estimated elasticities are higher than in Table 3.5 but as in Table 3.5, the quantity elasticity is less than unity and the own price elasticity is higher in absolute value than the cross price elasticity. As in Table 3.5, in the sample of all prices, the quantity sold by the high price stores is more strongly related to the quantity and price in the medium price stores.

Table 3.14 reports the regression estimates when using 5 bins. The quantity elasticities are higher but still less than unity and decreasing with the distance from the highest price bin. The cross price elasticity also decreases with the distance and the absolute value of the own price elasticity is greater than the cross price elasticity.

One bin	2004	2005	2004-2005
P	0.0765	0.0908	0.0836
X	0.3132	0.3166	0.3049

Two bins	2004	2005	2004-2005
P1	0.0641	0.0801	0.0722
P2	0.1036	0.1125	0.1108
X1	0.3131	0.3284	0.3184
X2	0.4396	0.3985	0.4099

Three bins	2004	2005	2004-2005
P1	0.0612	0.0750	0.0692
P2	0.0926	0.1036	0.0911
P3	0.1157	0.1166	0.1218
X1	0.3428	0.3460	0.3473
X2	0.3902	0.4077	0.4003
X3	0.5127	0.4273	0.4647

Five bins	2004	2005	2004-2005
P1	0.0597	0.0718	0.0679
P2	0.0708	0.0879	0.0786
P3	0.0844	0.1054	0.0931
P4	0.1025	0.1170	0.1140
P5	0.1275	0.1209	0.1303
X1	0.3902	0.3794	0.3865
X2	0.4228	0.4255	0.4321
X3	0.4480	0.4531	0.4626
X4	0.5024	0.4798	0.4994
X5	0.6049	0.4708	0.5270

Table 3.10: Standard deviations of the quantity sold for the average UPC by bin (all prices) accounting for store effect.

One bin	2004	2005	2004-2005
P	0.0241	0.0391	0.0320
X	0.1922	0.2054	0.2094

Two Bins	2004	2005	2004-2005
P1	0.0237	0.0367	0.0298
P2	0.0290	0.0462	0.0389
X1	0.2168	0.2409	0.2352
X2	0.2359	0.2391	0.2378

Three Bins	2004	2005	2004-2005
P1	0.0236	0.0358	0.0302
P2	0.0284	0.0437	0.0323
P3	0.0316	0.0481	0.0429
X1	0.2413	0.2639	0.2505
X2	0.2541	0.2719	0.2765
X3	0.2645	0.2600	0.2626

Five Bin	2004	2005	2004-2005
P1	0.0246	0.0356	0.0300
P2	0.0251	0.0395	0.0326
P3	0.0287	0.0443	0.0324
P4	0.0337	0.0483	0.0407
P5	0.0345	0.0500	0.0454
X1	0.2854	0.3070	0.2854
X2	0.3213	0.3216	0.3343
X3	0.3036	0.3172	0.3179
X4	0.3136	0.3198	0.3342
X5	0.3167	0.3009	0.2996

Table 3.11: The samples of regular prices accounting for store effect.

	2004	2005	2004-2005
	X1	X1	X1
Full sample	(1)	(2)	(3)
X2	0.7305***	0.7546***	0.7697***
	(0.0028)	(0.0021)	(0.0028)
P1	-2.3756***	-2.0344***	-2.1366***
	(0.0204)	(0.0152)	(0.0212)
P2	2.1127***	1.7960***	1.9594***
	(0.0205)	(0.0152)	(0.0206)
R^2	0.7701	0.7963	0.7879
n	34,580	56,368	33,696
	X1	X1	X1
Regular prices	(4)	(5)	(6)
X2	0.7081***	0.8580***	0.8967***
	(0.0081)	(0.0044)	(0.0101)
P1	-1.4920***	-1.6860***	-0.7566***
	(0.1010)	(0.0541)	(0.1556)
P2	1.2213***	1.6050***	0.6351***
	(0.1018)	(0.0533)	(0.1576)
R^2	0.8806	0.8639	0.8838
n	4,160	11,180	1,872

Table 3.12: Two bins regressions accounting for store effect.
Standard errors in parentheses. Category dummies are included in all regressions.

	2004		2005		2004-2005	
	X1	X1	X1	X1	X1	X1
Full sample	(1)	(2)	(3)	(4)	(5)	(6)
X2	0.7439***		0.7696***		0.7616***	
	(0.0031)		(0.0024)		(0.0032)	
P2	1.9836***		2.0645***		2.1473***	
	(0.0301)		(0.0205)		(0.0310)	
X3		0.6302***		0.6712***		0.6791***
		(0.0031)		(0.0024)		(0.0032)
P3		1.8253***		1.5818***		1.7519***
		(0.0194)		(0.0146)		(0.0197)
P1	-2.214***	-2.159***	-2.248***	-1.884***	-2.31***	-1.99***
	(0.0301)	(0.0193)	(0.0205)	(0.0147)	(0.0316)	(0.0205)
R^2	0.7411	0.6776	0.7668	0.7118	0.7432	0.6987
n	34,580	34,580	56,368	56,368	33,696	33,696
	X1	X1	X1	X1	X1	X1
Regular prices	(7)	(8)	(9)	(10)	(11)	(12)
X2	0.7387***		0.8075***		0.8653***	
	(0.0089)		(0.0048)		(0.0125)	
P2	1.1600***		1.6713***		0.6341**	
		(0.1531)		(0.0735)		(0.3219)
X3		0.6122***		0.8027***		0.8210***
		(0.0094)		(0.0052)		(0.0123)
P3		1.3228***		1.4705***		0.6226***
		(0.0887)		(0.0502)		(0.1522)
P1	-1.4607***	-1.6792***	-1.7698***	-1.5768***	-0.8800***	-0.8171***
	(0.1534)	(0.0873)	(0.0743)	(0.0513)	(0.3203)	(0.1498)
R^2	0.8734	0.8357	0.8230	0.8049	0.8427	0.8330
n	4,160	4,160	11,180	11,180	1,872	1,872

Table 3.13: Three bins regressions accounting for store effect.

Full sample	X1 (1)	X1 (2)	X1 (3)	X1 (4)	X1 (5)
X2	0.7102***				0.3752***
X3		0.6699***			0.2068***
X4			0.6199***		0.1095***
X5				0.5791***	0.1235***
P1	-2.0454***	-2.0671***	-1.8820***	-1.6689***	-2.4396***
P2	1.8549***				0.9778***
P3		1.8247***			0.9138***
P4			1.5756***		0.1376***
P5				1.3139***	0.2245***
R^2	0.6946	0.6617	0.6349	0.6179	0.7468
n	56,368	56,368	56,368	56,368	56,368

Regular prices	X1 (6)	X1 (7)	X1 (8)	X1 (9)	X1 (10)
X2	0.7700***				0.2952***
X3		0.7483***			0.2208***
X4			0.7342***		0.2223***
X5				0.7227***	0.1538***
P1	-1.9860***	-1.9842***	-1.8659***	-1.4147***	-2.2054***
P2	1.8611***				0.5169***
P3		1.8453***			0.5732***
P4			1.6952***		0.7013***
P5				1.2750***	0.3115***
R^2	0.7512	0.7404	0.7351	0.7098	0.8119
n	11,180	11,180	11,180	11,180	11,180

Table 3.14: Five bins regression accounting for store effect using the 2005 sample.
The first four columns report the results when using 3 explanatory variables. The last column is the results when using 9 explanatory variables. The first rows use the sample of all prices. The rows that follow use the sample of regular prices. Category dummies are included in all the regressions.

3.7 UPC-Specific Store Effect

A store may promote a specific UPC by placing it in a visible and easy to reach place. Therefore we allow for the store effect to vary across UPCs and run for each UPC the following regression.

$$\ln(P_{ijt}) = a_i + b_{ij}(store - dummy) + e^P_{itj} \tag{3.28}$$

$$\ln(x_{ijt}) = a_i + b_{ij}(store - dummy) + e^x_{ijt} \tag{3.29}$$

As before, we repeat the Tables after replacing $\ln(P)$ with the residuals e^P_{itj} and $\ln(x)$ with the residuals e^x_{ijt}. It turns out that the results are qualitatively similar to the case in which we control for non-UPC specific store effects, but there are some large differences in the magnitudes of the estimated elasticities. It thus makes a difference whether one controls for store effects or for UPC specific store effects.

Tables 3.15 and 3.16 are comparable to corresponding Tables 3.2 and Table 3.7. The bin sizes are the same as in Table 3.7 and so are the unconditional frequency of sales (the last column of Table 3.8). Instead of reporting the unconditional frequency of sales, we report now in the last column of Table 3.15, the percentage of weeks in which an average UPC is not on sale in any store. For example, in 2005 the average UPC was not on sale in 43% of the weeks. To appreciate this number we consider the case in which each store uses a mixed strategy to determine whether the item is on sale or not. In the 2005 sample there are on average 21 stores per UPC and the frequency of sale is 0.19. If stores use a mixed strategy as in Varian [82], the probability that there are no sales is: $(1-0.19)^{21} = 0.01$. This suggests no sales in only 1% of the weeks. Since no sales occur in 43% of the weeks, temporary sales are correlated across stores.

In the 2005 sample, the fraction of stores that has the item on sale fluctuates between 0 and 0.7. The average (over UPCs) standard deviation of this fraction is 0.2. There are thus substantial variations (over weeks) in the percentage of stores that has the item on sale. This may explain the conditional probabilities in Table 3.16. In the absence of variations over weeks we will have the item on sale in 19% of the stores in every week and the probability that a price is in bin 5 given

2 bins	bin 1	bin 2				No sales
2004	0.07	0.32				0.40
2005	0.09	0.31				0.43
2004-2005	0.08	0.35				0.38
3 bins	**bin 1**	**bin 2**	**bin 3**			**No sales**
2004	0.05	0.15	0.38			0.40
2005	0.06	0.18	0.35			0.43
2004-2005	0.05	0.18	0.41			0.38
5 bins	**bin 1**	**bin 2**	**bin 3**	**bin 4**	**bin 5**	**No sales**
2004	0.030	0.08	0.15	0.27	0.42	0.40
2005	0.05	0.10	0.17	0.27	0.39	0.43
2004-2005	0.04	0.09	0.17	0.31	0.44	0.38

Table 3.15: Frequency of temporary sales by bins accounting for UPC-specific store effect.
The frequency of temporary sales are the same as in Table 3.8. The last column is the percentage of weeks in which the average UPC is not on ”sale” in any store.

that it is a ”sale price” should be close to one. Instead we find that the conditional probabilities in Tables 3.15 and 3.16 are less than 0.5.

Tables 3.17 and 3.18 are comparable to the corresponding Tables 3.4, 3.5, 3.10, and 3.11. It shows the same pattern: The standard deviation across weeks is increasing with the index of the bin. This is not the case in the sample of regular prices.

Table 3.19 is comparable to Tables 3.4 and 3.12. There are large differences between the Tables. The quantity elasticities and the cross price elasticities are much lower. The quantity elasticity is about 38 to 55 percent of the quantity elasticities in Table 3.4 and 32 to 47 percent of the quantity elasticities in Table 3.12. But still the quantity elasticities are between zero and unity as predicted by the theory. The cross price elasticities are 38-51 percent of the cross price elasticities in Table 4 and 32-37 percent of the cross price elasticities in Table 3.12. The own price elasticities are higher. They are 140-149 percent of the own price elasticities in Table 3.4 and 124-130 percent of the own price elasticities in Table 3.12.

Note that in Table 3.19 the quantity elasticity is close to the cross price elasticity divided by the absolute value of the own price elasticity as implied by (3.25).

2 bins	bin 1	bin 2
2004	0.18	0.82
2005	0.23	0.77
2004-2005	0.2	0.8

3 bins	bin 1	bin 2	bin 3
2004	0.08	0.25	0.67
2005	0.11	0.29	0.6
2004-2005	0.09	0.26	0.64

5 bins	bin 1	bin 2	bin 3	bin 4	bin 5
2004	0.04	0.07	0.15	0.27	0.47
2005	0.05	0.10	0.17	0.27	0.42
2004-2005	0.05	0.08	0.15	0.27	0.45

Table 3.16: The probability that the price is in bin i given that it is a "sale price" accounting for UPC-specific store effect.

Table 3.20 is comparable to Tables 3.5 and 3.13. Also here the quantity elasticities and the cross price elasticities are much lower than in Tables 3.5 and 3.13 and the own price elasticity is higher in absolute value. Here, with UPC specific residuals, the distance matters more than with non-specific residuals.

Table 3.21 is comparable to Table 3.6 and Table 3.14. Relative to Table 3.14, the quantity elasticity and the cross price elasticities are considerably lower, suggesting that it makes a difference if we control for UPC specific store effects or just for store effects.

We calculated Table 3.21 for the other two samples (2004 and 04-05). Figures 3.7 and 3.8 are comparable to Figures 3.5 and 3.6 and describe the estimate of all three samples. As can be seen there is a considerable degree of consensus about the elasticities across the three samples, especially if we use the 9 variables regression that is reported in the last column of Table 3.21.

3.8 The Contribution of Temporary Sales to Price Flexibility

A first look at the data through the lens of sticky price models, suggests that prices move too much. But this is not the case once a distinction between regular and sale prices is introduced.

One bin	2004	2005	2004-2005
P	0.0765	0.0908	0.0836
X	0.3132	0.3166	0.3049
Two bins	**2004**	**2005**	**2004-2005**
P1	0.06	0.0912	0.0713
P2	0.1069	0.1122	0.1126
X1	0.2931	0.3211	0.2964
X2	0.4282	0.3868	0.4005
Three bins	**2004**	**2005**	**2004-2005**
P1	0.0582	0.0762	0.0692
P2	0.0787	0.1036	0.0906
P3	0.1214	0.1172	0.125
X1	0.3095	0.3287	0.3117
X2	0.3713	0.3919	0.3707
X3	0.4908	0.4099	0.4487
Five bins	**2004**	**2005**	**2004-2005**
P1	0.0573	0.0724	0.0672
P2	0.0659	0.0904	0.0789
P3	0.0904	0.1056	0.0926
P4	0.1041	0.1161	0.1135
P5	0.1345	0.1227	0.1336
X1	0.3415	0.349	0.3384
X2	0.3757	0.3952	0.3718
X3	0.4104	0.4262	0.4085
X4	0.4672	0.4467	0.4517
X5	0.5709	0.4467	0.4965

Table 3.17: Standard deviations of the quantity sold for the average UPC by bin accounting for UPC-specific store effect.

One bin	2004	2005	2004-2005
P	0.0211	0.0380	0.0312
X	0.1861	0.2017	0.2081

Two Bins	2004	2005	2004-2005
P1	0.0189	0.0385	0.0320
P2	0.0266	0.0430	0.0359
X1	0.2111	0.2232	0.2350
X2	0.2190	0.2328	0.2317

Three bins	2004	2005	2004-2005
P1	0.0204	0.0403	0.0343
P2	0.0174	0.0370	0.0286
P3	0.0315	0.0472	0.0410
X1	0.2335	0.2418	0.2547
X2	0.2329	0.2468	0.2492
X3	0.2451	0.2548	0.2539

Five bins	2004	2005	2004-2005
P1	0.0228	0.0425	0.0369
P2	0.0173	0.0387	0.0317
P3	0.0174	0.0372	0.0286
P4	0.0225	0.0410	0.0337
P5	0.0367	0.0522	0.0457
X1	0.2635	0.2682	0.2766
X2	0.2776	0.2812	0.3002
X3	0.2709	0.2772	0.2817
X4	0.2806	0.2826	0.2961
X5	0.2806	0.288	0.2833

Table 3.18: Standard deviations of the quantity sold for the average UPC by bin for the samples of regular prices accounting for UPC-specific store effect.

