# Mutual Interdependence across Consumers and Firms

by

Hee Mok Park

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business Administration)
in The University of Michigan
2013

Doctoral Committee:

Professor Puneet Manchanda, Chair
Professor Daniel A. Ackerberg
Professor Peter J. Lenk
Assistant Professor Yesim Orhun
Assistant Professor Srinivasaraghavan Sriram

# ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all of my dissertation committee members, Srinivasaraghavan Sriram, Yesim Orhun, Peter Lenk, and Daniel Ackerberg. I am blessed to have such an outstanding group of academics serving on my committee. I would also like to dedicate this work to my parents, June Shic Park and Kyung Il Shin; to my wife, Sun Young Ahn; and to my daughter, Dahnbi Park for their love and encouragement during the past years. Last but far from least, I would like extend my sincerest appreciation to my advisor Puneet Manchanda for his guidance, encouragement, and continuous support throughout the course of this work.

# TABLE OF CONTENTS


**ACKNOWLEDGEMENTS** . . . . . . . . . . . . . . . . . . . . . . . . . . ii

**LIST OF FIGURES** . . . . . . . . . . . . . . . . . . . . . . . . . . . v

**LIST OF TABLES** . . . . . . . . . . . . . . . . . . . . . . . . . . . vi

**ABSTRACT** . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . vii

**CHAPTER**

**I. General Introduction** . . . . . . . . . . . . . . . . . . . . . . . 1

**II. When Harry Bet with Sally: An Empirical Analysis of Peer Effects in Casino Gambling Behavior** . . . . . . . . . . . . . . . 4

2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . 4
2.2 Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
2.2.1 Model Development . . . . . . . . . . . . . . . . . . 9
2.2.2 Identification . . . . . . . . . . . . . . . . . . . . . 10
2.2.3 Estimation . . . . . . . . . . . . . . . . . . . . . . . 12
2.3 Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
2.3.1 Peer Identification . . . . . . . . . . . . . . . . . . 17
2.3.2 Evidence for Peer Effects in the Data . . . . . . . . 18
2.4 Model Specification . . . . . . . . . . . . . . . . . . . . . 20
2.5 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
2.5.1 Parameter Estimates . . . . . . . . . . . . . . . . . . 24
2.5.2 Heterogeneity and the Asymmetry in the Endogenous Peer Effect . . . . . . . . . . . . . . . . . . . . 25
2.6 Managerial Implications . . . . . . . . . . . . . . . . . . . 26
2.6.1 Tradeoff between the Endogenous and Exogenous Peer Effects . . . . . . . . . . . . . . . . . . . . . . . . . 26
2.6.2 Implication of Negative Peer Presence Effect . . . . 27
2.6.3 Implications for Resource Allocation . . . . . . . . . 27

2.7 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28
2.8 Appendix . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31

**III. Measuring the Agglomeration Effect on Consumers Store Choice** 45

3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
3.2 Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 48
3.2.1 The Industry . . . . . . . . . . . . . . . . . . . . . . . . . 48
3.2.2 Data Description . . . . . . . . . . . . . . . . . . . . . . 50
3.2.3 Data Patterns . . . . . . . . . . . . . . . . . . . . . . . . 50
3.3 Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53
3.4 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
3.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57

**IV. General Conclusion** . . . . . . . . . . . . . . . . . . . . . . . 63

**BIBLIOGRAPHY** . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

# LIST OF FIGURES

**Figure**


2.1 Casino Floor Plan . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42

2.2 Heterogeneity in Endogenous Peer Effect . . . . . . . . . . . . . . . . 43

2.3 Asymmetry in Endogenous Peer Effect . . . . . . . . . . . . . . . . . 43

2.4 Negative Peer Presence Effect . . . . . . . . . . . . . . . . . . . . . 44

3.1 Avg. No. of Visits by an household to Grocery Stores . . . . . . . . 60

3.2 Avg. No. of Visits by an household to Toy Stores . . . . . . . . . . 61

3.3 Avg. No. of Visits by an household to Electronic Stores . . . . . . . 61

3.4 Avg. No. of Visits by an household to Book Stores . . . . . . . . . 61

3.5 Avg. No. of Visits by an household to Multicategory Stores . . . . . 62

3.6 Avg. No. of Visits by an household to Office Product Stores . . . . 62

3.7 Avg. No. of Visits by an household to Home Improvement Stores . 62

# LIST OF TABLES

**Table**


2.1 Descriptive Statistics . . . . . . . . . . . . . . . . . . . . . . . . . 38

2.2 Play Zones within the Casino . . . . . . . . . . . . . . . . . . . . . . 39

2.3 Validity Check on the Excluded Variable . . . . . . . . . . . . . . . . 39

2.4 Parameter Estimates . . . . . . . . . . . . . . . . . . . . . . . . . . 40

2.5 Distribution of Endogeous Peer effect . . . . . . . . . . . . . . . . . 40

2.6 Distribution of Asymmetry of Endogenous Peer effect . . . . . . . . 40

2.7 Simulation Study Result 1 . . . . . . . . . . . . . . . . . . . . . . . 41

2.8 Simulation Study Result 2 . . . . . . . . . . . . . . . . . . . . . . . 41

3.1 Descriptive Statistics for households' shopping behavior . . . . . . . 58

3.2 Descriptive Statistics for households' shopping behavior across grocery and non-grocery shopping . . . . . . . . . . . . . . . . . . . . . 58

3.3 Number of grocery stores having each combination of non-grocery stores within a one-mile distance . . . . . . . . . . . . . . . . . . . 58

3.4 Coefficient Estimates . . . . . . . . . . . . . . . . . . . . . . . . . . 59

3.5 Comparison between models with and without store fixed effects . . 59

3.6 Brand names of stores considered for each category . . . . . . . . . 60

# ABSTRACT

Mutual Interdependence across Consumers and Firms

by

Hee Mok Park

Chair: Professor Puneet Manchanda

In markets where multiple agents coexist, decisions across agents can be interdependent. A consumer's decisions about what product to purchase or how much to consume could be influenced by the decisions his or her peers make (e.g., family members, friends, and neighbors). Likewise, when firms make decisions about actions such as entry, pricing, product development, and location, they also consider how other firms' behaviors could affect their sales. This research explores how to quantify these interdependences across both consumers and firms based on behavioral data (also known as "revealed choice data"), while also accounting for confounding factors using econometrics methods.

The first essay explores the interdependence across consumers. In this study, we extend the previous literature on consumption in various settings by accounting for exogenous factors that could change the peer's behavior (the exogenous peer effect) and whether the peer is present at the time of consumption but does not consume (the peer presence effect) in addition to the typically modeled endogenous peer effect (how one's behavior is influenced directly by the peer's behavior). We develop a simultaneous equation model that allows us to identify all three peer effects

simultaneously and apply it to behavioral data from a casino gambling setting.

In the second essay, we extend the context to measuring the interdependence among firms' decisions by focusing on the location choices of retailers. An agglomeration of retailers providing different goods can create positive spillovers by attracting multi-purpose shoppers. An interesting notion is that these multi-purpose shoppers may travel farther to visit a store with other retailers nearby than patronizing a standalone store. In the present study, we quantify the agglomeration effect as the increase in the catchment area for retailers. We develop a multinomial choice model and apply it to a consumer store choice data. We measure the increase in consumers' likelihood of visiting a particular grocery store during peak demand periods for non-grocery stores located in the vicinity of the grocer. Based on this measure, we then infer the increase in the catchment area that a retailer could enjoy by locating next to other types of stores.

# CHAPTER I

# General Introduction

In markets where multiple agents coexist, decisions across agents can be interdependent. A consumer's decisions about what product to purchase or how much to consume could be influenced by the decisions his or her peers make (e.g., family members, friends, and neighbors). Likewise, when firms make decisions about actions such as entry, pricing, product development, and location, they also consider how other firms' behaviors could affect their sales. Measuring this interdependence across consumers is important as it enables marketing managers to optimize marketing resources across consumers. For example, they can allocate more promotions to a consumer who is more influential than others. Also, quantifying how other firms' behaviors affect profit can lead managers to make decisions more effectively. For example, retailers (e.g., grocery store) might want to locate close to another retailer that carries different goods (e.g., electronic store) in order to attract multi-purpose shoppers.

Because many firms, including marketing research companies, collect data regarding both their customers and other firms, ample data sources are available to measure interdependence and use such information in decision making. Isolating such interdependence from behavioral data (also known as "revealed choice data") is difficult, however, because there could be many confounding factors. This research explores how to quantify interdependence across both consumers and firms based on behavioral data, while also accounting for confounding factors using statistical or econometrics methods. Based on the estimates of interdependence, we provide implications regarding how managers can use these measures to make effective marketing decisions.

The first essay of the dissertation explores the interdependence across consumers. In many consumption settings (e.g., restaurants, casinos, and online gaming etc.), individuals consume products either alone or with their peers (i.e., friends and family members). In this study, we extend the previous literature on consumption in various settings by accounting for exogenous factors that could change the peer's behavior (the exogenous peer effect) and whether the peer is present at the time of consumption but does not consume (the peer presence effect) in addition to the typically modeled endogenous peer effect (how one's behavior is influenced directly by the peer's behavior).

We develop a simultaneous equation structural model that decomposes overall peer effect into three: endogenous peer effect, exogenous peer effect and peer presence effect. We then estimate the model under Hierarchical Bayes framework while resolving "tricky" ratio estimation problem by applying the MELO (Minimum Expected Loss) approach. Our data comprise detailed gambling activities at the individual level for a single casino over a two-year period. Our results show that all three types of peer effects exist. The endogenous peer effect is positive while the other two are negative. These effects vary across individuals and there is considerable asymmetry in the size of these effects within peer groups. Our results suggest that accounting for these peer effects simultaneously and identifying them at an individual level could help marketing managers draw better guidelines for promotion policies.

