NORTHWESTERN UNIVERSITY

Essays on Industrial Organization

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Yongbae Lee

EVANSTON, ILLINOIS

December 2008

ABSTRACT

Essays on Industrial Organization

Yongbae Lee

In the first part of the dissertation, I investigate the nature of retail coupons, a popular tool for non-price competition. The widely expressed view that coupons are primarily a tool to allow price discrimination has received mixed empirical supports. I depart from the static framework of the price discrimination theory to explore what alternative roles the coupons may play in an environment where demand is dynamic. In Chapter 1, I examine the consumer-level panel data. I show that, while coupons themselves do not have any lasting effect on consumers' brand choice, they induce different responses from consumers with varying degree of consumption experience. The evidence implies that coupons may have promotional effects that reinforce consumers' decaying consumption experience. In Chapter 2, I examine the retailer-level sales data and investigate whether coupon availability constitutes a state variable for the retailers' pricing decision. I estimate a linear probability model to show that coupon availability does have an influence over retailers' sale decision even after accounting for accumulation of latent demand over time.

In the second part, I conduct an econometric exercise using the dynamic discrete choice model. The stage utility functions in dynamic discrete choice models are, in general, not nonparametrically identified even when the discount factor and the distribution of the unobservable state vector are known to the researcher. Aguirregabiria (2002) demonstrated that it is feasible to identify the counterfactual choice probabilities without evaluating the stage utility function, when the policy in question linearly modifies the stage utility function. I study a different type of policy implementation that results in a shift in transition probabilities, with which Aguirregabiria's results are not replicated. It is shown, however, that with a sufficient variation in transition probabilities, we can point identify the stage utility function when we are given the opportunity to observe the change in the agent's behavior following such a policy implementation.

Acknowledgements

I am deeply indebted to the members of my dissertation committee, Professors Igal Hendel, Aviv Nevo and Shane Greenstein for their continuous guidance and patience. Despite numerous mistakes I made over the course of years, they always judged me by my potential, not by my errors.

I am equally grateful to Professors Leemore Dafny, David Dranove and Alberto Salvo for letting me participate in their exciting research projects. They survived through many of my lame jokes and they were open-minded enough to admit that at least some of them were actually funny. There is no doubt that I learned much more from these projects than they benefited from my assistance. I would also like to thank Professors Martin Zelder and Robert Coen for sharing with me the joy of teaching.

I thank Kwanjeong Education Foundation for the graduate fellowship, that partially funded the research. Professor David Bell of University of Pennsylvania gave me his permission to use the data set analyzed in the first two chapters of this dissertation. Priscilla Medeiros kindly helped me to obtain the data set. Discussions with Xun Tang and Mateo Caronia were especially helpful in developing the research ideas.

Most importantly, I would like to express my gratitude to my parents and to my wife, Jeonghoo, to whom I dedicate this dissertation. My parents have never abandoned their faith in me even when I had doubts about myself. In their comforting words, I found new courage to endure the difficult times. Jeonghoo has been the best companion I could

hope for throughout this endeavour. Despite the hectic schedule of her own, she willingly shared my burden whenever I was one hand short. She inspired me and challenged me intellectually when I was laid back, and then gave me just enough distractions to keep me in one piece when I was stressed out.

Table of Contents

ABSTRACT 3

Acknowledgements 5

Chapter 1. Habit Persistence, Its Reversion and the Promotional Effect of Coupons 7

1.1. Introduction 7

1.2. Data 11

1.3. The Model 16

1.4. Estimation Results 23

1.5. Discussion 28

Chapter 2. The Impact of Coupon Availability on Retailers' Decision to Hold Sale 30

2.1. Introduction 30

2.2. Data 32

2.3. Empirical Analysis 38

2.4. Discussion 50

Chapter 3. Identification of Dynamic Discrete Choice Models with Variations in Transition Probabilities 55

3.1. Introduction 55

3.2. Model 57

3.3. Identification and Non-identification of Counterfactual Choice Probabilities 63

3.4. Identification with Additional Data 72

3.5. Discussion 75

3.6. Conclusion 79

References 80

List of Tables

1.1	Household Summary Statistics	13
1.2	Households' quantity and duration responses to marketing variables	17
1.3	Households' brand choice responses to marketing variables	18
1.4	Top 15 Brands	25
1.5	Estimation Results	26
2.1	Major Brands and Their Market Shares	34
2.2	Summary Stats of Key Variables	37
2.3	Impact of Coupon Availability on Normalized Price	41
2.4	Impact of Coupon Availability on Retailers Sale Decision	43
2.5	Duration Dependence	46
2.6	Correlation in Sale Across Brands Conditional on Store Sale	47
2.7	Summary Statistics for Store-Week Data	48
2.8	Impact of Coupon Availability on Store-wide Sale	50
2.9	Impact of Coupon Availability on Inter-Sale Duration	51

List of Figures

1.1	Redemption Patterns by Coupon Type	14
2.1	Price Path for General Mills' Cheerios in Store 1420	35
2.2	Evolution of Normalized Price	36

CHAPTER 1

Habit Persistence, Its Reversion and the Promotional Effect of Coupons

1.1. Introduction

The widely expressed view that coupons are primarily a tool to allow price discrimination has received mixed empirical support. An alternative hypothesis suggests that coupons induce consumers to try a new product or remind consumers of an existing product, making them more likely to purchase the product at full price in the future. Implicit in this hypothesis is the assumption that consumption experience leaves a lasting impression on consumer's preference and that this effect diminishes over time if not reinforced by another consumption experience, rendering the consumer more susceptible to sellers' promotional activities such as coupons. Using a discrete choice model that allows for an interaction between consumer's recent purchase experience and coupon availability, I attempt to verify whether evidence consistent with the repeat purchase hypothesis is present in household-level grocery purchase panel data.

First distributed in the late nineteenth century, coupons are now one of the most popular marketing tools adopted by consumer packaged good manufacturers. According to a recent survey by NCH Marketing Services, Inc., U.S. consumer packaged goods companies distributed some 279 billion coupons in 2006 with $1.18 average face value, of which 2.8 billion were redeemed.[1] The cost of designing, printing, distributing and processing coupons, along with the discounts offered, constitutes a major component of these companies' promotion budget. A single nation-wide coupon drop could easily represent a multi-million dollar investment from a manufacturer's point of view.

Economists traditionally regarded coupons as a canonical example of third-degree price discrimination. If consumers are heterogenous in terms of their price elasticities of demand and if consumer types are not directly observable, a monopolistic seller may set the regular price high and, at the same time, distribute coupons to let the consumers self-select in accordance to their price sensitivities, extracting more surplus than he would have had he set a uniform price provided that the distribution of consumers' tendency to clip and redeem coupons is positively correlated with the distribution of their price elasticities.[2] Early empirical studies, such as Teel, Williams & Bearden (1980), used household surveys to establish the demographic profile of coupon users and found that coupon users have significantly larger family sizes, larger incomes and are significantly younger than nonusers of coupons. Narasimhan (1984) also provided indirect evidence in support of the price discrimination hypothesis by showing that the users of coupons are more price elastic than nonusers of coupons.

Nevo and Wolfram (2002), on the other hand, examined the implication of price discrimination hypothesis directly using the data on shelf price and coupon availability for ready-to-eat cereals. They found that shelf prices are generally lower when coupons are

[1] The total face value of redeemed coupons is roughly \$2.8 billion.

[2] Corts (1998) demonstrated that, in a competitive industry, sellers' price discrimination may result in lower prices for all consumers if different type consumers have different brand preferences.

available, which is inconsistent with the predictions of the monopolistic price discrimination hypothesis under a broad range of assumptions. Furthermore, they found that the effect of a past coupon discount on volume is different from that of a past sale, which is also inconsistent with the hypothesis that coupons are used as a tool to implement Sobel-type (1984) intertemporal price discrimination. The fact that coupons, unlike sales, have a positive impact on volume sold in subsequent time periods, however, is consistent an alternative hypothesis suggesting that the sellers may be using coupons to induce trial of a new product or to remind consumers of an existing product, in hope that they will purchase the product again in the future at full price.

This hypothesis carries with it the following three underlying assumptions : (i) the current consumption experience has a lasting impact on the consumer's subsequent brand choices, (ii) the effect of consumption experience diminishes over time and it is eventually overcome by sellers' promotional activities and (iii) coupons have promotional effects that reinforce consumers' consumption experience, which is absent in simple price cuts.

The three assumptions of the repeat purchase hypothesis collectively yields an interesting prediction. If all three effects are present, then coupons should have different impacts on those consumers who recently purchased the product and those who did not. A coupon for a product is likely to have similar effects as a simple price cut to those who purchase the product recently since these consumers, through their consumption experience, are already well aware of the product's existence and its characteristics. The same coupon should have an additional promotional effect on those who did not purchase the product recently since their awareness of the product's existence and its characteristics have diminished over time and the coupon can convey such information to them.

The dynamic effect of consumption experience is well documented in both economics and marketing literature. A positive serial correlation in consumer's brand choice may arise due to a variety of reasons such as brand loyalty, consumer learning, switching cost and inertia, and it is frequently referred to as habit persistence. Empirical researches in marketing, Guadagni and Little (1983), for example, often report that habit persistence is the single most important determinant of consumer's brand choice. The reversion of such an effect in the absence of reinforcements, however, received little attention. Stigler and Becker's (1977) conceptualization of consumption capital that depreciates over time is largely in line with both habit persistence and its reversion. More recently, Villas-Boas and Villas-Boas (2006) investigated the implication of consumers' learning and forgetting in sellers' decision to hold sales. The promotional effect in (iii) is similar to the informational effect of advertising.[3] Ward and Davis (1978) first recognized that a coupon may have value to a consumer beyond the price discount embedded in it, as the coupon itself serves as a tangible reminder of the product's availability.[4]

In this paper, I take Nevo and Wolfram's (2002) analysis of dynamic demand effect as a starting point and attempt to verify whether patterns consistent with the implications of these three assumptions are present in the household-level grocery purchase data. The econometric model I adopt is close to that used in Ackerberg (2001) to distinguish the informational effect and the prestige effect of advertising. While Ackerberg relied on the existence of a newly introduced product to separate consumers into the experienced group and the inexperienced group, I take advantage of the reversion property in (ii) and classify

[3] See Bagwell (2005) for a complete survey the advertising literature.

[4] Ward and Davis (1978), however, focused on coupons' impact on quantity sold and did not address the incidence of brand switching (changes to market share) resulting from coupon drops.

consumers in accordance to whether they recently purchased a brand. In addition, I adopt the framework outlined in Erdem, Keane and Sun (1999) to account for the unobserved coupon availability problem.

This paper is organized as follows. In Section 2, I describe the data set used for the analysis and report the results of preliminary analyses. In Section 3, I propose a model of consumers' brand choice reflecting the assumptions outlined above. In Section 6, I present the empirical results. Finally, in Section 5, I discuss the implications and limitation of the analysis.

1.2. Data

For the empirical analysis in this paper, I use the household-level purchases taken from the Stanford Basket data set[5]. The data are drawn from two separate metro markets in a large U.S. city and cover a two-year period from June 1991 to June 1993. For each shopping trip a sample of households made during this period, I know the store visited, the UPC of the product purchased, the number of units purchased and, most importantly, the type and the value of coupons redeemed. I construct a list of brands based on the product information database accompanying the purchase data set and translate consumers' product choices into brand choices by matching the UPCs. Among the 24 different categories present in the full data set, I focus on laundry detergents and ready-to-eat cereals, which are the two product categories with the most intense coupon activities. In total, the household panels include 13,200 purchases of laundry detergents

[5]The data were originally collected by Information Resources, Inc., a marketing research company in Chicago, using scanners in nine different supermarkets. See Bell and Lattin (1988) for the details on the construction of the data set.

and 35,089 purchases of cereals generated by 1,013 and 1,026 households respectively. I aggregate the purchase entries by household-time-brand, by replacing all purchase entries involving the same brand for a given household for a given day by a single entry involving a purchase of multiple units.[6] The purchase entries involving different brands, on the other hand, are treated as separate observations.

I supplement the household panel further with the weekly prices and promotional activities available at the store-level. A well recognized problem in scanner panel research is that only prices of products the consumers purchased are recorded. In general, the prices consumers face for alternatives that they did not purchase are not available. Researchers often extrapolate prices from nearby weeks to fill in the missing prices. The nature of this missing price problem is similar to that of unobserved coupon availability problem, which I model explicitly in this paper. Since the set of products for which weekly prices are available at the store-level far exceeds that present in the household panel, I abstract away from the missing price problem and simply assume that I observe the entire price vector consumers face. Because I model consumers' brand choice and because a brand typically consists of a number of UPCs with different package sizes, I have to aggregate the prices of different UPCs to construct the brand-level price index. The price measure I use is price-per-oz, computed as total revenue a brand generated in a store-week divided by total weight sold.

The coupons in the data set fall into two categories - manufacturers' coupons and store coupons. In this paper, I am primarily interested in the manufacturers' coupons

[6]We aggregate 1,409 purchase entries in the cereal category and 289 purchase entries in the laundry detergent category in this process.

Table 1.1. Household Summary Statistics

Laundry Detergents					
Number of brands identified	64				
Number of households present in purchase panel	1,013				
Proportion of purchases on manufacturers coupons	19%				
Proportion of purchases on store coupons	27%				
(Household)	Min	Max	Median	Mean	Std.Dev.
Number of shopping trips	1.00	119.00	10.00	13.03	12.29
Number of brands purchased	1.00	20.00	4.00	4.56	3.17
Brand HHI (weight)	0.08	1.00	0.43	0.50	0.28
Proportion of purchases on manufacturers coupons	0%	100%	5%	18%	24%
Proportion of purchases on all coupons	0%	100%	18%	28%	29%
Cereals					
Number of brands identified	191				
Number of households present in purchase panel	1,026				
Proportion of purchases on manufacturers coupons	23%				
Proportion of purchases on store coupons	39%				
(Household)	Min	Max	Median	Mean	Std.Dev.
Number of shopping trips	1.00	229.00	22.50	34.20	33.47
Number of brands purchased	1.00	58.00	9.00	11.48	8.70
Brand HHI (weight)	0.03	1.00	0.20	0.26	0.21
Proportion of purchases on manufacturers coupons	0%	100%	15%	20%	21%
Proportion of purchases on all coupons	0%	100%	35%	36%	26%

distributed via Sunday supplements[7] (henceforth, "the manufacturers' coupons"), which account for a majority of all coupons redeemed in these categories. Figure 1.1 shows the redemption patterns of different types of coupons over time. The store coupon redemptions are much more temporally concentrated than the manufacturers' coupon redemptions. I am unable to determine whether the consumers are required to clip and store physical coupons in order to take advantage of the discounts coded as store coupon redemptions[8], hence I limit the scope of our analysis. However, when I include all types of coupons in

[7]These are also known as free standing inserts (FSIs). For laundry detergents, this class of coupons account for approximately 70% of all redeemed coupons. The corresponding figure is approximately 60% for cereals.

[8]Ackerberg (2001), for example, argues that store coupons are typically available and announced at the point of purchase.

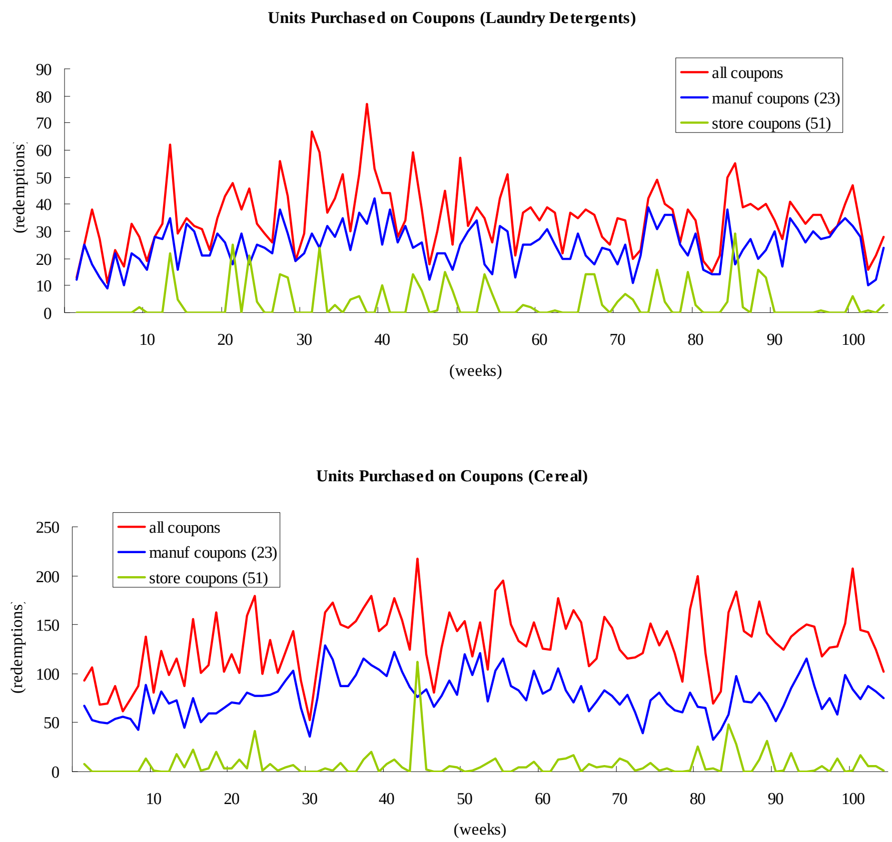

Figure 1.1. Redemption Patterns by Coupon Type

our analysis, the estimation results are qualitatively the same as what I present later in this paper. Summary statistics related to households' purchases are presented in Table 6. The consumers redeemed manufacturers' coupons on 19% of all laundry detergent purchases (24% of cereal purchases). The average face value of manufacturers' coupons redeemed is approximately 90 cents for laundry detergents and 70 cents for cereals.[9] A typical household made 13 laundry detergent purchases in 5 different brands (34 cereal purchases in 12 different brands) over the 104-week period.

[9]The statistics on coupon discount achieved are not reliable due to apparent miscoding. The figures presented are computed after screening out coupon values in excess of shelf prices.

A preliminary analysis suggests that the promotional effect described in Section 1 may be present in data. I first repeat the variance component analysis in Hendel and Nevo (2003) with coupon redemption in place of sale by estimating the following equations :

$$Q_{it} = \mu_1 + \kappa C_{it} + \xi_i + \omega_{it} \tag{1.1}$$

$$\overline{Q}_i = \mu_2 + \kappa \overline{C}_i + e_i \tag{1.2}$$

$$Q - \overline{Q}_i = \kappa(C_{it} - \overline{C}_i) + \nu_{it}, \quad \nu_{it} = \omega_{it} - \overline{\omega}_i \tag{1.3}$$

where Q_{it} are different quantity and duration measures for i's shopping occasion t, C_{it} is an indicator for whether the consumer purchased on sale[10] or redeemed a coupon for the transaction, ξ_i is a consumer-specific effect, ω_{it} is a disturbance term which is assumed to be uncorrelated with all other right-hand side variables and a bar denotes the average over a consumer's shopping occasions. Initially, I estimate (1.1) by OLS, assuming ξ_i is uncorrelated with C_{it}, to obtain the average without marketing instrument ($\widehat{\mu}_1$) and the total estimate for κ, which is inefficient but consistent if ξ_i and C_{it} are indeed uncorrelated. Then I estimate (1.1) by GLS obtaining the random effects estimate of κ, which is consistent and efficient provided that ξ_i is uncorrelated with C_{it}. I also estimate (1.2) by GLS to obtain between estimate of κ, using as the weight a diagonal matrix whose entries are inverse of the number of observations for each household. Finally, I estimate (1.3) to obtain the within estimate of κ, which is consistent even if ξ_i and C_{it} correlated but inefficient if they are not. There is, in fact, a good reason to expect that ξ_i is correlated with the consumer's coupon redemption. The empirical evidence presented in

[10] As in Hendel and Nevo (2003), we classify a transaction as "purchase on sale" if the price paid is at least 5% below the modal shelf price for the UPC at the same store over the 104 week period.