	2004	2005	2004-2005
	X1	X1	X1
Full sample	(1)	(2)	(3)
X2	0.2359***	0.3544***	0.3040***
	(0.0041)	(0.0035)	(0.0045)
P1	-2.9514***	-2.6542***	-2.6505***
	(0.0185)	(0.0123)	(0.0182)
P2	0.6822***	0.8430***	0.7886***
	(0.0168)	(0.0129)	(0.0170)
R^2	0.5216	0.6025	0.518
n	34,580	56,368	33,696
	X1	X1	X1
Regular prices	(4)	(5)	(6)
X2	0.4753***	0.5612***	0.6208***
	(0.0133)	(0.0079)	(0.0191)
P1	-2.4648***	-1.9469***	-2.3481***
	(0.1309)	(0.0448)	(0.1644)
P2	0.8285***	0.9728***	1.2196***
	(0.1013)	(0.0447)	(0.1363)
R^2	0.3141	0.4426	0.4292
n	4,160	11,180	1,872

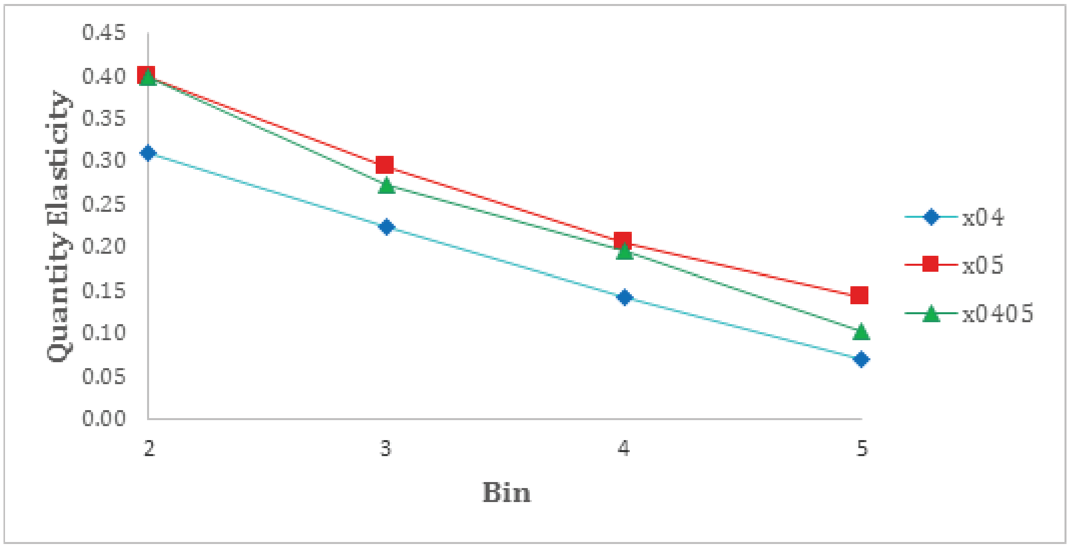
Table 3.19: Two bins regression accounting for UPC-specific store effect.
Standard errors in parentheses. Category dummies are included in all the regressions.

	2004		2005		2004-2005	
	X1	X1	X1	X1	X1	X1
Full sample	(1)	(2)	(3)	(4)	(5)	(6)
X2	0.3317***		0.4177***		0.3956***	
	(0.0050)		(0.0038)		(0.0051)	
P2	0.9280***		1.0959***		1.0987***	
	(0.0254)		(0.0168)		(0.0238)	
X3		0.1254***		0.2259***		0.1728***
		(0.0040)		(0.0036)		(0.0045)
P3		0.3594***		0.5044***		0.4481***
		(0.0160)		(0.0128)		(0.0166)
P1	-2.8147***	-2.8763***	-2.6612***	-2.6598***	-2.5979***	-2.6075***
	(0.0227)	(0.0191)	(0.0163)	(0.0128)	(0.0220)	(0.0188)
R^2	0.4928	0.4441	0.5662	0.5057	0.4919	0.4275
n	34,580	34,580	56,368	56,368	33,696	33,696
	X1	X1	X1	X1	X1	X1
Regular prices	(7)	(8)	(9)	(10)	(11)	(12)
X2	0.4535***		0.5425***		0.6062***	
	(0.0142)		(0.0083)		(0.0202)	
P2	-0.0394		0.8814***		0.8299***	
	(0.1869)		(0.0666)		(0.2093)	
X3		0.3533***	0.43556***		0.5160***	
		(0.0140)		(0.0086)		(0.0206)
P3		0.5490***		0.6682***		0.8143***
		(0.0928)		(0.0437)		(0.1242)
P1	-1.5125***	-2.2996***	-1.7748***	-1.8736***	-1.7624***	-2.1259***
	(0.1717)	(0.1304)	(0.0582)	(0.0447)	(0.2094)	(0.1583)
R^2	0.2726	0.2079	0.3963	0.3206	0.3897	0.3219
n	4,160	4,160	11,180	11,180	1,872	1,872

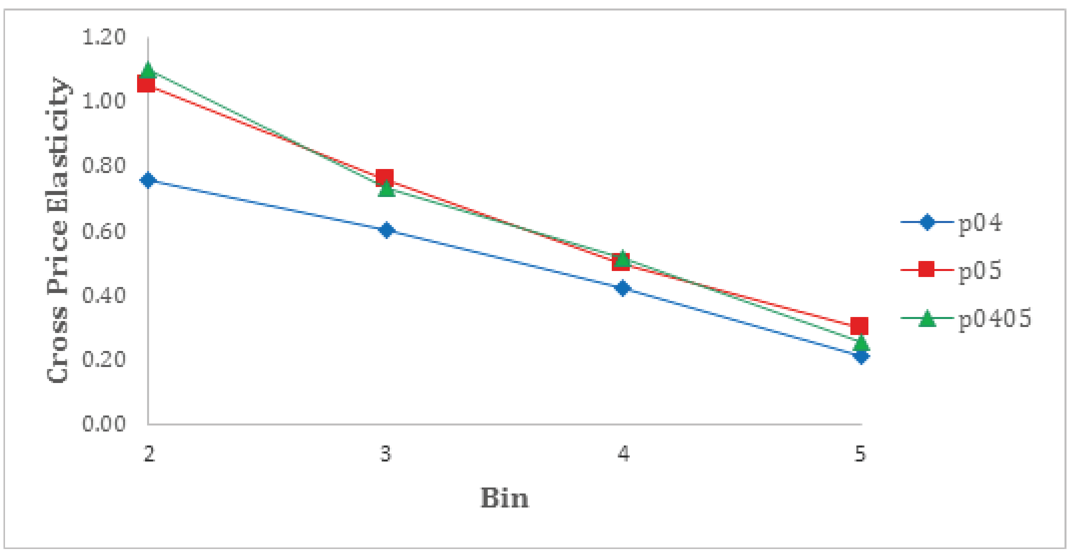
Table 3.20: Three bins regressions accounting for UPC-specific store effect.

Full sample	X1 (1)	X1 (2)	X1 (3)	X1 (4)	X1 (5)
X2	0.3982***				0.3235***
X3		0.2946***			0.1425***
X4			0.2053***		0.0517***
X5				0.1423***	0.0212***
P1	-2.6320***	-2.6298***	-2.6328***	-2.6397***	-2.6640***
P2	1.0487***				0.8685***
P3		0.7621***			0.4185***
P4			0.4971***	0.1064***	
P5				0.2977***	0.0425**
R^2	0.5009	0.4565	0.4312	0.4189	0.5182
n	56,368	56,368	56,368	56,368	56,368
Regular prices	**X1 (6)**	**X1 (7)**	**X1 (8)**	**X1 (9)**	**X1 (10)**
X2	0.4442***				0.2991***
X3		0.3948***			0.1752***
X4			0.3544***		0.1308***
X5				0.2949***	0.0733***
P1	-1.3769***	-1.6784***	-1.7796***	-1.8096***	-1.4170***
P2	0.2386***				-0.2194
P3		0.5126***			0.6979***
P4			0.5233***		0.3604***
P5				0.3685***	0.0328
R^2	0.3202	0.2699	0.2471	0.2184	0.3784
n	11,908	11,908	11,908	11,908	11,908

Table 3.21: Five bins regressions accounting for UPC-specific store effect using the 2005 sample. The first four columns report the results when using 3 explanatory variables. The last column is the results when using 9 explanatory variables. The first rows use the sample of all prices. The rows that follow use the sample of regular prices. Category dummies are included in all the regressions.

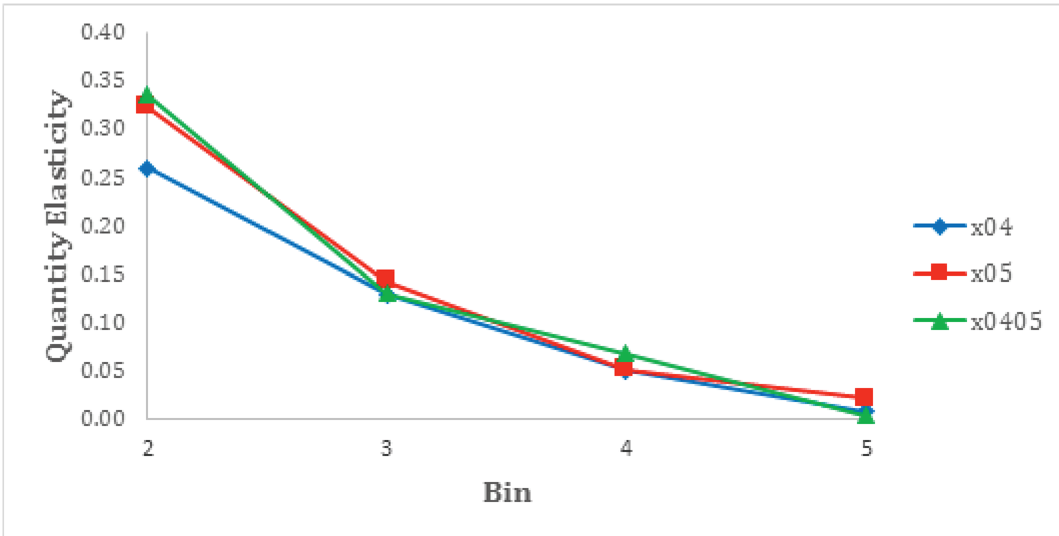


(a) Quantity elasticities

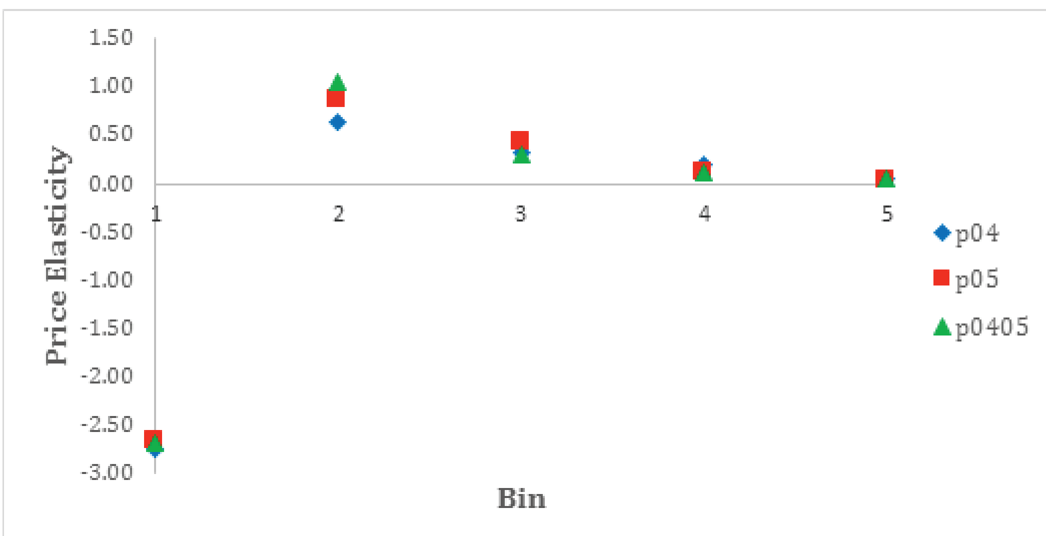


(b) Price elasticities

Figure 3.7: Elasticities based on a 3 explanatory variables regressions (using the samples of "all prices").



(a) Quantity elasticities



(b) Price elasticities

Figure 3.8: Elasticities based on a 9 explanatory variables regression (using the samples of "all prices").

For example, Kehoe and Midrigan [54] use a Calvo type model and assume that sometimes the store is allowed to make a regular price change and sometimes it is allowed to make a temporary price change that lasts for one period only. In this framework, the effect of money will depend primarily on the probability of making a regular price change and not on the probability of making a temporary price change.[17] An extreme view of this approach is that ”temporary sale prices” are not important for macro.

Looking at the data through the lens of the Bental-Eden model,“temporary sale prices” are important if they move the cross sectional price distribution. In Bental and Eden [9] demand shocks are but nevertheless they have a lasting effect on output. A high realization of demand is associated with high consumption and possibly high output in the current period. In the following period, inventories are low and the prices in all the hypothetical markets are relatively high. Production (that is determined by equating the marginal cost to the first market price) is relatively high but the increase in production is not enough to bring inventories back to “normal” in the current period. It takes time until inventories are back to normal and during this time production is relatively high. Thus a positive demand shock leads to an effect on output that may last for some time. For money to have a lasting effect on output, it is essential that prices react to changes in inventories. See Bental and Eden [10]. Since in the model, the prices in all markets react to changes in inventories, the cross sectional average price must fluctuate over time.

In the Bental-Eden model the behavior of the price at the individual store level is not important and so is the distinction between regular and sale prices. What is important is the behavior of the cross sectional price distribution which we will proxy by the cross sectional average price.[18] Here we examine the contribution of temporary sales to the variations in the cross sectional average price over weeks.

Table 3.7 calculates the ratio of the variations over weeks in the samples of regular prices to

[17] See also Nakamura and Steinsson [64] and Eichenbaum et al. [37].

[18] As was noted by Eden [32] and Head et al. [45], looking at the behavior of the average price is different from looking at the behavior of the price in an “average store”. It is possible for example that in each period x% of the stores put the item “on sale” and discount its price by y%. In this case, if the regular price does not change over time, the cross sectional average posted price will not change but the price in the “average store” will fluctuate.

variations over weeks in the samples of all prices. The columns 3.5/3.4 use Tables 3.4 and 3.5 to compute the ratio of the standard deviation in the sample of regular prices to the standard deviation in the sample of all prices. Similarly, the columns 3.11/3.10 use Tables 3.10 and 3.11 that control for store effects and the columns 3.18/3.17 use Tables 3.17 and 3.18 that control for UPC specific store effects. Average prices and quantities vary less overs weeks in the samples of regular prices. The effect of temporary sales on the variations in prices is larger than the effect on the variations in quantities. When looking at the entire sample (one bin) the standard deviation of the average price in the sample of regular prices is on average 37% of the standard deviation of the average price in the sample of all prices. The standard deviation of quantities in the sample of regular prices is 65% of the standard deviation in the sample of all prices. The effect of temporary sales is relatively large on the cheapest price bin. The standard deviation of the average regular price in the high price bin is on average 40% of the standard deviation in the sample of all prices. The percentage for the low price bin is 36. What may be somewhat of a surprise is that temporary sales affect the standard deviation in the high price bin. We find that this is also the case when using the 3 and 5 bins divisions (not reported in Table 3.7). One possible explanation relies on the observation that temporary sales are synchronized across stores (see the last column in Table 3.15). Sometimes an item is on sale in most stores and the average price in all bins goes down. This is consistent with Bental and Eden [9, 10] that assume increasing marginal cost. As was said above, in their model an increase in inventories leads to a decrease in all prices until inventories are back to "normal". It is therefore possible that all stores will have a "sale" at the same time.

Table 3.7 shows that removing temporary sales reduces our measure of price flexibility by a substantial amount. We may therefore conclude that from the point of view of the Bental-Eden model, temporary sale prices play an important role in moving the cross sectional price distribution over time and are therefore important for the propagation of monetary shocks.

	Original Price	(3.5/3.4)		Store Effect	(3.11/3.10)		UPC Specific	(3.18/3.17)		
One bin	2004	2005	04-05	2004	2005	04-05	2004	2005	04-05	Average
P	0.29	0.43	0.38	0.31	0.43	0.38	0.28	0.42	0.37	0.37
X	0.62	0.65	0.69	0.61	0.65	0.69	0.59	0.64	0.68	0.65
Two bins	2004	2005	04-05	2004	2005	04-05	2004	2005	04-05	Average
P1	0.35	0.45	0.40	0.37	0.46	0.41	0.32	0.42	0.45	0.40
P2	0.34	0.49	0.38	0.28	0.41	0.35	0.25	0.38	0.32	0.36
X1	0.70	0.76	0.71	0.69	0.73	0.74	0.72	0.70	0.79	0.73
X2	0.63	0.72	0.78	0.54	0.60	0.58	0.51	0.60	0.58	0.61

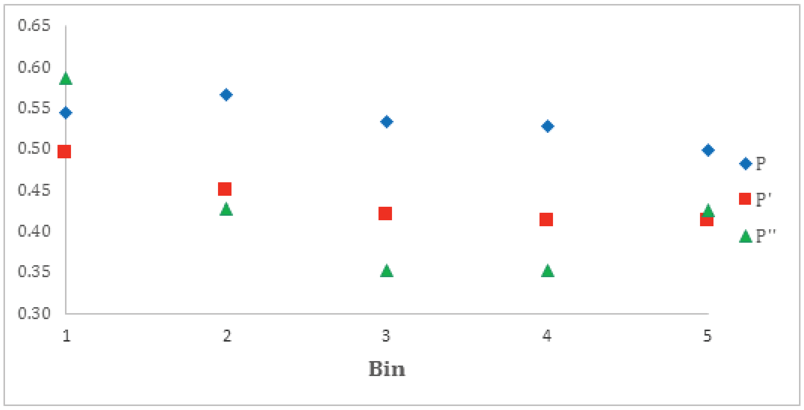
Table 3.7: The ratio of the standard deviation in the sample of regular prices to the standard deviation in the sample of all prices.
This Table computes the ratio of the standard deviation of the average price across week in the samples of regular prices to the standard deviation in the samples of all prices. The columns 3.5/3.4 use Tables 3.4 and 3.5 to compute the ratios. The columns 3.11/3.10 use Tables 3.10 and 3.11 and the columns 3.18/3.17 use Tables 3.17 and 3.18. The average in the last column is across the cells in the row.