In the second essay, we extend the context to measuring the interdependence among firms' decisions by focusing on the location choices of retailers. An important criterion for deciding where to locate a store is the nearby existence of other retailers that provide different types of products. A cluster of retailers that provides different goods can create positive spillovers by attracting multi-purpose shoppers. An interesting notion is that these multi-purpose shoppers may travel farther to visit a store with other retailers nearby than patronizing a standalone store. Retailers can thus increase the catchment area by locating next to a different type of retailer. In the present study, we quantify the agglomeration effect as the increase in the catchment area, which retailers can use to make better decision when locating new stores.

To investigate this objective, we develop and estimate an empirical model of consumers' store choice in the grocery industry. We use a panel dataset of household shopping behavior across multiple types of retailers. Using a multinomial store choice model, we measure the increase in consumers' likelihood of visiting a particular grocery store during peak demand periods for non-grocery stores located in the vicinity of the grocer. Based on this measure, we then infer the increase in the catchment area that a retailer could enjoy by locating next to other types of stores.

A challenge in empirically identifying the agglomeration effect is that retail clusters are likely to be located in areas that are intrinsically more attractive to businesses. We address this issue by controlling for location unobservables with store fixed effects. Another challenge is to infer the causal impact of demand for a store carrying non-grocery items on the demand for grocery stores. We resolve this problem by relying on the asymmetric increase in demand for non-grocery items (e.g., toys or electronics) during the peak demand period (e.g., the holiday shopping period) for that product as an exogenous shock to households' retail visits. Results show that a retailer could enjoy significant increase in the catchment area when they co-locate with other types of retailers.

# CHAPTER II

# When Harry Bet with Sally: An Empirical Analysis of Peer Effects in Casino Gambling Behavior

## 2.1 Introduction

In many consumption settings (e.g., restaurants, casinos, theme parks), individuals consume products either alone or with their peers (e.g., friends and/or family members). In such settings, it is likely that through social influence, a consumer's decision on what to purchase or how much to consume is influenced by the purchase or consumption decisions of her peers. There has been much research in marketing that documents the effect(s) of the peer's behavior on the focal consumer's behavior. Some recent examples are *Hartmann* (2010), *Zhang* (2010) and *Yang et al.* (2006). This document effect is the well known endogenous peer effect. The focus of this literature has been to provide methods to distinguish the true causal (endogenous) peer effect from other confounds such as endogenous group formation ("homophily"), correlated unobservables and simultaneity *Manski* (1993). However, a consumer could not only be affected by the peer's behavior but also by events that influence the peer to change his/her behavior. Specifically, suppose the peer gets a demand shock (e.g., a marketing promotion) that leads to her increasing her consumption behavior. While the endogenous peer effect will capture the influence of this change in the peer's consumption on the focal consumer's behavior, there is a possibility that the focal consumer's observing the peer getting the promotion could affect her (the focal consumer's) behavior directly. Another mechanism by which social influence could operate could be when the peer is physically present but does not engage in the behavior under question. In other word, the peer's mere presence could directly affect the focal consumer's consumption behavior.

In this paper, we take a deeper look at joint consumption by allowing for multiple types of peer effects that could influence the focal consumer. Besides the endogenous peer effect, we allow for the two other effects described above. We label the first as the *exogenous peer effect* and the second as the *peer presence effect.*[1] We develop a structural model that allows us to identify all three effects simultaneously and apply it to behavioral data from a casino setting. The model takes the form of a simultaneous equation model. Our data comprise detailed gambling activity for a panel of individuals at a single casino over a two-year period. Our results show that all three types of peer effects exist. These effects vary across individuals and exhibit considerable asymmetry within pairs of peers. The results also indicate that accounting for these peer effects simultaneously and identifying them at an individual level could help marketing managers draw up better guidelines for promotion policies.

There has been a recent surge of interest in documenting (endogenous) peer effects in the marketing literature using behavioral data at the individual level (see Hartmann et al. 2008 for an overview). As noted earlier, the focus of most of these papers is the identification of the endogenous peer effect at the individual level. Our research builds upon two of these papers in particular - *Yang et al.* (2006) and *Hartmann* (2010).[2]

[1]It is important to distinguish the peer presence effect from the mere presence effect documented in the consumer research literature *Argo et al.* (2005). The mere presence effect is based on the effect of the presence of a stranger. In our case, the peer presence effect is based on the presence of the peer (who is known to the focal consumer) when the peer is not engaging in the behavior in question.

[2]Some other relevant papers include *Yang and Allenby* (2003) and *Manchanda et al.* (2008). The first paper develops an autoregressive multivariate binomial model that allows the unobserved

*Yang et al.* (2006) build a simultaneous equation model using a spatial autoregressive structure to capture the interdependence of preferences among spouses in the domain of TV watching. Using aggregate data at the monthly level, they find an asymmetry in the watching behavior between spouses as husbands seem to have a bigger impact on their wives' behavior. *Hartmann* (2010) examines the decisions of golfers to visit a golf course together versus alone via a structural approach. Using the game theoretic framework proposed in the literature for market entry models *Bresnahan and Reiss* (1990), he estimates a discrete choice model with all possible combinations of visit outcomes for a pair of golfers included in the choice set. While both these papers (and others) have documented various strategies to identify peer effects, they have focused only on the endogenous peer effect. In our work, we provide a general framework to capture all types of peer effects - those that are related to the behavior in question as well as those that operate independent of the peer's behavior. In addition, we are able to capture heterogeneity in all three effects by exploiting the panel structure of our data and via the use of a specific estimator (details below). We do this while accounting for the common identification issues previously detailed in the literature *Nair et al.* (2010); *Manski* (1993).

Specifically, our empirical approach allows for the identification of all three peer effects described above while controlling for simultaneity, endogenous group formation (i.e., homophily) and the correlated unobservables confounds that make the identification of peer effects challenging. The approach is based on a simultaneous equation model, with each equation reflecting a consumer's consumption behavior. Our estimation strategy, formulated in a Hierarchical Bayesian framework, allows us to obtain individual level reduced form parameters. We then recover the structural parameters of the simultaneous equation system from these reduced form parameters.

demand shocks of consumers that exhibit demographic and geographic proximity to be correlated. The second papers documents contagion between physically proximate physicians in the context of adopion of a new drug.

The technical challenge inherent in the estimation is that the structural parameters are functions of ratios of the reduced form parameters. Prior literature, especially in marketing, on peer effects had not dealt directly with this challenge. We deal with it via the use of the Minimum Expected Loss (MELO) estimator, allowing us to obtain consistent estimates at the individual level.

As noted earlier, our research setting is that of the casino gambling. The main reason we choose this setting is that peers seemed to play significant role in affecting consumption behavior in this industry. This is based on previous research that documents that one of the most important motivations to visit a casino and play games is social i.e., being with others such as family and friends *Platz and Millar* (2001); *Lee et al.* (2006). Anecdotal feedback from industry sources also suggests that frequent visitors to the casino visit more often with others than alone. However, to the best of our knowledge, no estimates of peer effects in this industry exist. The secondary reason we choose this setting is that the casino industry is a major industry in the United States with revenues greater than that of sports teams and clubs, amusement parks and arcades, and museums (US Census Bureau data).[3] In terms of visits, more than a quarter of all Americans 21 and older visited a casino at least once during 2006 *AGA* (2007). Despite its prominence, research on this industry seems quite limited. Most of the research on gambling uses data from a laboratory (e.g., Gilovich, Vallone and Tversky 1985),[4] as opposed to a field setting. While some research focused on casino gambling based on field data is beginning to emerge *Croson and Sundali* (2005); *Narayanan and Manchanda* (2012), peer effects have been ignored in these studies. Overall, it seems that research on behaviors of the gambling population in

[3]Part of the growth in the industry can be attributed to the increased presence of casinos in the United States in recent years. In the 1980s, casino gambling was legal only in Nevada and Atlantic City in New Jersey. Today, however, casino gambling is legal in 29 states, generating total revenues of about $ 32 billion annually.

[4]While researchers can be quite versatile in terms of the questions that they address in a lab settings, these settings have their own limitations. For example, participants are many times asked not required to risk anything of value and they receive an unlimited supply of (gambling) tokens *Kassinove and Schare* (2001).

the United States has generally been neglected and, as a consequence, demands more attention.

We apply our model to a rich panel data set of casino gamblers who visit a single casino over a two-year period and gamble on slot machines. The data contain information on the gambling activity and the marketing promotions provided to each individual during her visit to the casino as well as demographics. We identify peer dyads via the use of temporal and geographic proximity in plays at the casino. Our dependent variable is the total amount of money bet by a consumer on slot machines on a given day.

Our results indicate the existence of all three peer effects on the amount bet on a given day – the endogenous peer effect, the exogenous peer effect and the peer presence effect. The endogenous peer effect is positive i.e., an increase in peer's amount bet leads to an increase in the focal consumer's amount bet. The exogenous peer effect is negative for promotions, suggesting that when the peer's amount bet is affected by a promotion, the focal consumer reduces the amount that she bets. Finally, the mere presence of the peer also leads the focal consumer to reduce the amount she bets i.e., the peer presence effect is negative as well. Our individual level approach allows us to quantify the variation in the size of this effect across individuals. More interestingly, we find that within pairs of peers, there is considerable asymmetry in these effects.

Our results are likely to be of interest to marketing managers trying to incorporate peer effects into marketing strategies. First, accounting for three peer effects allows a manager to obtain the complete picture with respect to pairwise interactions in consumption settings. Second, the asymmetry of the peer effect within a pair of peers can help managers identify the peer to focus on in terms of influencing joint behavior. Finally, our results suggest that leveraging peer effects to influence consumption needs to be carefully done as both the exogenous peer effect and the peer presence effect tend to work in the opposite direction of the endogenous peer effect - in fact, in our setting first two "cancel" out the third. The use of our estimates to carry out counterfactuals indicate that changes in promotion policies, especially with respect to targeting the more influential peer for promotions, is likely to result in economically significant gains for managers.