Narasimhan (1985)[11] show that coupon users tend to be more price elastic than nonusers of coupons, suggesting that some consumer-specific factors are potentially correlated with the consumer's coupon redemption decisions. I perform Hausman tests to verify whether the correlation assumption is justified.[12]

The estimation results are presented in Table 8. As documented in Hendel and Nevo (2003), retail sales have a pronounced effect on the quantity purchased and on the timing of the purchase. The patterns are consistent with consumer inventory behavior, the implication of which is discussed in detail in Hendel and Nevo (2006a, b). In contrast, the duration effect is nonexistent for coupon redemptions in both categories. Consumers tend to buy more cereal when they redeem coupons, possibly because coupons require purchase of multiple units, but such a quantity effect is not present in laundry detergent data.[13] The lack of duration effect is not unexpected because, while sales last only for a short spell of time, coupons usually remain effective for much longer than consumers' typical purchasing cycles. As a result, consumers do not have to deviate from their optimal purchasing cycles to take advantage of the discount embedded in coupons. This simple analysis demonstrates that the pattern in consumers' purchase behavior induced by coupons is different that induced by sales.

Next, using the same framework, I investigate the effect of coupons on consumers' brand choice. If the repeat purchase hypothesis in Section 1 is correct, coupons are more valuable to the consumers who did not purchase the brand recently due to the

[11] Also, Neslin (1990) states that the "deal proneness literature" describes the coupon users as being price sensitive / less brand loyal / venturesome. It is probably worth looking up a few such papers and cite in a footnote.

[12] For technical details of the variance component analysis, see Hsiao (1986).

[13] For an investigation of quantity effect of coupons, see Ward and Davis (1978), as well as Lee and Brown (1985).

promotional effect. As a result, the consumers who did not purchase the brand for a long period of time, in comparison to those who did purchase the brand recently, are more likely to purchase a brand when a coupon is available. Conditional on coupon redemption and after controlling for difference in repeat purchase rate across consumers, I should observe higher proportion of distant purchasers than I would if I conditioned on no coupon redemption. For each purchase occasion, I construct repeat purchase indicators which equals 1 if the consumer purchased the same brand within certain length of time in the past and 0 otherwise. The results presented in Table 8 show that coupon redemptions are associated with brand switching and that the effect is most significant when I use a narrow window to define consumers' recent purchases. As I extend the window and include the consumers who purchased the brand in distant past, the effect tends to diminish for both categories as predicted by the repeat purchase hypothesis. Sales do seem to have certain brand switching effects. In contrast to the brand switching effect of coupons, the brand switching effect of sales is the strongest when we adopt a broad windows to define consumers' recent purchases, implying that a sale tend to induce brand switching from those consumers who did not purchase the brand for a very long period of time.

These results are intuitive but hardly conclusive. There can be a number of problems that may complicate the interpretation of the estimated coefficients. In particular, I did not explicitly control for the variation in shelf prices which has an obvious implication on consumers' brand choice. On the other hand, if there is persistence in consumers' brand choices, the brand switching effect of coupons may be even more significant than this reduced form analysis may suggest. For further investigation, I now turn to the main economic model of the paper.

Table 1.2. Households' quantity and duration responses to marketing variables

Response to Coupons	Average	Difference with coupons				Hausman tests	
Variable	without coupons	Total	Within	Between	Random effects	Stat	p-value
Laundry Detergents							
Quantity (oz)	91.56 (0.90) ***	1.09 (2.02)	1.77 (2.09)	-0.22 (6.70)	1.82 (1.99)	0.01	0.94
Units	1.16 (0.01) ***	0.05 (0.02) *	0.04 (0.02) +	0.07 (0.05)	0.05 (0.02) *	0.44	0.51
Duration since last purchase	49.25 (0.75) ***	2.11 (1.68)	-0.76 (1.54)	7.72 (6.52)	-0.41 (1.52)	1.78	0.18
Duration till next purchase	42.01 (0.62) ***	1.33 (1.41)	-0.45 (1.39)	4.78 (4.90)	0.09 (1.36)	3.13	0.08 +
Cereals							
Quantity (oz)	22.51 (0.17) ***	7.45 (0.32) ***	6.82 (0.34) ***	9.06 (1.36) ***	6.97 (0.32) ***	2.86	0.09 +
Units	1.47 (0.01) ***	0.46 (0.02) ***	0.42 (0.02) ***	0.56 (0.08) ***	0.43 (0.02) ***	3.32	0.07 +
Duration since last purchase	29.16 (0.37) ***	-3.65 (0.71) ***	-0.36 (0.65)	-12.14 (3.80) **	-0.62 (0.65)	16.89	<0.01 ***
Duration till next purchase	25.35 (0.30) ***	-1.90 (0.59) **	0.09 (0.56)	-6.97 (2.86) *	-0.15 (0.56)	12.22	<0.01 ***

Response to Sales	Average	Difference during sale				Hausman tests	
Variable	during non-sale	Total	Within	Between	Random effects	Statistic	p-value
Laundry Detergents							
Quantity (oz)	80.11 (1.01) ***	30.54 (1.63) ***	21.93 (1.64) ***	51.72 (5.68) ***	23.73 (1.57) ***	14.88	<0.01 ***
Units	1.11 (0.01) ***	0.15 (0.02) ***	0.14 (0.02) ***	0.18 (0.05) ***	0.15 (0.02) ***	0.86	0.35
Duration since last purchase	47.33 (0.85) ***	6.11 (1.38) ***	-3.61 (1.22) **	30.03 (5.70) ***	-2.96 (1.21) *	11.25	<0.01 ***
Duration till next purchase	39.54 (0.70) ***	7.48 (1.16) ***	2.45 (1.11) *	19.37 (4.28) ***	2.99 (1.09) **	5.52	0.02 *
Cereals							
Quantity (oz)	23.03 (0.17) ***	5.79 (0.33) ***	5.10 (0.32) ***	8.75 (1.70) ***	5.22 (0.32) ***	3.09	0.08 +
Units	1.50 (0.01) ***	0.35 (0.02) ***	0.31 (0.02) ***	0.53 (0.10) ***	0.31 (0.02) ***	2.14	0.14
Duration since last purchase	28.22 (0.36) ***	-0.11 (0.72)	-1.18 (0.62) +	4.51 (4.76)	-1.16 (0.62) +	0.10	0.76
Duration till next purchase	24.55 (0.30) ***	1.16 (0.60) +	0.38 (0.54)	4.46 (3.56)	0.42 (0.54)	0.37	0.54

Note : Asterisks denote that the estimated coefficients are statistically significant at 99.9% (***), 99% (**), 95% (*) and 90% (+).

A purchase is classified as purchase on sale if the price paid is at least 5% below the modal shelf price for the UPC at the same store over the 104 week period. The estimated coefficients for laundry detergents do not necessarily coincide with those in Table 2 of Hendel and Nevo (2003) because the household panel was trimmed as described in Section 2.

Table 1.3. Households' brand choice responses to marketing variables

Response to Coupons	Average		Difference with coupons			Hausman tests	
Variable	without coupons	Total	Within	Between	Random effects	Stat	p-value
Laundry Detergents							
Purchased the brand within past 14 days	0.13 (0.00) ***	-0.07 (0.01) ***	-0.03 (0.01) ***	-0.15 (0.02) ***	-0.07 (0.01) ***	114.89	<0.01 ***
Purchased the brand within past 28 days	0.26 (0.00) ***	-0.10 (0.01) ***	-0.04 (0.01) ***	-0.22 (0.04) ***	-0.07 (0.01) ***	23.87	<0.01 ***
Purchased the brand within past 56 days	0.42 (0.01) ***	-0.09 (0.01) ***	-0.04 (0.01) **	-0.19 (0.04) ***	-0.05 (0.01) ***	3.69	0.05 +
Purchased the brand within past 84 days	0.50 (0.01) ***	-0.08 (0.01) ***	-0.03 (0.01) *	-0.19 (0.04) ***	-0.04 (0.01) **	4.56	0.03 *
Cereals							
Purchased the brand within past 14 days	0.14 (0.00) ***	-0.06 (0.00) ***	-0.03 (0.01) ***	-0.18 (0.02) ***	-0.04 (0.00) ***	43.97	<0.01 ***
Purchased the brand within past 28 days	0.27 (0.00) ***	-0.08 (0.01) ***	-0.03 (0.01) ***	-0.22 (0.03) ***	-0.04 (0.01) ***	9.47	<0.01 **
Purchased the brand within past 56 days	0.41 (0.00) ***	-0.05 (0.01) ***	-0.02 (0.01) **	-0.14 (0.04) ***	-0.02 (0.01) **	0.02	0.88
Purchased the brand within past 84 days	0.48 (0.00) ***	-0.02 (0.01) **	-0.01 (0.01)	-0.07 (0.04) +	-0.01 (0.01)	0.51	0.48
Response to Sales	Average		Difference during sale			Hausman tests	
Variable	during non-sale	Total	Within	Between	Random effects	Statistic	p-value
Laundry Detergents							
Purchased the brand within past 14 days	0.13 (0.00) ***	-0.03 (0.01) ***	-0.01 (0.01) +	-0.09 (0.02) ***	-0.03 (0.01) ***	80.02	<0.01 ***
Purchased the brand within past 28 days	0.27 (0.01) ***	-0.09 (0.01) ***	-0.03 (0.01) ***	-0.24 (0.03) ***	-0.06 (0.01) ***	54.58	<0.01 ***
Purchased the brand within past 56 days	0.44 (0.01) ***	-0.12 (0.01) ***	-0.04 (0.01) **	-0.33 (0.04) ***	-0.06 (0.01) ***	44.84	<0.01 ***
Purchased the brand within past 84 days	0.53 (0.01) ***	-0.12 (0.01) ***	-0.04 (0.01) ***	-0.34 (0.04) ***	-0.06 (0.01) ***	42.07	<0.01 ***
Cereals							
Purchased the brand within past 14 days	0.13 (0.00) ***	-0.04 (0.01) ***	-0.01 (0.01) *	-0.16 (0.03) ***	-0.02 (0.01) ***	46.20	<0.01 ***
Purchased the brand within past 28 days	0.27 (0.00) ***	-0.09 (0.01) ***	-0.03 (0.01) ***	-0.40 (0.04) ***	-0.04 (0.01) ***	95.47	<0.01 ***
Purchased the brand within past 56 days	0.43 (0.00) ***	-0.13 (0.01) ***	-0.04 (0.01) ***	-0.56 (0.05) ***	-0.06 (0.01) ***	107.90	<0.01 ***
Purchased the brand within past 84 days	0.51 (0.00) ***	-0.15 (0.01) ***	-0.06 (0.01) ***	-0.59 (0.04) ***	-0.07 (0.01) ***	105.87	<0.01 ***

Note : Asterisks denote that the estimated coefficients are statistically significant at 99.9% (***), 99% (**), 95% (*) and 90% (+).

A purchase is classified as purchase on sale if the price paid is at least 5% below the modal shelf price for the UPC at the same store over the 104 week period.

1.3. The Model

I model the brand choice decision a consumer makes after he walks into a store, s, with a portfolio of coupons, $c_{it} = (c_{i1t}, ..., c_{iJt})$. Each consumer is assumed to make a static decision taking the history of own brand choices and coupon redemptions as given. I specify consumer i's indirect utility from purchasing product j at shopping occasion t as follows[14] :

$$u_{ijt} = \alpha p_{jst} + x_j \beta_i + \gamma c_{ijt} + \delta y_{ijt} + \eta z_{jst} + \epsilon_{ijt} \tag{1.4}$$

where p_{jst} is the price of brand j in store s at time t, x_j is a vector of brand dummies, c_{ijt} is a vector of coupon variables, y_{ijt} is a vector of lagged purchase variables, z_{jst} is a vector of contemporaneous promotion variables such as display and feature and ϵ_{ijt} is an idiosyncratic utility shock that is observable to the consumer but not to the researcher.

The group of variables denoted c_{ijt} includes the current coupon availability and the coupon availability interacted with various measures of purchase experience. Its coefficient, γ, is the key to the analysis in this paper. I expect the estimated coefficients of the interaction terms to be negative and statistically significant, whereas we expect the contemporaneous effect of coupons on utility to be positive, if coupons indeed have differential promotional effects on the consumers who did and did not purchase the brand recently. I include the recent purchase indicators in y_{ijt} to capture the persistence in consumer's brand choice, which is one of the presumptions of the repeat purchase hypothesis.

[14]Since we do not model the timing of purchase, the set of shopping occasions is consumer specific. We denote the set of shopping occasions for consumer i as T_i.

If habit persistence is present in the data but the y_{ijt} term in (1.4) is omitted then the effect will be picked up by the interaction terms in c_{ijt} and bias our estimate of γ. I allow the coefficient β to depend on households' demographic characteristics such as income and family size in order to account for heterogeneity in consumers' taste.[15] Allowing a sufficient degree of heterogeneity is of particular importance due to the inclusion of past brand choices in our model. If I do not permit heterogeneity in our model, the permanent brand-consumer "fit" arising from the difference in preference across consumers will be reflected in the estimated coefficient for the lagged brand choice indicators, creating what Heckman (1981) describes as "spurious state dependence." In addition, I assume that the ϵ's are independently and identically distributed Type I Extreme Value random variables and describe the consumer's brand choice in terms of multinomial logit probabilities :

$$d_{ijt} = \begin{cases} 1 \quad if \quad u_{ijt} \geq u_{ikt} \quad \forall k \in J_{st} \\ 0 \qquad\qquad otherwise \end{cases}$$

$$\text{(1.5)} \qquad \Pr(d_{ijt} = 1 | p_{st}, c_{it}, y_{it}; \theta) = \frac{\exp(\alpha p_{jst} + x_j \beta_i + \gamma c_{ijt} + \delta y_{ijt} + \eta z_{jst})}{\sum_{k=1}^{J_{st}} \exp(\alpha p_{kst} + x_k \beta_i + \gamma c_{ikt} + \delta y_{ikt} + \eta z_{kst})}$$

where $p_{st} = (p_{1st}, ..., p_{J_{st}st})$, $c_{it} = (c_{i1t}, ..., c_{iJt})$, $y_{it} = (y_{i1t}, ..., y_{iJt})$ and $\theta = (\alpha, \beta, \gamma, \delta)$.

In this specifications, I let the consumer's most recent past purchase enter his indirect utility. In general, it is possible to specify an indirect utility that depends on the entire

[15] We can potentially do better by including household-specific coefficient and by specifying a parametric distribution. See Hartmann (2003) for a short discussion about implementing this using a method developed by Ackerberg (2001b) in the context of dynamic model. Ackerberg (2001a) simply integrates out the random term.

history of past brand choices and coupon redemptions. Guadagni and Little (1984) approximated such time dependence by introducing a loyalty variable which exponentially smooths past brand choices. In one of the specifications I estimate, I make the consumer's indirect utility a function of his brand choice in the previous category shopping occasion. I consider this utility specification as a limit case of Guadagni and Little (1984) where the smoothing parameter is set to 0.[16]

This specification is also similar to the indirect utility functions used in the advertising literature (e.g., Tellis (1988), Deighton, Henderson and Neslin (1994) and Ackerberg (2003)), in which advertising exposures are used in place of coupon availability of our model. The main difference lies in data observability. The typical data set adopted in the advertising literature includes the data generated by TV meters that keep track of households' advertising exposures. Such a data collection mechanism enables the researcher to observe each consumer's advertising exposure regardless whether the consumer eventually purchase the advertised brand or not. In contrast, the typical supermarket scanner data set includes coupon redemption instead of the set of coupons available to the consumer. In other words, I observe the c_{ijt} in our utility specification only for the brand the consumer purchases.[17] This observability problem gives rise to the classical missing data problem. Ignoring the unobserved coupon availability may create a self-selection bias in estimated coefficients. In order to obtain consistent estimates of the model coefficients in presence of

[16]This class of models is often referred to as the purchase event feedback type (Jones and Landwehr; 1988). Heckman (1983) proposed an alternative specification in which past utility enters directly in the random utility formulation. See Haaijer and Wedel (2001) for the comparison of the two.

[17]Some exceptions. Ward and Davis (1977) based on an experimentation in which coupons are distributed to a pre-determined group of consumers via direct mail. Hartmann (2003) used coupon as a source of price variation. The coupons are random distributed by emails / Hartmann obtained a record of coupon distribution and was able to identify the consumers who received the promotional emails.

unobserved coupon availability, I adopt the framework in Erdem, Keane and Sun (1999) and estimate this brand choice model jointly with a model for coupon process.

In order for a consumer to eventually use a coupon, the manufacturer has to issue the coupon and deliver it to the consumer and the customer has to clip and redeem the coupon that is made available to him. I simplify this coupon process using the following six assumptions :

Assumption 1 All coupons are issued at the beginning of each week and they expire at the end of each week.

Assumption 2 At most one type of coupon is available for each brand in each week.

Assumption 3 The availability of each coupon is exogenously determined and it is known to the researcher.

Assumption 4 All consumers have access to all available coupons in each week.

Assumption 5 Each consumer's decision to clip coupons in a particular week is characterized by an i.i.d. binomial random variable, $cclip_{it}$, with parameter λ. If a consumer decides to clip coupons in a week, he clips all available coupons.

Assumption 6 If a consumer chooses to clip coupons in a week, the consumer will use the coupon with certainty if he decides to purchase a product for which a coupon is available.

By adopting (Assumption 1), I assume away any intertemporal link in coupon availability and consumers' intertemporal decision with regard to coupon redemption. With (Assumption 1-3), I abstract away from manufacturers' strategic coupon decisions. In practice, I define the market-level coupon availability for time t as $cavail_t = (cavail_{1t}, ..., cavail_{Jt})$

where $cavail_{jt} = 1$ if there is at least one consumer who redeem a coupon for brand j at time t and $cavail_{jt} = 0$ otherwise. (Assumption 4) eliminates any randomness that may arise while the coupons are being delivered to the consumers. It also eliminates any heterogeneity in the set of coupons consumers face hence I obtain $cavail_{ijt} = cavail_{jt} \ \ \forall i, j, t$. I can describe this process as a common set of free standing inserts being delivered to each household at the beginning of each week. (Assumption 5) implies that, once a consumer takes delivery of the free standing inserts, he decides whether to go through this pile of free standing inserts and clip coupons by flipping a coin. In particular, I require that a consumer's coupon clipping decision is independent of the realized value of ϵ.[18] Under this assumption, I can describe the set of coupons a consumer has when he walks into a store as $c_{it} = (c_{i1t}, ..., c_{iJt})$ where $c_{ijt} = cclip_{it} \cdot cavail_{jt}$.[19] (Assumption 6) rules out a situation in which a consumer purchases a brand for which he has a coupon at hand yet he chooses not to redeem the coupon. I use this assumption to establish the following link between coupon redemption indicator in the data set and the set of coupons available to each consumer, which is unobservable to the researcher :

$$credmpt_{ijt} = c_{ijt} \cdot d_{ijt} \tag{1.6}$$

Given (Assumption 4-5), a consumer's coupon portfolio at t, c_{it}, is determined entirely by whether the consumer is a coupon clipper or not at t. Let $\pi_{ijt}^{cclip_{it}}$ be the probability of consumer i purchasing brand j on purchase occasion t, conditional on the product

[18]Alternatively, we may assume that the ϵ's are realized after the consumer makes the coupon clipping decision.

[19]Narasimhan (1994) argues that the decision to clip coupons is based on the tradeoff between costs of using coupons and the savings obtained. The implication of endogenous coupon clipping decision is left as a topic for discussion.

availability, the price vector and the set of coupons the consumer faces :

$$\begin{aligned}
(1.7) \qquad \pi^0_{ijt} &= \Pr(d_{ijt} = 1 | cclip_{ijt} = 0; \theta) \\
&= \frac{\exp(\alpha p_{jst} + x_j \beta_i + \delta y_{ijt} + \eta z_{jst})}{\sum_{k=1}^{J_{st}} \exp(\alpha p_{kst} + x_k \beta_i + \delta y_{ikt} + \eta z_{kst})} \\
\pi^1_{ijt} &= \Pr(d_{ijt} = 1 | cclip_{ijt} = 1; \theta) \\
&= \frac{\exp(\alpha p_{jst} + x_j \beta_i + \gamma c_{jt} + \delta y_{ijt} + \eta z_{jst})}{\sum_{k=1}^{J_{st}} \exp(\alpha p_{kst} + x_k \beta_i + \gamma c_{kt} + \delta y_{ikt} + \eta z_{kst})}
\end{aligned}$$

With these assumptions, if I observe consumer i redeeming a coupon when purchasing a brand at time t ($credmpt_{ijt} = 1$ for some $j \in J_{st}$), I can infer that the consumer is a coupon clipper at time t ($cclip_{it} = 1$). Likewise, if I observe consumer i purchasing without redeeming a coupon a brand at time t for which a coupon is available ($d_{ijt} = 1$ and $credmpt_{ijt} = 0$ for some $j \in J_{st}$ such that $cavail_{jt} = 1$), I can infer that the consumer is not a coupon clipper at time t ($cclip_{it} = 0$). If, on the other hand, I observe consumer i purchasing a brand at time t for which no coupon is available ($d_{ijt} = 1$ and $credmpt_{ijt} = 0$ for some $j \in J_{st}$ such that $cavail_{jt} = 0$), the consumer's choice does not give me any additional information about whether consumer i is a coupon clipper at time t or not.