3.8.1 The relative importance of "temporary sales" for the low price bin

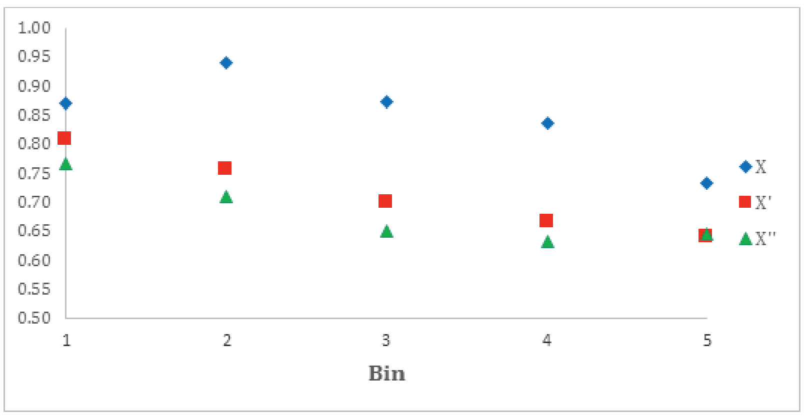
The version of the UST model in section 3.3.2 suggests that temporary sales are relatively more important for lower price bins. (In the example of Figure 3.2b removing temporary sale will affect the variances in the low price bin but will not affect the variances in the high price bin). To examine this prediction we look at the 5 bins division in the 2005 sample (not reported in Table 3.7). Figure 3.7a is the ratio of the standard deviation of the average regular price to the standard deviation of the average price (regular and sale). In the highest price bin (bin 1) the standard deviation of the average regular price is about 55% of the standard deviation of the average price. In the lowest price bin it is about 45%. Figure 3.7b is about quantities. In the highest price bin, the standard deviation of the average quantity in the sample of regular prices is about 85% of the standard deviation in the sample of all prices. In the lowest price bin the number is about 65%. Figure 3.7c use the average over samples and methods as the last column in Table 3.7. For price the average ratio is 0.47 for bin 1 and 0.37 for bin 5. It thus appears that removing temporary sale prices has a larger effect on the variation of the average price in the low price bin. This is consistent with the example in Figure 3.2b.

3.8.2 Variation over weeks by bins

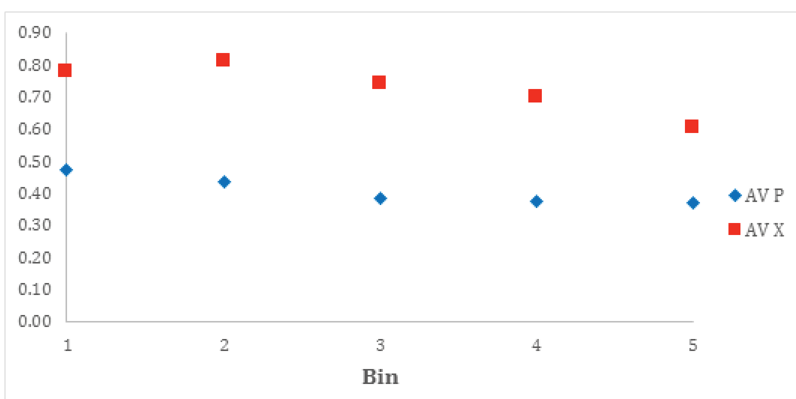
The example in Figure 3.2b also suggests that variations over weeks in the lowest price bin are larger than in the highest price bin. To examine this hypothesis, Table 3.8 computes the ratio of the average standard deviation in the lowest price bin to the average standard deviation in the highest price bin (averages across samples). The first column (P) reports the ratio of the standard deviations of prices when using the sample of all prices (Table 3.4). When using the 2 bins division, the standard deviation in the low price bin is 42% larger than the standard deviation in the high price bin. This difference is 52% when controlling for a store effect (Table 3.10) and 53% when controlling for a UPC specific store effect (Table 3.17). When using the 3 and 5 bins divisions the differences are larger. The percentage differences in the standard deviations are lower when using



(a) Ratios for SD prices in the 2005 sample (P uses the original variables, P' uses the residuals when controlling for store effects and P" uses the residuals when controlling for UPC specific store effects)



(b) Ratios for SD quantities in the 2005 sample (X uses the original variables, X' uses the residuals when controlling for store effects and X" uses the residuals when controlling for UPC specific store effects)



(c) Average over samples and methods

Figure 3.7: The ratio of the standard deviation (over weeks) of the cross-sectional average price in the sample of regular price to the standard deviation in the sample of all prices using the 5 bins division.

No Store Eff. Tables 3.3	P	RP	X	RX
2 bins	1.42	1.42	1.40	1.36
3 bins	1.65	1.57	1.59	1.35
5 bins	1.80	1.62	1.61	1.36
Store Eff. Tables 3.10, 3.11	P	RP	X	RX
2 bins	1.52	1.26	1.30	1.03
3 bins	1.74	1.37	1.36	1.04
5 bins	1.91	1.44	1.38	1.05
UPC specific Tables 3.17, 3.18	P	RP	X	RX
2 bins	1.53	1.22	1.34	1.02
3 bins	1.81	1.30	1.42	1.03
5 bins	2.01	1.36	1.47	1.05

Table 3.8: Ratio of the average standard deviation of the lowest price bin to the average standard deviation of the highest price bin.

The Table reports the ratio of the average standard deviation in the lowest price bin to the average standard deviation in the highest price bin. Averages are over samples. The first column (P) is the ratio of the standard deviations of prices in the samples of all prices. The second column (RP) is this ratio in the samples of regular prices. The third column (X) is the ratio of the standard deviations of quantities and the last column (RX) is this ratio in the sample of regular prices.

the sample of regular prices (RP in the second column).

The third column in Table 3.8 is the ratio of the standard deviation of the quantity sold by stores in the lowest price bin to the quantity sold by stores in the highest price bin. When using the 2 bins division, the standard deviation in the low price bin is 40% larger than the standard deviation in the high price bin. This difference is about 30% when controlling for a store effect and for a UPC specific store effect (the last rows of the Table). Also here the ratios are lower when using the sample of regular prices but the ratios are still greater than 1.

We may therefore conclude that variations over weeks are relatively large in the low price bin. Removing temporary sales observations tends to reduce the difference in variations especially when controlling for UPC specific store effects.

2 bins division	Pr 1	Pr 2
bin 2	0.65	0.68

3 bins division	Pr 1	Pr 2
bin 2	0.58	0.59
bin 3	0.77	0.81

5 bins division	Pr 1	Pr 2
bin 2	0.60	0.60
bin 3	0.71	0.71
bin 4	0.79	0.81
bin 5	0.86	0.89

Table 3.9: The probability of attracting shoppers by bin.
This Table uses the 2005 sample of all prices in Tables 3.19-3.21. The second column (Pr 1) is one minus the quantity elasticity. The third column (Pr 2) is one minus the ratio of the cross price elasticity to the absolute value of the own price elasticity.

3.9 Probabilities

We have examined two predictions of (3.25): the quantity elasticity is between zero and unity and the cross price elasticity is less than the absolute value of the own price elasticity. Equation (3.25) has an additional prediction: The ratio of the cross price elasticity to the absolute value of the own price elasticity (C/O) is the same as the quantity elasticity (QE). This prediction can be examined by comparing two alternative calculations of the probability of attracting shoppers by the low price stores. The two alternative methods do not yield similar numbers when using the original variables or the store effect residuals. But they yield similar numbers when using the UPC specific residuals. Here we present the calculations for the UPC specific residuals.

Table 3.9 uses the estimated elasticities for the samples of all prices in Tables 3.19-3.21 and 2 methods for computing the probability. The first (Pr 1) is 1-QE where QE is the quantity elasticity. The second (Pr 2) is 1 - C/O, where C/O is the ratio of the cross price elasticity to the absolute value of the own price elasticity. There is a substantial agreement between the two methods and the probability of attracting shoppers by stores in lower price bins is higher.

3.10 Concluding Remarks

We provide results about elasticities within UPC-week cells, variations over weeks within UPC and the role of temporary sales.

The results about elasticities are obtained by dividing the stores in each UPC-week cell into bins and running (3.1). Our main findings are: (a) The elasticity of the quantity sold by stores in the high price bin with respect to the quantity sold by stores in a low price bin (the quantity elasticity) is between zero and unity ($0 < b_2 < 1$); (b) This quantity elasticity is higher when the explanatory variables are from bins closer in price to the highest price bin; (c) The elasticity of the quantity sold by stores in the high price bin with respect to the price in a low price bin (the cross price elasticity) is positive but less than the absolute value of the own price elasticity ($0 < b_3 < -b_1$); (d) The cross price elasticity is higher when the explanatory variables are from bins closer in price to the highest price bin.

Observations (b) and (d) say that for the purpose of predicting the quantity sold by stores in the highest price bin, the quantity and price in medium price stores is more relevant than the quantity and price in low price stores.

We computed the average quantity sold and the cross sectional average price for each UPC-week-bin cell and found that: (e) Variations over weeks in the average price and quantity are lower for higher price bins. We also make the following observations about temporary sales: (f) The fraction of stores that offer an item on sale fluctuates over weeks in a way that is not consistent with the mixed strategy hypothesis; (g) Temporary sales contribute substantially to variations over weeks in the average posted price and the quantity sold; (h) The contribution of temporary sales to variations over weeks is large for all bins and somewhat larger for lower price bins.

Using our largest 2005 sample and controlling for UPC specific store effects the following estimates support the above conclusions:

[a] The quantity elasticity is 0.35 in the sample of all prices and 0.56 in the sample of regular prices (Table 3.19, 2 bins).

[b] The elasticity of the quantity sold by stores in bin 1 with respect to the quantity sold by stores in bin 2 is 0.4 while the elasticity with respect to the quantity sold by stores in bin 3 is 0.2 (Table 3.20, 3 bins).

[c] The cross price elasticity is 0.84 and the absolute value of the own price elasticity is 2.65 (Table 3.19, 2 bins).

[d] The elasticity with respect to the price posted by stores in bin 2 is 1.0 while the elasticity with respect to the price posted by stores in bin 3 is 0.5 (Table 3.20, 3 bins).

[e] The standard deviation of the average price over weeks in the low price bin is 23% higher than the standard deviation of the average price in the high price bin. The number for quantity is 20% (Table 3.17, 2 bins).

The following numbers (again, using the UPC specific 2005 sample) support the conclusions about temporary sales:

[f] For the average UPC, none of the stores have temporary sales in 43% of the weeks. The mixed strategy hypothesis implies that this number should be around 1%.

[g] The standard deviation of the (cross sectional) average price over weeks in the sample of regular prices is only 41.9% of the standard deviation in the sample of all prices (Tables 3.17 and 3.18, 1 bin). The standard deviation of the quantity sold over weeks in the sample of regular prices is 64% of the standard deviation in the sample of all prices (Tables 3.17 and 3.18, 1 bin).

[h] For bin 1, the standard deviation of the average price over weeks in the sample of regular prices is 42.2% of the standard deviation in the sample of all prices. The number for bin 2 is 38.3% (Tables 3.17 and 3.18, 2 bins). For bin 1, the standard deviation of the quantity sold over weeks in the sample of regular prices is 70% of the standard deviation in the sample of all prices. This number is 60% for bin 2 (Tables 3.17 and 3.18, 2 bins).

The main findings are consistent with a UST model that allows for storage and non-shoppers but are not consistent with the simple version of the UST model in section 3.3.1 and the simple version of the monopolistic competition model in section 3.2.

The intuition for the observations about elasticities ([a]-[d]) is as follows. The amount sold by stores depends on a random shock that is common to all buyers and on the number of shoppers (buyers who shop across stores). From the point of view of predicting the quantity sold by stores in the highest price bin the common shock is relevant and the number of shoppers is a "noise". Therefore shoppers' activity reduces the quantity elasticity. Since in the absence of shoppers the quantity elasticity is unity, we find a quantity elasticity that is less than one. The quantity sold by stores in the medium price bin is less influenced by shoppers' activity and therefore it provides a better signal for the common shock. As a result the elasticity with respect to the quantity sold by stores in the medium price bin is higher than the elasticity with respect to the quantity sold by stores in the low price bin.

We have assumed that the demand of shoppers is less stable than the demand of non-shoppers. This may be the result of storage activity by shoppers as in Hendel and Nevo [46]. Shoppers who find the item at a low price buy a large quantity and store most of it. They then stay out of the market until the level of inventories at home is low. We think that explicitly modeling this behavior will lead to the result that the demand of shoppers is relatively unstable and this will lead to result [e] about variations over week because shoppers are more important for low price stores.

In section 3.3.2 temporary sales occur when demand in the previous period was low and stores that post the high regular price accumulate inventories. The prevalence of weeks with no temporary sales ([f]) is consistent with this model.

Observation [g] is relevant to the question of whether temporary sales are important for price flexibility. The measure of price flexibility (or price rigidity) depends on the underlying model. In UST models with storage the behavior of the entire cross sectional distribution is important but the behavior of prices at the store level is not. We have focused on the average cross sectional price and found that temporary sales have a large effect on its variability over weeks. Thus, from the point of view of UST models, temporary sales are important for macro. Observation [g] is relevant to the question of whether temporary sales are important for price flexibility. The measure of price flexibility (or price rigidity) depends on the underlying model. In UST models with storage the behavior of the entire cross sectional distribution is important but the behavior of prices at the store level is not. We have focused on the average cross sectional price and found that temporary sales have a large effect on its variability over weeks. Thus, from the point of view of UST models, temporary sales are important for macro.

Observation [h] complements observation [f] in suggesting that temporary sales are correlated across stores. The UST model in Bental and Eden [9] may account for this observation. In this model an increase in inventories depresses all prices including the prices in the top of the distribution. Over time the level of inventories go back to "normal" and prices go back to their "regular level".

APPENDIX

A.1 Robustness Checks Using Unemployment Measure of Inattention

The positive correlation between 99-cent endings and unemployment withstands various alternative estimation procedures. Omitting sale prices, omitting the lowest price for each good, and excluding the anomalous 2001 data set all fail to reverse the positive and statistically significant relationship between unemployment and 99-cent endings.

A.1.1 Normal vs. Sale Prices

One concern regarding chapter 1's finding is that the positive relationship could be the result of temporary sale prices ending in 99 cents more often than non-sale, or normal, prices. To address this issue, I drop all observations in the original dataset that have the price reduction flag as calculated by IRI.[19] The aggregation to the Month-Market level is then performed on the remaining observations and the regressions in equations (1.6) and (1.7) are then run using the sample of only normal prices with the results reported in Table A.10, Columns (1) and (2).

One interesting observation is that the difference in coefficients between prices posted and units sold is now negligible when using the sample of only normal prices, with the difference between the coefficients falling to 0.0002 from 0.0015. Perhaps this suggests that the stronger positive relationship for units sold is a result of temporary sales and shoppers who overwhelmingly buy goods that coincidentally end in 99 cents. While this may be the case, the positive relationship between 99-cent endings and unemployment is still statistically significant and the proportion of prices ending in 99 cents is marginally higher when omitting sale prices. However, this observation does open the door to analyzing how the relationship may change when ignoring the most attentive shoppers who buy at the lowest price.

[19] While the PR flag is a somewhat rough measure of sales since it only requires a 5% temporary price reduction, it has some merit over other measures of temporary sales by not requiring uninterrupted weekly observations for each UPC-store combination.

Robustness Check	(1) No Sales $P_{i,t}$	(2) No Sales $Q_{i,t}$	(3) No Lowest $P_{i,t}$	(4) No Lowest $Q_{i,t}$	(5) No 2001 $P_{i,t}$	(6) No 2001 $Q_{i,t}$
Unemployment	0.0067***	0.0069***	0.0077***	0.0088***	0.0080***	0.0096***
	(0.0002)	(0.0003)	(0.0002)	(0.0003)	(0.0002)	(0.0003)
Time Trend	-0.00006***	-0.00041***	-0.00043***	-0.00028***	-0.00006***	-0.00037***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Market Dummies	Y	Y	Y	Y	Y	Y
Month Dummies	Y	Y	Y	Y	Y	Y
R^2	0.7353	0.6216	0.6907	0.5353	0.7388	0.6156
N	6,204	6,204	6,204	6,204	5,604	5,604
Mean 99-Cent	0.2388	0.1854	0.2112	0.1611	0.2323	0.1719
Calculated ε	0.17	0.23	0.22	0.34	0.29	0.25

Table A.10: Regression results and calculated elasticities when using only normal prices, omitting the lowest priced observations, or omitting 2001 data.
$*-$ 10 % significance level, $**-$ 5% significance level,$***-<$ 1% significance level.