The rest of the paper is organized as follows. We develop our model, discuss the identification of parameters and provide details on the estimation procedure in §2.2. In §2.3, we provide details on the institutional setting, the data, peer identification and provide some preliminary evidence for peer effects in the data. We describe the model specification in §2.4 and the results in §2.5. We discuss the managerial implications of our results in §2.6. We conclude in §2.7.

## 2.2 Model

### 2.2.1 Model Development

We model the amount of money bet by customer $A$ in time $t$, $q_{At}$, as follows:

$$q_{At} = \theta_{A1}^{'} X_{At} + \theta_{A2}^{'} X_{At} I_{Bt} + \theta_{A3} q_{Bt} I_{Bt} + \theta_{A4}^{'} X_{Bt}^{s} I_{Bt} + \varepsilon_{At} \tag{2.1}$$

Here $X_{At}$ represents exogenous variables (including the fixed effect) that influence customer $A$'s level of consumption (money bet) at time $t$. $\theta_{A2}$ represents the peer presence effect which captures whether the peer's presence increases or decreases the impact of the exogenous factors on customer impact on customer $A$'s level of consumption. $q_{Bt}$ represents consumer $B$'s level of consumption during time $t$. $I_{Bt}$ is a binary variable that indicates whether peer $B$ visited with the focal agent ($I_{Bt} = 1$) or not ($I_{Bt} = 0$). $\theta_{A3}$ represents the endogenous peer effect i.e., the impact of the peer's level of consumption on customer $A$'s consumption behavior. $\theta_{A4}$ represents the exogenous peer effect i.e., the effect on customer $A$'s consumption when customer $B$'s consumption is influenced by an exogenous demand shock. $X_{Bt}^{s}$ (i.e., subset of

$X_{Bt}$) represents exogenous factors of customer $B$ that also could affect customer $A$. Note that the peer's consumption and exogenous variables are multiplied by $I_{Bt}$ since they are observed by the focal agent only when the peer is present. Finally, $\varepsilon_{At}$ is the demand shock. All upper case letters (e.g., $X_{At}$) except $I_{At}$ and $I_{Bt}$ represent vectors.

Similarly, the amount of money bet by customer $B$ in time $t$ can be represented as:

$$q_{Bt} = \theta'_{B1}X_{Bt} + \theta'_{B2}X_{Bt}I_{At} + \theta_{B3}q_{At}I_{At} + \theta'_{B4}X^{s}_{At}I_{At} + \varepsilon_{Bt} \tag{2.2}$$

### 2.2.2 Identification

The identification of peer effects using behavioral data can be tricky due to multiple confounds. We therefore describe how we control for three possible confounds - endogenous group formation, correlated unobservables and simultaneity. Endogenous group formation occurs because consumers with similar tastes could have a tendency to form social groups and this correlation could be mistaken as casual effect of one's behavior on another *Manski* (1993). Econometrically, this would mean that there is a portion of the error term $\varepsilon_{At}$ (that represents the focal consumer's intrinsic preference) is correlated with that of peer $B$'s intrinsic preference. This would lead to a positive correlation between $\varepsilon_{At}$ and $\varepsilon_{Bt}$ which in turn would lead to a positive correlation between $\varepsilon_{At}$ and $q_{Bt}$. If the model does not account for this, it could result in an upward bias in $\theta_{A3}$ (and $\theta_{B3}$). We control for this correlation via the use of individual fixed effects *Nair et al.* (2010). The fixed effect included in $X_{At}$ will account for all factors that are unique to $A$. Similarly, the intercept term in $X_{Bt}$ will account for all factors that are unique to $B$.

Second, correlation in behavior within the peer group could also arise due to correlated unobservables (such as common demand shocks). Not controlling for these correlated unobservables could also result in an upward bias in $\theta_{A3}$ (and $\theta_{B3}$). We control for correlated unobservables by allowing a free correlation structure between $\varepsilon_{At}$ and $\varepsilon_{Bt}$ as follows:

$$[\varepsilon_{At}, \varepsilon_{Bt}]' = MVN(0, \Sigma) = MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_A^2 & \rho\sigma_A\sigma_B I_{At} I_{Bt} \\ \rho\sigma_A\sigma_B I_{At} I_{Bt} & \sigma_B^2 \end{bmatrix}\right) \tag{2.3}$$

Third, we need to control for simultaneity as with behavioral data it is hard to distinguish causally $A$'s influence on $B$ from the other way around *Manski* (1993). We resolve this problem by imposing an exclusion restriction condition.[5] Specifically, we introduce variables $Z_{At}$ and $Z_{Bt}$ such that $Z_{At}$ only affects $A$'s bet amount ($q_{At}$) but not the peer's bet amount ($q_{Bt}$) and $Z_{Bt}$ only affects $B$'s bet amount ($q_{Bt}$) but not the peer's bet amount ($q_{At}$) as follows:

$$q_{At} = \theta'_{A1} X_{At} + \theta'_{A2} X_{At} I_{Bt} + \theta_{A3} q_{Bt} I_{Bt} + \theta'_{A4} X^s_{Bt} I_{Bt} + \theta_{A5} Z_{At} + \varepsilon_{At} \tag{2.4}$$

$$q_{Bt} = \theta'_{B1} X_{Bt} + \theta'_{B2} X_{Bt} I_{At} + \theta_{B3} q_{At} I_{At} + \theta'_{B4} X^s_{At} I_{At} + \theta_{B5} Z_{Bt} + \varepsilon_{Bt} \tag{2.5}$$

In addition, note that many studies have assumed there are no exogenous peer effects ($\theta_{A4} = \theta_{B4} = 0$) and treated exogenous factors ($X^s_{At}$ and $X^s_{Bt}$) as excluded variables *Moffitt* (2001). This, however, will also result in biased estimates of structural parameters if assumption of no exogenous peer effect is incorrect. This is because in such a case the focal consumer's exogenous factors used as excluded variables will not be independent of the peer's behavior. For example, if consumer $B$ gets a large de-

[5]This exclusion restriction also allows us to solve the simultaneous equation model as the model is then exactly identified *Moffitt* (2001). Technical details with respect to the identification of our simultaneous equation system are available on request.

mand shock (e.g., an exogenous promotion) and that is used as an excluded variable, it might bias endogenous peer effect $\theta_{B3}$ because consumer $A$ could have not only changed his/her behavior due to a change in consumer $B$'s behavior but also due to observing consumer $B$'s receiving a promotion.

### 2.2.3 Estimation

We begin with the equation (2.6) which is the reduced form of our model (equations 2.4 and 2.5):

$$q_{At} = \varphi_{A1}[\theta_{A\cdot}^{'} Y_{At}] + \varphi_{A2}[\theta_{B\cdot}^{'} Y_{Bt}] + \upsilon_{At}$$

$$q_{Bt} = \varphi_{B1}[\theta_{B\cdot}^{'} Y_{Bt}] + \varphi_{B2}[\theta_{A\cdot}^{'} Y_{At}] + \upsilon_{Bt} \tag{2.6}$$

$$[\upsilon_{At}, \upsilon_{Bt}]^{'} \sim MVN(0, \Omega)$$

where $Y_{At} = [X_{At}, X_{At}I_{Bt}, X_{Bt}^{s}I_{Bt}, Z_{At}]^{'}$, $Y_{Bt} = [X_{Bt}, X_{Bt}I_{At}, X_{At}^{s}I_{At}, Z_{Bt}]^{'}$, $\theta_{A\cdot} = [\theta_{A1}, \theta_{A2}, \theta_{A4}, \theta_{A5}]^{'}$, $\theta_{B\cdot} = [\theta_{B1}, \theta_{B2}, \theta_{B4}, \theta_{B5}]^{'}$, $\varphi_{A1} = 1/(1-\theta_{A3}\theta_{B3})$, $\varphi_{A2} = \theta_{A3}/(1-\theta_{A3}\theta_{B3})$, $\varphi_{B1} = 1/(1-\theta_{A3}\theta_{B3})$, $\varphi_{B2} = \theta_{B3}/(1-\theta_{A3}\theta_{B3})$, $\Omega = (I-W)^{-1}\Sigma(I-W)^{-1}$, and $W = \begin{pmatrix} 0 & \theta_{A3} \\ \theta_{B3} & 0 \end{pmatrix}$.

The structural parameters, $\theta_{A3}$ and $\theta_{B3}$, can be expressed as functions of the reduced form parameters as in equation (2.7) below. Note that we do not necessarily have to estimate system of equations (2.6) jointly since the covariates in each equation are identical. From Kruskal's theorem *Davidson and MacKinnon* (1993); *Amemiya* (1985), when each equation contains the same set of regressors, the estimates from the joint system are numerically identical to the equation-by-equation OLS estimates.