The likelihood contribution of individual i is

$$\begin{aligned}
(1.8) \qquad L_i(\theta) &= \Pr((d_{ijt}, credmpt_{ijt})_{j,t} | (p_{st}, cavail_t)_{s,t}; \theta) \\
&= \sum_{t \in T_i} \sum_{j \in J_{st}} [\lambda (1 - (1 - credmpt_{ijt}) \cdot cavail_{jt}) \cdot d_{ijt} \cdot \pi^1_{ijt} \\
&\quad + (1 - \lambda)(1 - credmpt_{ijt}) \cdot d_{ijt} \cdot \pi^0_{ijt}]
\end{aligned}$$

where T_i is the set of time periods at which consumer i shops. I estimate θ by maximizing the log likelihood function :

$$L(\theta) = \sum_{i \in I} \log L_i(\theta) \tag{1.9}$$

1.4. Estimation Results

The large number of brands identified in these categories poses a computational challenge as I intend to allow heterogeneity in consumers' preference for each brand. The approach I take involves grouping the households in accordance to the family size (3 groups) and income (5 groups) and estimating coefficients for the interaction between brand dummies and the demographic characteristic dummies. As a result, the total number of β's I have to estimate is 7 times the number of brands. In order to limit the computing time, I model consumers' choice over top five brands in each category, aggregating all the other brands into a hypothetical "outside good." I use data on ten largest brands included in the outside good to compute promotion variables, such as brand price, feature and display. The top 15 brands used in the analysis are presented in Table 1.4.

This approach relies on the assumption that family size and income sufficiently explain the heterogeneity in preference across households. In other words, the utility household i generates from physical characteristics of product j can be described as :

$$\beta_{ij} = \beta_j^{FS} FS_i + \beta_j^{I} I_i + \zeta_{ij} \tag{1.10}$$

Table 1.4. Top 15 Brands

	Laundry Detergents				
	Brand	Type	Company	Share (weight)	Prop Sold on Coupons
1	TIDE	LIQUID	PROCTER & GAMBLE	15.2%	27.8%
2	TIDE	POWDER	PROCTER & GAMBLE	11.3%	24.8%
3	ALL	LIQUID	UNILEVER	10.7%	11.4%
4	WISK	LIQUID	UNILEVER	8.1%	25.6%
5	SOLO	LIQUID	PROCTER & GAMBLE	7.2%	14.9%
6	PUREX	LIQUID	THE DIAL CORPORATION	6.4%	12.6%
7	ARM & HAMMER	POWDER	CHURCH & DWIGHT CO INC	4.2%	23.0%
8	CHEER	LIQUID	PROCTER & GAMBLE	3.3%	26.5%
9	ARM & HAMMER	LIQUID	CHURCH & DWIGHT CO INC	3.2%	7.2%
10	AJAX	LIQUID	COLGATE PALMOLIVE	3.2%	1.5%
11	CHEER	POWDER	PROCTER & GAMBLE	3.0%	23.7%
12	YES	LIQUID	DOW CHEMICAL CO	2.9%	4.7%
13	SURF	LIQUID	UNILEVER	2.9%	39.2%
14	ERA	LIQUID	PROCTER & GAMBLE	2.7%	15.2%
15	ALL	POWDER	UNILEVER	1.5%	28.1%
	Total Number of Brands				64
	HHI (weight)				0.072
	5-Brand Concentration Ratio				0.525
	15-Brand Concentration Ratio				0.856
	Cereals				
	Brand	Type	Company	Share (weight)	Prop Sold on Coupons
1	KLLGGS FROSTED FLAKE	PRESWEETENED	KELLOGG CO	6.2%	15.3%
2	KLLGGS CORN FLAKES	REGULAR	KELLOGG CO	5.6%	16.3%
3	GENRL MLLS CHEERIOS	REGULAR	GENERAL MILLS	3.8%	26.8%
4	KLLGGS RAISIN BRAN	REGULAR	KELLOGG CO	3.5%	19.5%
5	POST GRAPE NUTS	REGULAR	PHILIP MORRIS CO INC	3.1%	27.8%
6	GENRL MLLS HNY NT CH	PRESWEETENED	GENERAL MILLS	3.0%	28.9%
7	QUAKER 100% NATURAL	REGULAR	QUAKER OATS COMPANY	3.0%	34.5%
8	QUAKER CAP N CRUNCH	PRESWEETENED	QUAKER OATS COMPANY	2.6%	21.4%
9	KLLGGS RICE KRISPIES	REGULAR	KELLOGG CO	2.5%	17.4%
10	KLLGGS FROSTED MINI	PRESWEETENED	KELLOGG CO	2.3%	20.5%
11	KLLGGS FROOT LOOPS	PRESWEETENED	KELLOGG CO	1.7%	19.7%
12	POST RAISIN BRAN	REGULAR	PHILIP MORRIS CO INC	1.6%	32.3%
13	KLLGGS SPECIAL K	REGULAR	KELLOGG CO	1.5%	24.9%
14	GENRL MLLS WHEATIES	REGULAR	GENERAL MILLS	1.5%	25.1%
15	KLLGGS MUESLIX	REGULAR	KELLOGG CO	1.5%	30.6%
	Total Number of Brands				191
	HHI (weight)				0.020
	5-Brand Concentration Ratio				0.222
	15-Brand Concentration Ratio				0.433

where FS_i and I_i are vectors of dummy variables describing household i's demographic characteristics and ζ_{ij} is household-specific preference for brand j which is unexplained by the family size and income and which is assumed to be negligible. It is feasible to refine the analysis by specifying a joint distribution for $\zeta_i = (\zeta_{i1}, ..., \zeta_{iJ})$, denoted $f(d\zeta_i|\theta)$, and integrate out, possibly by simulation method, to form the likelihood contribution of each household in (1.8). For now, I leave this as a future refinement of the paper.

The estimated coefficients of the model in Section 4 are reported in Table 1.4. In columns (1) through (4), I used different classifications to indicate whether a consumer is a recent purchaser (whether the consumer purchased the brand within 14, 28, 56 and 84 days in the past). In column (5), consumers are classified in accordance to whether they purchased the same brand in the previous category shopping occasion. In order to compare the estimated coefficients in the first four columns, I estimated another specification in which all recent purchase indicators used in columns (1) through (4) are used together. The results are presented in column (6).

Taking another look at Table 1.4 in Section 3, I find that the second past purchase indicator gives us the clearest contrast in behavior between two groups of consumers, hence I take the second column of Table 1.4 as the base case. The coefficient on recent purchase experience (δ_2) is positive and significant, indicating strong habit persistence in consumers' brand choice. The first coupon coefficient, denoted γ_0, is the value of a coupon to the consumers who did not purchase the brand within 28 days in the past. This should, according to the implication of the repeat purchase hypothesis, reflect both the price discount effect and the promotional effect of the coupon. Given the magnitude of the price coefficient, α, this translates to \$3.72 per 96 oz package, much higher than the average face value of coupons for laundry detergents, which is approximately 90 cents. The second coupon coefficient, denoted γ_2, is the differential value of a coupon to the consumers who recently purchased the brand. A negative coefficient implies that coupons

Table 1.5. Estimation Results

		(1)	(2)	(3)	(4)	(5)	(6)
Prob. of Coupon Clipping	lambda	0.203 ***	0.203 ***	0.205 ***	0.206 ***	0.205 ***	0.207 ***
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Price	alpha	-0.302 ***	-0.298 ***	-0.316 ***	-0.307 ***	-0.328 ***	-0.309 ***
		(0.011)	(0.011)	(0.011)	(0.011)	(0.012)	(0.011)
Coupon	gamma	1.170 **	1.154 **	0.924 **	0.877 **	1.188 ***	0.820 **
		(0.428)	(0.385)	(0.312)	(0.282)	(0.353)	(0.276)
Coupon*Experience (14 days)	gamma1	-0.930 ***					-0.391 *
		(0.128)					(0.171)
Coupon*Experience (28 days)	gamma2		-0.592 ***				-0.134
			(0.087)				(0.139)
Coupon*Experience (56 days)	gamma3			-0.410 ***			-0.117
				(0.076)			(0.146)
Coupon*Experience (84 days)	gamma4				-0.455 ***		-0.197
					(0.075)		(0.126)
Coupon*Experience (previous)	gamma5					-0.366 ***	
						(0.059)	
Experience (14 days)	delta1	1.784 ***					0.272 ***
		(0.061)					(0.079)
Experience (28 days)	delta2		1.852 ***				0.385 ***
			(0.046)				(0.070)
Experience (56 days)	delta3			1.964 ***			0.405 ***
				(0.041)			(0.072)
Experience (84 days)	delta4				1.983 ***		1.460 ***
					(0.041)		(0.063)
Experience (previous)	delta5					1.732 ***	
						(0.032)	
Past Coupon (14 days)	zeta1	-0.032					-0.231
		(0.122)					(0.162)
Past Coupon (28 days)	zeta2		-0.053				-0.261 *
			(0.080)				(0.124)
Past Coupon (56 days)	zeta3			0.043			0.055
				(0.059)			(0.108)
Past Coupon (84 days)	zeta4				0.275 ***		0.277 ***
					(0.052)		(0.084)
Past Coupon (previous)	zeta5					-0.006	
						(0.059)	
Display	eta1	0.341 ***	0.366 ***	0.393 ***	0.398 ***	0.373 ***	0.396 ***
		(0.034)	(0.035)	(0.036)	(0.037)	(0.037)	(0.037)
Feature	eta2	0.344 ***	0.380 ***	0.378 ***	0.393 ***	0.403 ***	0.403 ***
		(0.035)	(0.036)	(0.037)	(0.037)	(0.038)	(0.038)
logL		-15193.9	-14606.7	-13896.4	-13600.8	-13652.4	-13496.4
Algorithm		Nelder-Mead					

Note : Asterisks denote that the estimated coefficients are statistically significant at 99.9% (***), 99% (**), 95% (*) and 90% (+).

are less valuable to this group of consumers, which is consistent with the implication of the repeat purchase hypothesis. To these consumers, a coupon is worth $\gamma_0 + \gamma_2$, which translates to roughly $1.81 per 96 oz package. Moving from column (1) to column (4), as I expand the time window to define a recent purchase, the magnitude of the promotional effect is decreasing as predicted by the repeat purchase hypothesis.

There can be a few explanations for why the estimated value of a coupon even to the recent purchasers is much higher than the face value of a coupon. First, there is so called "smart shopper" phenomenon.[20] Consumers may derive additional satisfaction from taking advantage of an exclusive offer. Second, a coupon communicates the existence of a deal much more clearly than "the price displayed is 50 cents off the regular price," displayed on the shelf. In this regard, the estimated price coefficient may underestimate consumers' sensitivity to a price discount and inflate the value of the coupon we calculate. Finally, the price measure used is an average of shelf prices of different package sizes. Since, when a brand is on sale, typically only a subset of the associated UPCs is on sale. As a result, to induce 1 cent reduction in the average price, the price of the UPC on sale must fall by more than 1 cent. Conversion using the price coefficient based on the average price may inflate the implied value of a coupon.

In column (6), δ_4 indicates the current utility value of consumption experience for all consumers who purchased the brand within 84 days in the past. δ_3, δ_2 and δ_1 are the incremental utility those consumers who purchased the brand more recently (54, 28 and 14 days in the past) derive from their consumption experience. Positive and significant coefficients indicate that the influence of the consumption experience is the strongest to

[20]See Blattberg and Neslin (1990).

the consumers who purchased the brand most recently and that this effect declines as time passes. The γ's can be interpreted analogously. γ_4 in column (6) being statistically insignificant implies that the value of a coupon to the consumers who purchased a brand in the past 84 days is hardly different from the value of the same coupon to those who did not purchase a brand in the past 84 days. On the other hand, γ_1 is negative and statistically significant, which implies that the value of the coupon to those who purchased the band very recently is substantially lower.

Contemporaneous feature and display also have positive and significant effects, while past coupon redemptions do not have any meaningful impact on a consumer's current brand choice. Blattberg and Neslin (1990) suggested the possibility that consumers may attribute their purchase to the availability of the coupon rather than to the intrinsic merits of the brand. In this vein, the promotion enhancement literature[21] argues that promotions reduce brand loyalty. However, such an effect is not present in the data set we examine.

1.5. Discussion

In this paper, I demonstrated that coupons have additional promotional effects on those consumers who did not purchase the brand recently. I also verified that the positive intertemporal effects of consumption experience on consumers' subsequent brand choices tend to decline over time. While these findings are consistent with the implication of the repeat purchase hypothesis, they do not completely rule out the price discrimination hypothesis. They do suggest, however, that the sellers of a mature product category

[21] See Bridges, Briesch and Yim (2006) for a short survey.

may use coupons to remind consumers of the existence and the characteristics of their products.

This analysis has an immediate implication on the profitability of coupon promotions.[22] The presence of habit persistence implies that the calculation based on static demand may underestimate the true profitability of coupons, while the reversion of habit persistence implies that the sellers can improve profitability of coupon promotion by appropriately targeting the consumers.[23]

The primary limitation of this analysis stems from the strong assumptions imposed on the coupon process. In particular, the independence between consumer's coupon clipping decision and his brand choice seems demanding given the conclusions drawn in previous empirical researches highlighting the correlation between consumers' demographic profile and their tendency to use coupons.[24] It seems feasible to allow each consumer's coupon clipping probability to depend on the consumer's demographic profile, D_i, and potentially on unobserved consumer-specific fixed effects, ξ_i^d, :

$$\lambda_i = \frac{\exp(D_i\beta^d + \xi_i^d)}{1 + \exp(D_i\beta^d + \xi_i^d)}$$

where D_i may include FS_i and I_i, that we used to control for heterogeneity in the consumer's preference.[25] As long as ϵ's are assumed to be independent of other observed and unobserved covariates, the form of conditional choice probabilities in (1.7) can be maintained. By specifying a joint distribution for (ξ_i^d, ζ_i), I can still integrate out consumer-specific fixed effects and obtain the likelihood contribution in (1.8).

[22] See Blattberg and Neslin (1990) and Neslin (1990).
[23] This is also relevant to the timing of coupon drops.
[24] Teal, Williams and Bearden (1980) and Narasimhan (1984) to name a few.
[25] This specification is motivated by Ellison (1994).

The endogeneity of price is also a concern. While there are reasons to believe that the negative correlation between the shelf price and coupon availability Nevo and Wolfram (2002) found in aggregate quarterly data may have a limited implication on the high frequency data I use in this analysis[26], it seems feasible to take advantage of the fact that the data are drawn from two separate sub-markets and construct Nevo-style (2000) price instruments.

Finally, I assumed that past consumption experience has an impact on consumer's current indirect utility I but did not specify what may give rise to such an intertemporal effect. Further inference may be possible if I put this analysis in the context of brand loyalty or learning and forgetting.

[26]The top brand, for example, had coupons available in almost all weeks.

CHAPTER 2

The Impact of Coupon Availability on Retailers' Decision to Hold Sale

2.1. Introduction

Economists traditionally regarded retail coupons as a tool to price discriminate different types of consumers. If consumers are heterogeneous in terms of their price elasticities of demand and if consumer types are not directly observable, a monopoly seller may set the shelf price high and distribute coupons to allow the consumers to self-select in accordance to their willingness to clip and redeem coupons. To the extent that the consumers' tendency to use coupons is positively correlated with their price elasticities, the monopoly seller can extracting more surplus from the consumers than he would have had he set a uniform price. Narasimhan (1984) proposed a consumer model and showed that when individuals face a trade-off between costs of using coupons and the fixed savings from coupon redemption, coupon usage would indeed separate consumers in the way that makes price discrimination profitable.

One important implication of the static price discrimination theory is that the observed retail price should be higher in periods when coupons are made available to consumers, controlling for other factors that could potentially influence the seller's pricing decision. In contrast to this prediction, the empirical analysis of Nevo and Wolfram (2001) showed that the weighted average shelf prices for ready-to-eat cereals are negatively correlated with coupon availability, suggesting that the static theory comes short of explaining the role coupons may play in the market. Due to the data aggregation, Nevo and Wolfram did not elaborate on the source of low shelf price associated with the coupon availability.

In this paper, I examine the impact of coupon availability on retailers' decision to hold sale using weekly store-level data. Temporary price reductions, or sales, are widely observed in retail price data across many product categories. Researchers pointed out that sales occur with sufficient regularity that it is hard to attribute their occurrence to purely random realizations of sellers' costs or demand. Some researchers[1] proposed a static model in which competing sellers randomize in equilibrium, leading to a dispersion in prices over time. Some others[2] suggested models based on dynamic demand in which sellers can enhance profit by periodically cut price to clear latent demand, whose measure increase over time. With Sobel-type dynamic demand model in mind, I attempt to verify whether coupon availability has any power in explaining retailers' sale decision controlling for accumulation of latent demand that motivate temporary price reductions in this class of models. By restating the price data to a simpler binary indicator for sale I do lose useful variations contained in shelf prices. This approach, on the other hand, enables me to focus on an important source of price variation that is different in characteristic from the minor fluctuations in what retailers post as regular prices.

The results suggest that the dynamic demand effects are quite complex potentially due to the presence of many different types of latent consumers and that coupon availability does seem to have an influence over retailers' sale decision even after accounting for accumulation of demand over time. This exercise is most useful in the sense it helps

[1]Shilony (1977), Rosenthal (1980) and Varian (1980) to name a few.
[2]In particular, Conlisk, Gerstner and Sobel (1984) and Sobel (1984).

to identify a niche area where a theoretical improvement can make a contribution. An extension of model for retail sales in which a multi-product seller periodically offers a subset of available products at discount may generate testable implications where coupon availability can be used as a source of observable demand shocks.

2.2. Data

The empirical analysis of this paper is based on the data taken from the Stanford Basket data set[3]. The data are collected from two separate metro-markets in a large U.S. city and cover a two-year period from June 1991 to June 1993. The two components of the data set I used extensively in this paper are the store-level sale data and the household-level purchase data. The store-level data consist of weekly shelf price and number of units sold for each product in each store, identified by UPC. The store-level data also include information on whether the product was on special in-store display and whether the product was featured in the store's fliers in each week. The entries in the store data are matched with the entries in a separate product information database using UPCs to obtain such details as company name, brand name, product type and package size. The household-level data consist of a list of purchases individual households made during the two-year period. For each shopping trip a sample of households made, I know the store visited, the UPC of products purchased, the number of units purchased and, most importantly, the type and the value of coupons redeemed. Among the 24 product categories present in the data set, I focus on regular ready-to-eat cereals, which is one of the categories with the most intense coupon activities.

[3]See Bell and Lattin (1988) for the details on the construction of the data set.

A brand in my analysis is a set of products that share the common physical characteristics. Each brand consists of multiple UPCs that differ only in terms of package size. Typically, cents off coupons are made available at the brand level and they can be redeemed for purchase of any package size. In some cases, consumers are required to purchase particular package size in order to take advantage of the discount offered through coupons. Information on such restrictions is not available in the data set used. In deriving the coupon availability from consumers' coupon redemption data, I simply assume that the observed coupons are valid for any package size of a brand and that each consumers' package size choice conditional on brand choice is independent of the consumer's decision to redeem coupons.

To construct brand-level data, I need to aggregate the price and quantity variables across different package sizes. The derivation of brand-level price information is discussed in Section 2.1. A number of different quantity measures can be constructed based on the number of purchases, the number of units sold and total weight sold. As different products that constitute a brand share the common physical characteristics, it is natural to aggregate quantity using package size (weight in oz) of each product. The market shares, on the other hand, are defined in terms of total revenue (prevailing price for each package size multiplied by total number of units sold).