A.1.2 Lowest-Priced Goods

In light of the observation that omitting temporary price reductions weakens the positive relationship between 99-cent endings and unemployment, another possible explanation for the main regressions results could be that the lowest-priced goods are more likely to end in 99 cents. If there are more shoppers when unemployment rises and such shoppers succeed at finding their desired good at the lowest price, then the positive relationship between unemployment and 99-cent endings could be the result of the lowest priced products coincidentally ending in 99. I test this hypothesis by eliminating all observations in the original scanner data that post the lowest price for each UPC-week-market combination, meaning a UPC-store-week observation is dropped if it posts the lowest price in its market for that particular UPC in that week, and then aggregating to the market-month level to perform the regressions described in equations (1.6) and (1.7).

The results in Table A.10, columns (3) and (4) suggest that the lowest priced items can only account for a minor part of the observed positive relationship. While the percentage of 99-cent prices is the lowest among all of the robustness checks, suggesting that 99-cent endings are indeed over-represented in the lowest prices, the strong positive relationship between unemployment and the proportion of prices ending in 99 cents remains. In fact, after accounting for the lower proportion of 99-cent endings, the calculated elasticity for units sold is actually higher than the elasticity calculated using the base regression.

A.1.3 Omitting 2001 Data

As can be seen in Figure 1.2, the 2001 recession and Great Recession seem to exhibit different behaviors with respect to the fraction of prices ending in 99 cents. However, the visual disparity could merely be the result of a time trend. In addition, the data from 2001 does not include a significant category, carbonated beverages, due to data anomalies so performing the analysis on only data from 2002 to 2011 is not unreasonable even if the resulting sample only covers one recession. As shown in columns (5) and (6) of Table A.10, omitting 2001 actually strengthens the positive relationship between unemployment and the fraction of prices ending in 99 cents, suggesting that the different treatment of the 2001 data is a non-issue. In fact, the coefficients and calculated elasticities reach their highest estimates across all of the specifications.

A.2 Robust, Clustered Standard Errors Regressions

Using heteroskedastic robust standard errors that are clustered at the market level does not change the main conclusion of this paper but the results are included for completeness.

A.3 Additional Tables Using Household-Level Data

Table A.14 presents summary statistics for the expenditure by category for the 2,034 households used for the household-level analysis.

Table A.15 reports the proportion of prices ending in 99 cents for each category split by market.

	$P_{i,t}$	$Q_{i,t}$	$P_{i,t}$	$Q_{i,t}$	$P_{i,t}$	$Q_{i,t}$
	(1)	(2)	(3)	(4)	(5)	(6)
Unemployment Rate	0.0079***	0.0094***				
	(0.0012)	(0.0011)				
Labor Force to Pop. %			0.0015	0.0003		
			(0.0022)	(0.0023)		
Employment to Pop. %					-0.0062***	-0.0074***
					(0.0016)	(0.0018)
Time Trend	Linear	Linear	Linear	Linear	Linear	Linear
Month Dummies	Y	Y	Y	Y	Y	Y
Fixed Effect	Market	Market	Market	Market	Market	Market
Within R^2	0.2019	0.1871	0.0514	0.0163	0.1191	0.0959
Between R^2	0.0512	0.0213	0.1430	0.1030	0.3069	0.1801
Overall R^2	0.1013	0.0950	0.0155	0.0001	0.2410	0.1235
N	6,204	6,204	6,204	6,204	6,204	6,204

Table A.11: Results for regression of proportion of 99-cent endings on labor market measures with market fixed effects and robust standard errors clustered at market level, 2001-2011.
$*-$ 10 % significance level, $**-$ 5% significance level,$***-$ $<1\%$ significance level.

	$P_{i,t}$	$Q_{i,t}$	$P_{i,t}$	$Q_{i,t}$	$P_{i,t}$	$Q_{i,t}$
	(1)	(2)	(3)	(4)	(5)	(6)
Hourly Wages	-0.0013	0.0004				
	(0.0032)	(0.0024)				
Weekly Earnings			-9.13e-06	-4.03e-05		
			(5.85e-05)	(4.49e-05)		
Weekly Hours Worked					0.0015	-0.0028**
					(0.0019)	(0.0012)
Time Trend	Linear	Linear	Linear	Linear	Linear	Linear
Month Dummies	Y	Y	Y	Y	Y	Y
Fixed Effect	Market	Market	Market	Market	Market	Market
Within R^2	0.4704	0.4725	0.4698	0.4736	0.4706	0.4751
Between R^2	0.1274	0.0295	0.1126	0.0396	0.0114	0.0176
Overall R^2	0.0787	0.1982	0.1166	0.1495	0.1226	0.1742
N	2,820	2,820	2,820	2,820	2,820	2,820

Table A.12: Results for regression of percentage of 99-cent endings on wage and earning measures with market fixed effects and robust standard errors clustered at market level, 2007-2011.
$*-$ 10 % significance level, $**-$ 5% significance level,$***-<$ 1% significance level.

	$P_{i,t}$	$Q_{i,t}$
	(1)	(2)
Poverty Rate	-0.3280	-0.3164
	(0.2798)	(0.2706)
Year Dummies	Y	Y
Fixed Effect	Market	Market
Within R^2	0.3655	0.4396
Between R^2	0.0302	0.1468
Overall R^2	0.0535	0.0846
N	517	517

Table A.13: Results for regression of percentage of 99-cent endings on poverty rate with market fixed effects and robust standard errors clustered at market level, 2001-2011.
$*-$ 10 % significance level, $**-$ 5% significance level,$***-<$ 1% significance level.

Category	Mean	S.D.	Min	Max
Diapers	0.0038	0.0148	0	0.2215
Razors	0.0008	0.0050	0	0.2007
Photographic Film	0.0009	0.0038	0	0.0704
Laundry Detergent	0.0338	0.0279	0	0.1946
Beer	0.0457	0.0891	0	0.8114
Coffee	0.0471	0.0477	0	0.3469
Blades	0.0037	0.0079	0	0.0834
Toilet Tissue	0.0527	0.0372	0	0.2625
Shampoo	0.0058	0.0081	0	0.0948
Hot Dogs	0.0177	0.0165	0	0.2150
Paper Towels	0.0253	0.0286	0	0.3380
Toothbrushes	0.0049	0.0117	0	0.3797
Sugar Substitutes	0.0031	0.0089	0	0.1565
Deodorant	0.0050	0.0079	0	0.1150
Salty Snacks	0.0902	0.0566	0	0.4177
Toothpaste	0.0091	0.0115	0	0.1216
Frozen Pizza	0.0470	0.0469	0	0.3137
Household Cleaners	0.0103	0.0121	0	0.0869
Peanut Butter	0.0154	0.0152	0	0.1603
Cold Cereal	0.0762	0.0571	0	0.4442
Margarine & Butter	0.0216	0.0262	0	0.2196
Milk	0.1240	0.0775	0	0.5419
Carbonated Beverages	0.1131	0.0956	0	0.7694
Spaghetti Sauce	0.0197	0.0183	0	0.1625
Mustard & Ketchup	0.0092	0.0065	0	0.0745
Mayonnaise	0.0159	0.0124	0	0.1228
Facial Tissues	0.0175	0.0206	0	0.1887
Frozen Dinners	0.0590	0.0685	0	0.7090
Cigarettes	0.0122	0.0612	0	0.8807
Soup	0.0558	0.0385	0	0.4122
Yogurt	0.0536	0.0543	0	0.4169

Table A.14: Share of household expenditures by category, ordered from highest to lowest proportion of 99-cent price endings (n = 2,034).

Category	Pittsfield Proportion of Prices	Eau Claire Proportion of Prices
Diapers	0.54	0.44
Razors	0.55	0.43
Photographic Film	0.58	0.54
Laundry Detergent	0.51	0.44
Beer	0.66	0.40
Coffee	0.50	0.39
Blades	0.49	0.38
Toilet Tissue	0.49	0.40
Shampoo	0.46	0.34
Hot Dogs	0.46	0.40
Paper Towels	0.43	0.34
Toothbrushes	0.46	0.35
Sugar Substitutes	0.42	0.23
Deodorant	0.44	0.33
Salty Snacks	0.40	0.35
Toothpaste	0.43	0.37
Frozen Pizza	0.44	0.40
Household Cleaners	0.41	0.33
Peanut Butter	0.42	0.34
Cold Cereal	0.39	0.35
Margarine & Butter	0.44	0.38
Milk	0.38	0.31
Carbonated Beverages	0.39	0.38
Spaghetti Sauce	0.41	0.33
Mustard & Ketchup	0.36	0.30
Mayonnaise	0.44	0.35
Facial Tissues	0.39	0.34
Frozen Dinners	0.44	0.31
Cigarettes	0.12	0.09
Soup	0.35	0.29
Yogurt	0.33	0.21

Table A.15: Fraction of posted prices ending in 99 cents by category and market.

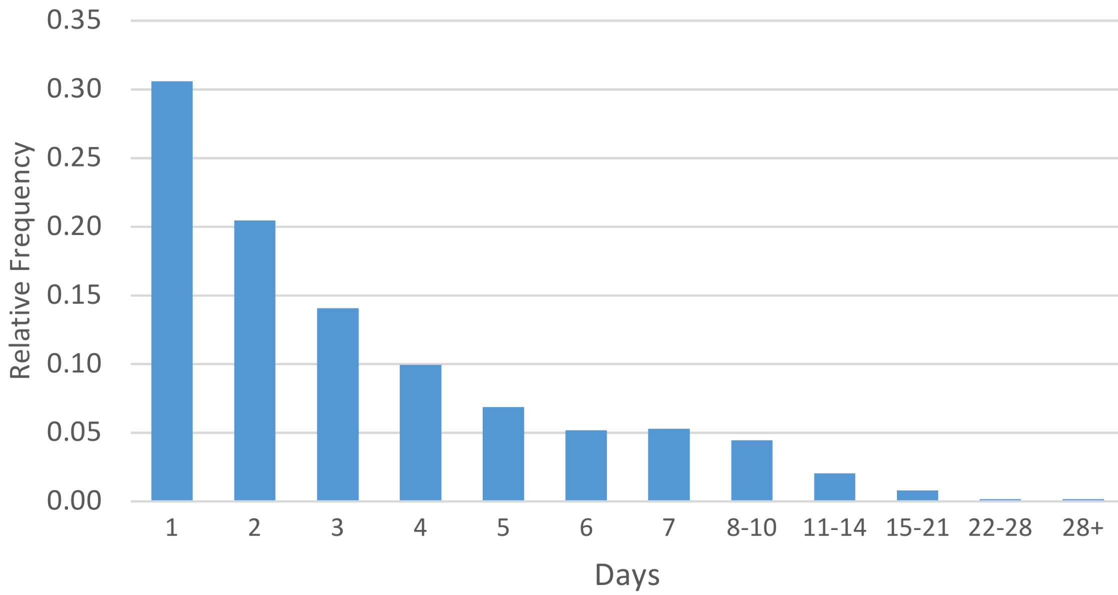


Figure A.8: Days between shopping days for household-trip observations, 2008-2011 (n = 859,509).

A.4 Days Between Shopping Days

Figure A.8 plots the distribution of the number of days between shopping days for all household-trip days in the sample of data used in chapter 2.[20] In addition, Figure A.9 shows the distribution of average days between shopping days across households. The fact that over 30% of shopping days are followed by another shopping day (Figure A.8) seems to imply that a significant portion of shopping does not occur at regularly spaced intervals. This is further supported by the average number of days between shopping days for households being distributed mostly between 4 to 6 days, which does not correspond to averages one might expect from regularly planned shopping intervals, such as twice a week, meaning 3.5 days between shopping days, or once a week, meaning an average of 7 days between shopping days (Figure A.9).

[20] I define a shopping day as a day in which the household is observed making a purchase in the data.

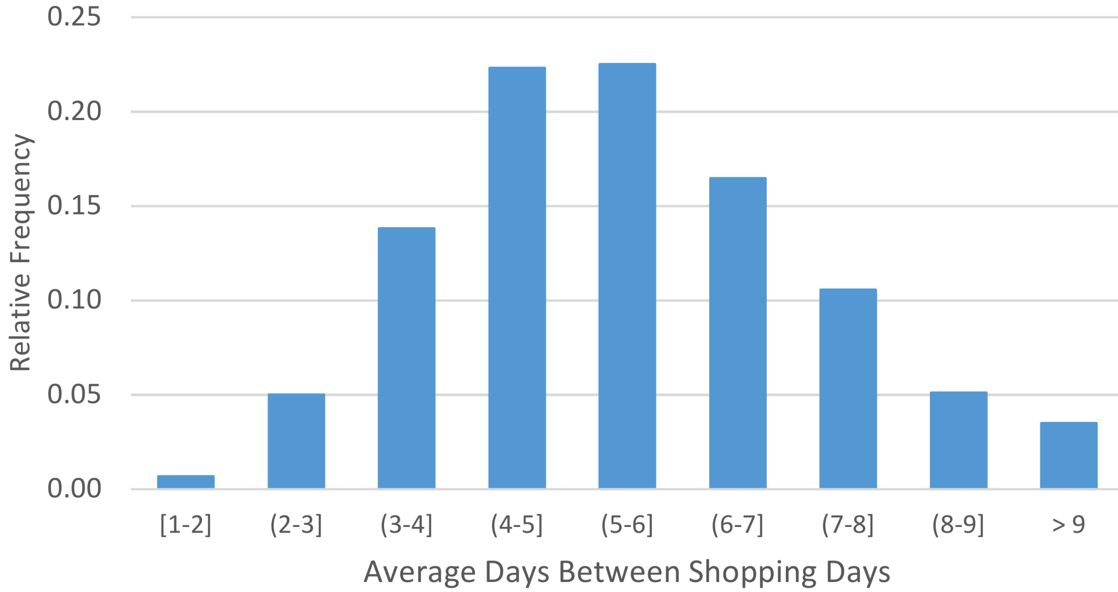


Figure A.9: Average days between shopping days by household, 2008-2011 (n = 2,034).

A.5 Additional Demographic Description

Figures A.10 to A.13 illustrate the distribution of demographic characteristics across the sample used for chapter 2. These figures show that the vast majority of households are headed by a male and female but there is also a sizable proportion of households led by single female (Figure A.10). Furthermore, the education level of the primary head of household in the sample is somewhat low, with over a third of heads of households not graduating from high school and fewer than 15% receiving any college education (Figure A.12). However, nearly 20% attended a technical school after high school, which may explain why households are as likely to have incomes between \$75,000 and \$99,999 as they are to have incomes between \$25,000 and \$34,999 (Figure A.13).

On top of the skewness in the age demographics (Figure A.14), one concern is how well the two markets, Eau Claire and Pittsfield, represent the general American populace in other charactertistics and what impact the differences may have on the external validity of the observed results of this study. Table A.16 provides a general comparison using demographic data from the U.S. Census' American Community Survey (ACS) and the 2010 decennial census.[21]

[21] Specific years used are reported in the table and are intended to be as close to 2012 as possible.