$$\theta_{A3} = \varphi_{A2}/\varphi_{B1}$$

$$\theta_{B3} = \varphi_{B2}/\varphi_{A1} \tag{2.7}$$

Conventional estimators like Least Squares Estimator (LS) for $\theta_{A3}$ and $\theta_{B3}$ would be simply expressed as $\widehat{\theta}_{A3}^{LS} = \widehat{\varphi}_{A2}/\widehat{\varphi}_{B1}$and $\widehat{\theta}_{B3}^{LS} = \widehat{\varphi}_{B2}/\widehat{\varphi}_{A1}$. Note, however, that these estimators are ratios of two random variables. Under general conditions, the ratio of two random variables has Cauchy (distribution) like tails resulting in no finite moments *Diebold and Lamb* (1997). Furthermore, it is well known that the distribution of the ratio is typically multimodal *Zellner* (1978). Therefore, with conventional estimators we would not be able to obtain the (Bayesian) estimates of interest especially the posterior means. To overcome this ratio estimation problem, we implement the Minimum Expected Loss (MELO) approach *Zellner* (1978). This approach guarantees finite first and second moments. The MELO estimator is defined as follows:

$$\widehat{\theta}_{A3}^{MELO} = \frac{E(\varphi_{A2})}{E(\varphi_{B1})} \cdot \frac{1 + cov(\varphi_{A2}, \varphi_{B1})/E(\varphi_{A2})E(\varphi_{B1})}{1 + var(\varphi_{B1})/E(\varphi_{B1})^2} = \frac{E(\varphi_{A2})}{E(\varphi_{B1})} \cdot F$$

$$\widehat{\theta}_{B3}^{MELO} = \frac{E(\varphi_{B2})}{E(\varphi_{A1})} \cdot \frac{1 + cov(\varphi_{A1}, \varphi_{B2})/E(\varphi_{A1})E(\varphi_{B2})}{1 + var(\varphi_{A1})/E(\varphi_{A1})^2} = \frac{E(\varphi_{B2})}{E(\varphi_{A1})} \cdot F \tag{2.8}$$

where $E(.)$ denotes the posterior expectation, $var(.)$ denotes the posterior variance, $cov(.)$ denotes the posterior covariance and $F$ denotes the "shrinkage factor." A simple rearrangement of the terms in equation 2.8 yields an expression that gives a better intuition how $F$ works *Diebold and Lamb* (1997):

$$F > 1 \quad if \quad \frac{cov(\varphi_{A2}, \varphi_{B1})}{var(\varphi_{B1})} > \frac{E(\varphi_{A2})}{E(\varphi_{B1})} = \widehat{\theta}_{A3}^{LS}$$

$$F = 1 \quad if \quad \frac{cov(\varphi_{A2}, \varphi_{B1})}{var(\varphi_{B1})} = \frac{E(\varphi_{A2})}{E(\varphi_{B1})} = \widehat{\theta}_{A3}^{LS} \tag{2.9}$$

$$F < 1 \quad if \quad \frac{cov(\varphi_{A2}, \varphi_{B1})}{var(\varphi_{B1})} < \frac{E(\varphi_{A2})}{E(\varphi_{B1})} = \widehat{\theta}_{A3}^{LS}$$

Equation (2.9) shows how the shrinkage factor adjusts the LS estimator toward $\frac{cov(\varphi_{A2}, \varphi_{B1})}{var(\varphi_{B1})}$. Note that when the posterior pdf's of the variables, $\varphi_{A1}, \varphi_{A2}, \varphi_{B1}, \varphi_{B2}$, are very tight (e.g., when sample sizes are very large), $F$ converges to one because the posterior variance and covariance terms will converge to zero. Thus, the MELO estimator will be most effective for small sample sizes. Thus the use of the MELO estimator is perfectly suited to our situation as we would like to obtain individual level estimates (the sample size for each individual is quite small). We confirm that our situation warrants the use of the MELO estimator via a simulation study (discussed in the Appendix).

Given the estimated parameters, $\widehat{\theta}_{A3}^{MELO}$ and $\widehat{\theta}_{B3}^{MELO}$, we then estimate the remaining parameters $\theta_{A\cdot}$ and $\theta_{B\cdot}$. Let us first consider $\theta_{A\cdot}$. To account for the correlation structure between $\varepsilon_{At}$ and $\varepsilon_{Bt}$ , we first derive the conditional distribution of $\varepsilon_{At} \mid \varepsilon_{Bt}$ as follows:

$$\begin{aligned} q_{At} - \widehat{\theta}_{A3} q_{Bt} I_{Bt} &= \theta'_{A\cdot} Y_{At} + \varepsilon_{At} \mid \varepsilon_{Bt} \\ &= \theta'_{A\cdot} Y_{At} + \rho \varepsilon_{Bt} + v_{A|B} \end{aligned} \tag{2.10}$$

where $var(v_{A|B}) \equiv \sigma^2_{A|B} = \sigma^2_A(1 - \rho^2)$. We now perform a conventional regression analysis. Note that in our setting $\varepsilon_{At}$ and $\varepsilon_{Bt}$ are assumed to be uncorrelated when a consumer visits alone. Therefore, for those observations, the joint estimation method described above does not apply. However, our estimation strategy (details provided in the Appendix) allows us to augment the “missing” error term of the peer *Zeithammer and Lenk* (2006). This allows to follow the same joint estimation strategy for all observations.

We cast our model under a Hierarchical Bayesian framework and use Markov chain Monte Carlo methods to obtain the parameter estimates (details are provided in the Appendix).

## 2.3 Data

Our data were obtained from a gaming and gambling company operating a single casino property located in the northwestern United States. The property is located in a small town and the nearest casino is 160 miles away so our casino can be considered a local monopoly. The casino uses a loyalty program to facilitate the building of relationship with its customers. Customers are encouraged to sign up for the loyalty card (at no cost to the consumer) and to use the card whenever they engage in any activity at the casino. Typically, the customer swipes the card when s/he begins to gamble at a station. All gambling activity between the swiping the card and the exit from a station (e.g., a slot machine) is uniquely linked to that customer's account and is identified as a “play.”

Our data consist of a panel data set where the customer activity is recorded for a two year period (July 2005 to June 2007). The data contain information about the games that is played, the amount that is bet, the amount that is won or lost, the start and end times (calendar time) of each play and the identity of the slot machine (if the station is a slot machine). In addition to the activity information, the data also contain information on marketing policies of the casino. Based on feedback from the company, we learned that these policies take two forms at this property - “comps” and "promotion." Comps refer to incentives or rewards that casino gives out to make customers play longer on a given trip. The nature of comps ranges from free key chains all the way up to free suites with all meal and beverages included. In our data, the comps are recorded in dollar equivalents. In general, comps are based on a tiered classification based on total dollars bet in the previous calendar year, where customers who bet more in a given year receive higher comps in the following year.[6] Promotions, on the other hand, are given during a visit randomly to create excitement and engagement among customers on that day. Promotion is recorded as a binary variable indicating whether a customer got a promotion or not on a given day. If a customer receives the promotion for that day, the customer's activity in terms of the amount of money bet is given double credit toward the comp tier classification for the next year.[7] Finally, the data also contain basic demographic information, such as age and gender, on each customer.

The total number of customers in the data is 44,732 accounting for 7,110,376 total plays. The average number of visits in a year to the casino for our panelists was 5.9. A survey of American gamblers by Harrah's/Caesar's in 2006 found that the average number of visits to a casino was 6.1.[8] The average time spent by a customer within a day was about 155 minutes. In terms of bet activity, the mean total amount bet and total money earned was $ 999 (median $ 425.20) and $ -204 (median $ -55.90)

---

[6]Note that this could lead to an acceleration in amounts bet towards the end of the calendar year. However, when we regress the average daily amount bet in the last three months of the year on the average daily amount bet in the first nine months of the year, we find that the slope coefficient is 0.94, suggesting that there is no evidence for acceleration for the customers in our data.

[7]Excutives at the casino told us that these promotions were given randomly at the entrance every day and because of this, customers do not have any prior expectation of receiving a promotion on a particular day. We were able to verify that these promotions were indeed delivered randomly. Specifically, we carried out an analysis to see if more "valuable" customers were more likely to be given a promotion. The casino values customers based on the total amount bet in a calendar year. We therefore estimated a discrete choice model for the customers in our data with the dependent variable being whether a customer received a promotion in 2007 and the independent variable being the total amount bet in 2006. We found that the total amount bet in 2006 was not predictive of promotion. This lends credence to the casino's assertion to us that promotions were delivered randomly.

[8]The survey results are at http://www.caesars.com/images/PDFs/Profile_Survey_2006.pdf.

respectively on a typical visit (the negative sign on the total money earned denotes a loss). Customers were given a promotion, on average, once every 25 visits to the casino. Our sample is evenly balanced in terms of gender (47% male and 50% female; gender data was missing for 3% of our sample). The mean age across panelists was 56 years with a standard deviation of 15.5 years.

We restrict our sample to slot machine players. Slot machine play accounts for the vast majority of play dollars in our data set (over 90%). Consumer preferences in a casino are also overwhelmingly in favor of slot machines - 71% of Americans prefer slot machines while only 14% prefer table games *AGA* (2007). Finally, slot machines represent "games of chance" (as opposed to table games which are considered "games of skill") where outcomes (wins/losses) are determined purely randomly. This is an attractive feature that our modeling approach leverages for identification.

#### 2.3.1 Peer Identification

We follow the approach in *Hartmann* (2010) for identifying peer groups. Specifically, in that paper, peers were identified based on temporal proximity in terms of beginning play on a golf course on at least two occasions. In our case, we carry out a similar analysis for frequent visitors to the casino (those that visit the casino at least five times a year) - a total of 8,870 customers - to identify peers. For these customers, we identified how many times each customer visited the casino with another customer. We defined visiting together as starting the first game on a visit within five minutes of each other within the same geographic area (defined as a bank of machines in the casino). This resulted in a matrix of size 8,870 by 8,870, with the $(i\ ,\ j)$ th element of this matrix representing the number of visits that customer $i$ made with customer $j$ during the entire two year period. From this matrix, we then identified the column with the highest element for each row, resulting in a matrix of size 8,870 by 1. This identified a customer who visited the casino most frequently with the focal customer as the most probable peer for him/her relative to all customers in our data.[9] In order to rule out peer group identification based on chance, we required that for each identified pair there were at least four occasions that they visited together.[10] This resulted in total of 1,626 dyads. Table 2.1 provides descriptive statistics for our entire sample, our base sample of frequent visitors (visited more than five times a year) and the estimation sample consisting of the chosen dyads. In terms of activity at the casino, frequent gamblers accounted for 87% of the total amount bet across the time period of our data and our chosen sample accounted for 47% of the total amount bet. However, in terms of activity, with the exception of a couple of measures, there isn't much difference between all gamblers, frequent gamblers and our estimation sample (see Table 2.1).

### 2.3.2 Evidence for Peer Effects in the Data

We now provide some basic evidence in the data for peer effects. In our estimation sample of 1,626 dyads, on average, each customer visited the casino alone 22 times and with a peer 48 times over the two year period. The mean amount bet when a customer visited with a peer was \$ 1,076 (with a standard deviation of \$ 1653) while it was \$ 1,046 (with a standard deviation of \$ 1862) when s/he visited alone. While the amount bet with peer is higher on average, it is not statistically different. However, we expect that there could be significant heterogeneity across customers - customers who primarily come alone could be different from customers that come primarily with the peer - and we need to account for that. So we run a simple OLS model with the dependent variable being the amount bet and the independent variable a dummy variable denoting whether the customer visited alone or with the peer. To account

[9] It is of course possible that the size of peer group be larger than two. We found that the number of such peer groups was very low. We therefore dropped any customer that was part of such peer groups and restricted our analysis to customers that were in unique dyads.