The data set in question is extremely detailed and it is a challenge to aggregate the data set to a manageable level. The cereal category contains 35,089 purchases of cereals by 1,026 households. At the store level, a total of 147,682 entries are available for the cereal category, spanning 534 UPCs sold at 9 different stores. Even after brand-level aggregation,

Table 2.1. Major Brands and Their Market Shares

Serial No.	Brand	Company	Unit Share	Qty Share	Rev Share
11	GENRL MLLS CHEERIOS	GENERAL MILLS	4.27%	3.82%	4.17%
2	KLLGGS CORN FLAKES	KELLOGG CO	5.05%	5.58%	3.21%
3	KLLGGS RAISIN BRAN	KELLOGG CO	2.62%	3.50%	2.81%
4	KLLGGS RICE KRISPIES	KELLOGG CO	2.92%	2.54%	2.48%
21	QUAKER 100% NATURAL	QUAKER OATS CO	2.05%	2.97%	2.45%
18	POST GRAPE NUTS	PHILIP MORRIS CO INC	2.40%	3.10%	2.35%
7	KLLGGS SPECIAL K	KELLOGG CO	1.69%	1.52%	1.82%
8	KLLGGS MUESLIX	KELLOGG CO	1.54%	1.50%	1.81%

Note : All market shares are calculated as proportion to total units sold, total quantity sold and total revenue generated for the cereal category.

the cereal market is very much fragmented with HHI equal to 0.017 [4]. For the analysis, I trimmed the data to include the only the brands with market share exceeding 1.5%[5]. This reduces the product space for regular cereals down to 8 major brands that account for 37.3% of regular cereal market[6] in these two sub-markets. A list of major brands used for analysis can be found in Table 26. General Mill's Cheerio is the leading brand in this sub-category with 4.17% market share, followed by Kellogg's Corn Flakes, which has 3.21% market share.

2.2.1. Construction of Regular Prices and Sale Indicators

Previous researches on the patterns of retail shelf prices commonly recognized that the retail price tends to stay at the same level over an extended period of time followed by a temporary but significant price reduction, commonly known as a "sale." The same pattern is present in the price path of a popular cereal product depicted in Figure 2.1[7].

[4]The HHI is 0.028 if we restrict to regular cereal sub-category.

[5]The revenue share for entire cereal category.

[6]21.1% of total cereal market.

[7]Figure 2.1 depicts the evolution of the weekly shelf price of General Mills' Cheerios 15-oz package in store 1420. Cheerios has the highest market share among 114 brands of regular ready-to-eat cereals present in

In order to formalize the notion of sale, I first need to define what non-sale or "regular" price is. Eyeballing through the data, I recognize that there is general appreciation in what appears to be non-sale price over time, probably due to inflation. Different researchers used different methods to construct regular price series and each method has advantages and disadvantages. Using a single unadjusted modal price for each UPC tends to erroneously indicate the earlier observations as being on sale. Using inflation-adjusted shelf prices tends to complicate the computation of the modal price as appreciations of individual product's shelf price do not necessarily coincide with the frequency at which price indices are available.

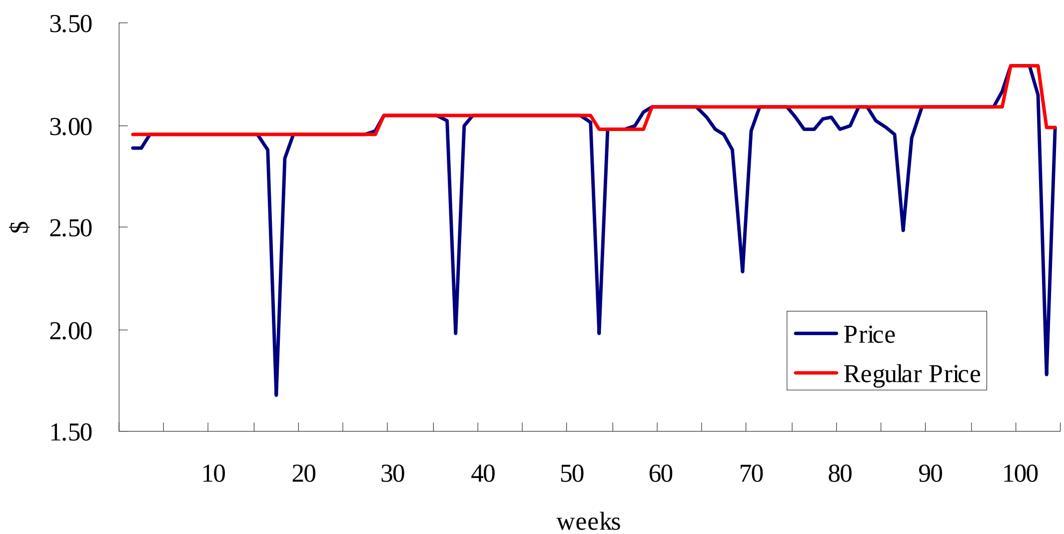

Figure 2.1. Price Path for General Mills' Cheerios in Store 1420

To avoid these problems, I use the following algorithm to construct the regular price series. First, I compile the raw weekly shelf price series for each UPC at each store. At

the data set, with the 15-oz package being the most popular item. Store 1420 also has the highest market share among 9 retail stores present in the data set.

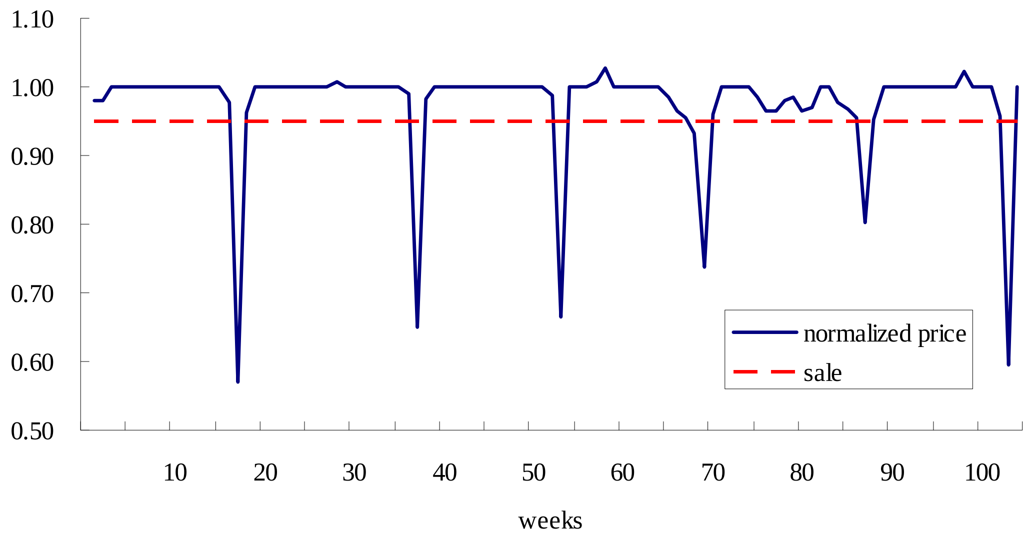

Figure 2.2. Evolution of Normalized Price

week t, assuming that I already have well defined regular price for weeks $t+1$ and beyond, I examine the shelf prices for a fixed length window leading up to week t. If all prices in this window are uniformly below (uniformly above) the regular price for $t+1$, I recognize that the regular price has increased (decreased) to a new level and begin computing a new modal price. Starting from the end of the data period, this process is repeated until I reach the beginning of the price data. For the initial regular price, I use the maximum of raw weekly shelf price series. The regular price thus constructed for the sample product in Figure 2.1 is represented as the solid red line.

Once I define the regular price, I derive the normalized shelf price for each UPC as the proportion of the weekly shelf price to the prevailing regular price and, finally, indicate the product as being on sale if the normalized price falls below a fixed threshold. In the analysis, I chose 5% threshold to construct sale indicators to take into account the

Table 2.2. Summary Stats of Key Variables

	Obs	Min	Max	Median	Mean	Std
Weighted Average Price (15 oz)	6468	0.5515	4.1427	2.5839	2.6866	0.6812
Normalized Price	6468	0.2276	1.0373	1.0000	0.9764	0.0944
Sale	6468	0.0000	1.0000	0.0000	0.0748	0.2631
Coupon Availability	6468	0.0000	1.0000	0.0000	0.3063	0.4610
Coupon Redemption	6468	0.0000	11.0000	1.0000	1.1620	1.3610
Feature	6468	0.0000	1.0000	0.0000	0.0606	0.2386
Display	6468	0.0000	1.0000	0.0000	0.0951	0.2934
Duration (store-brand)	4679	1.0000	82.0000	9.0000	13.1800	13.0252
Duration (store)	6260	1.0000	16.0000	3.0000	3.9669	2.8065
Duration (brand-market)	5679	1.0000	37.0000	5.0000	6.8924	5.7463

fact that the variation in the normalized price, if it rises above 100%, is limited to 5% range. Changing the sale thresholds to 10, 15 or 20% made little difference[8]. I define the brand-level normalized price for each store-week as the minimum of the normalized prices for the associated UPCs. I define a sale for brand analogously. Effectively, a brand is indicated as being on sale if any one of the associated UPCs is on sale in a given store-week. This construction reflects the implicit assumption that different packages within a brand classification are near perfect substitutes controlling for package-specific fixed effects.

The summary statistics of sale indicators are presented in Table 27. Among the 6,468 observations in the data set, approximately 7.5% are classified as being on sale. Note that other forms of promotional activities, such as display and feature, are closely related to temporary price reductions. The frequency of display and feature are approximately the same as the frequency for sales. The correlation between sale and display and the correlation between sale and feature are 0.50 and 0.66 respectively.

[8]See Figure 2.2 for normalized price and sale threshold for General Mills' Cheerios at Store 1420.

2.2.2. Coupon Availability

There are largely two different types of coupons present in the data set : manufacturers' coupons and store coupons. Figure 1.1 shows the redemption patterns of different types of coupons over time. The store coupon redemptions are much more temporally concentrated than the manufacturers' coupon redemptions. There is no information available to determine whether the promotions recorded as redemptions of store coupons actually require clipping of physical coupons. As I am primarily interested in the interaction between the manufacturers and the retailers, I focus on the manufacturers' coupons distributed via Sunday newspaper supplements (henceforth, "the manufacturers' coupons"), which also account for a majority of all coupons redeemed[9].

Since the data set does not contain explicit information on whether manufacturers made coupons available for particular brands at particular time period, I have to derive coupon availability from consumers' coupon redemption activities. I indicate that coupon for a brand is available in a sub-market in a week if the number of coupon redemptions for the brand in that sub-market in the week exceeds a certain threshold[10]. This is a noisy measure of coupon availability and it contrasts the data used in Nevo and Wolfram (2001), where they supplemented the consumer data with coupon issue data collected by Promotion Information Management (PIM). In the analysis, I assume that the process with which coupons become available is exogenous to the retailers' sale decision. The implication of the use of this measure is discussed briefly in Section 4.

[9]These are also known as free standing inserts (FSIs). This class of coupons account for approximately 60% of all redeemed coupons for cereal category.
[10]The threshold used is 2 redemptions per brand per week per market.

The summary statistics for coupon availability and coupon redemptions is presented in Table 2.2. In the trimmed data set, coupons are available for approximately 30% of brand-store-weeks.

2.3. Empirical Analysis

Nevo and Wolfram (2001) showed that, under a fairly general set of conditions, a retailer's profit maximizing shelf price should be higher when there are coupons available than the uniform price it would set in the absence of coupons, when coupons are used as a price discrimination device[11]. I express the retailer's static price decision as a function of coupon availability, *cavail*, and a set of other factors, x:

$$np_{bst} = f(cavail_{bt}, x_{bst}) \tag{2.1}$$

where np is the normalized price and the subscripts, b, s and t indicate brand, store and time period respectively. The econometric model of interest is

$$np_{bst} = \beta_{np} cavail_{bt} + \delta_{np} z_{bst} + \gamma_{bst} + \eta_{bst} \tag{2.2}$$

where z is a set of observable state variables and γ is a set of fixed effects that capture the influence of unobservable factors over retailer's price decision. It is possible that the issuers of coupons, the manufacturers, condition their coupon decisions on factors that are correlated with the retailers' state variables. Provided that the fixed effects sufficiently control for the variations in unobserved state variables, leaving η uncorrelated

[11]See Proposition 1 in Nevo and Wolfram (2001)

with $cavail$, I can estimate the coefficient β_{np} consistently using OLS. The static monopoly price discrimination theory of coupons predict that the estimated β_{np} be positive.

I start by replicating the main result in Nevo and Wolfram (2001) using the weekly shelf price data in the data set. The empirical model used is

$$np_{bst} = \beta_{np} cavail_{bt} + \gamma_{bs} + \gamma_{st} + \gamma_{tb} + \eta_{bst} \tag{2.3}$$

where the γ's are various fixed effects. The specification differs from equation 1 in Nevo and Wolfram only in terms of the nature of variables used and in terms of frequency of observations. The use of normalized shelf price enables me to focus purely on the retailers' pricing decision, purging the effect of consumers' package choice that influences the weighted average shelf price. On the other hand, the derivation of coupon availability from observed coupon redemptions creates a noise in the explanatory variables and can potentially complicate the interpretation of the estimated coefficients. I also repeat the same regression using $crdmpt_{bmt}$ in place of $cavail_{bst}$. $crdmpt_{bmt}$ measures the number of coupon redemptions for brand-market-week and I use this as a proxy for how widely the coupons for the brand was available to the consumers, albeit the obvious limitation that this measure is derived from consumers' coupon redemptions and, unless there is a reason to believe that coupon redemptions occur at random, the measure is influenced by consumers' coupon redemption decisions.

The results are presented in Table 29. The estimates in different columns vary because of different sets of fixed effects used. The estimated coefficients indicate that there is a negative correlation between the shelf price and coupon availability, which is in line with the results in Nevo and Wolfram. The coefficients for the lagged coupon availability are

Table 2.3. Impact of Coupon Availability on Normalized Price

	[1]	[2]	[3]	[4]	[5]
Coupon Availability(t)	-0.0151 ***	-0.0091 ***	-0.0091 ***	-0.0105 ***	-0.0049
	(0.0025)	(0.0027)	(0.0027)	(0.0028)	(0.0040)
adjusted R-squqred	0.0053	0.0566	0.0539	0.0473	0.2375
Coupon Redemption(t)	-0.0070 ***	-0.0052 ***	-0.0053 ***	-0.0057 ***	-0.0041 **
	(0.0009)	(0.0010)	(0.0010)	(0.0010)	(0.0014)
adjusted R-squqred	0.0100	0.0593	0.0567	0.0505	0.2385
Number of Observations	6,468	6,468	6,468	6,468	6,468
Controls	constant	brand, store and week	brand-store and week	brand-store and store-week	brand-store, store-week and brand-week

Note : Asterisks denote that the estimated coefficients are statistically significant at 99.9% (***), 99% (**), 95% (*) and 90% (+).

Table 2.3 Impact of Coupon Availability on Normalized Price (continued)

	[6]	[7]	[8]	[9]	[10]
Coupon Availability(t)	-0.0143 ***	-0.0107 ***	-0.0108 ***	-0.0122 ***	-0.0055
	(0.0026)	(0.0028)	(0.0028)	(0.0029)	(0.0041)
Coupon Availability(t-1)	-0.0030	0.0001	0.0000	-0.0006	0.0015
	(0.0026)	(0.0028)	(0.0028)	(0.0029)	(0.0041)
Coupon Availability(t-2)	-0.0053 *	-0.0016	-0.0018	-0.0019	0.0011
	(0.0026)	(0.0028)	(0.0028)	(0.0029)	(0.0041)
Coupon Availability(t-3)	-0.0064 *	-0.0036	-0.0038	-0.0036	-0.0040
	(0.0026)	(0.0028)	(0.0028)	(0.0029)	(0.0042)
adjusted R-squqred	0.0077	0.0563	0.0539	0.0480	0.2361
Coupon Redemption (t)	-0.0060 ***	-0.0055 ***	-0.0056 ***	-0.0060 ***	-0.0041 **
	(0.0009)	(0.0010)	(0.0010)	(0.0010)	(0.0015)
Coupon Redemption (t-1)	-0.0017 +	-0.0005	-0.0006	-0.0008	0.0000
	(0.0009)	(0.0010)	(0.0010)	(0.0010)	(0.0015)
Coupon Redemption (t-2)	-0.0014	-0.0003	-0.0004	-0.0006	0.0004
	(0.0009)	(0.0010)	(0.0010)	(0.0010)	(0.0015)
Coupon Redemption (t-3)	-0.0016 +	-0.0004	-0.0005	-0.0005	-0.0012
	(0.0009)	(0.0010)	(0.0010)	(0.0010)	(0.0015)
adjusted R-squqred	0.0119	0.0588	0.0566	0.0511	0.2369
Number of Observations	6,300	6,300	6,300	6,300	6,300
Controls	constant	brand, store and week	brand-store and week	brand-store and store-week	brand-store, store-week and brand-week

Note : Asterisks denote that the estimated coefficients are statistically significant at 99.9% (***), 99% (**), 95% (*) and 90% (+).

all insignificant, raising suspicion that the effect of coupons on retailers' pricing, if any, may be limited to the period in which the coupons are actually present in the market. It is worth noting that explanatory power of these specifications are rather low. When I include the most flexible form of dummy variables (column 5), the negative correlation between the normalized price and coupon availability seems to vanish, as was the case in Nevo and Wolfram.

For the main model of the section I consider a slightly different type of retailers' decision - the decision to hold sale. As described in Section 2, sale indicators are constructed as :

$$sale_{bst} = \begin{cases} 1 & if \quad np_{bst} < \overline{np} \\ 0 & otherwise \end{cases} \tag{2.4}$$

with $\overline{np}$ being a predefined sale threshold. By replacing the retailers' pricing decision with a simpler bivariate sale indicator, I am essentially throwing away some information in the data. This, on the other hand, enables me to take advantage of the predictions in the literature on retail sale. Combining (2.3) and (2.4) gives rise to an inverted threshold crossing model, and the monopoly price discrimination theory predicts that a retailer should hold a sale with a higher probability when there are coupons available than it would in the absence of coupons, controlling for other factors that influence the retailer's pricing decision.

One approach to bring this implication to the data is to postulate a functional relationship between a latent variable[12] that dictates the retailer's decision to hold sale and

[12]This latent variable should not be confused with the normalized price used to generate the sale indicator.

a set of observed and unobserved state variables :

$$y^*_{bst} = \beta cavail_{bt} + \delta z_{bst} + \gamma_{bst} + \epsilon_{bst} \tag{2.5}$$

Imposing different distributional assumptions on the unobserved state variable, ϵ, gives me familiar discrete choice models, such as probit or logit. An alternative approach is to specify a linear probability model :

$$sale_{bst} = \beta cavail_{bt} + \delta z_{bst} + \gamma_{bst} + \epsilon_{bst} \tag{2.6}$$

Unlike the discrete choice model, estimation of the parameters only requires a set of moment conditions, namely $E[\epsilon_{bst}|X_{bst}] = 0$ and $E[\epsilon_{bst} \cdot X_{bst}] = 0$ where $X_{bst} = [cavail_{bt}, z_{bst}, \gamma_{bst}]$. The estimated parameters have a straightforward interpretation as the covariate's contribution to the probability of retail sale.

I estimate a number of different versions of equation (2.6) by including different combinations of explanatory variables and fixed effects. The results are presented in Table **??**. The raw correlation between brand sale and coupon availability, presented in column 1 is 0.0428. If we take equation (2.6) as representing the true causal relationship between coupon availability and brand sale, then this estimate implies that the presence of coupons increase the sale probability by approximately 4%. It is difficult to imagine, however, that the manufacturers would distribute the coupons randomly across different markets and across different time periods. It is beyond the scope of this paper to model the manufacturers' couponing strategy, although it is a topic of considerable interest. Instead, I experiment with different sets of fixed effects, hoping that these fixed effects would appropriately control for the manufacturers' behavior so that the residual terms is uncorrelated with coupon availability. I perform a sequence of Hausman tests to determine which set of dummy variables to include in the regression. The store-brand, store-week and brand-week fixed effects, when added to the model with coupon availability only all rejected the null hypothesis of equivalence between fixed effect model and random effect model. Once I include the store-brand fixed effects, the null hypothesis was, again, rejected when I added store-week fixed effects whereas the null hypothesis was not rejected when I added brand-week fixed effects. For this reason, I take the results that include store-brand and store-week fixed effects as the base case.

The estimated coefficients show that the presence of coupons is positively correlated with retailers' decision to hold sale and that the influence of coupons on the retailers' sale decision is statistically significant. In the base case, after controlling for the store-brand and store-week fixed effects, the probability of a brand being on sale is approximately 3.1% higher in a week where coupons for the brand are available in the market. Unlike the previous regression of normalized price over coupon availability, the coefficient for coupon availability remains statistically significant at 95% level even when I include the most flexible form of fixed effects. This positive correlation is not surprising given the previous result using equation (2.1) and the stylized fact that a bulk of variation in retail prices comes in the form of a sale. The coefficients obtained using coupon redemptions in equation (2.6) instead of coupon availability are, in general, more statistically significant as was the case in Table **??**. Again, if the number of coupon redemptions can be used as a proxy for how widely coupons are distributed, the results in Table **??** seems to suggest that the extent of coupon availability is positively correlated with the retailers' sale decisions.