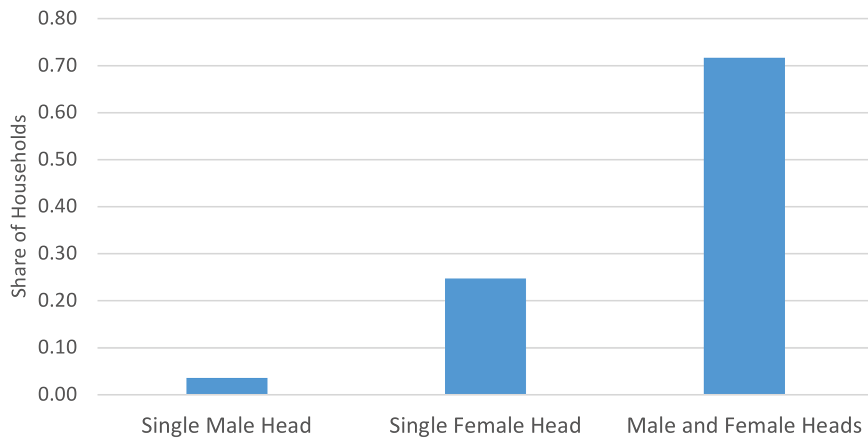


Figure A.10: Distribution of households by type

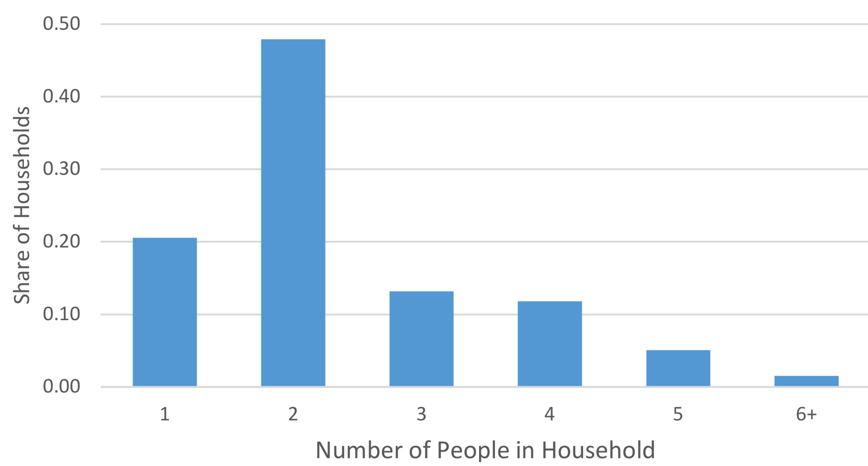


Figure A.11: Distribution of number of people in household

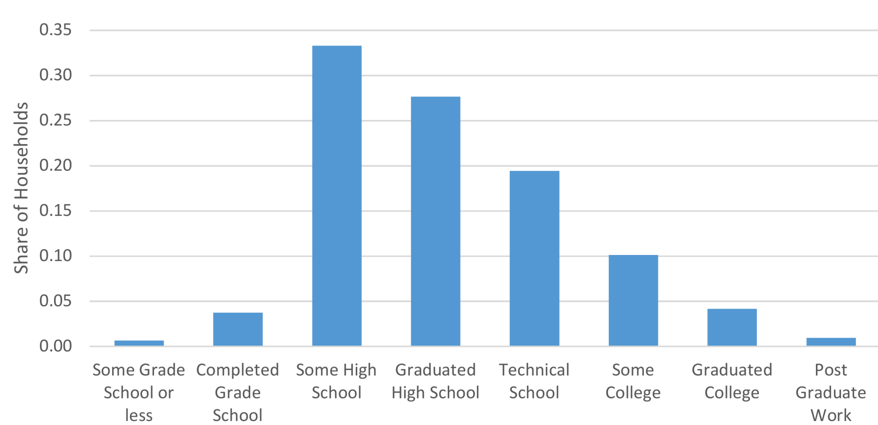


Figure A.12: Distribution of head of household eduction level

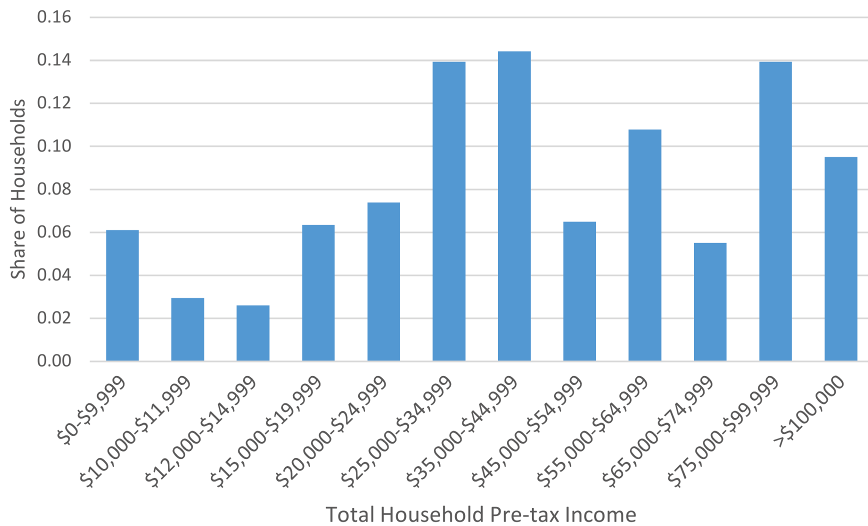


Figure A.13: Distribution of total annual pre-tax household income

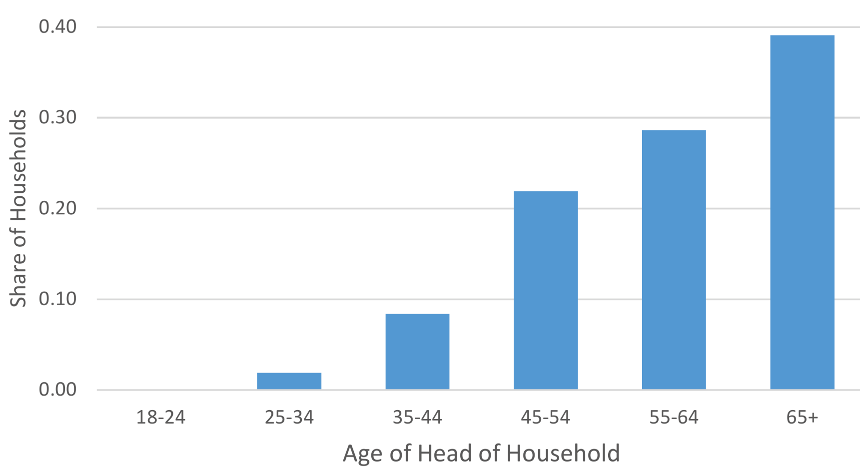


Figure A.14: Distribution of head of household age

While there are subtle differences in the age distribution of both IRI markets, with a lower proportion of young people in both markets, a more noticeable divergence from the general U.S. population is the high proportion of White alone, not Hispanic or Latino persons in both Pittsfield and Eau claire. At 90.60% and 92.10% for Pittsfield and Eau Claire, respectively, both markets have populations that have a 50% higher proportion of White persons at the cost of all other races. Also worth noting is that both markets have a higher percentage of high school graduates and people with a Bachelor's or higher degree compared to the total U.S. population. Despite the higher level of education, median income in both markets is about 8% lower than the national median income level.

Using the 2011 American Housing Survey (AHS) provided by the U.S. Census allows for a better comparison to the IRI dataset's demographics by reporting demographics measured at the household level for the U.S., similar to the demographic data from IRI. Figures A.15 to A.20 report the distribution of various demographics for the entire United States and the sample of households in the IRI dataset divided into the Pittsfield and Eau Claire markets after restricting the sample to households satisfying the minimum reporting requirements as described in section 1.4 (n = 2,034). Of particular note is that the IRI sample of households deviates not only from the overall American population, but also the demographics for the counties containing the Pittsfield and Eau Claire markets.[22]

Figure A.15 shows that the sample of households in the IRI data are more often 2-person households with fewer households of all other sizes compared to the national distribution. Furthermore, over half of households in the Eau Claire sample are 2-person households, much higher than the national percentage of 33% and the Pittsfield sample's 43%. A more dramatic divergence from the national distribution is the age of householders. Both markets exhibit significantly older householders compared to the rest of the United States (Figure 2.5 is replicated here for convenience as Figure A.16). Almost twice as many households in the IRI markets are headed by someone 65 years or older and the proportion of householders between 45 and 54 or 55 and 64 are also higher

[22] In order to accurately compare the national data from the U.S. Census and the data from the IRI panelists, a few variables are aggregated using both the IRI and Census categories.

Age & Gender (Census April 1, 2010)	USA	Pittsfield	Eau Claire
Persons under 5 years, percent	6.50%	4.70%	5.90%
Persons under 18 years, percent	24.00%	19.50%	21.10%
Persons 65 years and over, percent	13.00%	18.60%	12.60%
Female persons, percent	50.80%	51.90%	51.00%
Race (Census April 1, 2010)			
White alone, not Hispanic or Latino, percent	61.30%	90.60%	92.10%
Black or African American alone, percent	12.60%	2.70%	0.90%
Hispanic or Latino, percent	16.30%	3.50%	1.80%
Asian alone, percent	4.80%	1.20%	3.30%
Education (ACS 2011-2015)			
High school graduate or higher, percent of age 25+	86.70%	90.70%	93.20%
Bachelor's degree or higher, percent of age 25+	29.80%	31.60%	31.10%
Other Demographics (ACS 2011-2015)			
Persons per household	2.64	2.22	2.43
In civilian labor force, total, percent of age 16+	63.30%	63.10%	69.30%
Median household income (in 2015 dollars)	$53,889	$49,956	$49,513
Persons in poverty, percent	13.50%	14.30%	13.60%

Table A.16: Comparison of selected demographics for the entire U.S., Berkshire County (Pittsfield market), and Eau Claire County (Eau Claire market).

than the national levels. This increase in older households comes at the cost of householders under 25 or between 25 and 34 in particular. While the proportion of householders between the ages of 35 and 44 in the IRI sample is about half as high as the national proportion, the percentage of householders between 25 and 34 is 2.5% for Pittsfield and 1.1% for Eau Claire compared to the national level of 17% and the percentage of householders under 25 years old is less than 0.1% for each market compared to the 4.7% national percentage. Given the older distribution of age in the IRI sample, the higher proportion of child-less households in the IRI sample should not be surprising as children reach adulthood and leave their parents' household (Figure A.17).

As indicated in Table A.16, the two markets are overwhelmingly White in terms of race according to Census data. This is corroborated by Figure A.18, which shows the distribution for the race of the head of household in the IRI data. The proportion of households headed by a

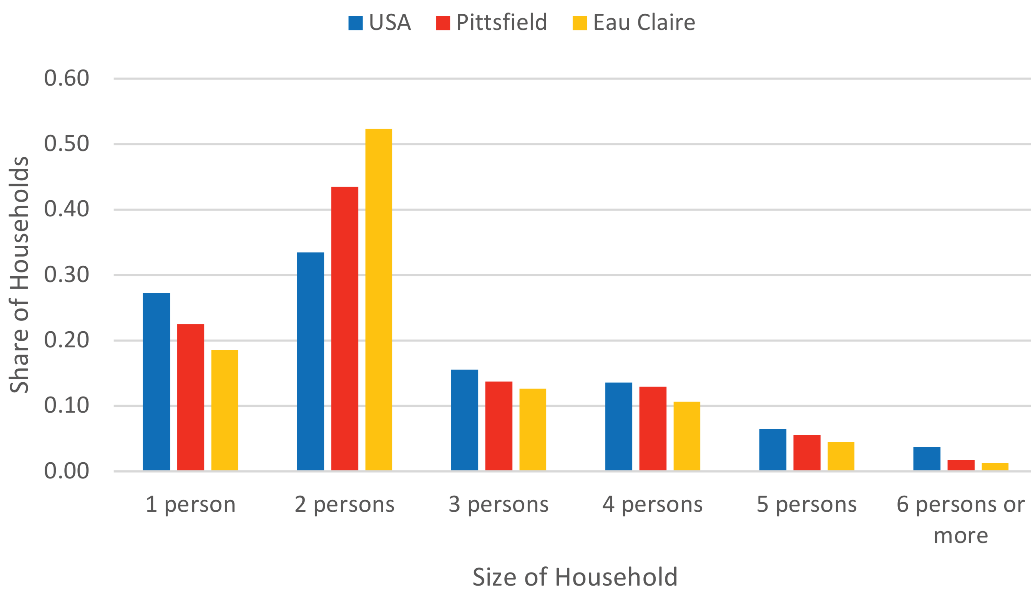


Figure A.15: Distribution of household size for the United States versus Pittsfield and Eau Claire Markets, $n_{Pittsfield}$ = 1,021, $n_{EauClaire}$ = 1,013.

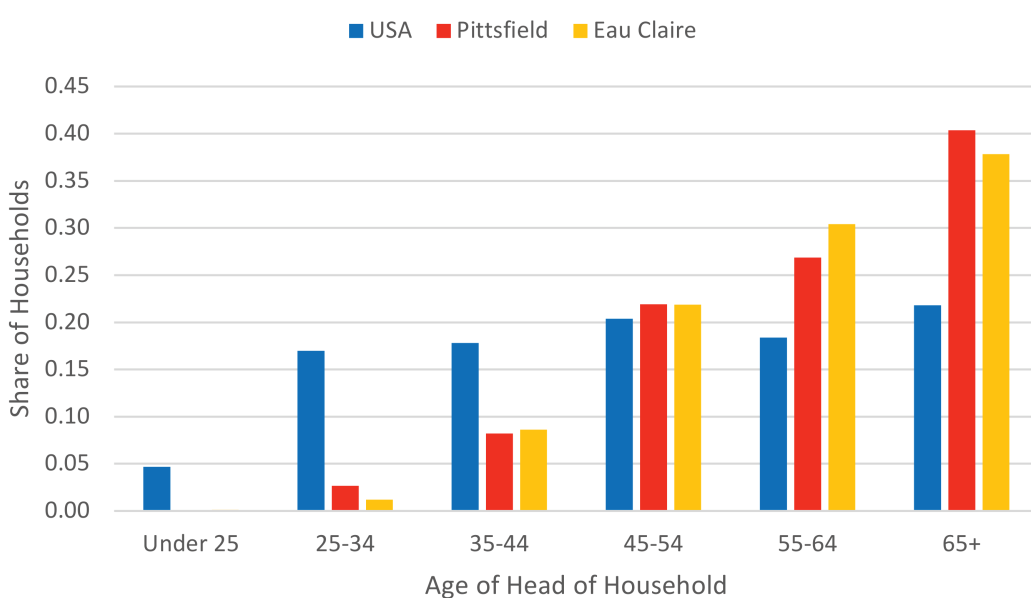


Figure A.16: Distribution of age of head of household for the United States versus Pittsfield and Eau Claire Markets, $n_{Pittsfield}$ = 1,013, $n_{EauClaire}$ = 1,010*.
*8 households in Pittsfield and 3 households in Eau Claire did not report age.

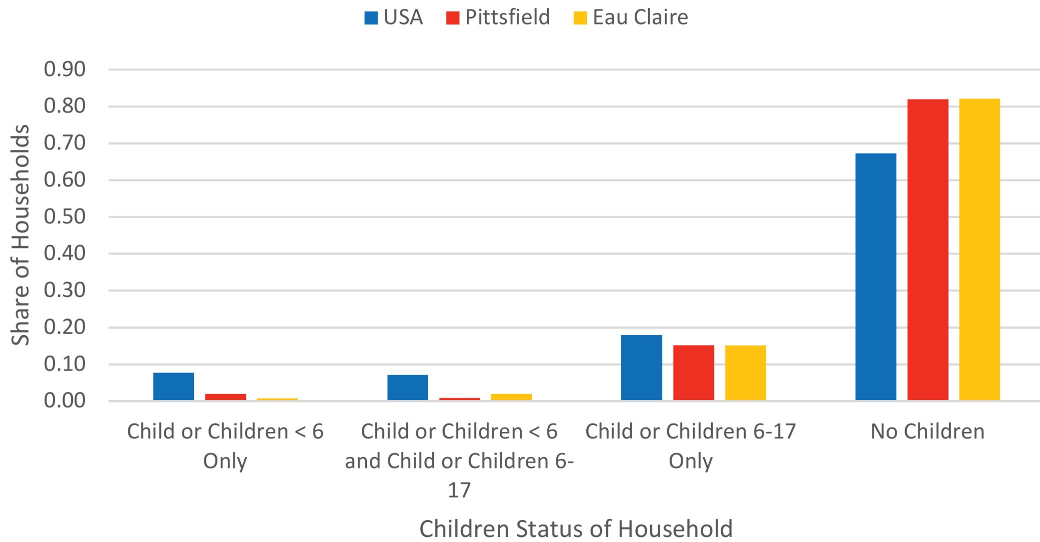


Figure A.17: Distribution of households' children status for the United States versus Pittsfield and Eau Claire Markets, $n_{Pittsfield}$ = 1,021, $n_{EauClaire}$ = 1,013.

White, non-Hispanic person is 96% for Pittsfield and 98% for Eau Claire, which is even higher than the proportion reported by the Census for the respective counties. The number of non-White households is effectively nonexistent, especially for the Hispanic race.[23]

In terms of education, Table A.16 suggested that the counties containing the Pittsfield and Eau Claire markets were more educated than the general United States in terms of high school graduation and obtaining a Bachelor's degree or higher. However, the sample of households in the IRI dataset is drastically different from the U.S. Census data. While a larger proportion of households in the Pittsfield and Eau Claire market have a high school diploma as their highest level of education compared to the national statistic, the percentage of householders attaining an educational level of some college or higher is significantly lower and not reflective of the statistics reported by the Census. Table A.17 presents an alternative statistic measuring education levels in terms of the percent with a high school diploma or bachelor's degree or higher. For both measures, the two IRI markets have a lower proportion than the national value, indicating that the increased percentage of households having a high school diploma as the householder's highest level of education comes at

[23]There are only 7 Hispanic householders in the entire sample across both markets.