[10] In separate analyses (not reported here), we checked the robustness of our results to the criteria we used to identify peers. We found no qualitative change in our results.

for heterogeneity, we include individual fixed effects in the analysis. We find that a customer spends \$ 138 more when s/he visits with a peer relative to visiting alone. This difference is statistically significant (standard error 6.87, $t$-stat 19.69).

Note, however, that the purpose of this study is to investigate various mechanisms that could have caused this difference in customer spending when a customer visits with a peer relative to visiting alone. We therefore try to find evidence of all three peer effects in the data without imposing any structure on it. In order to do so, we run an OLS model with the dependent variable being the amount bet. For the independent variables, the peer's bet amount was included to capture the endogenous peer effect. The peer's wins or losses on the previous visit, extra-large wins for the peers (jackpots) from the previous visit and marketing promotions given to the peer on the current visit were included to capture the exogenous peer effect and a dummy variable denoting whether the peer visited or not was included to capture the peer presence effect. The results suggest that there is a positive relationship between focal customer's bet amount and that of the peer's (coefficient 0.43, standard error 0.002, $t$-stat 181.70). We also find that the peer's wins or losses on the previous visit has a positive effect (coefficient 0.09, standard error 0.007, $t$-stat 13.18) but that peer's jackpot from the previous visit has a negative effect (coefficient -140, standard error 21.04, $t$-stat -6.70) on the focal consumer's bet amount. Interestingly, there is a negative effect on the focal customer's betting when the peer gets a promotion (coefficient -165, standard error 15.24, $t$-stat -10.82).

Finally, peer presence also has a negative effect (coefficient -215, standard error 8.79, $t$-stat -24.56) on the focal consumer's bet amount. Note that the peer presence effect is identified from visits when the focal consumer visited together with the peer but the peer did not bet (or bet very small amounts). Patterns in the data also support the negative peer presence effect that we found in the OLS analysis. The data suggest that when peers bet a very small amount, less than \$ 10, the focal consumer typically bets an average of \$ 346 (with a standard deviation of \$ 932). This amount is much lower than the average amount bet when s/he visited alone (\$ 1046).

Taken together, the above suggests that there is enough variation in the data to identify all three peer effects. Obviously, the statistics/results presented above are only indicative and specification of a full structural model is required to pin down the the significance and magnitude of the peer effects.

## 2.4 Model Specification

We now describe the model specification. Our dependendent variable is $ln(q_{At}+1)$, the log transformation of the total amount of money bet by a customer during a day of a visit (indexed by $t$).[11] We divide the factors that could influence a focal customer's betting behavior into own factors, peer factors and environmental factors. For the own factors, we focus on four factors - state-dependence, "irrational" beliefs, extra-large wins ("jackpots") and marketing promotions.

$$X_{At} = \begin{pmatrix} 1 & ln(q_{At-1}+1) & Earn_{At-1} \, Jackpot\_self_{At-1} & Promo_{At} \end{pmatrix}$$

$q_{At-1}$ indicates the total bet amount by the consumer in the previous time period. This variable was included to account for state-dependence in the amount bet. Previous research has characterized positive state-dependence as evidence for "addiction" *Pollak* (1970); *Becker and Murphy* (1988). Specifically, this research has defined addiction as the positive effect of past consumption on the marginal utility of current consumption with the reduced form test for addiction being a positive relationship

[11]We carried out a log transformation of the dependent variable in order to be consistent with our assumption of a normally distributed error term. Also, as a few observations were zero, we added one to every observation.

between past and current consumption*Becker and Murphy* (1988). More recent studies *Guryan and Kearney* (2008); *Narayanan and Manchanda* (2012) have used this reduced form test and found evidence for addiction in gambling settings.

$Earn_{At-1}$ indicates the total money won or lost by the consumer in the previous visit. Previous literature has documented that wins and/or losses can affect future betting behavior in a variety of contexts. Often this is due to irrational beliefs - the "gambler's fallacy" and the "hot hand myth" - held by customers *Gilovich et al.* (1985); *Guryan and Kearney* (2008). The gambler's fallacy is the belief that two consecutive independent outcomes are negatively correlated. Therefore, customers who hold this belief will bet less money if they won money in the previous visit and bet more money otherwise. The hot hand myth is the belief that two consecutive independent outcomes are positively correlated i.e., it is the exact opposite of the gambler's fallacy. Thus, customers who hold this will bet more money if they earned money in the previous visit and bet less money otherwise. Thus, the sign of the coefficient of this variable will provide us information on whether customers in our sample hold such beliefs - a negative estimate suggests a belief in the gambler's fallacy while a positive estimate suggests a belief in the hot hand myth (a zero estimate suggests that consumer do not hold either of these irrational beliefs).

$Jackpot\_self_{At-1}$ is a count variable indicating how many times the customer experienced winning a jackpot during the previous visit. A jackpot is a rare event and represents a larger than usual win[12] and therefore accounted for separately as a count variable (the dollar value of the jackpot win is represented by $Earn_{At-1}$). We expect that the number of jackpots won will influence a customer in a similar manner to dollar wins and losses (as above).

$Promo_{At}$ is a binary variable indicating whether the customer was given a marketing promotion during the visit or not. As described earlier, a promotion allows a

[12] The dollar value of a jackpot win was about 18 times larger than a regular win. Customers won a jackpot, on average, once every 100 visits to the casino.

customer to accrue activity points faster (typically at twice the normal rate). Our hypothesis is that this promotion will act like a typical marketing promotion i.e., induce customer to bet higher amounts of money.

We next focus on peer factors that could affect the focal customer's betting behavior. The first is the log transformation of total bet amount by the peer during the visit plus one, $ln(q_{Bt}+1)$. This represents the direct impact of the peers behavior on the focal customer's behavior and is the endogenous peer effect. The peer presence effect is captured by the interaction of $I_{Bt}$ (the binary variable indicating the peer's visit) and the intercept term in $X_{At}$. The interaction betwen $I_{Bt}$and rest of variables in $X_{At}$ (excluding the intercept) captures the difference in the focal customer's response to own factors when s/he was with peer compared to when s/he was alone. For the exogenous peer effect $X^s_{Bt}$, we focus on three factors - the peer's wins or losses on the previous visit, extra-large wins for the peers (jackpots) from the previous visit and marketing promotions given to the peer on the current visit. We do not have a clear prediction of the effect of the peer's wins or losses and jackpots on the behavior of the focal customer. Based on previous research *Darke and Dahl* (2003); *Feinberg et al.* (2002), we expect that effect of promotion given to the peer will lead to decrease in the focal customer's bet amount.

Finally, in order to control for unobserved environmental factors that could affect the behavioral of all gamblers (e.g., a day with unepxectedly good weather that could change the demand patterns for the casino), we use the average amount bet per customer across all customers (excluding the peer group in question) who visited the casino in time period $t$. We label this variable $EnvCtrl_{(-A)(-B)t}$.

As noted earlier, we use an excluded variable to break the simultaneity confound. The excluded variable we use, $Jackpot\_stranger_{At}$, is a count variable indicating how many times the focal customer observed other customers (excluding the peer) experiencing a jackpot. The key to the exclusion restriction is that these jackpot wins are observed by the focal consumer but not his/her peer. We try and ensure this by including only jackpot wins experienced by other customers in physical proximity to the focal customer but at some distance from the peer. We divide the casino floor into contiguous areas (below) and count a jackpot only if the peer was playing in any area that was not contiguous to the area that the focal customer was playing in (see figure 2.1 and table 2.2 for details). For example, if the focal customer was playing in Area 1 and experienced a stranger winning a jackpot in that area, the peer had to be playing in Area 3 or farther in order for the jackpot to be counted as part of the excluded variable. Based on this definition, the mean $Jackpot\_stranger_{At}$ for a focal consumer was 0.037.

To ensure that the $Jackpot\_stranger_{At}$ variable is a valid excluded variable, we ran a regression analysis as in table 2.3. This is analogus to the first stage regression in the two stage least square estimation method. As can be seen from table 2.3, the F statistic is high enough to reject the null hypothesis that excluded variable is not valid.

The final specification is as follows:

$$\begin{gathered}
ln(q_{At}+1) = \theta_{A1}\begin{pmatrix} 1 & ln(q_{At-1}+1) & Earn_{At-1}\, Jackpot_self_{At-1} & Promo_{At} \end{pmatrix} + \\
\theta_{A2}\begin{pmatrix} 1 & ln(q_{At-1}+1) & Earn_{At-1}\, Jackpot_self_{At-1} & Promo_{At} \end{pmatrix} I_{Bt} + \\
\theta_{A3} ln(q_{Bt}+1) I_{Bt} + \theta_{A4}\left(Earn_{Bt-1}\, Jackpot_self_{Bt-1}\, Promo_{Bt}\right) I_{Bt} + \\
\theta_{A5}\left(Jackpot_stranger_{At}\right) + \theta_{A6} ln(EnvCtrl_{-(A)-(B)t}) + \varepsilon_{At}
\end{gathered} \tag{2.11}$$

## 2.5 Results

### 2.5.1 Parameter Estimates

We first focus on the own parameters that could affect the amount bet by the focal customer.[13] The coefficient for $ln(q_{At-1}+1)$ was positive (albeit "small" at 0.18). It means that 1 percent increase in previous bet amount leads to 0.18% increase in current bet amount, suggesting evidence for positive state dependence (on average). This result is different from that in *Narayanan and Manchanda* (2012). However, it is difficult to pinpoint the reason for difference in the result as the models differ in the level of aggregation and specification. The coefficient for $Earn_{At-1}$ was positive but not significantly different from zero. Customers, however, tended to bet more (21%) when they won jackpot previously. Lastly, promotion had a positive effect on the focal consumer's bet amount with an increase by 75% on the days when the focal customer was given a promotion, thus attesting to the impact of the casino's marketing programs.