Table 2.4. Impact of Coupon Availability on Retailers Sale Decision

	[1]	[2]	[3]	[4]	[5]
Coupon Availability(t)	0.0428 ***	0.0233 **	0.0233 **	0.0313 ***	0.0234 *
	(0.0071)	(0.0076)	(0.0076)	(0.0077)	(0.0114)
adjusted R-squqred	0.0055	0.0597	0.0673	0.0896	0.2198
Coupon Redemption(t)	0.0208 ***	0.0141 ***	0.0143 ***	0.0167 ***	0.0159 ***
	(0.0024)	(0.0026)	(0.0027)	(0.0027)	(0.0041)
adjusted R-squqred	0.0115	0.0624	0.0702	0.0930	0.2216
Number of Observations	6,468	6,468	6,468	6,468	6,468
Controls	constant	brand, store and week	brand-store and week	brand-store and store-week	brand-store, store-week and brand-week

Note : Asterisks denote that the estimated coefficients are statistically significant at 99.9% (***), 99% (**), 95% (*) and 90% (+).

Table 2.4 Impact of Coupon Availability on Retailers' Sale Decision (continued)

	[6]	[7]	[8]	[9]	[10]
Coupon Availability(t)	0.0392 ***	0.0270 ***	0.0272 ***	0.0353 ***	0.0239 *
	(0.0073)	(0.0077)	(0.0077)	(0.0078)	(0.0116)
Coupon Availability(t-1)	0.0156 *	0.0028	0.0030	0.0089	0.0179
	(0.0073)	(0.0077)	(0.0077)	(0.0078)	(0.0116)
Coupon Availability(t-2)	0.0182 *	0.0058	0.0062	0.0069	0.0076
	(0.0073)	(0.0077)	(0.0077)	(0.0078)	(0.0115)
Coupon Availability(t-3)	0.0143 *	0.0033	0.0038	0.0057	0.0023
	(0.0073)	(0.0077)	(0.0077)	(0.0078)	(0.0117)
adjusted R-squqred	0.0085 ***	0.0592 ***	0.0674 ***	0.0916 ***	0.2221 ***
Coupon Redemption (t)	0.0165	0.0142	0.0145	0.0168	0.0149
	(0.0025)	(0.0027)	(0.0027)	(0.0028)	(0.0041)
Coupon Redemption (t-1)	0.0095 ***	0.0052 +	0.0055 *	0.0070 *	0.0080 +
	(0.0026)	(0.0027)	(0.0027)	(0.0028)	(0.0041)
Coupon Redemption (t-2)	0.0062 *	0.0028	0.0031	0.0036	0.0034
	(0.0026)	(0.0027)	(0.0027)	(0.0028)	(0.0041)
Coupon Redemption (t-3)	0.0044 +	0.0001	0.0005	0.0013	0.0018
	(0.0025)	(0.0027)	(0.0027)	(0.0028)	(0.0041)
adjusted R-squqred	0.0165	0.0627	0.0711	0.0960	0.2240
Number of Observations	6,300	6,300	6,300	6,300	6,300
Controls	constant	brand, store and week	brand-store and week	brand-store and store-week	brand-store, store-week and brand-week

Note : Asterisks denote that the estimated coefficients are statistically significant at 99.9% (***), 99% (**), 95% (*) and 90% (+).

In contrast, the lagged coupon availability, the estimated coefficients of which are presented in Table **??** columns 6 through 10, appear to have little or no impact on the retailers' decision to hold sale. If the lagged coupon availabilities are somehow correlated with the residual term of the regression, then the estimated coefficients are biased toward zero. In an unreported regression, I used the fact that there are two sub-markets in the data set to run IV regression using the (lagged) coupon availability of the other market as instruments. The coefficients on lagged coupon availabilities are still statistically insignificant, suggesting that coupons do not have strong intertemporal effect on retailers' sale decision. It compares well with the results in Lee (2007), in which it is suggested that coupons have strong contemporaneous effect on individual consumers' brand choice but do not appear to have significant intertemporal effect.

As before, the adjusted R-squared of these regressions are relatively low. Much of the variation in whether retailers hold sale for particular brand in a particular week remains unexplained by coupon availability. Researchers proposed various theories to explain the patterns in retail pricing. I examine two factors that could potentially influence retailers' decision to hold sale.

The first such factor is the dynamics in demand. Conlisk, Gerstner, and Sobel (1984) describes a monopoly seller's intertemporal pricing decision when facing a steady inflow of different types of consumers. It is shown that, when the low valuation consumers are willing to postpone purchase until the price drops to below their reservation price, the monopolist can enhance profit by serving only the high valuation consumers but occasionally cut the price (or hold a sale) to clear accumulated latent demand from low valuation consumers, rather than using a static pricing rule. In contrast to the price discrimination theory proposed by Varian (1980), in which no temporal pattern in price is suggested as sellers randomize in order to differentiate between informed and uninformed consumers, Conlisk, Gerstner and Sobel's model predicts a distinct intertemporal pattern in demand. Namely, the probability of a sale increases the longer the retailer keeps the price at the high "regular" price as the latent low valuation demand accumulates and increases the profitability of holding a sale over time[13].

To examine the presence of the such a dynamic demand effect, I modify equation (2.6) to include various measures of duration since last sale as explanatory variables. The duration measures include the duration since last sale of the brand in the same store, the duration since last sale of any brand in the same store and the duration since last sale of the brand in any store in the sub-market. These duration measures are intended to capture the process with which the latent demand accumulates during inter-sale periods, if it accumulates at all.

The estimation results are presented in Table **??**. As predicted by the theory of Conlisk, Gerstner and Sobel, the duration measures are, in general, positively correlated with the probability of sale. The patterns of duration dependence are, however, quite complex. The estimated coefficients on duration measures indicate that the probability of sale for a brand in a store is positively correlated with the duration since last sale

[13]The model is also based on a number of simplifying assumptions on how consumers behave. In particular, Conlisk, Gerstner and Sobel assume that consumers will permanently exit the market once they make a purchase. This contrasts the typical purchasing pattern of the consumers, which involves repeated purchase of the same category over time. However, such an assumption seems inocuous if the measure of individual consumers is negligible and if the distribution of different types of consumers entering the market in each period remain more or less stationary, as retailers pricing decision is based on the aggregate demand of the market rather than the behavior of individual consumers.

Table 2.5. Duration Dependence

	[1]	[2]	[3]	[4]	[5]	[6]
Coupon Availability(t)	0.0249 **	0.0249 **	0.0234 *	0.0242 **	0.0238 *	0.0224 *
	(0.0093)	(0.0093)	(0.0093)	(0.0092)	(0.0092)	(0.0093)
Competitors' Coupon(t)			-0.0393 *			-0.0362 *
			(0.0166)			(0.0164)
Duration (store-brand)	0.0001	0.0014 **	0.0014 **	0.0034 ***	0.0051 ***	0.0051 ***
	(0.0004)	(0.0004)	(0.0004)	(0.0010)	(0.0010)	(0.0010)
Duration (store)	0.0160 ***	0.0157 ***	0.0157 ***	0.0365 ***	0.0366 ***	0.0364 ***
	(0.0017)	(0.0017)	(0.0017)	(0.0050)	(0.0050)	(0.0050)
Duration (brand-market)	0.0085 ***	0.0079 ***	0.0078 ***	0.0157 ***	0.0145 ***	0.0144 ***
	(0.0009)	(0.0009)	(0.0009)	(0.0021)	(0.0021)	(0.0021)
Duration Sq (store-brand)				-0.0001 ***	-0.0001 ***	-0.0001 ***
				(0.0000)	(0.0000)	(0.0000)
Duration Sq (store)				-0.0019 ***	-0.0020 ***	-0.0020 ***
				(0.0004)	(0.0004)	(0.0004)
Duration Sq (brand-market)				-0.0004 ***	-0.0003 ***	-0.0003 ***
				(0.0001)	(0.0001)	(0.0001)
adjusted R-squqred	0.0991	0.1084	0.1094	0.1106	0.1201	0.1209
Number of Observations	4,654	4,654	4,654	4,654	4,654	4,654
Controls	brand, store, and week	brand-store, and week	brand-store, and week	brand, store, and week	brand-store, and week	brand-store, and week

Note : Asterisks denote that the estimated coefficients are statistically significant at 99.9% (***), 99% (**), 95% (*) and 90% (+).

of the brand in the store, as well as with the duration since last sale of the brand in any store in the sub-market. This dependence, of course, is the motivation behind the empirical analysis of Pesendorfer (2000), in which three major ketchup brands are modeled as engaged in intertemporal price discrimination in face of accumulating latent demand and in face of competition from the same brand in different stores. An implicit assumption in Pesendorfer's analysis is that the evolution of latent demand is brand-specific hence a retailer's decision to hold a sale for one brand has no impact on the profit the retailer generates from another brand.

The estimated coefficients for duration since last sale of any brand in store seems to suggest a different scenario. The magnitude of the estimated coefficients for this particular duration measure is dominant compared to the estimated coefficients on brand-specific duration measures[14]. If, as in Pesendorfer, the evolution of demand for each brand is independent of each other, we should expect to see the second duration measure having little or no power to explain the retailer's sale decision. The fact that all duration measures have certain degree of explanatory power, in the context of the theory by Conlisk, Gerstner and Sobel, seems to suggest that the low valuation consumers can potentially be further classified into multiple types, depending on their loyalty to different brands or to different stores. The coefficients for coupon availability remain statistically significant after inclusion of duration measures. Including squared duration measures to capture the potential nonlinearity in duration dependence[15] has only a marginal impact on estimated coefficients.

The second factor to consider is potential competition across different brands in store. Conlisk, Gerstner and Sobel's theory described the price path of a monopolist seller that sells a single product. Sobel's 1984 sequel introduced competition across different retailers, but the retailers in this model sells a single homogeneous product. Effectively, the retailers' action space in this model is single dimensional. If retailers' sale decisions for different brands are indeed driven only by brand-specific considerations, we should expect to see no correlation in observed brand sales.

[14]Since the duration measures are not normalized, it would be erroneous to simply compare the magnitude of estimated coefficients. The means of three duration measures are 13.2413, 3.8911 and 6.7361, respectively. If we casually normalize the duration measures so that their means are equalized, the duration since last sale of any brand in store is still the most influential factor in the retailer's sale decision.

[15]Results presented in Table 5 columns 4 through 5.

Table 2.6. Correlation in Sale Across Brands Conditional on Store Sale

		Brands							
		2	3	4	7	8	11	18	21
Brands	2	1.0000	-0.0832	-0.0595	0.0193	-0.0723	-0.2788	-0.1353	-0.1050
	3	-0.0832	1.0000	0.0314	0.0778	0.0004	-0.0690	-0.0713	-0.0385
	4	-0.0595	0.0314	1.0000	-0.0576	-0.0599	-0.1603	-0.1353	-0.0656
	7	0.0193	0.0778	-0.0576	1.0000	-0.0298	-0.1653	0.0634	-0.0327
	8	-0.0723	0.0004	-0.0599	-0.0298	1.0000	-0.0859	0.1192	0.1448
	11	-0.2788	-0.0690	-0.1603	-0.1653	-0.0859	1.0000	-0.1271	-0.0682
	18	-0.1353	-0.0713	-0.1353	0.0634	0.1192	-0.1271	1.0000	0.1117
	21	-0.1050	-0.0385	-0.0656	-0.0327	0.1448	-0.0682	0.1117	1.0000

Table **??** summarizes the correlation of sale indicators for different brands. Conditional on there being a sale in a store, brand sales are generally negatively correlated indicating that when one brand is on sale, the other brands tend to be sold at regular prices[16].

There can be a number of different reasons why a retailer may and may not want to offer multiple brands on sale at the same time. First and foremost, if different brands are close substitutes and the low valuation consumers are equally attracted either to one brand or another so long as at least one brand is on sale, the existence of menu cost will induce the retailer to place only a small number of brands (if not one brand) on sale at a time. There may also be an unobserved factor that commonly influence both coupon availability and the shelf price. The presence of unobserved factor that is correlated with coupon availability may cause a bias in coefficients and render the correlations in Table **??** unreliable. To capture this relationship, I include competitors coupon availability as an explanatory variable in equity (2.6). As anticipated, there is an inverse relationship between the estimated coefficients on competitors coupon availability and sale.

[16]There are observations with multiple sale brands. The mean number of major brands on sale is 1.39 whereas the median is 1.

Table 2.7. Summary Statistics for Store-Week Data

	Obs	Min	Max	Median	Mean	Std
Store-wide Sale	809	0	1	0	0.4302	0.4954
Sale on Coupon Brand	809	0	1	0	0.2188	0.4137
Coupon Availability (brand 2)	809	0	1	0	0.4326	0.4957
Coupon Availability (brand 3)	809	0	1	0	0.2312	0.4218
Coupon Availability (brand 4)	809	0	1	0	0.3053	0.4608
Coupon Availability (brand 7)	809	0	1	0	0.2213	0.4154
Coupon Availability (brand 8)	809	0	1	0	0.1941	0.3957
Coupon Availability (brand 11)	809	0	1	1	0.5278	0.4995
Coupon Availability (brand 18)	809	0	1	0	0.2373	0.4257
Coupon Availability (brand 21)	809	0	1	0	0.2991	0.4582
Duration (store)	793	1	15	2	2.6974	2.2966

The regression results suggest that the state space that influence a retailer's sale decision can be quite complex. I now investigate whether I can separate the retailer's choice of sale brand from its dynamic decision to hold a sale. I restate the data set to make the store-week the primary unit of observation and derive indicators for whether the retailer holds a sale for any of the major brands in each week and whether there were coupons available for each brand in that week. The summary statistics are presented in Table **??**. An important issue associated with this transformation is that it reduces the number of observations substantially to the number of stores times the number of weeks. In the state-week data set, I only have 809 observations[17], and it becomes a challenge to estimate the model coefficients with good enough precision.

With this data set, I estimate the following reduced form models :

$$sale_{st} = \beta_1 cavail_t + \beta_3 duration_{bst} + \gamma_s + \gamma_t + \epsilon_{st} \tag{2.7}$$

[17] I can only use 793 observations when I include duration since last sale.

$$duration_{st} = sale_{st} * cavail_t + +\gamma_s + \gamma_t + \eta_{st} \tag{2.8}$$

where the γ's are the usual dummy variables to control for store and week fixed effects and where ϵ_{st} and η_{st} are unobserved determinants of store sale and and unobserved determinants of the length of no sale periods, which are assumed to be independent of the other explanatory variables. The estimation results are presented in the Tables **??** through **??**. Due to the limited number of observations, most of the estimated coefficients are statistically significant. One implication I should mention is that coupon availability for different brands has different impact on sale probability. Coupon availability for Kellogg's Corn Flakes, for example, is more highly correlated with the store-wide sale for regular cereal category than the coupon availability for other brands is. Such a difference could arise if consumers respond differently to seemly identical coupons issued by different manufacturers or coupons for different brands are inherently different in their characteristics such as the face value, the extent of distribution and the mode of distribution. This result, although the explanatory power of these regressions is extremely low, contrasts the simplification I used in deriving the coupon variables - it was implicitly assumed that all coupons are identical except that each can only be used for a particular brand.

Equation (2.8) attempts to verify whether the coupon availability has an impact not only on the retailer's choice of sale brand but also on the retailer's sale timing decision. If coupons make it more profitable for a retailer to hold a sale for a particular brand, unless coupons for a particular brand has an adverse impact on profitability of other brands, the profitability of holding a sale weakly increases and we should expect to see an acceleration

Table 2.8. Impact of Coupon Availability on Store-wide Sale

	[1]	[2]
Coupon Availability (any brand)	-0.0839	0.0021
	(0.0692)	(0.0818)
Coupon Availability (Brand 2)	-0.0814 *	0.2290
	(0.0376)	(0.1886)
Coupon Availability (Brand 3)	0.1510 ***	0.0944
	(0.0418)	(0.1886)
Coupon Availability (Brand 4)	0.0232	0.4021 *
	(0.0387)	(0.1886)
Coupon Availability (Brand 7)	0.0396	0.2386
	(0.0433)	(0.1886)
Coupon Availability (Brand 8)	-0.0429	0.1809
	(0.0448)	(0.1886)
Coupon Availability (Brand 11)	0.0043	0.2293
	(0.0382)	(0.1831)
Coupon Availability (Brand 18)	0.1047 *	0.2389
	(0.0425)	(0.1831)
Coupon Availability (Brand 21)	-0.0969 *	0.3700 *
	(0.0388)	(0.1846)
adjusted R-squqred	0.0296	0.1430
Number of Obs	809	809
Controls	constant	store and week

Note : Asterisks denote that the estimated coefficients are statistically significant at 99.9% (***), 99% (**), 95% (*) and 90% (+).

of sale when the retailer holds a sale for a couponed brand. The estimated coefficients are all negative but, due to very limited number of observations I can use, they are not statistically significant. The only statistically significant estimate is the coefficient on sale dummy for Kellogg's Corn Flake. The inter-sale duration is noticeably shorter when the retailer holds a sale for this brand.

2.4. Discussion

This paper documents the impact of coupon availability on retailers' sale decision. It is motivated by the empirical findings in Nevo and Wolfram (2001), where they demonstrated

Table 2.9. Impact of Coupon Availability on Inter-Sale Duration

	[1]	[2]	[3]	[4]	[5]
Coupon Availability (any brand)	-0.2143	-0.3260	-0.1656	-0.0974	-0.0996
	(0.2402)	(0.2446)	(0.3410)	(0.3542)	(0.3551)
Coupon Availability (Brand 2)				0.2093	0.2230
				(0.4307)	(0.4380)
Coupon Availability (Brand 3)				-0.4419	-0.4547
				(0.4771)	(0.4831)
Coupon Availability (Brand 4)				-0.7804 *	-0.7755 +
				(0.3952)	(0.3970)
Coupon Availability (Brand 7)				0.1944	0.1955
				(0.5835)	(0.5848)
Coupon Availability (Brand 8)				-0.9035	-0.8846
				(0.9878)	(0.9952)
Coupon Availability (Brand 11)				-0.5101	-0.5040
				(0.3661)	(0.3683)
Coupon Availability (Brand 18)				-0.1057	-0.0942
				(0.5255)	(0.5303)
Coupon Availability (Brand 21)				0.4027	0.3951
				(0.7632)	(0.7660)
Past Sale on Coupon Brand					-0.0588
					(0.3184)
adjusted R-squqred	-0.0006	0.0373	0.0753	0.0766	0.0727
Number of Obs	340	340	340	340	340
Controls	constant	store	store and week	store and week	store and week

Note : Asterisks denote that the estimated coefficients are statistically significant at 99.9% (***), 99% (**), 95% (*) and 90% (+).

that the weighted average shelf price for ready-to-eat cereals is negative correlated with the coupon availability. In recognition that sales account for a bulk of variation in retail shelf prices, I considered a natural variation of this analysis by examining how retailers' sale decisions are correlated with coupon availability. The positive correlation I reported is not surprising given the results in Nevo and Wolfram (2001) and given the stylized fact that quantity sold increases substantially during sale periods, and yet it is at odd with the static monopoly price discrimination of coupons. The assumptions of the price discrimination theory, however, come short of describing the structure of the market since the manufacturers of the product category in question are not monopoly sellers and since the manufacturers issue coupons but they do not set shelf prices.

There are different explanations for retail sales in the literature. Shilony (1977), Rosenthal (1980) and Varian (1980), for example, advanced static models in which sellers pursue mixed strategies, leading to dispersion of price across stores and, potentially, across time. One implication from this class of models is that prices are not predictable, yet the data seem to suggest that there is substantial regularity in price reductions. Conlisk, Gerstner and Sobel (1984) and Sobel (1991), on the other hand, proposed a model with constant inflow of new consumers, in which periodic sales enhance seller's profit over repetition of static pricing strategies. The main empirical analysis of this paper shows that, in addition to the duration since last sale that approximate the accumulated latent demand, coupon availability has some power in explaining retailers' sale decision.