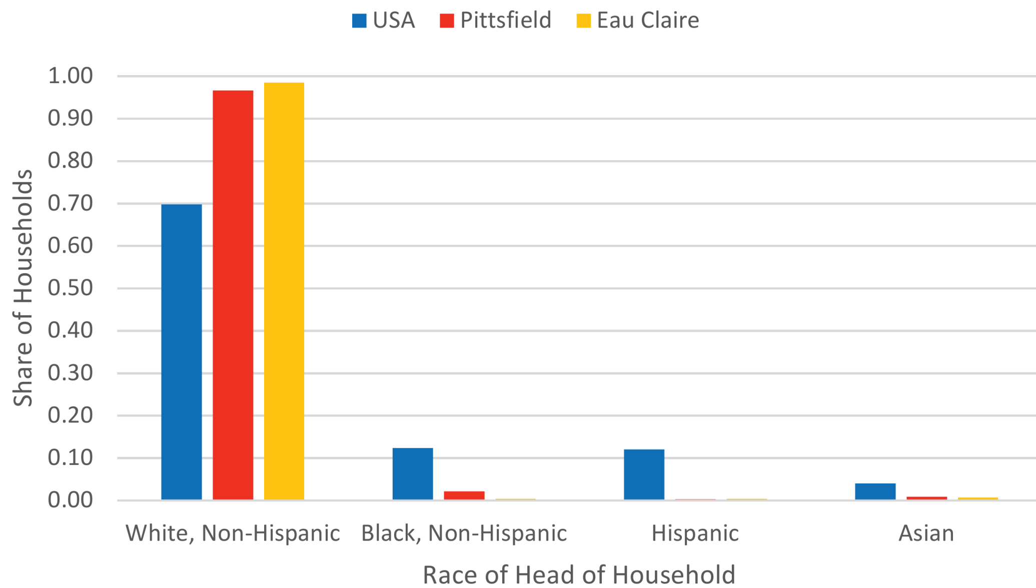


Figure A.18: Distribution of race of head of household for the United States versus Pittsfield and Eau Claire Markets, $n_{Pittsfield}$ = 1,014, $n_{EauClaire}$ = 1,010*.
*7 households in Pittsfield and 3 households in Eau Claire did not report race.

	USA	Pittsfield	Eau Claire
Percent of householders, high school graduate or higher	0.87	0.62	0.64
Percent of householders, bachelor's degree or higher	0.31	0.06	0.06

Table A.17: Additional household educational attainment demographics (2011 AHS and IRI Data).

the cost of households attaining higher education. Instead, a large proportion of households in both IRI markets fail to complete high school or go to a technical school, with the proportion being as much as 5 times higher than the national proportion. Possibly related to the difference in education, the distribution of household income for the IRI sample is more clustered around middle incomes. Around half as many households earn over $100,000 in the Pittsfield and Eau Claire samples, but there is also a lower percentage of households with less than $10,000 in annual income compared to the entire United States (Figure A.20).

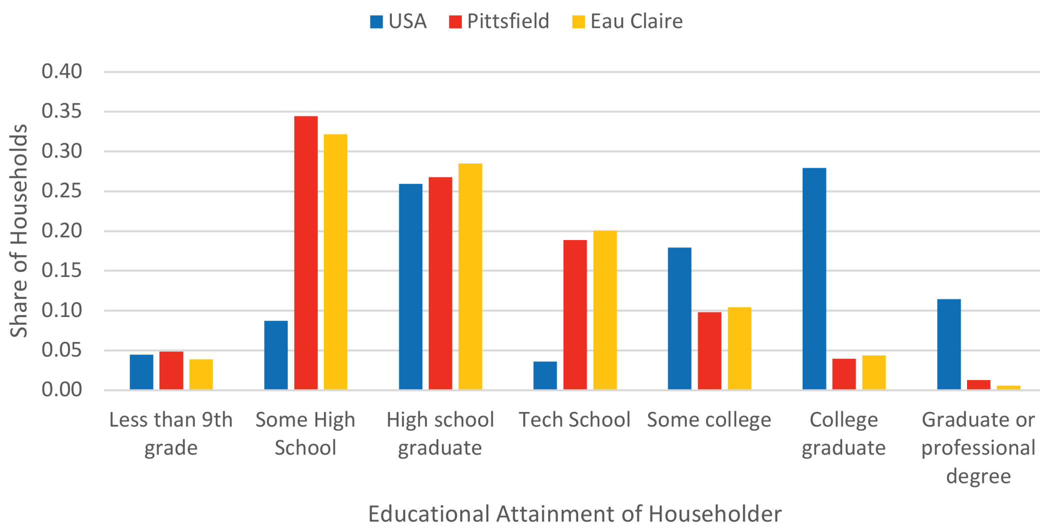


Figure A.19: Distribution of head of household's educational attainment for the United States versus Pittsfield and Eau Claire Markets, $n_{Pittsfield}$ = 1,008, $n_{EauClaire}$ = 1,007*.
*13 households in Pittsfield and 6 households in Eau Claire did not report age.

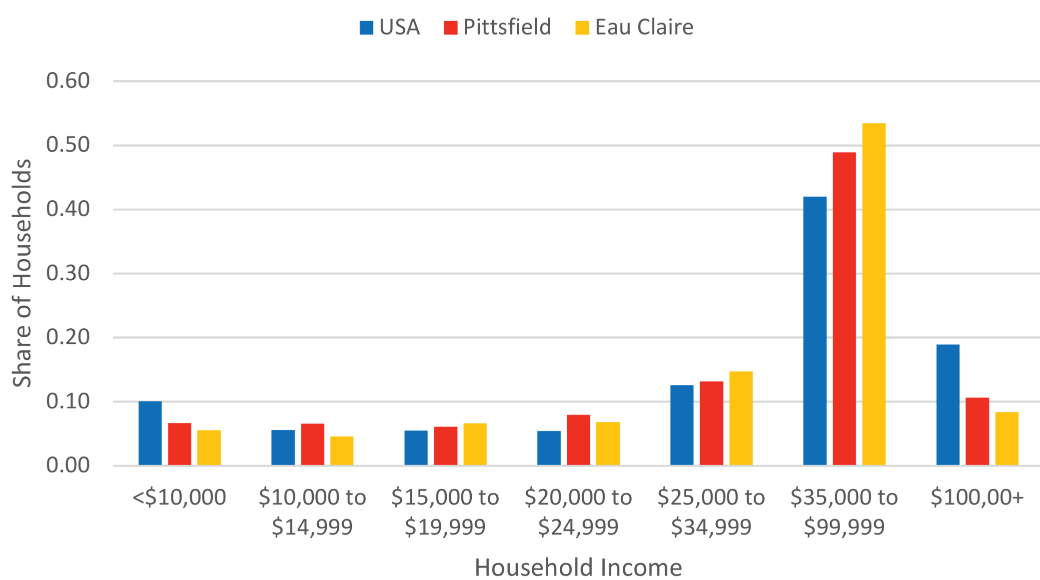


Figure A.20: Distribution of household income for the United States versus Pittsfield and Eau Claire Markets, $n_{Pittsfield}$ = 1,018*, $n_{EauClaire}$ = 1,013.
*3 households in Pittsfield did not report income.

A.6 Price Index Example

The price index, $P_{i,t}$, is constructed using the equation

$$P_{i,t} = \frac{\sum\limits_{j \in J_{i,t}} p_{i,j,k,t} \cdot q_{i,j,k,t}}{\sum\limits_{j \in J_{i,t}} \overline{p}_{j,k,t} \cdot q_{i,j,k,t}}$$

where i indexes the household, j indexes the individual good at the UPC level, $J_{i,t}$ is the set of goods purchased by household i in week t, k is the market (Eau Claire or Pittsfield), t indexes the week, $p_{i,j,k,t}$ is the price household i living in market k paid for good j in week t, $q_{i,j,k,t}$ is the quantity of good j purchased in week t by the household, and $\overline{p}_{j,k,t}$ is the average price paid for good j in market k for week t across all consumers who purchased good j that week.

Figure A.21 illustrates the construction of the price index measurement. In this example, there are two households, A and B, and two goods, 1 and 2. Household A purchases one unit of Good 1 at a price of \$1 and one unit of Good 2 at a price of \$2, so household A's total spending for the week is \$3, $p_{A,1,k,t} \cdot q_{A,1,k,t} = \$1 \cdot 1$ and $p_{A,2,k,t} \cdot q_{A,2,k,t} = \$2 \cdot 1$. Household B also purchases one unit of Good 1 but instead pays \$3 and then pays \$2 for one unit of Good 2, leading to a total basket cost of \$5, $p_{B,1,k,t} \cdot q_{B,1,k,t} = \$3 \cdot 1$ and $p_{B,2,k,t} \cdot q_{B,2,k,t} = \$2 \cdot 1$. For Good 1, the average price paid is \$2, $\overline{p}_{1,k,t} = \frac{\$1+\$3}{2}$, and the average price of Good 2 is also $\overline{p}_{2,k,t} = \$2$. This means that the average basket price for household A is \$4, $\overline{p}_{1,k,t} \cdot 1 + \overline{p}_{2,k,t} \cdot 1 = \$2 \cdot 1 + \$2 \cdot 1$, and the average basket price for household B is also \$4. The price index measure for household A, $P_{A,t}$, is then \$3/\$4 = 0.75 while $P_{B,t}$ is \$5/\$4 = 1.25.

A.7 Instrumental Variable Regressions Using Age

To perform an instrumental variable estimation, I average weekly frequency variables over the entire four year sample period in a similar manner as Aguiar and Hurst [1].[24] Hausman tests for

[24] Each observation is therefore a household. The n in the regression results is slightly lower than the total number of households in the full dataset due to missing age data for the instrument.

Household	Good	Units Bought, $q_{i,j,k,t}$	Price Paid, $p_{i,j,k,t}$	Week's Avg. Price, $\overline{p}_{j,k,t}$
A	1	1	\$1	\$2
A	2	1	\$2	\$2
A Total			\$3	\$4
B	1	1	\$3	\$2
B	2	1	\$2	\$2
B Total			\$5	\$4

Figure A.21: Illustration of price index construction.

endogeneity indicate that the measures of shopping behavior are not endogenous when averaged over the sample time period for all instruments considered, suggesting that an instrumental variable approach may not be necessary. Furthermore, Stock-Yogo tests for weak instruments indicate that education level of the head of household, household income, family size, the number of retired heads of household, and the number of not working heads of household are weak instruments for shopping behavior. Only age of the head of household as an instrument for average weekly trips could be considered a strong instrument based on the statistical tests.

Table A.18 reports the result from this instrumental variable estimation. While the estimate for the effect of additional weekly shopping trips is higher than all other estimates, roughly 2.5 times higher than the OLS estimates, the estimated elasticity of 1.35% is similar to the estimated effect for increasing the number of trips.[25] The higher estimate does suggest that there may be some concern about the OLS estimates being biased downwards due to reasons besides endogeneity, but the IV estimate is still on par with "a penny saved".

A.8 Alternative Definition of a Good

Rather than treat a good as a UPC, Kaplan and Menzio [52] utilize an alternative definition of a good that keeps brand, size, and product characteristics the same while allowing the exact UPC barcode to vary, which they call "generic brand aggregation". This allows UPCs with slight

[25] Results for a similar instrumental variable estimation using the number of stores is included for completeness but age is a weak instrument for the number of stores with a purchase. The store elasticity is higher, as was seen in the OLS and Fixed Effect estimation, but the difference in elasticity between trips and stores is much smaller.

	$ln\overline{P_i}$	$ln\overline{P_i}$
	(1)	(2)
ln(Average Number of Weekly Trips)	-0.0135***	
	(0.0050)	
ln(Average Number of Weekly Stores)		-0.0163***
		(0.0051)
Controls	Basket, Market	Basket, Market
Instrument	Age Dummy	Age Dummy
R^2	0.0730	0.1414
n	2,023	2,023

Table A.18: Instrumental variable regression results relating average weekly price index to average weekly shopping behavior.
Standard errors in parentheses, * = 10% significance, ** = 5% significance, *** $<$ 1% significance.

variations in packaging or vendor-specific packaging to be considered the same good.[26] Tables A.19 and A.20 present the results of equation (2.6) when using OLS or a household fixed effect, respectively, after calculating the price index variable, $P_{i,t}$, using the generic brand aggregation definition of a good. Under this wider definition of a good, the estimated effects from increased weekly shopping frequency are roughly 10% higher in magnitude compared to using UPCs to define goods. However, the overall estimates are still economically small, peaking at an elasticity of -0.0091 for the OLS estimate of the number of stores, implying that doubling the number of stores shopped at in a week only yields a 0.91% reduction in a household's price index.

A.9 An Existence Proof

Claim A1: If $(ka_t)^{-1}\Pi_t$ is small, there exists a unique solution to (3.5) that satisfies: $\lambda_t \leq P_{qt} < P_{2t} \cdots < P_{mt}$.

[26]The data from IRI does not always include how a product is packaged so there may be some error in assuming that UPCs that satisfy this definition of a good are comparable, especially since only total volume is reported. For example, this aggregation would hypothetically treat ten, 12-ounce aluminum cans of a specific soda as the same good as six, 20-ounce plastic bottles of that soda because they are both 120 ounces of that soda in total. People shopping for soda may not consider these two different packages of soda as equivalent.

	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$
	(1)	(2)	(3)	(4)
ln(Number of Trips)	-0.0047***		-0.0061***	
	(0.0002)		(0.0002)	
ln(Number of Stores)		-0.0084***		-0.0091***
		(0.0002)		(0.0002)
Controls	Market	Market	Market & Basket	Market & Basket
R^2	0.0021	0.0050	0.0171	0.0193
n	348,283	348,283	348,283	348,283

Table A.19: OLS results relating price index to shopping behavior using generic brand aggregation definition for a good.
Standard errors in parentheses, * = 10% significance, ** = 5% significance, *** $<$ 1% significance.

	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$	$lnP_{i,t}$
	(5)	(6)	(7)	(8)
ln(Number of Trips)	-0.0008***		-0.0024***	
	(0.0002)		(0.0003)	
ln(Number of Stores)		-0.0023***		-0.0034***
		(0.0003)		(0.0003)
Controls	None	None	Basket	Basket
Fixed Effect	Household	Household	Household	Household
R^2 Within	0.0000	0.0002	0.0133	0.0134
R^2 Between	0.0703	0.1349	0.0568	0.0753
R^2 Overall	0.0021	0.0049	0.0155	0.0167
n	348,283	348,283	348,283	348,283

Table A.20: Household fixed effect regressions results relating price index to shopping behavior using generic brand aggregation definition for a good.
Standard errors clustered at the household level, * = 10% significance, ** = 5% significance, *** $<$ 1% significance.

Proof. Let

$$A(P_{it}) = \frac{(k+1)\lambda_t}{k+q_i} + \frac{\Pi_t}{(k+q_i)a_t P_{it}^{\theta}} = C + BP_{it}^{-\theta}$$

We look for a fixed point: $P_{it} = A(P_{it})$. Since $\theta = \dfrac{1}{\gamma - 1} < 0$ and $1 + \theta = \dfrac{\gamma}{\gamma - 1} < 0$, we can sign the following derivatives:

$$A'(P_{it}) = -\theta BP_{it}^{-\theta-1} = -\theta BP_{it}^{-(1+\theta)} > 0$$

and

$$A''(P_{it}) = \theta(1+\theta)BP_{it}^{-(1+\theta)-1} > 0.$$

Note that $A'(0) = 0$. Since $B \leq (ka_t)^{-1}\Pi_t$ is small, the function $A(P_{it})$ will intersect the 45 degree line twice. We choose the lower intersection in Figure A.22. Note that the function $A(P_{it})$ is decreasing in q_i and therefore an increase in q_i will shift the curve downward and the fixed point will be lower. Therefore: $P_{1t} < P_{2t} < \cdots < P_{mt}$. Since $C \geq \lambda_t$ it follows that $P_{it} \geq \lambda_t$. $\square$

The monopoly price charged by the non-advertisers is:

$$P_t^m = \arg\max_p ka_t P^{\theta}(P - \lambda_t) = \frac{\theta \lambda_t}{\theta + 1} = \frac{\lambda_t}{\gamma} \tag{A1}$$

The monopoly profits are:

$$\Pi_t^m = ka_t(\lambda_t)^{\theta+1}(1-\gamma)\gamma^{-(1+\theta)} \tag{A2}$$

We require that advertisers will make the same profits as non-advertisers:

$$\Pi_t = \Pi_t^m \tag{A3}$$

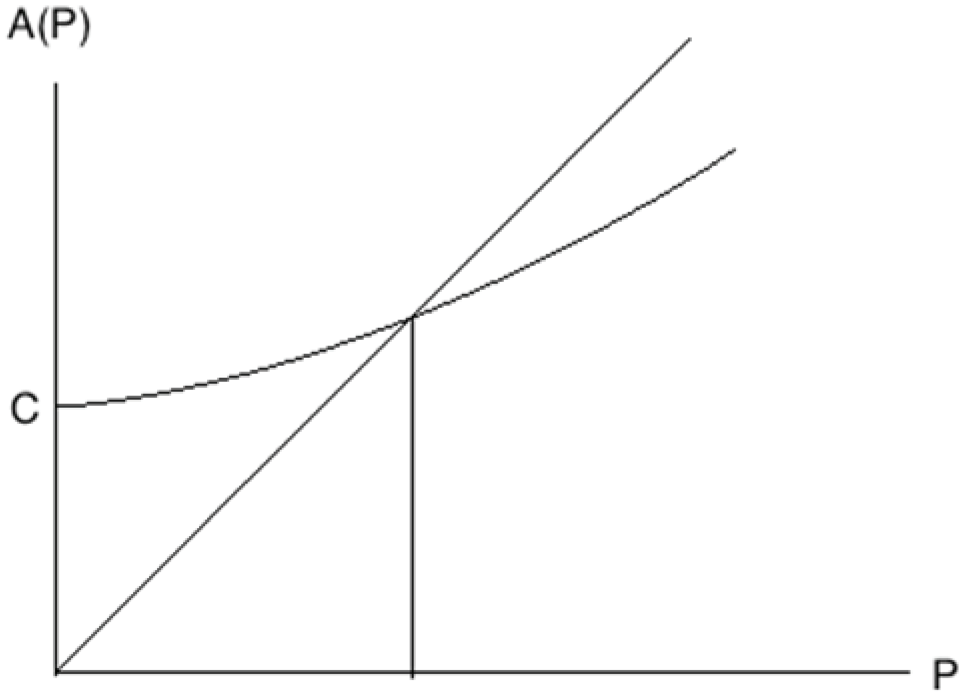


Figure A.22: Illustration of the lower fixed point of $A(P_{it})$.