We now turn to the peer effects. The coefficient for the mean endogenous peer effect was 0.78 implying that that the focal customer increases his/her bet amount by 0.78% when the peer increases his/her bet money by 1%. In terms of the exogenous peer effect, we find that when the peer was given a promotion - the focal consumer reduced his/her amount bet by 23% - confirming the findings from the previous literarture *Darke and Dahl* (2003); *Feinberg et al.* (2002). In terms of the peer presence effect, the main effect (measured via the intercept) was negative i.e., the presence of the peer when the peer was not betting reduced the amount bet by the focal customer. The presence of the peer also induced negative state dependence (though for a modest amount, a drop of 0.11% from 0.18% of state dependence when s/he was without a

[13]We ran the sampler for 20,000 iterations and obtained the posterior mean and standard deviation for all parameters from the last 5,000 draws. Note also that we divided $Earn_{At-1}$ by 100 for computational tractability.

peer), leading to less addicitive behavior. Finally, we also find that the coefficient for $Promo_{At} \times I_{Bt}$ was negative (-0.24), suggesting that the focal consumer responds less to promotion in the presence of peer. The result seems to indicate that the presence of peer dampens the response to marketing promotion.

### 2.5.2 Heterogeneity and the Asymmetry in the Endogenous Peer Effect

Our individual level approach allows us to quantify the variation in the size of peer effects across consumers. Specifically, the use of the MELO approach enables us to estimate the all peer effects, including the endogenous peer effect, at the individual level. Figure 2.2 shows the heterogeneity in the estimated endogenous peer effect across individuals. The size of the endogenous peer effect for a majority of the customers (83%) is between 0 and 1 implying that when the focal consumer increases his/her bet amount, the peer also increases his/her bet amount but less than the amount increased by the focal customer. Of these (83%) customers, 48% had an endogenous peer effect coefficient that is significantly different from zero. Of the remaining 17% of customers, a majority - 15% - of peers increase their bet amount more than the amount increased by the focal customer with 94% of these customers having the coefficient significantly different from zero. Finally, the endogenous peer effect was negative for the remaining 2% of the peers i.e., they move in the opposite direction of the focal customer. However, the estimated peer effect for all these customers is not significantly different from zero.

The correlation patterns between the individual coefficients also provide interesting insights. For example, we find a high and negative correlation coefficient (-0.87) between individual level intercepts and the endogenous peer effect coefficients. This suggest that customers that tend to bet more on average are less susceptible to being influenced by peer behavior. We also find that customers that show high state dependence (more “addictive” behavior) are also less susceptible to the peer’s behavior (the correlation coefficient is -0.62).

We turn to an examination of the asymmetry of the endogenous peer effect within a peer group. We compute the asymmetry as the absolute difference in the endogenous peer effects within of the two individual peer effects in each peer group. Figure 2.3 documents the distribution of this difference across the 1,626 dyads in our estimation data set. The figure shows that there is considerable asymmetry in this effect across groups with the difference between the endogenous peer effects being 0.2 (i.e., absolute difference in elasticity of endogenous peer effect is 0.2%) on average. The asymmetry of the peer effect within a pair of peers can help managers identify the peer to focus on in terms of influencing joint behavior.

## 2.6 Managerial Implications

We investigate the implications of our results for managers. First, we compute the total response to a promotion delivered to a customer by quantifying the tradeoff between the endogenous and exogenous peer effects. Second, we demonstrate the impact on bet behavior arising from the peer presence effect. Finally, we carry out two counterfactuals to show how managers could benefit from better resource allocation.

### 2.6.1 Tradeoff between the Endogenous and Exogenous Peer Effects

We use the estimates of the endogenous and exogenous peer effects to compute the overall spillover effect induced by a promotion given to an individual customer. Assume that a promotion was given to customer A. Assuming all else being equal, the estimates suggest that the promotion increases the amount bet by customer A by 36% (=exp(0.56-0.24)). This results in a 27% (= exp(0.78*(0.56-0.24))) increase in the amount bet by customer B via the endogenous peer effect. However, via the exogenous peer effect, the promotion leads to a simultaneous decrease of 23% (=exp(-0.25)) in the amount bet by customer B. The net spillover effect on customer B is, therefore approximately *zero* (=1.27*0.77). This suggests that managers cannot take a positive spillover for granted when they deliver a promotion to customers.

### 2.6.2 Implication of Negative Peer Presence Effect

As noted earlier, we find that peer presence has a negative effect on the focal customer's behavior. This implies that, all else being equal, if a peer is present but not consuming, the focal customer lowers her/his betting amounts to a level even below that when s/he visits alone. To illustrate the magnitude of this effect, we plot the relationship between the amount bet by customers A and B in Figure 2.4. From the plot, it can be seen that the negative impact of the peer presence effect can be balanced out at a spend level of $ 106 by the peer. The casino therefore needs to think about mechanisms that can incentivize play to at least this level.

### 2.6.3 Implications for Resource Allocation

The above analysis suggests that in order to have the most effective allocation of resources, managers need to eliminate factors that lead to a negative exogenous peer effect. In addition, resource allocation can be optimized by leveraging the asymmetry in the endogenous peer effects. We document the impact of both these strategies via counterfactual analysis. For the first analysis, we fix all the variables except the $Promo_{At}$ and $Promo_{Bt}$ at the average value for each customer. We then calculate the total amount money bet by customers using the estimated parameters. Here we restrict our observations to when at least one person in a group gets a promotion. By adding the amount bet over these observations, we obtain the baseline amount i.e., the total amount bet under the current promotion strategy. We model a scenario where the casino can eliminate the negative exogenous peer effect by setting the parameter for this effect to zero. We then compute the total amount bet under this scenario. Our results show that the elimination of the negative exogenous peer effect results in an increase of 18% relative to the baseline (assuming that elimination of the exogenous peer effect is costless to the firm). The recommendation to the firm is therefore to develop a mechanism to deliver a promotion to customer A in a manner that it is not observed by customer B e.g., via e-mail or on a smartphone. Of course, this will not preclude situations where customer A informs customer B about the promotion. But it may be worth it for the casino to develop and test such mechanisms.

In the second counterfactual, as noted earlier, we focus on leveraging the asymmetry in the endogenous peer effects within a pair. We followed a similar strategy for constructing the baseline as above, but restrict ourselves only to those occasions when only one of the two peers gets a promotion. If the peer who received the promotion had a smaller effect on the other peer, we reversed the promotion (i.e., the peer with the bigger peer effect received the promotion). This resulted in us reversing the recipient of the promotion on 47% of the occasions (when only one peer got a promotion). We then used our estimates to compute the change in the total amount bet. We found that the amount of money bet increased 15% relative to the baseline condition (as before, we assumed that the reallocation of promotion was costless to the firm).

Both counterfactual analyses described above suggest that there is considerable upside to developing better promotional mechanisms and targeting the more influential peer within a peer group.

## 2.7 Conclusion

Our paper adds to the small but growing body of research that investigates individual level peer effects in marketing settings. Our main contribution is to provide a general framework for measuring these effects via the inclusion of peer effects that are based on the behavior in question as well as peer effects that operate independent of that behavior. Specifically, our approach allows to verify the existence of and measure the endogenous peer effect, the exogenous peer effect and the peer presence effect. We choose casino gambling as our setting as both the academic literature and the industry suggest that peer effects play a large role in the affecting consumer behavior, but no estimates of such effects seem to exist. The casino industry also represents a large and significant industry, with broad participation by American adults, that has not been studied much by economists and marketers. Methodologically, we show how the use of the MELO estimator in simultaneous equation setting allows researchers to obtain consistent individual level estimates with the typical amount of data available in marketing panel data sets. Our results, for our specific setting, suggest that the endogenous peer effect is positive but the other two are negative.

For managers, our approach is likely to be of interest whenever and wherever there is interest in leveraging peer effects in marketing strategies. First, accounting for three peer effects allows a manager to obtain the complete picture with respect to pairwise interactions in consumption settings. Second, the asymmetry of the peer effect within a pair of peers can help managers identify the peer to focus on in terms of influencing joint behavior. Our results suggest that leveraging peer effects to influence consumption needs to be carefully done as both the exogenous peer effect and the peer presence peer effect tend to work in the opposite direction of the endogenous peer effect. The use of our estimates to carry out counterfactuals indicate that their is likely to be high upside for managers, especially when they target the more influential peer for promotions.

Our research suffers from some limitations. Our data come from one industry and from one casino property in particular. Our analysis and conclusions also apply to the "heavy half" of all casino customers. Our method for identifying peers is based on data and not actual knowledge of peer groups. Finally, we do not observe peer groups larger than two in our data (based on our peer identification strategy) and therefore our model only accounts for dyadic relationship. We hope that future research can help address these limitations.

## 2.8 Appendix

### The Hierarchical Model

In this section we specify the hierarchy for the individual-level parameters. We assume that individual level parameters follow normal distributions as follows:

$$\varphi_i \sim N\left(\bar{\varphi}, \pi\right)$$

$$\theta_{i\cdot} \sim N\left(\bar{\theta}, \psi\right)$$

The prior distributions for the population-level, ($\bar{\varphi}$, $\pi$, $\bar{\theta}$, $\psi$, $\Sigma$), parameters are as follows

$$\bar{\varphi}|\pi \sim N\left(\varphi_0, \pi A^{-1}\right)$$

$$\pi \sim \text{Inverse Wishart}\left(\mu_\pi, S_\pi\right)$$

$$\bar{\theta}|\psi \sim N\left(\theta_0, \psi B^{-1}\right)$$

$$\psi \sim \text{Inverse Wishart}\left(\mu_\psi, S_\psi\right)$$

$$\Sigma \sim \text{Inverse Wishart}\left(\mu_\Sigma, S_\Sigma\right)$$

where $\theta_0 = [0, 0, 0, ...0]$, $\varphi_0 = [0, 0]$, $A = 0.01$, $B = 0.01$, $\mu_\pi = N_\pi + 3$,$S_\pi = \mu_\pi I_\pi$, $\mu_\psi = N_\psi + 3$,$S_\psi = \mu_\psi I_\psi$ , $\mu_\Sigma = N_\Sigma + 3$,$S_\Sigma = \mu_\Sigma I_\Sigma$ .$N_\pi$ is the number of parameters for the reduced form model. $N_\psi$ is the number of parameters for the structural model excluding endogenous peer effect parameter. $N_{\Sigma}$ is the number of structural parameters.