The estimated coefficients for duration measures imply that the dynamic demand effects are quite complex suggesting that there may be different types of consumers that constitute the latent demand, with the store-loyal consumers playing a more significant role than the brand-loyal consumers. This exercise is useful in the sense it points to a way to construct a model of retailers' sale decision. I envision a Sobel-style model in which a monopolistic retailer that carry multiple products periodically puts one or more brands on sale to accommodate accumulated low valuation consumers, with the brand's coupon availability having an impact on retailers' profit when the brand is on sale. In principle, I need to keep track of the measures of accumulated low valuation consumers for each brand and it would be challenging to estimate an econometric model with such a large state space where the retailer's decision is inherently dynamic. By adopting the assumptions outlined in Hendel and Nevo (2006), I may be able to separate the retailers sale brand choice as a static decision from the retailer's sale decision that takes the total measure of accumulated low valuation consumers as a scalar state variable. This structural model of retailer's sale decision can be used for a counter-factual simulation to examine what impact withdrawing coupons from the market may have on the frequency and brand choice of retail sales. This is an interesting exercise in light of the history that some major consumer packaged product manufacturers attempted to discontinue their coupon programs in mid-1990s but had to reinstate them soon afterwards due to the consumers' protests.

While the positive correlation between coupon availability and sale seems to suggest that retailers find it profitable to put the couponed brand on sale, the proposed econometric model itself is silent about the source of this profitability. Although it is a norm is for the coupon issuer to offer a small premium over the face value per redeemed coupon to cover the retailer's coupon handling expenses, it is hard to imagine that these fees constitute a significant source of profit to the retailers. Coupons, on the other hand, can be an effective instrument to change the nature of demand[18]. Lee (2007), for example, examined individual consumers' purchasing behavior and concluded that the coupons induce brand switching. In the absence of a direct intertemporal effects, the presence of coupons may seem like an isolated positive demand shock from the retailers' perspective, and we are required to consider what mode of retailer's behavior would create a positive correlation between sale and a positive demand shock for a brand.

The association of a positive demand shock with a low price is similar to the phenomenon described in Warner and Barsky (1995), MacDonald (2000) and Chevalier, Kashyap and Rossi (2003). The presence of coupons amounts to an observable demand shock,

[18]See Blattberg and Neslin (1990) for survey of the literature.

which we can take advantage of instead of the seasonality in aggregate demand that the aforementioned researches are based on. The idea that the retailers have multiple instruments to implement price cuts is also similar to the idea in Hosken and Reiffen (2004a, 2004b). Hosken and Reiffen extended Varian's model to multi-product retailers and described how a retailer circulate sale categories. In this model, retailers aim to offer enough surplus to induce consumers to shop by cutting prices for a small number of categories. The existence of transportation cost creates an incentive for the consumers to shop in one outlet, hence different categories in this model are effectively complements to the retailers. In contrast, different brands within one product category are substitutes hence introducing multiple sale instruments may yield different predictions.

The empirical analyses in this paper face a number of limitations that could potentially complicate the interpretation of estimated coefficients. The most obvious originates from the way the key variables are constructed. Coupon availability, for instance, is defined at the sub-market level and I required a fixed number of coupon redemptions in a week in a sub-market to indicate that coupons were available for a particular brand-market-week. If coupons are randomly distributed and coupon redemptions follow an independent random process then this classification may correctly approximate the extent of coupon availability. However, as purchase volume is, in general, higher during sale periods, the data construction method could create a positive correlation between sale and coupon availability through consumers' coupon redemption choice even when the process with which the manufacturers make coupons available is totally unrelated to retailers' sale decision. Ideally, the store sale data in this data set should be supplemented by the true coupon availability data such as the PIM data used in Nevo and Wolfram (2001).

CHAPTER 3

Identification of Dynamic Discrete Choice Models with Variations in Transition Probabilities

3.1. Introduction

The structural dynamic discrete choice model has become a popular tool in analyzing the behavior of forward looking economic agents. Empirical applications appeared in studies of fertility (Wolpin (1984), Hotz & Miller (1993)), labor force participation (Eckstein & Wolpin (1982)), patent renewal (Pakes (1986)), engine replacement (Rust (1987)), valuation of option embedded securities (McConnell & Singh (1994)), resource allocation (Timmins (2002)) and product portfolio choice (Benkard (2004)). The dynamic model permits the researchers to model behaviors that are difficult to justify with the assumption of rational but myopic agents. The structural specification, on the other hand, enables the researchers to not only describe economic agents' behaviors, but also evaluate policy effects and conduct various counterfactual experiments. The key challenges to this line of research are the heavy computational burden and the non-identification of the primitives. As a result, the researchers are often confined to rigid parametric specifications with few parameters to estimate.

The identification of the primitives in the static discrete choice model is well established (Manski (1975, 1985), Matzkin (1992, 1993)). The identification problem for the dynamic discrete choice model is three-folded. First, without any restriction on model primitives, any dynamic model can be equally described as a static model (Manski (1993)), and the researchers are forced to take the discount factor as determined outside the model. Second, even after fixing the discount factor, the dynamic model is not nonparametrically identified when the stage utility function and the error distribution function are unrestricted (Rust (1994)). Finally, even with the knowledge of the error distribution, the researcher cannot typically separate the agent's stage utility from the agent's expectation for the future (Magnac & Thesmar (2002)).

Some possibilities remain. Aguirregabiria (2005a, 2005b) describes how the counterfactual choice probabilities are nonparametrically identified when a particular type of policy is implemented, even when the stage utility function is not identified. Aguirregabiria's analysis, however, is limited to the policy that modifies the stage utility linearly.

This paper examines whether Aguirregabiria's result can be extended to a different type of policy implementation. Based on the non-identification result, it then asks the question whether identification of the stage utility function can be achieved when the researcher is given an opportunity to observe the change in the agent's behavior following such a policy implementation.

The paper is organized in the following way. In section 2, we introduce the standard single-agent dynamic discrete choice model, state the assumptions, and define the notion of observational equivalence and of identification. In section 3, we show that the counterfactual choice probabilities resulting from a shift in transition probabilities are not identified and contrast the result with that of Aguirregabiria's (2005). Based on this result, we propose an identification strategy and outline conditions for identification in section 4. Finally, we discuss the potential extensions and limitations of the approach in section 5.

3.2. Model

Consider a single-agent model in discrete time, analogous to the engine replacement example of Rust (1987), in which the decision maker observes a vector of state variables, $s_t \in S$, and chooses an alternative ("action") from a discrete set, $A = \{0, \cdots, J-1\}$, at each period. The agent has preference defined over a sequence of states and action choices, $\{(a_{t+\tau}, s_{t+\tau})\}_{\tau=0}^{\infty}$, and it is summarized by an additively time-separable utility function of the form, $\sum_{\tau=0}^{\infty} \beta^{\tau} U_{t+\tau}(a_{t+\tau}, s_{t+\tau})$, where $\beta \in [0,1)$ is the discount factor, and $U_t(.,.)$ is the stage utility function. For the present purposes, we do not place any restriction on the shape of this stage utility function. The agent's action choice at time t has an impact not only on the contemporaneous stage utility, but also on the outcome of the future states. The agent faces an uncertainty with respect to the evolution of the future states. As a result, the agent, after observing a history of states and of action choices, $H_t = (a_0, a_1, \cdots, a_{t-1}, s_0, s_1, \cdots s_t)$, chooses a sequence of actions to maximize the expected utility

$$\max_{a_t, \{a_{t+\tau}(H_{t+\tau})\}_{\tau=1}^{\infty}} \widetilde{E}\left[\sum_{\tau=0}^{\infty} \beta^{\tau} U_{t+\tau}(a_{t+\tau}, s_{t+\tau}) | a_t, H_t\right] \tag{3.1}$$

where $\widetilde{E}[.]$ is the expectation taken with respect to the agent's subjective belief about the evolution of the future states. We follow Rust (1994) and impose the following assumptions :

Assumption 1 (Markovian Transition) The evolution of the state variables is governed by a sequence of Markov transition functions, $\Pi_t(s_{t+1}|a_t, s_t)$.

Assumption 2 (Rational Expectation) The agent's subjective belief about the outcome of the future states given the history, H_t, is rational in the sense it coincides with the true transition probability of the state vector.

The combined implications of Markovian transition and rational expectation are that (i) the expectation in (3.1) is the mathematical expectation based on the true transition probability functions, $\Pi_t(s_{t+1}|a_t, s_t)$, (ii) the expectation can be computed conditional on the current state and the current action choice only, instead of the entire history of states and action choices, and (iii) the optimal decision rule is a function of the current state only.

Assumption 3 (Additive Separability) The stage utility function has an additively separable form

$$\forall a_t \in A, \qquad U_t(a_t, x_t, \epsilon_t) = u_t(a_t, x_t) + \epsilon_t(a_t) \tag{3.2}$$

Alternatively to making such an assumption, we may consider a decomposition of the vector of state variables into two components - one that is observable to the researcher, denoted x_t, and one unobservable to the researcher, denoted ξ_t. Without loss of generality, we can also decompose the stage utility function into two components :

$$U_t(a_t, s_t) = u_t(a_t, x_t) + \epsilon_t(a_t, x_t, \xi_t) \tag{3.3}$$

where $u_t(a_t, x_t) := E[U_t(a_t, s_t)|a_t, x_t]$ and $\epsilon_t(a_t, x_t, \xi_t) := U_t(a_t, s_t) - u_t(a_t, x_t)$. By construction, ϵ_t is mean independent of x_t, and has zero mean conditional on (a_t, x_t). We may take the vector of ϵ's as our unobservable state variables, instead of ξ_t.

Assumption 4 (Conditional Independence) The cumulative transition probability of the state variables factors as $\Pi_t(s_{t+1}|a_t, s_t) = F_{t+1}(\varepsilon_{t+1})\pi_t(x_{t+1}|a_t, x_t)$.

As Rust (1994) points out, the conditional independence assumption implies that the unobservable state vector, ϵ_t, is essentially a noise term superimposed on the main dynamics embodied in the evolution of x_t.

Under these assumptions, the agent's problem can be rewritten as :

$$\max_{a_t(s_t), \{a_{t+\tau}(s_{t+\tau})\}_{\tau=1}^{\infty}} E\left[\sum_{\tau=0}^{\infty} \beta^{\tau}(u_{t+\tau}(a_{t+\tau}(s_{t+\tau}), x_{t+\tau}) + \epsilon_{t+\tau}(a_{t+\tau}, (s_{t+\tau})))|a_t, s_t\right] \tag{3.4}$$

and, by Bellman's principle of optimality, we may define a sequence of value functions in a recursive equation :

$$\begin{aligned} V_t(s_t) &= \max_{a_t \in A}\{U_t(a_t, s_t) + \beta E[V_{t+1}(s_{t+1})|a_t, s_t]\} \\ &= \max_{a_t \in A}\{u_t(a_t, x_t) + \epsilon_t(a_t) + \beta \int V_{t+1}(s_{t+1}) dF_{t+1}(\epsilon_{t+1}) d\pi_t(x_{t+1}|a_t, x_t)\} \qquad (3.5) \end{aligned}$$

Related to the value function thus defined, we can also define a set of conditional value functions as

$$(3.6)\quad v_t(a_t, x_t) := u_t(a_t, x_t) + \beta \int \max_{j\in A}\{v_{t+1}(j, x_{t+1}) + \epsilon_{t+1}(j)\} dF_{t+1}(\epsilon_{t+1}) d\pi_t(x_{t+1}|a_t, x_t)$$

which is the expected discounted utility if the agent takes action a_t in the current period. The value function can be written in terms of the conditional value functions :

$$(3.7)\qquad V_t(s_t) = \max_{a_t\in A}\{v_t(a_t, x_t) + \varepsilon_t(a_t)\}.$$

Assume that F_t is absolutely continuous with respect to the Lebesgue measure, then the

agent's optimal choice set given s_t,

$$(3.8)\qquad \alpha(s_t) = \arg\max_{a_t\in A}\{v_t(a_t, x_t) + \varepsilon_t(a_t)\}$$

is singleton almost everywhere.

Assumption 5 (Stationarity) The stage utility function, the distribution function of the unobservable state vector and the transition function for the observable state vector are invariant of time.

With stationarity, the agent faces the same optimization problem at each period, and we can drop the time subscripts from the expressions above.

Assumption 6 (Discrete Support) The observable state vector, x, has a discrete support, $X = \{x^1, \cdots, x^K\}$.

With discrete support, the transition function for the observable states, $\pi(x\prime|a,x)$, can be represented by a $K \times K$ transition matrix, whose $(i,j)^{th}$ element is $\Pr(x\prime = x^i|a, x = x^j)$. The stage utility function given an action choice, a, can be represented by a K-dimensional vector, $u(a,X)$, where we adopted, for an arbitrary function, f, the expression $f(X)$ to be the vector :

$$f(X) = \begin{pmatrix} f(x^1) \\ f(x^1) \\ \vdots \\ f(x^K) \end{pmatrix}$$

Model Structure, Reduced Form and the Data

What we define to be the model's structure depends on the nature of the data and on what assumption we impose on the model. In a typical single-agent model, for a data set with a relatively short length, the model structure consists of $(\beta, (u_t, F_t, \pi_t)_{t=1}^T, v_T)$. With a longer data set, v_T can be determined endogenously as a function of the other structural components, in which case, the model structure is reduced to $(\beta, (u_t, F_t, \pi_t)_{t=1}^\infty)$. By imposing the stationarity assumption, this set can be further reduced to (β, u, F, π).

As the researcher does not have a priori knowledge of the utility function and does not observe ε, the agents' optimizing behavior is described by the choice probability function that indicates what proportion of the agents chose action a when faced with a state involving the observable component x :

$$P(a|x) = \Pr\left(a \in \arg\max_{j \in A}\{v(j,x) + \epsilon(j)\}\right) \tag{3.9}$$

Let $\widetilde{v}(a,x) = v(a,x) - v(0,x)$ and $\widetilde{\epsilon}(a) = \epsilon(a) - \epsilon(0)$. By construction, $\widetilde{v}(0,x) = 0$ and $\widetilde{\epsilon}(0) = 0$. Equation (3.9) can be restated in terms of $\widetilde{v}(a,x)$:

$$P(a|x) = \Pr\left(a \in \arg\max_{j \in A}\{\widetilde{v}(j,x) + \widetilde{\epsilon}(j)\}\right) \tag{3.10}$$

Given F, (**??**) maps $\{\widetilde{v}(a,x) : \forall a \in A\}$ into $\{P(a|x) : \forall a \in A\}$. By Hotz & Miller (1993), this mapping is invertible :

$$v(a,x) = v(0,x) + q(a,x,P;F) \tag{3.11}$$

As we describe $v(a,x)$ as a fixed point of the Bellman equation in (3.6), (**??**) permits us to determine $u(a,x)$ as a function of $u(0,x)$ and the optimal choice probabilities, P.

Consider a population of identical individuals behaving in accordance to the model described above. For individual n and for time period t, the researcher observes a triplet $(x, a, x\prime)$, that is the observable state, x, the action choice of the individual, a, and the next period state , x',that resulted from (a,x). As in Magnac & Thesmar (2002), we denote the model structure as $b = (\beta, u(0,X), F, \pi)$, and take the conditional choice probability function, P_b, and the observed transition probabilities, π_b^o, to be the reduced for of the model. We adopted the subscript, b, to indicate that the optimal choice probabilities and the observed transition probabilities are generated by the structure b. Now let us define the notion of observational equivalence and of identification in the usual way (as in Magnac & Thesmar (2002)) :

Definition 1. *(Observational Equivalence; Magnac & Thesmar (2002)) Let B be the set of structures, and let $\overset{o}{\Longleftrightarrow}$ denote observational equivalence. For two structures, $b_1, b_2 \in B$,*

$$b_1 \overset{o}{\Longleftrightarrow} b_2 \qquad \textit{if and only if} \qquad (P_{b_1}, \pi^o_{b_1}) = (P_{b_2}, \pi^o_{b_2}) \tag{3.12}$$

Definition 2. *(Identification; Magnac & Thesmar (2002)) The model structure is identified if and only if*

$$\forall b_1, b_2 \in B \qquad (b_1 \overset{o}{\Longleftrightarrow} b_2) \longrightarrow (b_1 = b_2) \tag{3.13}$$

As we study identification with a large sample, we let the size of the panel to approach infinity in both dimensions. We also require that the reduced form of the model - the optimal choice probabilities and the transition probabilities - is fully identified from the data, which, in turn, requires the assumption that the sample has variability over the whole support of $(a, x, x\prime)$. We have to keep in mind that some of the states may occur with lower frequency than others. This implies that with finite sample, the estimates at such a state of the choice and transition probabilities may be inaccurate, and this may have an impact as we derive a finite-sample estimation strategy. However, for now, we assume that the researcher observes an infinite sample, hence he can identify the reduced form of the model from the data without an error.

The standard example that fit into this specification is the engine replacement model of Rust (1987). At each time period, the decision maker, named Harold Zurcher, observes the accumulated mileage since last replacement of each bus, denoted x_t, and other characteristics specific to each bus, denoted ε_t, and decides whether to replace the engine. At each period, only the accumulated mileage and his replacement decisions are recorded. If the engine is not replaced, the bus incurs a relatively high maintenance cost as it accumulates more miles. In contrast, if the engine is replaced, the bus incurs a one-time replacement cost, but its mileage is set to zero.

3.3. Identification and Non-identification of Counterfactual Choice Probabilities

The identification for discrete choice model in static environment ($\beta = 0$) was first established in Manski (1975, 1985), in which a semi-parametric model was assumed. In Matzkin (1992, 1993), it was shown that the utility differences can be nonparametrically identified under mild restrictions such as homogeneity, which are often implied by economic theories. The attempts to extend these results to dynamic models were initially deterred by the result in Manski (1993), which states that any dynamic problem has an observationally equivalent static representation. It became a common practice to take $\beta \in [0, 1)$ as being determined outside the model.

Even after assuming that the researcher observes the discount factor determined outside the model, identification of dynamic discrete choice models remained problematic.

Rust (1994) finds that, for any arbitrary decision rule (i.e. the choice probabilities), we can always find a set of primitives that rationalize the given decision rule within the framework of an agent behaving in according to the Bellman equation. Magnac & Thesmar (2002) provided a detailed exposition of the topic in a somewhat simplified setting, confirmed the non-identification result and showed the exact degree of underidentification.

In summary, the researcher faces two difficulties. In contrast to the result developed for static discrete choice models (Matzkin; 1992, 1993), when the distribution of the unobservable state vector is not known to the researcher, it is difficult to find a set of restrictions on the primitives (i.e. on the stage utility function) that would enable the researcher to separate the variation in the deterministic component of the value function from the variation due to the randomness of the unobservable state. Such a decomposition would require placing normalizing restrictions on the value function. However, since the value function is determined implicitly as the fixed point of the Bellman equation, it typically does not inherit any restriction we place on the model primitives, in particular, on the stage utility function. Even when the distribution of the unobservable state vector is known to the researcher, hence the differences in conditional value functions are identified from the data, the researcher cannot, in general, separate the agent's stage utility from the agent's expectation about the future. The non-identification results, combined with the heavy computational burden in estimating the dynamic models, have led the researchers to adopt relatively simple parametric specifications.

A different approach to this identification problem can be found in Aguirregabiria (2005), in which a nonparametric identification result was developed for the counterfactual choice probabilities, rather than for the usual model primitives. It was shown that, even when the model primitives remain unidentified, it is possible to evaluate the change in the agents' behavior following a class of policy implementation that brings about an additive change in the stage utility. An example of this class of policy implementation is a subsidy given to engine replacement that alleviates the replacement cost of the decision maker.

Following Aguirregabiria's result, and keeping in mind that the non-identification result stemmed from the inability to separate the stage utility from the agent's expectation, we can envision a different type of policy implementation that leaves the stage utility unchanged but alters the transition probabilities that govern the evolution of observable state variables. In the engine replacement example, this type of policy implementation amounts to the firm's decision to operate the buses more intensively. Under this intensive use regime, the transition from a low mileage to a high mileage will be accelerated when the engines are left unreplaced, causing a shift in the transition probability matrix.

In order to state this result more rigorously, we alter our previous definition of the structure and the reduced form of the model. As we are interested in identification of the counterfactual choice probabilities, we take $b = (\beta, F, u(0, X), \pi, \psi, P^*)$ as the model's structure, where $(\beta, F, u(0, X), \pi)$ is as before, ψ is the description of the policy implementation being considered, and P^* is the optimal choice probability following the policy implementation, ψ. The reduced form of the model, (P_b, π_b^o), remains unchanged. With this modified structure, the previous definitions for observational equivalence and of identification apply without further modification. In Aguirregabiria's model, ψ was the function, $\tau : A \times X \longrightarrow R$ that described the additive modification to the stage utility function. In this alternative policy implementation, $\psi = \pi^*$ where π^* is the new transition function for the observable state vector following the policy implementation.