Equilibrium.

We define equilibrium as a vector $(P_{it},\ldots,P_{mt};\Pi_t,\Pi_t^m)$ that satisfies (A1)-(A3) and $\lambda_t \le P_{1t} < P_{2t}\cdots < P_{mt}$.

Claim A2: There exists an equilibrium if

$$\frac{\Pi_t^m}{ka_t} = \frac{ka_t(\lambda_t)^{\theta+1}(1-\gamma)\gamma^{-(1+\theta)}}{ka_t} = (1-\gamma)\left(\frac{\gamma}{\lambda_t}\right)^{\frac{\gamma}{1-\gamma}} \tag{A4}$$

is small.

To show this Claim, note since $\Pi_t = \Pi_t^m$, (A4) insures that the condition in Claim A1 is satisfied for all q_i. The expression (A4) is decreasing in $\gamma$and in λ_t. Therefore the sufficient condition is satisfied when either γ is large or λ is large.

Figure A.23 provides a numerical example. In the figure there are three curves. The curve labeled "lambda/q" describes the standard UST prices which corresponds to the case $k = \Pi = 0$.

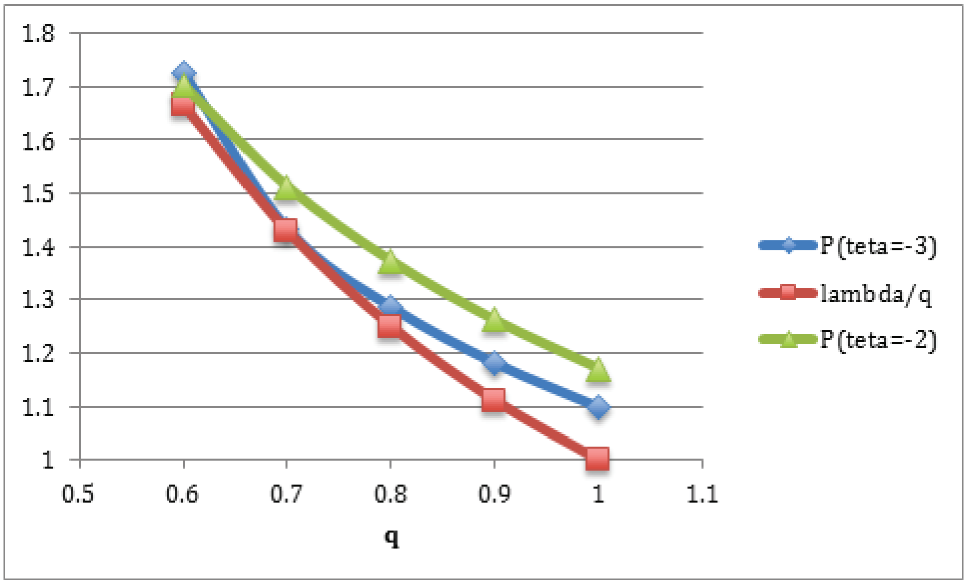


Figure A.23: The relationship between equilibrium price and probability of making a sale to shoppers.
The probability of making a sale to shoppers (q) is on the horizontal axis. Equilibrium prices are on the vertical axis. All three curves assume $\lambda = a = 1$. The curve lambda/q $\left(\frac{\lambda}{q}\right)$ is the prices of the standard UST model with $k = \Pi = 0$. The curves P assume $k = 1$. The curve P(teta=-3) assumes $\gamma = 2/3$, $\theta = -3$ and $\Pi_t = \Pi_t^m = (1-\gamma)\gamma^{-(1+\theta)} = 0.15$. The curve P(teta=-2) assumes $\gamma = 1/2$, $\theta = -2$ and $\Pi_t = \Pi_t^m = (1-\gamma)\gamma^{-(1+\theta)} = 0.25$.

The curve labeled P(teta=-3) computes the equilibrium prices under the assumption that the own price elasticity is $\theta = -3$ and the curve labeled P(teta=-2) computes the equilibrium prices under the assumption $\theta = -2$. As can be seen having non-shoppers in the model increase prices and profit. In general when θ is lower in absolute value there is more monopoly power and prices are higher. This seems to be the case when the probability of selling to shoppers is high. In our example, the price when $q = 0.6$ is an exception to the rule. The monopoly price is 1.5 when $\theta = -3$ and 2 when $\theta = -2$. In this example, the monopoly price is the highest price when $q \geq 0.7$.

Deviation from equilibrium behavior.

Suppose that an advertiser increases his price from say P_{it} to $P_{it} + \varepsilon < P_{i+1t}$. In this case a non-advertiser will fill the gap in the equilibrium price distribution by advertising P_{it}. As a result the deviant advertiser will be able to sell only if $\tilde{s} > i$ with probability q_{i+1}. Will he increase his expected profits? To answer this question, let $\Pi(q,P) = (k+q)aP^{1+\theta} - (k+1)\lambda aP^{\theta}$ denote the expected profits as a function of the probability of making a sale to shoppers and the price. The function $\Pi(q,P)$ is increasing in q. Under the assumption that $P \geq \dfrac{\lambda}{q}$ and q is large the function $\Pi(q,P)$ is also increasing in P. When $\Pi(q,P)$ increases in both its arguments: $\Pi(q_{i+1}, P_i + \varepsilon < P_{i+1}) < \Pi(q_{i+1}, P_{i+1}) = \pi(q_i, P_i)$ and the above deviation reduces expected profits.

A.10 Generalizing the Results in Claim 2

Changing the dependent variable makes a big difference in this case. Substituting $\ln(a_t) = -\ln(k) + \ln(x_t^m) - \theta \ln(P_t^m)$ in (3.22) leads to:

$$\ln(x_{it}) = \ln(\tilde{\omega}_t) - \ln(k) + \ln(x_t^m) - \theta \ln(P_t^m) + \theta \ln(P_{it}) \tag{B1}$$

Here there is no correlation between the quantity sold by the monopoly $\ln(x_t^m)$ and the number of buyers served by the advertiser $\ln(\tilde{\omega}_t)$ and therefore when running a regression based on (B1) we expect the quantity elasticity to equal unity and the cross price elasticity to equal the absolute value of the own price elasticity.

<u>The relationship between the quantities sold by advertisers.</u>

We now turn to study the relationship between the quantities sold by stores in two different bins that are occupied by advertisers.

We start with two stores indexed $j < i \leq m$. Subtracting the quantity sold by advertiser j, $\ln(x_{jt}) = \ln(\tilde{\omega}_j) + \ln(a) + \theta \ln(P_{jt})$, from (3.22) leads to:

$$\ln(x_{it}) = \ln(x_{jt}) + \theta \ln(P_{it}) - \theta \ln(P_{jt}) + D_{ij} \tag{B2}$$

where $D_{ij} = \ln(\tilde{\omega}_i) - \ln(\tilde{\omega}_j)$ is the difference in the number of shoppers. Since $j < i$, when $s < j$ the number of shoppers is zero for both stores and $D_{ij} = 0$. It is also zero when $s \geq i$ because in this case the number of shoppers is 1 for both stores. The difference D_{ij} is negative when $j \leq s < i$

because in this case a shopper arrives in the low index store but does not arrive in the high index store. Thus,

$$D_{ij} = \ln(k) - \ln(1+k) = \begin{cases} \ln(k) - \ln\left(\frac{x_{jt}}{aP_{jt}^{\theta}}\right) & \text{if} \quad j \leq s < i \\ 0 & \text{otherwise} \end{cases} \tag{B3}$$

Since $q_i = \sum_{s=i}^{m} \pi_s$ and $q_j = \sum_{s=j}^{m} \pi_s$,

$$Prob(j \leq s < i) = \sum_{s=j}^{i-1} \pi_s = q_j - q_i \tag{B4}$$

The expected difference in the number of shoppers is:

$$E(D_{ij}) = \psi_{ij} + (q_j - q_i)\left(-\ln(s_{jt}) + \theta \ln(P_{jt})\right) \tag{B5}$$

where $\psi_{ij} = (q_j - q_i)\left[\ln(k) + E(\ln(\tilde{a}))\right]$. We can therefore write:

$$D_{ij} = \psi_{ij} + (q_j - q_i)\left(-\ln(x_{jt}) + \theta \ln(P_{jt})\right) + \varepsilon_{ijt} \tag{B6}$$

where ε_{ijt} is an i.i.d. random variable with zero mean.

Substituting (B6) in (B2) leads to:

$$\ln(x_{it}) = \psi_{ij} + (1 + q_i - q_j)\ln(x_{jt}) + \theta \ln(P_{it}) - (1 + q_i - q_j)\theta \ln(P_{jt}) + \varepsilon_{ijt} \tag{B7}$$

Since $q_i < q_j$, $0 < 1 + q_i - q_j < 1$, the quantity elasticity is between zero and unity and the own price elasticity is higher in absolute value than the cross price elasticity.

Claim B1: (a) the quantity elasticity is between zero and unity, (b) the own price elasticity is higher in absolute value than the cross price elasticity, (c) the quantity elasticity and the cross price elasticity are decreasing with the distance between the bins.

Parts (a) and (b) follow from $0 < 1 + q_i - q_j < 1$. To show (c) note that the absolute value of the difference in the probabilities of selling to shoppers ($|q_i - q_j|$) is larger for bins that are further apart.

We now turn to show that Claim 4 holds also for the case: $j > i$.

When $j > i$, $D_{ij} = 0$ if $s \geq j$ or if $s < i$. Thus,

$$D_{ij} = \ln(1+k) - \ln(k) = \begin{cases} \ln(1+k) - \ln\left(\frac{x_{jt}}{a_t P_{jt}^{\theta}}\right) & \text{when} \quad i \leq s < j \\ 0 & \text{otherwise} \end{cases} \tag{B8}$$

The expected value of (B8) is:

$$E(D_{ij}) = \psi_{ij}^{*} - (q_i - q_j)\left(\ln(x_{jt}) - \theta \ln(P_{jt})\right) \tag{B9}$$

where $\psi_{ij}^{*} = (q_i - q_j)\left[\ln(1+k) + E(\ln(\tilde{a}))\right]$. We can therefore write:

$$D_{ij} = \psi_{ij}^{*} + (q_j - q_i)\left(\ln(x_{jt}) - \theta \ln(P_{jt})\right) + \varepsilon_{ijt} \tag{B10}$$

Substituting (B10) in (B2) leads to:

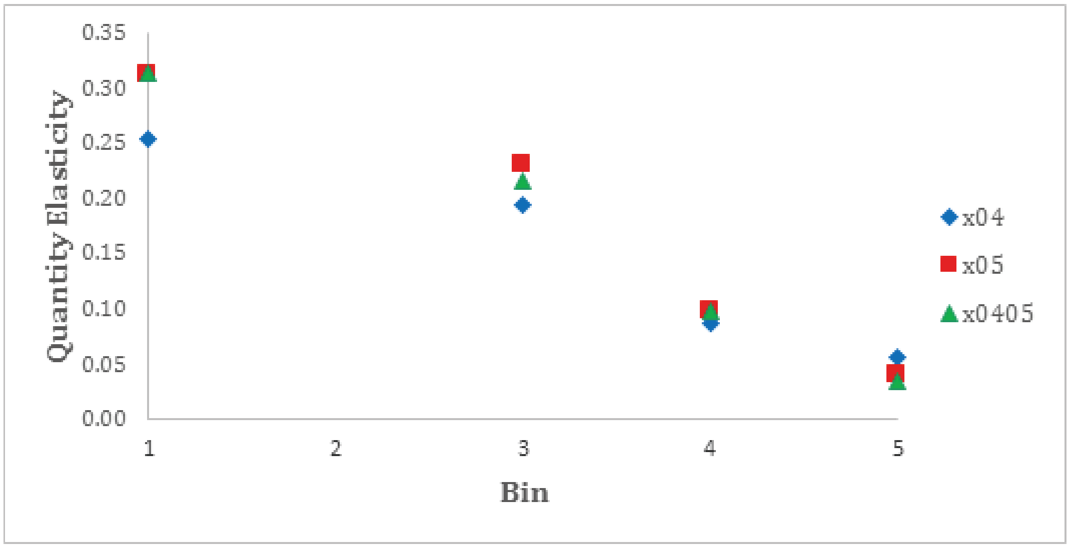
$$\ln(x_{it}) = \psi_{ij}^{*} + (1 + q_j - q_i)\ln(x_{jt}) + \theta \ln(P_{it}) - (1 + q_j - q_i)\theta \ln(P_{jt}) + \varepsilon_{ijt} \tag{B11}$$

Since $q_j < q_i$, Claim B1 holds also for this case.

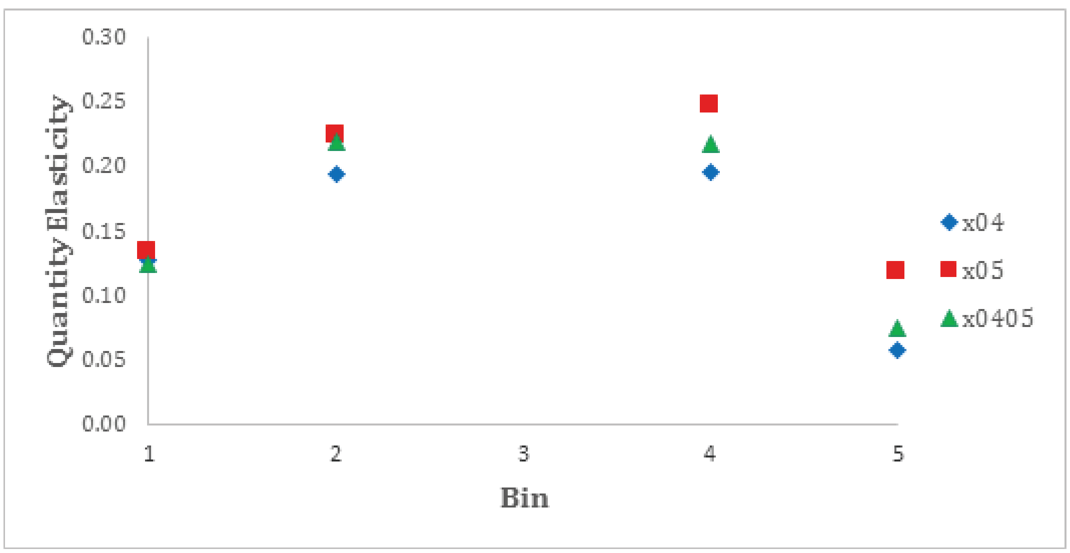
The theory says that the quantity elasticity and the cross price elasticity declines with the index of the explanatory bin. We have seen that this is the case when the dependent variable is the quantity sold by stores in the first bin. Does it holds for the case where the dependent variable is X2 or X3?

We ran X2 on 9 explanatory variables: X1, X3, X4, X5, P1, P2, P3, P4, P5. We expect that the coefficient of X3 will be larger than the coefficient of X4 and X5. We also expect that the coefficient of P3 will be larger than the coefficient of P4 and P5. We also ran X3 on X1, X2, X4, X5, P1, P2, P3, P4, P5. In this regression we expect that the coefficient of X4 is larger than the coefficient of X5 and the coefficient of P4 is larger than the coefficient of P5.

We ran these regressions using the original variables, the variables net of store effects and the variables net of UPC specific store effects. The results support the above hypotheses. Figures A.24 and A.25 describe the results when using the variables net of UPC specific store effects.