**The Markov Chain Monte Carlo Algorithm**

- Generating $\theta^k_{A3}$ and $\theta^k_{B3}$ (Endogenous Peer Effect Parameter)

$\varphi^k_i = \left[\varphi^k_{i1}, \varphi^k_{i2}\right] \sim MVN(U^k_i, S^k_i)$

where

$S^k_i = \left[\sum_{t=1}^{T^k} D^{k\prime}_{it} D^k_{it} + \pi^{-1}\right]^{-1}$

$U^k_i = S^k_i \left[\sum_{t=1}^{T^k} D^{k\prime}_{it} q^k_{it} + \pi^{-1}\bar{\varphi}\right]$

and

$D^k_{it} = \left[\theta^{\prime}_{i\cdot} Y^{\prime}_{it}, \theta^{\prime}_{j\cdot} Y^{\prime}_{jt}\right]$

and

$i = A, B \quad\quad j = B, A$

We recover $\theta^k_{A3}$ and $\theta^k_{B3}$ from

$$\widehat{\theta}^k_{A3} = \frac{E(\varphi^k_{A2})}{E(\varphi^k_{B1})} \cdot \frac{1 + cov(\varphi^k_{A2}, \varphi^k_{B1})/E(\varphi^k_{A2})E(\varphi^k_{B1})}{1 + var(\varphi^k_{A2})/E(\varphi^k_{B1})^2}$$

$$\widehat{\theta}^k_{B3} = \frac{E(\varphi^k_{B2})}{E(\varphi^k_{A1})} \cdot \frac{1 + cov(\varphi^k_{A1}, \varphi^k_{B2})/E(\varphi^k_{A1})E(\varphi^k_{B2})}{1 + var(\varphi^k_{A1})/E(\varphi^k_{B2})^2}$$

Note that since regressors in equations for A and B in equation (6) are exactly the same, we estimate the parameters equation-by-equation rather than estimating jointly. From Kruskal's theorem*Davidson and MacKinnon* (1993); *Amemiya* (1985), when each equation contains the same set of regressors, the estimators are numerically identical to equation-by-equation OLS estimates.

- Generating $\bar{\varphi}$

$\bar{\varphi} \sim MVN\left(L, K\right)$

where

$K = \left( ((\pi^{-1}/(2N))^{-1} + 0.01I \right)^{-1}$

$L = K \left( \pi^{-1} \sum_{k=1}^{N} \sum_{i}^{A,B} \varphi_i^k + 0.01I \times \varphi_0 \right)$

where $N$ is the total number of dyads.

- Generating $\pi$

$$\pi \sim IW \left( \sum_{k=1}^{N} \sum_{i}^{A,B} \left( \varphi_i^k - \bar{\varphi} \right) \left( \varphi_i^k - \bar{\varphi} \right)' + S_\pi, 2N + \mu_\pi \right)$$

- Generating $\theta_{i\cdot}$

$$\theta_{i\cdot}^k \sim MVN\left(M_i, N_i\right)$$

where

$N_i = \left[ \sum_{t=1}^{N} Y_{it}^{k'} Y_{it}^{k} + \psi^{-1} \right]^{-1}$

$M_i = N_i \left[ \sum_{t=1}^{N_1} Y_{it}^{k'} (I - W) q_{it}^k + \psi^{-1} \bar{\theta} \right]$

where

$Y_{it}^{k*} = Y_{it}^{k} / \sqrt{\sigma_i^2 (1 - \rho^2)}$, $q_{it}^{k*} = (q_{it}^k - \theta_{i3} q_{jt}^k I_{jt} - \rho \varepsilon) / \sqrt{\sigma_i^2 (1 - \rho^2)}$ , and $W = \begin{pmatrix} 0 & \theta_{A3} \\ \theta_{B3} & 0 \end{pmatrix}$

- Generating $\bar{\theta}$

$$\bar{\theta} \sim MVN(H, W)$$

where

$W = ((\psi^{-1}/(2N))^{-1} + 0.01I)^{-1}$

$H = W(\psi^{-1} \sum_{k=1}^{N} \sum_{i}^{A,B} \theta_{i\cdot}^k + 0.01I \times \psi_0)$

- Generating $\psi$

$$\psi \sim IW \left( \sum_{k=1}^{N} \sum_{i}^{A,B} (\theta_{i\cdot}^k - \bar{\theta})' (\theta_{i\cdot}^k - \bar{\theta}) + S_\psi,\ 2N + \mu_\psi \right)$$

- Generating $\Sigma$

In our model, the covariance term ($\rho_{AB}\sigma_A\sigma_B$) is defined across joint consumption occasions while the variance terms ($\sigma_A^2, \sigma_B^2$) are defined across all consumption occasions. This results in different observation number to estimate the covariance term and the variance term. One consequence of this unbalanced observation number (or absent dimensions) is that the standard Bayesian analysis of the multidimensional covariace sturuture becomes difficult (i.e. the full conditional distribution for $\Sigma$ is no longer inverted Wishart distribution). To overcome this difficulty, we implement *Zeithammer and Lenk* (2006) method. The basic idea is to augment *Tanner and Wong* (1987) the residuals for the absent dimensions.

First, we draw absent residual as follows. $R_{jt}^{kp}$ is residual for the observation when customer j visited.

$R_{it}^{ka} \mid R_{jt}^{kp}, \Sigma, \theta_j$ ~ $N(F_{it}, G_{it})$

where

$F_{it} = (\rho_{ij}\sigma_i\sigma_j)/\sigma_j^2 \times R_{jt}^{kp}$

$G_{it} = \sigma_i^2 - (\rho_{ij}\sigma_i\sigma_j)/\sigma_j^2 \times (\rho_{ij}\sigma_i\sigma_j)$

$R_{jt}^{kp} = q_{jt}^k - \theta_{j\cdot}^{\prime} Y_{jt}$

Together with the present residuals, the augmentation produces full-dimensional residual vectors (i.e. $R_A^{k'} = [R_{A1}^{ka}, R_{A2}^{kp}, R_{A3}^{kp}, R_{A4}^{ka} \ldots\ldots R_{At-1}^{kp}, R_{At}^{kp}]$) that can be used as pseudo-observations in the conditional posterior draw of $\Sigma$ as if there were no absent dimensions as follows.

$$\Sigma \sim IW\left(\sum_{k=1}^{N} R^k R^{k'} + S_\Sigma,\ 2N + \mu_\Sigma\right)$$

where

$R^k = [R_A^k, R_B^k]$

**Simulation Study**

The simulated data set comprises 1,000 dyads. We first generated 70 observations per consumer - 50 observations as joint visits and 20 observations as single visits.

This observation number is similar to what we have in our estimation data set. The following model was estimated:

$$q^k_{At} = \beta^k_{A1} + \beta^k_{A2}x^k_{At} + \beta^k_{A3}I^k_{Bt} + \beta^k_{A4}q^k_{Bt}I^k_{Bt} + \beta^k_{A5}x^k_{Bt}I^k_{Bt} + \beta^k_{A6}z^k_{At} + \varepsilon^k_{At}$$

$$q^k_{Bt} = \beta^k_{B1} + \beta^k_{B2}x^k_{Bt} + \beta^k_{B3}I^k_{At} + \beta^k_{B4}q^k_{At}I^k_{At} + \beta^k_{B5}x^k_{At}I^k_{At} + \beta^k_{B6}z^k_{Bt} + \varepsilon^k_{At}$$

where$x^k_{At}$ and $x^k_{Bt}$ are exogenous factors that influence the focal customer as well as peer (through exogenous peer effect), $I^k_{At}$ and $I^k_{Bt}$ represent the peer's presence, and $z^k_{At}$ and $z^k_{Bt}$ are the excluded variables. This model replicates the main feautres of our model. $\beta^k_{A3}$ and $\beta^k_{B3}$ represent the peer presence effect, $\beta^k_{A4}$ and $\beta^k_{B4}$ represent the endogenous peer effect and $\beta^k_{A5}$ and $\beta^k_{B5}$ represent the exogenous peer effect. The $k$ subscript indicates the $k$th dyad. $x^k_{At}$, $x^k_{Bt}$,$z^k_{At}$ and $z^k_{Bt}$ were generated from standard normal distribution and variance and covariance for error terms were set as 1 and 0.5 respectively. The specification of true values are as follows:

$[\beta^k_{j1}, \beta^k_{j2}, \beta^k_{j3}, \beta^k_{j5}, \beta^k_{j6}]' \sim MVN([\beta_1, \beta_2, \beta_3, \beta_5, \beta_6]', \Psi)$

$\beta^k_{j4} \sim MVN(\beta_4, \Omega)$

$[\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6]' = [1, 1, -0.3, 0.9, 0.5, 2]$.

$\Psi$ is a matrix with diagonal elements set to 0.1 and all other elements set to 0.05. $\Omega = 1$ and $j = A, B$.

Table 2.7 shows the result of the simulation. First, the use of the MELO estimator allows us to recover all the true values in an unbiased manner. Relative to the LS estimator, the MELO estimator performs much better than the LS estimator approach in terms of generating consistent etimates and especially in terms of efficiency. The large posterior standard deviations for the LS estimators show that when the sample size is small, the (individual level) estimates are not consistent.