Consider a binary choice case, $A = \{0, 1\}$, and define the differential value function as $\widetilde{v}(x) = v(1, x) - v(0, x)$. The optimal choice probabilities are derived as

$$
\begin{aligned}
P(x) &= \Pr(v(1,x) + \varepsilon(1) \geq v(0,x) + \varepsilon(0)) \\
&= \Pr(\varepsilon(0) - \varepsilon(1) \leq v(1,x) - v(0,x)) \\
&= \Pr(\widetilde{\varepsilon} \leq \widetilde{v}(x)) \\
&= F_{\widetilde{\varepsilon}}(\widetilde{v}(x))
\end{aligned} \tag{3.14}
$$

Since we assume that the researcher knows F, and can identify $P(x)$ from the data, $\widetilde{v}(x)$ is identified. The counterfactual choice probabilities are generated in the same way

$$
P^*(x) = F_{\widetilde{\varepsilon}}(\widetilde{v}^*(x)) \tag{3.15}
$$

As the researcher knows F, and since F is absolutely continuous with respect to the Lebesgue measure, identification of the counterfactual choice probabilities is equivalent to identification of the counterfactual differential value function, $\widetilde{v}^*$. A result in Aguirregabiria (2002) enables us to decompose the differential value function in two different components :

Lemma 3. *(Proposition 2 in Aguirregabiria) The optimal choice probability function, P, is the unique fixed point of the mapping, $\Psi(P)$, where*

$$
\Psi(P)(x) := F_{\widetilde{\varepsilon}}(\widetilde{\varphi}(x) + \widetilde{\delta}(x, P)) \tag{3.16}
$$

and (1) , where $\widetilde{\varphi}(x) = \varphi(1,x) - \varphi(0,x)$*, and* $\varphi(a,x)$ *is the value of choosing alternative* a *today and then select alternative 0 forever in the future; and (2)* $\widetilde{\delta}(x,P) = \delta(1,x,P) - \delta(0,x,P)$*, where* $\delta(a,x,P)$ *is the value of behaving optimally in the future minus the value of choosing always alternative 0, given that the current choice is* a*. These functions can be obtained recursively as follows :*

$$\varphi(a,x) = u(a,x) + \beta \int \varphi(0,x')d\pi(x'|a,x) \tag{3.17}$$

and

$$\delta(a,x,P) = \beta \int (G(x',P) + \delta(0,x',P))d\pi(x'|a,x) \tag{3.18}$$

where $G(x,P)$ *is McFadden's surplus function that is defined as* $\int \max\{0; \widetilde{v}(x) - \widetilde{\varepsilon}\} dF_{\widetilde{\varepsilon}}(\widetilde{\varepsilon})$*, and it can be represented as a function of the optimal choice probability* $P(x)$ *as*

$$G(x,P) = P(x)F_{\widetilde{\varepsilon}}^{-1}(P(x)) - \int_{-\infty}^{F_{\widetilde{\varepsilon}}^{-1}(P(x))} \widetilde{\varepsilon} dF_{\widetilde{\varepsilon}}(\widetilde{\varepsilon}) \tag{3.19}$$

Proof. See Aguirregabiria (2002) □

It is important to note that, given the optimal choice probabilities, P, the function δ is determined independent of the structural component $u(0,X)$. Any normalization involving $u(0,X)$ will only have an impact on φ. Furthermore, $\varphi(0,x)$ is determined as the unique fixed point of (3.15) given $u(0,X)$ and π.

Lemma 4. *If $\widetilde{\varphi}_1(x) \neq \widetilde{\varphi}_2(x)$, then $P_1(x) \neq P_2(x)$, where $P_1(x)$ and $P_2(x)$ are the fixed points of the mapping $\Psi(P)(x)$ associated with $\widetilde{\varphi}_1(x)$ and $\widetilde{\varphi}_2(x)$, respectively.*

Proof. Suppose $P_1(x) = P_2(x)$. Then $F_{\widetilde{\varepsilon}}(\widetilde{\varphi}_1(x) + \widetilde{\delta}(x, P_1)) = F_{\widetilde{\varepsilon}}(\widetilde{\varphi}_2(x) + \widetilde{\delta}(x, P_2))$. Since $\widetilde{\delta}(x, P_1) = \widetilde{\delta}(x, P_2)$, and $F_{\widetilde{\varepsilon}}$ is invertible, we have $\widetilde{\varphi}_1(x) = \widetilde{\varphi}_2(x)$, which is a contradiction. □

Given Aguirregabiria's decomposition, the optimal choice probabilities, P^*, associated with the new transition probabilities, π^*, is the unique function satisfying $P^*(x) = F_{\widetilde{\varepsilon}}(\widetilde{\varphi}^*(x) + \widetilde{\delta}^*(x, P^*))$ with $\widetilde{\varphi}^*(x)$ and $\widetilde{\delta}^*(x, P^*)$ defined analogously as above, with π replaced by π^*. Lemma 4 implies that a necessary condition for identification of the counterfactual choice probabilities is that we obtain the same $\widetilde{\varphi}^*(x)$ regardless of our choice of $u(a, x)$. Rewrite (3.15) as

$$\begin{aligned}
(3.20) \qquad \widetilde{\varphi}(x) &= \widetilde{u}(x) + \beta \left(\int \varphi(0, x') d\pi(x'|1, x) - \int \varphi(0, x') d\pi(x'|0, x) \right) \\
&= F_{\widetilde{\epsilon}}^{-1}(P(x)) - \widetilde{\delta}(x, P)
\end{aligned}$$

As in Hotz & Miller (1993), this binds $\widetilde{u}(x)$ as a function of $u(0, x)$ and the data. Substitute (3.18) into the analogous equation for $\widetilde{\varphi}^*(x)$:

$$\begin{aligned}(3.21)\qquad \widetilde{\varphi}^*(x) &= \widetilde{u}(x) + \beta\left(\int \varphi^*(0,x')d\pi^*(x'|1,x) - \int \varphi^*(0,x')d\pi^*(x'|0,x)\right) \\ &= \widetilde{\varphi}(x) - \beta\left(\int \varphi(0,x')d\pi(x'|1,x) - \int \varphi(0,x')d\pi(x'|0,x)\right. \\ &\quad \left. + \int \varphi^*(0,x')d\pi^*(x'|1,x) - \int \varphi^*(0,x')d\pi^*(x'|0,x)\right)\end{aligned}$$

As we place no restriction on π^* or $u(0,x)$, the bracketed term in (3.19) is, in general, non-zero. By Lemma 4, the resulting counterfactual choice probabilities are, in general, not point identified.

In order to provide a more precise condition, we now impose the discrete support assumption.

Lemma 5. *For a $K \times K$ Markov transition probability matrix, π, and $\beta \in [0,1)$, the matrix $I - \beta\pi$ is invertible, where I is the $K \times K$ identity matrix.*

Proof. For $\beta = 0$, invertibility trivially follows since the matrix in question is simply the $K \times K$ identity. For $\beta \in (0,1)$, following the proof in Hotz & Miller (1993), decompose

$I - \beta\pi$ into two matrices :

$$D_1 = \begin{bmatrix} 1-\beta\pi_{1,1} & 0 & \cdots & 0 \\ 0 & 1-\beta\pi_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1-\beta\pi_{K,K} \end{bmatrix}$$

$$D_2 = \begin{bmatrix} 0 & \beta\pi_{1,2} & \cdots & \beta\pi_{1,K} \\ \beta\pi_{2,1} & 0 & \cdots & \beta\pi_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \beta\pi_{K,1} & \beta\pi_{K,2} & \cdots & 0 \end{bmatrix}$$

where $\pi_{i,j}$ is the $(i,j)^{th}$ element of the matrix π. By construction, $I - \beta\pi = D_1 - D_2$. Since π is a Markov transition probability matrix, each of its elements is non-negative and is at most 1. As a result, with $\beta < 1$, the diagonal elements of D_1 are strictly positive, hence D_1 is invertible. Rewrite the matrix as

$$I - \beta\pi = D_1 - D_2 = D_1(I - D_1^{-1}D_2)$$

Let m_{ij} : the $(i,j)^{th}$ element of $D_1^{-1}D_2$, then we have

$$m_{ij} = \begin{cases} 0 & if\ i = j \\ \beta\pi_{i,j}/(1-\beta\pi_{i,i}) & if\ i \neq j \end{cases}$$

Since $\beta\pi_{i,j} \leq \beta(1-\pi_{i,i}) \leq 1 - \beta\pi_{i,i}$, we have $m_{i,j} \in [0,1]\ \forall(i,j)$, and

$$\sum_{j=1}^{J} m_{ij} = \frac{1}{(1-\beta\pi_{i,i})}\sum_{j=1}^{J}\beta\pi_{i,j} = \frac{\beta(1-\pi_{i,i})}{(1-\beta\pi_{i,i})} < 1$$

By Hadley (1973, p.118), $(I - D_1^{-1}D_2)$ is invertible. This, combined with the invertibility of D_1, imply the invertibility of $I - \beta\pi$. $\square$

The invertibility of $I - \beta\pi$ permits us to invert (3.15) and obtain $\varphi(0, X)$ as an explicit function of $u(0, X)$.

Proposition 6. *Assume discrete support, and define*

$$\Delta\Pi_1 := (\pi(1) - \pi(0))(I - \beta\pi(0))^{-1} - (\pi^*(1) - \pi^*(0))(I - \beta\pi^*(0))^{-1} \tag{3.22}$$

where $\pi(a)$ *is the transition matrix whose* $(i, j)^{th}$ *element is* $\pi(x' = x^i | a, x = x^j)$. *If* $\Delta\Pi_1$ *has any non-zero element, the counterfactual choice probabilities following a policy implementation shifting the transition probabilities is not point identified in the* $[0, 1]^K$ *space.*

Proof. Note, any choice of K-dimensional vector, $u(0, X)$, is consistent with the observed optimal choice probabilities, P. It remains to show that two different choices of $u(0, X)$ lead to different optimal choice probabilities under π^*. With discrete support assumption, we can rewrite $\varphi(a, X)$ as

$$\varphi(a, X) = u(a, X) + \beta\pi(a)\varphi(a, X)$$

and

$$\varphi(0, X) = (I - \beta\pi(0))^{-1}u(0, X)$$

The bracketed term in (3.19) becomes

$$
\begin{aligned}
& \int \varphi(0,x')d\pi(x'|1,x) - \int \varphi(0,x')d\pi(x'|0,x) \\
& + \int \varphi^*(0,x')d\pi^*(x'|1,x) - \int \varphi^*(0,x')d\pi^*(x'|0,x) \\
= \; & \Delta\Pi_1 u(0,X)
\end{aligned}
$$

Suppose the $(i,j)^{th}$ element of $\Delta\Pi_1$ is non-zero. Let u_1 be the K-dimensional zero vector and u_2 be the K-dimensional vector whose elements are all zeros except the j^{th} element, which is set to 1. It follows from (3.18) and (3.19) that

$$
\begin{aligned}
\widetilde{\varphi}_1^*(X) &= F_{\widetilde{\epsilon}}^{-1}(P(X)) - \widetilde{\delta}(X,P) - \beta\Delta\Pi_1 u_1 \\
&= F_{\widetilde{\epsilon}}^{-1}(P(X)) - \widetilde{\delta}(X,P) \\
&\neq F_{\widetilde{\epsilon}}^{-1}(P(X)) - \widetilde{\delta}(X,P) - \beta\Delta\Pi_1 u_2 = \widetilde{\varphi}_2^*(X)
\end{aligned}
$$

Since two different choices of $u(0,X)$ lead to two different vectors, $\widetilde{\varphi}_1^*(X)$ and $\widetilde{\varphi}_2^*(X)$, by Lemma 4, they are consistent with two different vectors of optimal choice probabilities, P_1^* and P_2^*. Given β, G, π, π^* and P, two structures, $b_1 = (\beta, F, u_1, \pi, \pi^*, P_1^*)$ and $b_2 = (\beta, F, u_2, \pi, \pi^*, P_2^*)$ are observationally equivalent where $P_1^* \neq P_2^*$. Therefore, the counterfactual choice probabilities are not point identified in the $[0,1]^K$ space. $\square$

This result does not imply that for any shift in transition probability that yields non-zero $\Delta\Pi_1$, we can span the entire $[0,1]^K$ space for the counterfactual choice probabilities by selecting $u(0,x)$ appropriately. However, this contrasts to Aguirregabiria's result, in which the counterfactual choice probabilities are point identified. Aguirregabiria's success is partly due to the fact that the impact of the additive change to the stage utility can be isolated, leaving the separation issue of stage utility and expectation unaddressed. In the policy implementation that shifts the transition probabilities, such an isolation was not possible, except for such special cases as $\pi = \pi^*$ (no shift in transition probabilities) or $\pi(1) = \pi(0)$ and $\pi^*(1) = \pi^*(0)$ (transition of the observable state vector is independent of the action choice before and after the shift in transition probabilities).

While it is clear that a non-zero $\Delta\Pi_1$ accompanies a shift in transition probabilities, the interpretation of this matrix is not straight forward. A more intuitive interpretation can be applied when the transition probabilities shift in such a way that $\pi(0)$ is preserved (i.e. $\pi(0) = \pi^*(0)$). In this case, from the definition (3.20), the condition that $\Delta\Pi_1 \neq 0$ is translated directly to $\pi(1) \neq \pi^*(1)$. The engine replacement example falls into this category, as the decision to replace the engine always resets the mileage to zero. The shift in transition probabilities occurs only when the agent decides not to replace.

3.4. Identification with Additional Data

Although the non-identification result in section 3 may be considered a set back, it leaves an opportunity to identify the model primitives in an entirely different way. In Aguirregabiria (2005b), the counterfactual choice probabilities were point identified in the $[0, 1]^K$ space without placing any constraint on the stage utility function. This implies that, as long as the data set is generated in accordance to the model, when the researcher observes the change in behavior once the policy is in place, the new choice probabilities will be consistent with the prediction made in Aguirregabiria, which, in turn, provides us with no additional information about the primitives of the model, particularly about the stage utility function. In contrast, if the policy implementation involves a shift in transition probabilities, observing a change in behavior will lead to a restriction on the stage utility function because, as examined in Proposition 6, different choices for $u(0, X)$ will lead to different post-policy implementation choice probabilities.

In section 2, we emphasized that Proposition 6 does not necessarily imply that we can span the entire P^* space by choosing $u(0, X)$ appropriately. In this context, the relevant question is, when we observe arbitrary P and P^* both in $[0, 1]^K$, whether and under what conditions we can find unique $u(0, X)$ that is consistent with the model specification.

For the sake of fomality, we redefine the model structure as $b = (\beta, F, u(0, X), \pi, \pi^*)$ and the reduced form of the model as (P, P, π^o, π^{o*}). The definitions for observational equivalence and of identification can be modified accordingly.

Proposition 7. *Suppose assumptions 1 through 6 hold and suppose (P, π) and (P^*, π^*) are identified from the data. Take $\Delta\Pi_1$ to be as in (3.20). Given the discount factor, β, and the distribution of the unobservable states, F, if*

$$Rank(\Delta\Pi_1) = K \tag{3.23}$$

then the stage utility function, $u(a, x)$, is identified.

Proof. Since we have the representation of $\widetilde{u}$ in (3.18), the identification of u is reduced to identification of $u(0, X)$. Fix $u(0, X)$. Given $P, P^* \in [0, 1]^K$, the utility differences consistent with P and P^* are respectively

$$\begin{aligned}\widetilde{u}(X) &= F_{\widetilde{\epsilon}}^{-1}(P(X)) - \widetilde{\delta}(X,P) - \beta(\pi(1)-\pi(0))(I-\beta\pi(0))^{-1}u(0,X) \\ \widetilde{u}^*(X) &= F_{\widetilde{\epsilon}}^{-1}(P^*(X)) - \widetilde{\delta}(X,P^*) - \beta(\pi^*(1)-\pi^*(0))(I-\beta\pi^*(0))^{-1}u(0,X)\end{aligned}$$

where the function $\widetilde{\delta}$ is determined independently of $u(0,X)$, as in (3.16) and (3.17). Since the agent's stage utility function is assumed unchanged, we require $\widetilde{u}(X) = \widetilde{u}^*(X)$, or, equivalently,

$$\Delta\Pi_1 u(0,X) = D$$

where $D = F_{\widetilde{\epsilon}}^{-1}(P(X)) - \widetilde{\delta}(X,P) - F_{\widetilde{\epsilon}}^{-1}(P^*(X)) + \widetilde{\delta}(X,P^*)$ is a constant vector obtained from the data observed. $Rank(\Delta\Pi_1) = K$ is necessary and sufficient to point identify $u(0,X)$. □

The proposition states that with enough variation in transition probabilities, the stage utility function is point identified, with the notion "enough" being described as the rank condition on the matrix we denoted $\Delta\Pi_1$. As before, for a shift in transition probabilities that leaves $\pi(0)$ unchanged, the condition (**??**) becomes

$$Rank(\pi(1) - \pi^*(1)) = K \tag{3.24}$$

While the interpretation of this condition is fairly straight forward, the requirement that the transition probability matrix shift in such a way that yields the difference in two transition matrices to have full rank may seem demanding. This requirement, however, remains invariant when depart from the binary choice case and introduce more actions.

Corollary 8. *Let* $A = \{0, \cdots, J-1\}$*. For each action choice,* $j \in \{1, \cdots, J-1\}$*, define*

$$\Delta\Pi_j := (\pi(j) - \pi(0))(I - \beta\pi(0))^{-1} - (\pi^*(j) - \pi^*(0))(I - \beta\pi^*(0))^{-1} \tag{3.25}$$

and

$$\Delta\Pi = \begin{bmatrix} \Delta\Pi_1 \\ \Delta\Pi_2 \\ \vdots \\ \Delta\Pi_{J-1} \end{bmatrix} \tag{3.26}$$

Given the discount factor, β*, and the distribution of the unobservable states,* F*, if*

$$Rank(\Delta\Pi) = K \tag{3.27}$$

then the stage utility function, $u(a, x)$*, is identified.*

Proof. It follows from (**??**) that

$$v(a, x) = v(0, x) + q(a, x, P; F) \qquad \textit{and} \qquad v^*(a, x) = v^*(0, x) + q(a, x, P^*; F) \tag{3.28}$$

Let

$$R(z; F) := \int \max_{j \in A}(z + \epsilon_j - \epsilon_0) dF(\epsilon) \tag{3.29}$$

From the definition of the conditional value function in (3.6), we obtain :

$$
\begin{aligned}
(3.30) \qquad v(a,x) &= u(a,x) + \beta E\left[\max_{j\in A}\{v(j,x') + \epsilon'(j)\}|a,x\right] \\
&= u(a,x) + \beta E\left[\{v(0,x') + \epsilon'(0) + \max_{j\in A}\{\widetilde{v}(j,x') + \widetilde{\epsilon}'(j)\}|a,x\right] \\
&= u(a,x) + \beta E\left[v(0,x')|a,x\right] + \beta E[R(\widetilde{v}(j,x');F)|a,x] \\
&= u(a,x) + \beta E\left[v(0,x')|a,x\right] + \beta E[R(q(j,x',P;F);F)|a,x]
\end{aligned}
$$

Combining (**??**) and (**??**), we get :

$$
\begin{aligned}
(3.31) \qquad \widetilde{u}(a,x) &= q(a,x,P;F) \\
&\quad -\beta\{E[R(q(j,x',P;F);F)|a,x] - E[R(q(j,x',P;F);F)|0,x]\} \\
&\quad -\beta\{E[v(0,x')|a,x] - E[v(0,x')|0,x]\}
\end{aligned}
$$

where $q(a,P;F) - \beta\{E[R(q(j,x',P;F);F)|a,x] - E[R(q(j,x',P;F);F)|0,x]$ is obtained from the observed P, independently of $u(0,X)$.