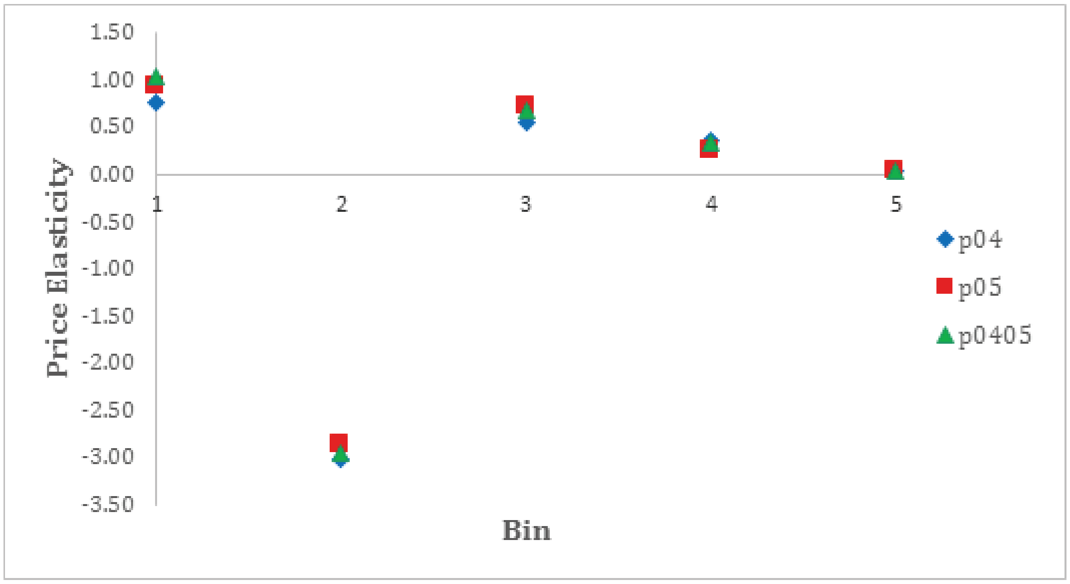


(a) Dependent variable = X2

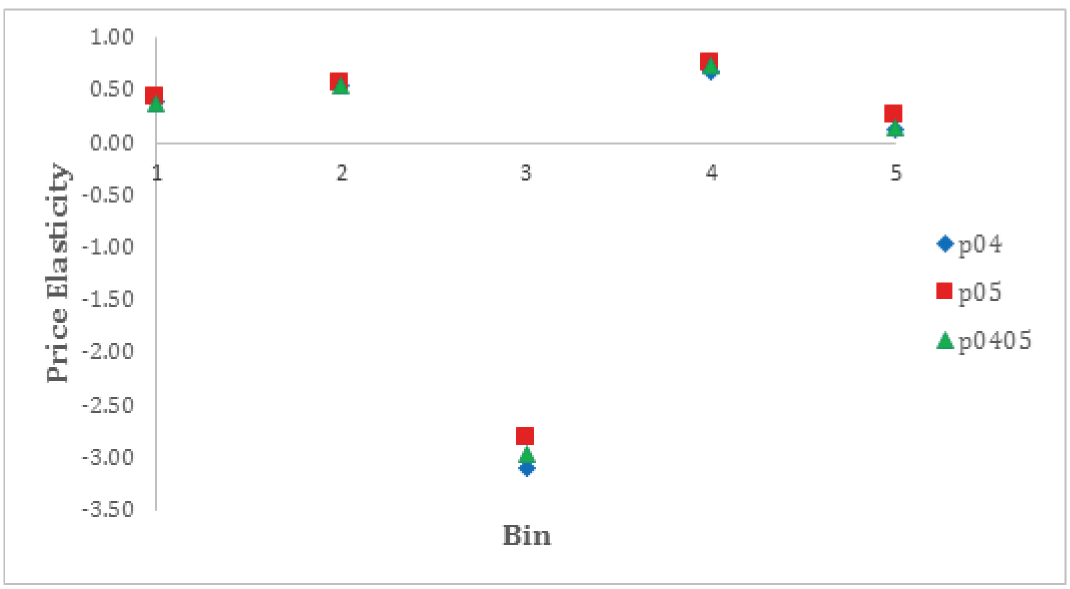


(b) Dependent variable = X3

Figure A.24: Quantity elasticities based on 9 explanatory variable regressions with UPC-specific store effects.



(a) Dependent variable = X2



(b) Dependent variable = X3

Figure A.25: Price elasticities based on a regression with 9 explanatory variables with UPC-specific store effect.

BIBLIOGRAPHY

[1] Aguiar, M. and Hurst, E. (2007). Life-cycle prices and production. The American Economic Review, 97(5):1533–1559.

[2] Aguirregabiria, V. (1999). The dynamics of markups and inventories in retailing firms. Review of Economic Studies, 66(2):275 – 308.

[3] Ahlin, C. and Shintani, M. (2007). Menu costs and markov inflation: A theoretical revision with new evidence. Journal of Monetary Economics, 54(3):753 – 784.

[4] Anderson, E. T. and Simester, D. (2013). Advertising in a competitive market: The role of product standards, customer learning, and switching costs. Journal of Marketing Research, 50(4):489 – 504.

[5] Baharad, E. and Eden, B. (2004). Price rigidity and price dispersion: Evidence from micro data. Review of Economic Dynamics, 7(3):613 – 641.

[6] Banerjee, A. V. and Mullainathan, S. (2008). Limited attention and income distribution. American Economic Review, 98(2):489 – 493.

[7] Basu, K. (1997). Why are so many goods priced to end in nine? and why this practice hurts the producers. Economics Letters, 54(1):41 – 44.

[8] Basu, K. (2006). Consumer cognition and pricing in the nines in oligopolistic markets. Journal of Economics and Management Strategy, 15(1):125 – 141.

[9] Bental, B. and Eden, B. (1993). Inventories in a competitive environment. Journal of Political Economy, 101(5):863 – 886.

[10] Bental, B. and Eden, B. (1996). Money and inventories in an economy with uncertain and sequential trade. Journal of Monetary Economics, 37(3):445 – 459.

[11] Bergen, M., Kauffman, R. J., and Lee, D. (2004). Store quality image and the rational inattention hypothesis: As empirical study of the drivers of $9 and 9¢ price-endings among internet-based sellers. In INFORMS Conference on Information Systems and Technology.

[12] Blinder, A. S., Canetti, E. R., Lebow, D. E., and Rudd, J. B. (1998). Asking About Prices: A New Approach to Understanding Price Stickiness. Russel Sage Foundation, New York.

[13] Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. Journal of Political Economy, 121(5):803 – 843.

[14] Bordalo, P., Gennaioli, N., and Shleifer, A. (2016). Competition for attention. Review of Economic Studies, 83(2):481 – 513.

[15] Bordalo, P., Gennaioli, N., and Shleifer, A. (Working Paper). Memory, attention, and choice. NBER Working Paper No. w23256.

[16] Bronnenberg, B. J., Kruger, M. W., and Mela, C. F. (2008). Database paper: The IRI marketing data set. Marketing Science, 27(4):745 – 748.

[17] Burdett, K. and Judd, K. L. (1983). Equilibrium price dispersion. Econometrica, 51(4):955–969.

[18] Burdett, K., Shi, S., and Wright, R. (2001). Pricing and matching with frictions. Journal of Political Economy, 109(5):1060 – 1085.

[19] Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1997). Sticky price and limited participation models of money: A comparison. European Economic Review, 41(6):1201 – 1249.

[20] Coibion, O., Gorodnichenko, Y., and Hong, G. H. (2015). The cyclicality of sales, regular and effective prices: Business cycle and policy implications. American Economic Review, 105(3):993 – 1029.

[21] Conlon, C. T. and Rao, N. (2016). Discrete prices and the incidence and efficiency of excise taxes. Working Paper.

[22] Cornia, M., Gerardi, K. S., and Shapiro, A. H. (2012). Price dispersion over the business cycle: Evidence from the airline industry. Journal of Industrial Economics, 60(3):347 – 373.

[23] Dana, James D., J. (1998). Advance-purchase discounts and price discrimination in competitive markets. Journal of Political Economy, 106(2):395 – 422.

[24] Dana, James D., J. (1999). Equilibrium price dispersion under demand uncertainty: The roles of costly capacity and market structure. RAND Journal of Economics, 30(4):632 – 660.

[25] Dana, James D., J. (2001). Monopoly price dispersion under demand uncertainty. International Economic Review, 42(3):649 – 670.

[26] de Clippel, G., Eliaz, K., and Rozen, K. (2014). Competing for consumer inattention. Journal of Political Economy, 122(6):1203 – 1234.

[27] Deneckere, R. and Peck, J. (2012). Dynamic competition with random demand and costless search: A theory of price posting. Econometrica, 80(3):1185 – 1247.

[28] Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. American Economic Review, 67(3):297 – 308.

[29] Dixon, D. F. and McLaughlin, D. J. (1971). Shopping behavior, expenditure patterns, and inner-city food prices. Journal of Marketing Research, 8(1):96–99.

[30] Dow, J. (1991). Search decisions with limited memory. Review of Economic Studies, 58(1):1 – 14.

[31] Eden, B. (1990). Marginal cost pricing when spot markets are complete. Journal of Political Economy, 98(6):1293 – 1306.

[32] Eden, B. (1994). The adjustment of prices to monetary shocks when trade is uncertain and sequential. Journal of Political Economy, 102(3):493 – 509.

[33] Eden, B. (2001a). Inflation and price adjustment: An analysis of microdata. Review of Economic Dynamics, 4(3):607 – 636.

[34] Eden, B. (2001b). Inventories and the business cycle: Testing a sequential trading model. Review of Economic Dynamics, 4(3):562 – 574.

[35] Eden, B. (forthcoming). Price dispersion and demand uncertainty: Evidence from us scanner data. International Economic Review.

[36] Eden, B., Eden, M., and Yuen, J. (Working Paper). Inside the price dispersion box: Evidence from us scanner data. Vanderbilt University Department of Economics Working Papers. VUECON-16-00017.

[37] Eichenbaum, M., Jaimovich, N., and Rebelo, S. (2011). Reference prices, costs, and nominal rigidities. American Economic Review, 101(1):234 – 262.

[38] Escobari, D. (2012). Dynamic pricing, advance sales and aggregate demand learning in airlines. Journal of Industrial Economics, 60(4):697 – 724.

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

[40] Gerardi, K. S. and Shapiro, A. H. (2009). Does competition reduce price dispersion? new evidence from the airline industry. Journal of Political Economy, 117(1):1 – 37.

[41] Ginzberg, E. (1936). Customary prices. American Economic Review, 26:296.

[42] Glandon, P. (2016). Sales and the (mis)measurement of price level fluctuations. Available at SSRN: https://ssrn.com/abstract=2047127 or http://dx.doi.org/10.2139/ssrn.2047127.

[43] Griffith, R., Leibtag, E., Leicester, A., and Nevo, A. (2009). Consumer shopping behavior: How much do consumers save? The Journal of Economic Perspectives, 23(2):99–120.

[44] Hackl, F., Kummer, M. E., and Winter-Ebmer, R. (2014). 99 cent: Price points in e-commerce. Information Economics and Policy, 26:12 – 27.

[45] Head, A., Liu, L. Q., Menzio, G., and Wright, R. (2012). Sticky prices: A new monetarist approach. Journal of the European Economic Association, 10(5):939 – 973.

[46] Hendel, I. and Nevo, A. (2013). Intertemporal price discrimination in storable goods markets. American Economic Review, 103(7):2722 – 2751.

[47] Herrman, R., Moeser, A., and Weber, S. A. (2005). Price rigidity in german grocery-retailing sector: Scanner data evidence on magnitude and casues. Journal of Agricultural and Food Industrial Organization, 3(1):1–35.

[48] Herrmann, R. and Moeser, A. (2006). Do psychological prices contribute to price rigidity? evidence from german scanner data on food brands. Agribusiness, 22(1):51–67.

[49] Hoeberichts, M. and Stokman, A. (2010). Price setting behaviour in the Netherlands: Results of a survey. Managerial and Decision Economics, 31(2-3):135 – 149.

[50] Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2):263 – 291.

[51] Kalyanaram, G. and Little, J. D. C. (1994). An empirical analysis of latitude of price acceptance in consumer package goods. Journal of Consumer Research, 21(3):408 – 418.

[52] Kaplan, G. and Menzio, G. (2015). The morphology of price dispersion. International Economic Review, 56(4):1165 – 1205.

[53] Kaplan, G. and Menzio, G. (2016). Shopping externalities and self-fulfilling unemployment fluctuations. Journal of Political Economy, 124(3):771–825.

[54] Kehoe, P. and Midrigan, V. (2015). Prices are sticky after all. Journal of Monetary Economics, 75:35 – 53.

[55] Khaw, M. W., Stevens, L., and Woodford, M. (2016). Discrete adjustment to a changing environment:experimental evidence. Working Paper.

[56] Kling, J. R., Mullainathan, S., Shafir, E., Vermeulen, L. C., and Wrobel, M. V. (2012). Comparison friction: Experimental evidence from medicare drug plans. Quarterly Journal of Economics, 127(1):199 – 235.

[57] Lacetera, N., Pope, D. G., and Sydnor, J. R. (2012). Heuristic thinking and limited attention in the car market. American Economic Review, 102(5):2206 – 2236.

[58] Lach, S. (2002). Existence and persistence of price dispersion: An empirical analysis. Review of Economics and Statistics, 84(3):433 – 444.

[59] Lal, R. and Matutes, C. (1994). Retail pricing and advertising strategies. Journal of Business, 67(3):345 – 370.

[60] Lambert, Z. V. (1975). Perceived prices as related to odd and even price endings. Journal of Retailing, 51(3):13–22; 78.

[61] Lloyd, R. and Jennings, D. (1978). Shopping behavior and income: Comparisons in an urban environment. Economic Geography, 54(2):157–167.

[62] Mankiw, N. G. and Reis, R. (2002). Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. Quarterly Journal of Economics, 117(4):1295 – 1328.

[63] Mullainathan, S., Schwartzstein, J., and Shleifer, A. (2008). Coarse thinking and persuasion. Quarterly Journal of Economics, 123(2):577 – 619.

[64] Nakamura, E. and Steinsson, J. (2008). Five facts about prices: A reevaluation of menu cost models. Quarterly Journal of Economics, 123(4):1415 – 1464.

[65] Poltrock, S. E. and Schwartz, D. R. (1984). Comparative judgments of multidigit numbers. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10(1):32–45.

[66] Prescott, E. C. (1975). Efficiency of the natural rate. Journal of Political Economy, 83(6):1229 – 1236.

[67] Reinganum, J. F. (1979). A simple model of equilibrium price dispersion. Journal of Political Economy, 87(4):851–858.

[68] Reinsdorf, M. (1994). New evidence on the relation between inflation and price dispersion. American Economic Review, 84(3):720 – 731.

[69] Reis, R. (2006a). Inattentive consumers. Journal of Monetary Economics, 53(8):1761 – 1800.

[70] Reis, R. (2006b). Inattentive producers. Review of Economic Studies, 73(3):793 – 821.

[71] Salop, S. and Stiglitz, J. E. (1977). Bargains and ripoffs: A model of monopolistically competitive price dispersion. Review of Economic Studies, 44(3):493 – 510.

[72] Schindler, R. M. and Kibarian, T. M. (1996). Increased consumer sales response through use of 99-ending prices. Journal of Retailing, 72(2):187–199.

[73] Schindler, R. M. and Kirby, P. N. (1997). Patterns of rightmost digits used in advertised prices: Implications for nine-ending effects. Journal of Consumer Research, 24(2):192 – 201.

[74] Shilony, Y. (1977). Mixed pricing in oligopoly. Journal of Economic Theory, 14(2):373 – 388.

[75] Shy, O. (2000). Why 99 cents? Working Paper.

[76] Simon, H. A. (1955). A behavioral model of rational choice. The Quarterly Journal of Economics, 69(1):99–118.

[77] Sims, C. A. (2003). Implications of rational inattention. Journal of Monetary Economics, 50(3):665 – 690.

[78] Sorensen, A. T. (2000). Equilibrium price dispersion in retail markets for prescription drugs. Journal of Political Economy, 108(4):833 – 850.

[79] Stigler, G. J. (1961). The economics of information. Journal of Political Economy, 69(3):213–225.

[80] Stiving, M. and Winer, R. S. (1997). An empirical analysis of price endings with scanner data. Journal of Consumer Research, 24(1):57 – 67.

[81] Tang, Z., Smith, M. D., and Montgomery, A. (2010). The impact of shopbot use on prices and price dispersion: Evidence from online book retailing. International Journal of Industrial Organization, 28(6):579 – 590.

[82] Varian, H. R. (1980). A model of sales. The American Economic Review, 70(4):651–659.