Now, assume discrete support. By Lemma 5, (**??**) can be written as

$$v(0,X) = u(0,X) + \beta\pi(0)v(0,X) + \beta E[R(q(j,x',P;F);F)|0,X]$$

or

$$v(0,X) = (I - \beta\pi(0))^{-1}\{u(0,X) + \beta E[R(q(j,x',P;F);F)|0,X]\}$$

It follows that

$$\begin{aligned}
\widetilde{u}(a, X) &= q(a, X, P; F) \\
&-\beta\{E[R(q(j, x', P; F); F)|a, X] \\
&-E[R(q(j, x', P; F); F)|0, X]\} \\
&+\beta(\pi(a) - \pi(0))(I - \beta\pi(0))^{-1} \\
&\times\{u(0, X) + \beta E[R(q(j, x', P; F); F)|0, X]\}
\end{aligned}$$

Analogously :

$$\begin{aligned}
\widetilde{u}^*(a, X) &= q^*(a, X, P^*; F) \\
&-\beta\{E^*[R(q^*(j, x', P^*; F); F)|a, X] \\
&-E^*[R(q^*(j, x', P^*; F); F)|0, X]\} \\
&+\beta(\pi^*(a) - \pi^*(0))(I - \beta\pi^*(0))^{-1} \\
&\times\{u(0, X) + \beta E^*[R(q^*(j, x', P^*; F); F)|0, X]\}
\end{aligned}$$

For notational simplicity, let combine all the terms that are determined from the data independently of $u(0, X)$ as

$$\begin{aligned} D(a, X, P, P^*; F) &= q^*(a, X, P^*; F) - q(a, X, P; F) \\ &-\beta\{E^*[R(q^*(j, x', P^*; F); F)|a, X] \\ &-E^*[R(q^*(j, x', P^*; F); F)|0, X]\} \\ &+\beta\{E[R(q(j, x', P; F); F)|a, X] \\ &-E[R(q(j, x', P; F); F)|0, X]\} \\ &+\beta(\pi^*(a) - \pi^*(0))(I - \beta\pi^*(0))^{-1} \\ &\times\beta E^*[R(q^*(j, x', P^*; F); F)|0, X] \\ &-\beta(\pi(a) - \pi(0))(I - \beta\pi(0))^{-1} \\ &\times\beta E[R(q(j, x', P; F); F)|0, X] \end{aligned}$$

The requirement that $\widetilde{u}(a, X) = \widetilde{u}^*(a, X)$ for all $a \in A$ yields

$$((\pi(a) - \pi(0))(I - \beta\pi(0))^{-1} - (\pi^*(a) - \pi^*(0))(I - \beta\pi^*(0))^{-1})u(0, X)$$

$$= D(a, X, P, P^*; F)$$

or

$$\Delta\Pi_j u(0, X) = D(j, X, P, P^*; F) \qquad \forall j \in \{1, \cdots, J-1\}$$

The system of linear equations can be represented as

$$\Delta\Pi u(0,X) = \begin{bmatrix} D(1,X,P,P^*;F) \\ D(2,X,P,P^*;F) \\ \vdots \\ D(J-1,X,P,P^*;F) \end{bmatrix} \tag{3.32}$$

A unique vector, $u(0,X)$, satisfying this system of linear equations can be recovered if $Rank(\Delta\Pi) = K$. $\square$

Intuitively stated, for each of the action $j \in \{1,\cdots,J-1\}$, we obtain K restrictions as in the binary choice model. In total, we are restricted by $(J-1)\times K$ linear equations, where the number of unknowns to be identified remains fixed at K. For the purpose of identification, having multiple action choices is equivalent to having multiple shifts in transition probabilities in the binary choice model.

3.5. Discussion

The application of this approach to the single-agent dynamic optimization model is limited in the sense the identification result requires that the researcher observe a very specific exogenous change in the environment in which the agent is situated. However, the approach is still interesting as it bears certain implication on the identification of the multiple agent model.

In a typical multiple agent dynamic discrete choice model, there are I agents, indexed by $i = 1,\cdots,I$, playing a stage game with its rivals in each period. The outcome of the stage game is determined by the stage vector, $s = (x,\epsilon)$, and the profile of action choices,

$a = (a^i)_{i=1}^I$. Each agent has an additively time-separable utility function defined over a sequence of state vectors and action profiles, $\{(a_{t+\tau}, s_{t+\tau})\}_{\tau=0}^{\infty}$:

$$\sum_{\tau=0}^{\infty} \beta^\tau U^i(a_{t+\tau}, s_{t+\tau}) \tag{3.33}$$

where $\beta \in [0, 1)$ is the common discount factor, and $U^i(a, s)$ is agent i's payoff from the stage game given the state s and the action profile a. In contrast to the single agent model, the evolution of the state vector is governed by the action profile, a, rather than the action choice of any single agent. In addition to the assumptions 1 through 6 for the single agent model, we assume that agents are identical and drop the i superscript. Note that there is no private information. At each period, the realization of ϵ, as well as x, is commonly known to the agents, albeit unobservable to the researcher.

The decision making process of each agent is not different from that of the single-agent model. Taking the rivals' strategies, a^{-i}, and the current state vector, s_t, as given, agent i chooses his own action sequence to maximize the expectation of (3.24). We restrict to optimal strategies that are dependent only on the payoff-relevant histories. Since the stage utility, and the evolution of the state vector depends only on the current state, we are effectively confined to the class of Markov strategies. A profile of Markov strategies, $(a^1, \cdots, a^I)$, forms a Markov Perfect Equilibrium (MPE)[1], if given that the rivals' behave in accordance to a^{-i} at each period, a^i maximizes i's expected intertemporal utility at any time assuming that agent i, himself, will behave in accordance to a^i in the future. In practice, we restrict our attention to the class of Symmetric Markov Perfect Equilibrium.

[1]The justification for the use of Markov Perfect Equilibrium appears in Maskin & Tirole (1988).

One of the main hurdles in adopting the notion of MPE is the potential multiplicity of equilibria. In contrast to the single-agent model, the stage utility and the evolution of the state vector depends on the rivals' strategies. As a result, each agent's optimal action choice depends on the agent's expectation about the behavior of his rivals. Equilibrium multiplicity arises even though there is no variation in the stage utility function and the process governing the evolution of the state vector given the current state and the action profile, because there can be different expectations about rivals' behaviors and the optimal strategy resulting from such an expectation that are consistent with the concept of MPE.

In practice, it is often assumed that the transition probability function, $F(x\prime|a, x)$, is identified from the data, which consist of observations obtained from a cross section of markets. Implicit in this practice is the assumption that the same equilibrium is played in each of the markets that we acquire data from. If, on the other hand, different equilibria are played in different markets, the transition probability function estimated using the observations aggregated across markets playing different equilibria has no clear meaning. This problem is first recognized in Pakes (????). Berry & Tamer (2006) suggests that we resolve this issue by including the equilibrium selection function, which is supplied from outside of the model.

The results outlined in section 4 proposes an alternative approach. Assuming that the agents in different markets are identical (i.e. they have the same stage utility function), multiple equilibria are analogous to the agents with the same utility function behaving differently when faced with different transition probabilities that govern the evolution of the state variables. The variation in transition probabilities across markets playing different equilibria is obtained because (i) the transition probabilities are the joint product of the common underlying process governing the evolution of the state vector given the current state and the profile of actions and the agents' expectation about the rivals' behaviors, and (ii) the agents in different markets have different expectations about their rivals' behaviors. In short, we obtain a shift in transition probabilities due to the variation in expectation, without an exogenous change in the transition process. With enough variation in transition probabilities across markets, the result in section 4 applies. As a consequence, we can take advantage of the equilibrium multiplicity to identify the structural component - the stage utility function - which was previously not identified.

Limitations and Potential Extensions

We may encounter a practical problem when we attempt to implement this approach with a finite sample. In the multiple agent model, it is important to identify which equilibrium is being played in each of the markets. Under the assumption of stationarity, this is done by estimating the transition probabilities separately in each market. With an infinite sample, we do identify the transition probabilities exactly for each market, and we can attribute the variation in transition probabilities across markets to the variation in agents' expectations in different markets. With a finite sample, however, the transition probabilities can only be estimated with some error. As a result, when we observe a variation in transition probabilities across two different markets, we cannot be sure whether the variation is a result of agents' playing different equilibria in two markets, or simply a by-product of the estimation error. On the upside, the hypothesis that the agents are playing two different equilibria should be testable with the data at hand.

The same problem may arise in the single-agent model as well. In fact, the conditions for the propositions in sections 3 and 4 are stated in terms of the rank of a particular matrix, which is, in turn, a combination of transition matrices. When the estimation error is associated only with the observed choice probabilities (i.e. the transition probabilities are identified exactly), then propositions 6 and 7 follow without modification. It is possible that in corollary 8, we have $rank(\Delta\Pi) = K$ and $rank([\Delta\Pi \vdots D]) = K + 1$, but, in this case, we can envision an estimator for $u(0, X)$ that minimizes the distance between $\Delta\Pi u(0, X)$ and $D(X, P, P^*; F)$ in (**??**). If, on the other hand, there is an estimation error associated with the observed transition probabilities, then how to interpret the fact that $\Delta\Pi$ calculated using the estimated transition probabilities has rank K or less than K is not straight forward at this stage. These problems may be less severe when we have a "long" panel (i.e. rich in time series data).

As a potential extension, we should consider whether the change in behavior following a shift in transition probabilities has an identifying power when the distribution of unobservable state vector is unknown. It is conjectured that certain shape restriction on the stage utility function is required in order to separate the unknown distribution from the differential value function. The analysis may be complicated by the fact that such a restriction could potentially have an impact on the separation of the stage utility function from the agent's expectation, which has been the main topic of our discussion in this paper.

Within the multiple agent framework, it is worth noting that the variation in transition probabilities is not unrestricted (i.e. it has to be consistent with the notion of MPE). This contrasts to the exogenous shift in the single-agent model, in which we had the full freedom to set the new transition probabilities. This requirement that transition probabilities be consistent with the notion of MPE could potentially help us to determine whether two markets are playing different equilibria when we can only estimate the transition probabilities with some error.

3.6. Conclusion

In this paper, we have examined whether and under what conditions we can identify the stage utility function of the dynamic discrete choice model when we are given the opportunity to observe the change in the agent's behavior following a known policy implementation. A sufficient variation in transition probabilities, rather a variation in stage utility function, is needed in order to identify the stage utility function using this additional data, where the notion of "a sufficient variation" can be summarized as the matrix, $\Delta\Pi$, having the full rank. This result can be applied to the multiple agent models, in which agents play different equilibria in a cross section of structurally identical markets because they have different expectations about how their rivals behave. The limitations and potential extensions were also discussed.

References

[1] Aguirregabiria, V., "The Dynamics of Markups and Inventories in Retailing Firms," *The Review of Economic Studies*, Vol. 66, No. 2. (Apr., 1999), pp. 275-308

[2] ————, "Nonparametric Identification of Behavioral Responses to Counterfactual Policy Interventions in Dynamic Discrete Decision Processes," *Economics Letters*, Vol. 87, (2005), pp. 393-398

[3] ————, "Another Look at the Identification of Dynamic Discrete Decision Processes," working paper, (2005)

[4] Aguirregabiria, V. & P. Mira, "Swapping the Nested Fixed Point Algorithm : A Class of Estimators for Markov Decision Models," *Econometrica*, Vol. 70, No. 4, (Jul., 2002), pp. 1519-1543

[5] Bell, D.R. & J.M. Lattin, "Shopping Behavior and Consumer Preference for Store Price Format: Why 'Large Basket' Shoppers Prefer EDLP," *Marketing Science*, Vol. 17, No. 1 (1998), pp. 66-88

[6] Benkard, C. L., "A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft," *Review of Economic Studies*, Vol. 71, No. 3, (Jul., 2004), pp. 581-611

[7] Berry, S. & E. Tamer, "Identification in Models of Oligopoly Entry," working paper, (2006)

[8] Blattberg, R.C. & S.A. Neslin, *Sales Promotions: Concepts, Methods, and Strategies.* Englewood Cliffs, N.J.: Prentice Hall, 1990

[9] Chevalier, J., A. Kashyap and P. Rossi, "Why Don't Prices Rise during Periods of Peak Demand? Evidence from Scanner Data," *American Economic Review*, Vol. 93, No. 1, (Mar., 2003), pp. 15-37

[10] Chiang, J., "Competing Coupon Promotions and Category Sales," *Marketing Science*, Vol. 14, No. 1. (1995), pp. 105-122

[11] Conlisk, J., E. Gerstner & J. Sobel, "Cyclic Pricing by a Durable Goods Monopolist," *The Quarterly Journal of Economics*, Vol. 99, No. 3. (Aug., 1984), pp. 489-505

[12] Corts, K.S., "Third-Degree Price Discrimination in Oligopoly: All-Out Competition and Strategic Commitment," *The RAND Journal of Economics*, Vol. 29, No. 2. (Summer, 1998), pp. 306-323

[13] Deighton, J., C.M. Henderson & S.A. Neslin, "The Effects of Advertising on Brand Switching a nd Repeat Purchasing," *Journal of Marketing Research*, Vol. 31, No. 1. (Feb., 1994), pp. 28-43

[14] Eckstein, Z. & K. Wolpin, "Dynamic Labor Force Participation of Married Women and Endogenous Work Experience," *Review of Economic Studies*, Vol. 56, No. 3, (Jul. 1982), pp. 375-390

[15] Erdem, T., M.P. Keane & B. Sun, "Missing Price and Coupon Availability Data in Scanner Panels: Correcting for the Self-Selection Bias in Choice Model Parameters," *Journal of Econometrics*, Vol. 89 (1999) pp. 177-196

[16] Gonul, F. & K. Srinivasan, "Estimating the Impact of Consumer Expectations of Coupons on Purchase Behavior: A Dynamic Structural Model," *Marketing Science*, Vol. 15, No. 3. (1996), pp. 262-279

[17] Hartman, W.R., "Intertemporal Effects of Consumption and Their Implications for Demand Elasticity Estimates," *Quantitative Marketing and Economics*, Vol. 4, No. 4. (2006) pp. 325-349

[18] Hausman, J., "Specification Tests in Econometrics," *Econometrica*, Vol. 46, No. 6., (Nov., 1978), pp. 1251-1271

[19] Hendel, I. & A. Nevo, "The Post-Promotion Dip Puzzle: What Do the Data Have to Say?" *Quantitative Marketing and Economics*, Vol. 1, No. 4. (2003) pp. 409-424

[20] ————, "Sales and Consumer Inventory," *The RAND Journal of Economics*, Vol. 37, No. 3. (Autumn, 2006), pp. 541-561

[21] ————, "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, Vol. 74, No. 6. (Nov., 2006), pp. 1637–1673

[22] Hosken, D. & D. Reiffen, "How Retailers Determine Which Products Should Go on Sale : Evidence From Store-Level Data," *Journal of Consumer Policy*, Vol. 27, No. 2. (Jun. 2004), pp. 141-177.

[23] ————, "Pricing Behavior of Multiproduct Retailers," working paper (Dec. 2004)

[24] Hotz, J. & R. Miller, "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, Vol. 60, No. 3, (Jul., 1993), pp. 397-429

[25] Inman, J.J. & L. McAlister, "Do Coupon Expiration Dates Affect Consumer Behavior?" *Journal of Marketing Research*, Vol. 31, No. 3. (Aug., 1994), pp. 423-428

[26] Kanetkar, V., C.B. Weinberg & D.L. Weiss, "Price Sensitivity and Television Advertising Exposures: Some Empirical Findings," *Marketing Science*, Vol. 11, No. 4. (Autumn, 1992), pp. 359-371

[27] Lee, Yongbae, "Habit Persistence, Its Reversion and the Promotional Effect of Coupons," memeo (Nov. 2007)

[28] Lee, J. & M.G. Brown, "Coupon Redemption and the Demand for Frozen Concentrated Orange Juice: A Switching Regression Analysis," *American Journal of Agricultural Economics*, Vol. 67, No. 3. (Aug., 1985), pp. 647-653

[29] MacDonald, J., "Demand, Information, and Competition : Why Do Food Prices Fall at Seasonal Demand Peaks?" *Journal of Industrial Economics*, Vol. 48, No. 1, (Mar., 2000) pp. 27-45

[30] Magnac, T. & D. Thesmar, "Identifying Dynamic Discrete Decision Processes," *Econometrica*, Vol. 70 No. 2, (Mar., 2002), pp. 801-816

[31] Manski, C., "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, Vol. 3, No. 3, (1975), pp. 205-228

[32] ————, "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, Vol. 27, No. 3, (1985), pp. 313-333

[33] ————, "Dynamic Choice in Social Settings: Learning from the Experiences of Others," *Journal of Econometrics*, Vol. 58, No. 1-2, (1993), pp. 121-136

[34] Matzkin, R., "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models," *Econometrica*, Vol. 60, No. 2, (Mar., 1992), pp. 239-270

[35] ————, "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics*, Vol. 58, No. 1-2, (Jul., 1993), pp. 137-168

[36] Matzkin, R. & N. Roussanov, "Nonparametric Identification" in Heckman, J.J., Leamer, E.E. (Eds.), *Handbook of Econometrics*, Vol. 6. The Netherlands, North-Holland, 2006

[37] Narasimhan, C., "A Price Discrimination Theory of Coupons," *Marketing Science*, Vol. 3, No. 2. (Spring, 1984), pp. 128-147

[38] Neslin, S.A., "Market Response Model for Coupon Promotions," *Marketing Science*, Vol. 9, No. 2. (Spring, 1990), pp. 125-145

[39] Neslin, S.A. & R.W. Shoemaker, "A Model for Evaluating the Profitability of Coupon Promotions," *Marketing Science*, Vol. 2, No. 4. (Autumn, 1983), pp. 361-388

[40] Nevo, A., L. Einav & E. Leibtag, "Errors in Self-Reported Data: A Cross-Validation of Homescan Data," memio, Northwestern University

[41] Nevo, A. & C. Wolfram, "Why Do Manufacturers Issue Coupons? An Empirical Analysis of Breakfast Cereals," *The RAND Journal of Economics*, Vol. 33, No. 2. (Summer, 2002), pp. 319-339

[42] Pakes, A. (1986) "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54, 755-784

[43] Pakes, A., M. Ostrovsky & S. Berry (2005) "Simple Estimators of the Parameters in Discrete Dynamic Games (with entry/exit)," working paper

[44] Pesendorfer, Martin, "Retail Sales: A Study of Pricing Behavior in Supermarkets," *Journal of Business*, Vol. 75, No. 1, (Jan., 2002), pp. 33-66.

[45] Rosenthal, Robert, "A Model in Which an Increase in the Number of Sellers Leads to a Higher Price," *Econometrica*, Vol. 48, No. 6, (Nov., 1980), pp. 1575-1579

[46] Rotemberg, Julio and Garth Saloner, "A Supergame-Theoretic Model of Price Wars during Booms," *American Economic Review*, Vol. 76, No. 3 (Jun. 1986) pp. 390-407

[47] Rust, J., (1987) "Optimal Replacement of GMC Bus Engines; An Empirical Model of Harold Zurcher," *Econometrica*, 55 (No. 5) pp. 999-1033

[48] ————, (1994) "Structural Estimation of Markov Decision Processes," in Engle, R.E., McFadden, D. (Eds.), *Handbook of Econometrics*, vol. 4. Amsterdam, The Netherlands, North-Holland

[49] Shilony, Yuval, "Mixed Pricing in Oligopoly," *Journal of Economic Theory*, Vol. 14, No. 2 (1977) pp. 373-388

[50] Sobel, J., "The Timing of Sales," *The Review of Economic Studies*, Vol. 51, No. 3. (Jul., 1984), pp. 353-368

[51] ————, "Durable Goods Monopoly with Entry of New Consumers," *Econometrica*, Vol. 59, No. 5. (Sep., 1991), pp. 1455-1485

[52] Sun, B., "Promotion Effect on Endogenous Consumption," *Marketing Science*, Vol. 24, No. 3, (Summer 2005), pp. 430-443

[53] Sun, B., S. Neslin & K. Srinivasan, "Measuring the Impact of Promotions on Brand Switching Under Rational Consumer Behavior," *Journal of Marketing Research*, Vol 40. (2003), pp. 389-405

[54] Sweeting, A. (2004) "Coordination Games, Multiple Equilibria, and the Timing of Radio Commercials," working paper

[55] Varian, Hal R., "A Model of Sales," *American Economic Review*, Vol. 70, No. 4, (Sept., 1980), pp. 651-659

[56] Ward R.W. & J.E. Davis, "A Pooled Cross-Section Time Series Model of Coupon Promotion," *American Journal of Agricultural Economics*, Vol. 60, No. 3. (Aug., 1978), pp. 393-401

[57] Warner, Elizabeth and Robert Barsky, "The Timing and Magnitude and Retail Store Markdowns: Evidence from Weekends and Holidays," *Quarterly Journal of Economics*, Vol. 110, No. 2, (May, 1995), pp. 321-52