Next, we ran another simulation study to verify whether the LS estimator generates similar results to the MELO estimator as the sample size gets larger. We therefore allow for five times the number of observations per individual - a total of 350 observations per individual with 250 observations as joint visits and 100 observations as individual visits. Not surprisingly, the LS estimator does much better in such situations (table 2.8).

The results based on the above simulation(s) seem to provide strong support for the use of the MELO estimator in our setting.

**Glossary**

- **Correlated Unobservables**: The effect of common factors not observed to the researcher that makes a group of peers to behave in a similar manner. For example, peers may consume more coffee when a specific barista is making the drinks. So consumption for both a peer and focal consumer goes up, not because one is influencing the other, but because both are influenced by the presence of the barista. The researcher may erroneously attribute this to a peer effect when, in reality, a factor that is not observed by the researcher - the presence of the barista - is driving the joint consumption.

- **Endogenous peer effect**: The effect of peer's behavior on a focal consumer's behavior. An example is when the peer's internally driven consumption of coffee (say) influences the focal consumer's consumption of coffee.

- **Endogenous group formation (a.k.a Homophily)**: The social phenomenon by which consumers with similar tastes and preferences tend to form social group(s).

- **Exogenous peer effect**: The effect of an independent or external influence on a peer's behavior that ends up in turn influencing the focal consumer's behavior.

An example is when a peer gets a coupon to buy coffee. The focal consumer observes that the peer gets the coupon and that impacts her consumption.

- **Demand shocks**: A random event that could affect consumers and firm behavior. For example, an unexpected drought in South America could increase raw coffee prices leading to firms' adjusting their prices and consumers' changing their consumption as a result.

- **Mere presence effect**: The effect of others' physical presence on the focal consumer's behavior when there is no interaction between the focal consumer and others. This may occur, for example, when a focal consumer consumes less coffee when the coffee shop is more crowded.

- **Peer presence effect**: The effect of peer's physical presence without any engagement in the behavior of interest on the focal consumer's behavior. An example is if the focal consumer drinks less coffee in the presence of the peer (when the peer is not consuming coffee) than when she is alone.

- **Simultaneity**: The phenomenon by which the group of peers affect each other's behavior in the same temporal period. If one observes higher coffee consumption by the peer and the focal consumer on a given day, it is hard to determine whether the peer influenced the focal consumer or vice versa.

- **Spillover effect**: A secondary effect that could influence consumers who are not part of the peer group under consideration to change their behavior (either through observation or interaction with the peer group).

Table 2.1: Descriptive Statistics

| | Population | | Frequent visitor | | Sample of dyads | |
|---|---|---|---|---|---|---|
| | N=44,732 | | N=8,870 | | N=1,626*2 | |
| | Mean | S.D. | Mean | S.D. | Mean | S.D. |
| Number of days played during two years | 11.84 | 31.95 | 48.44 | 58.79 | 70.01 | 73.18 |
| Time Spent per day | 153.28 | 136.68 | 162.13 | 142.49 | 166.89 | 136.83 |
| Money bet per day(dollar) | 975.58 | 1,803.51 | 1,038.64 | 1,882.09 | 975.15 | 1533.03 |
| Money won per day(dollar) | 790.10 | 2,108.04 | 844.99 | 2,271.71 | 821.76 | 1,317.88 |
| Number of Jackpots per day | 0.01 | 0.14 | 0.01 | 0.16 | 0.01 | 0.13 |
| Number of Promotions per day | 0.04 | 0.19 | 0.04 | 0.19 | 0.04 | 0.19 |

Table 2.2: Play Zones within the Casino

| Focal Consumer Location | Peer Location |
|---|---|
| 1 | All except 2 |
| 2 | All except 1, 3 |
| 3 | All except 2, 4 |
| 4 | All except 3, 5, 8 |
| 5 | All except 4, 6, 7, 8 |
| 6 | All except 5, 7 |
| 7 | All except 5, 6, 8 |
| 8 | All except 4, 5, 7, 9 |
| 9 | All except 8, 10 |
| 10 | All except 9, 11 |
| 11 | All except 10 |

Table 2.3: Validity Check on the Excluded Variable

| Variables | On only excluded variable | | | First Stage Regression | | |
|---|---|---|---|---|---|---|
| | Parameter | S.E. | t-stat | Parameter | S.E. | t-stat |
| $ln(q_{At-1}+1)$ | | | | 0.09*** | 0.01 | 29.31 |
| $Earn_{At-1}$ | | | | 0.17** | 0.01 | 2.44 |
| $Jackpot\_self_{At-1}$ | | | | 0.27*** | 0.02 | 12.90 |
| $Jackpot\_stranger_{At}$ | 0.32*** | 0.01 | 26.01 | 0.39*** | 0.01 | 34.64 |
| $Promo_{At}$ | | | | 0.32*** | 0.01 | 20.67 |
| $ln(EnvCtrl_{-(A)-(B)t})$ | | | | 0.08*** | 0.01 | 5.39 |
| $I_{Bt}$ | | | | -2.23*** | 0.02 | -96.87 |
| $ln(q_{Bt}+1)$ | | | | 0.43*** | 0.01 | 121.27 |
| $Earn_{Bt-1}$ | | | | 0.05*** | 0.01 | 7.47 |
| $Jackpot\_self_{Bt-1}$ | | | | 0.13*** | 0.02 | 5.70 |
| $Promo_{Bt}$ | | | | -0.19*** | 0.02 | -11.16 |
| $F$ | 61.66 | | | 84.70 | | |
| $R^2$ | 0.56 | | | 0.63 | | |
| *p<0.05, **p<0.01, ***p<0.001 | | | | | | |
| Note that individual fixed effects are not reported. | | | | | | |

Table 2.4: Parameter Estimates

| Parameter Description | Parameters | Mean | SD | 2.5% | Median | 97.5% |
|---|---|---|---|---|---|---|
| Own factors | $Intercept$ | 3.98 | 0.18 | 3.62 | 3.98 | 4.34 |
| | $ln(q_{At-1}+1)$ | 0.17 | 0.01 | 0.15 | 0.18 | 0.20 |
| | $Earn_{At-1}$ | 0.02 | 0.04 | -0.05 | 0.02 | 0.11 |
| | $Jackpot\_self_{At-1}$ | 0.19 | 0.09 | 0.00 | 0.19 | 0.39 |
| | $Promo_{At}$ | 0.56 | 0.06 | 0.44 | 0.56 | 0.68 |
| Endogenous peer effect | $ln(q_{Bt}+1)$ | 0.78 | 0.01 | 0.76 | 0.78 | 0.80 |
| Exogenous peer effect | $Earn_{Bt-1}$ | 0.01 | 0.02 | -0.03 | 0.01 | 0.06 |
| | $Jackpot\_self_{Bt}$ | -0.06 | 0.05 | -0.18 | -0.06 | 0.03 |
| | $Promo_{Bt}$ | -0.24 | 0.07 | -0.39 | -0.24 | -0.11 |
| Peer presence effect | $Intercept$ | -3.64 | 0.10 | -3.83 | -3.64 | -3.44 |
| | $ln(q_{At-1}+1) \times I_{Bt}$ | -0.11 | 0.01 | -0.14 | -0.11 | -0.08 |
| | $Earn_{At-1} \times I_{Bt}$ | 0.03 | 0.04 | -0.06 | 0.03 | 0.12 |
| | $Jackpot\_self_{At-1} \times I_{Bt}$ | -0.04 | 0.12 | -0.30 | -0.04 | 0.19 |
| | $Promo_{At} \times I_{Bt}$ | -0.24 | 0.07 | -0.39 | -0.24 | -0.11 |
| Excluded variable | $Jackpot\_stranger_{At}$ | 0.33 | 0.02 | 0.30 | 0.33 | 0.37 |
| Environmental factor | $EnvCtrl_{-(i)-(j)t}$ | 0.07 | 0.02 | 0.02 | 0.07 | 0.12 |

Table 2.5: Distribution of Endogeous Peer effect

| 1st Quantile | Median | Mean | 3rd Quantile |
|---|---|---|---|
| 0.68 | 0.86 | 0.79 | 0.96 |

Table 2.6: Distribution of Asymmetry of Endogenous Peer effect

| 1st Quantile | Median | Mean | 3rd Quantile |
|---|---|---|---|
| 0.06 | 0.13 | 0.20 | 0.26 |

Table 2.7: Simulation Study Result 1

| Parameters | True Value | MELO | | LS | |
|---|---|---|---|---|---|
| | | Posterior Mean | Posterior S.D. | Posterior Mean | Posterior S.D. |
| $\beta_1$ | 1 | 1.00 | 0.01 | 0.95 | 1.06 |
| $\beta_2$ | 1 | 1.00 | 0.01 | 1.02 | 0.57 |
| $\beta_3$ | 0.9 | 0.90 | 0.004 | 0.88 | 0.63 |
| $\beta_4$ | -0.3 | -0.31 | 0.03 | -0.24 | 2.11 |
| $\beta_5$ | 0.5 | 0.50 | 0.01 | 0.50 | 0.42 |
| $\beta_6$ | 2 | 2.02 | 0.013 | 2.05 | 1.03 |

Table 2.8: Simulation Study Result 2

| Parameters | True Value | MELO | | LS | |
|---|---|---|---|---|---|
| | | Posterior Mean | Posterior S.D. | Posterior Mean | Posterior S.D. |
| $\beta_1$ | 1 | 0.98 | 0.01 | 0.98 | 0.05 |
| $\beta_2$ | 1 | 1.00 | 0.01 | 1.02 | 0.03 |
| $\beta_3$ | 0.9 | 0.90 | 0.001 | 0.90 | 0.09 |
| $\beta_4$ | -0.3 | -0.27 | 0.01 | -0.25 | 0.10 |
| $\beta_5$ | 0.5 | 0.50 | 0.007 | 0.51 | 0.03 |
| $\beta_6$ | 2 | 2.02 | 0.008 | 2.02 | 0.06 |

Figure 2.1: Casino Floor Plan

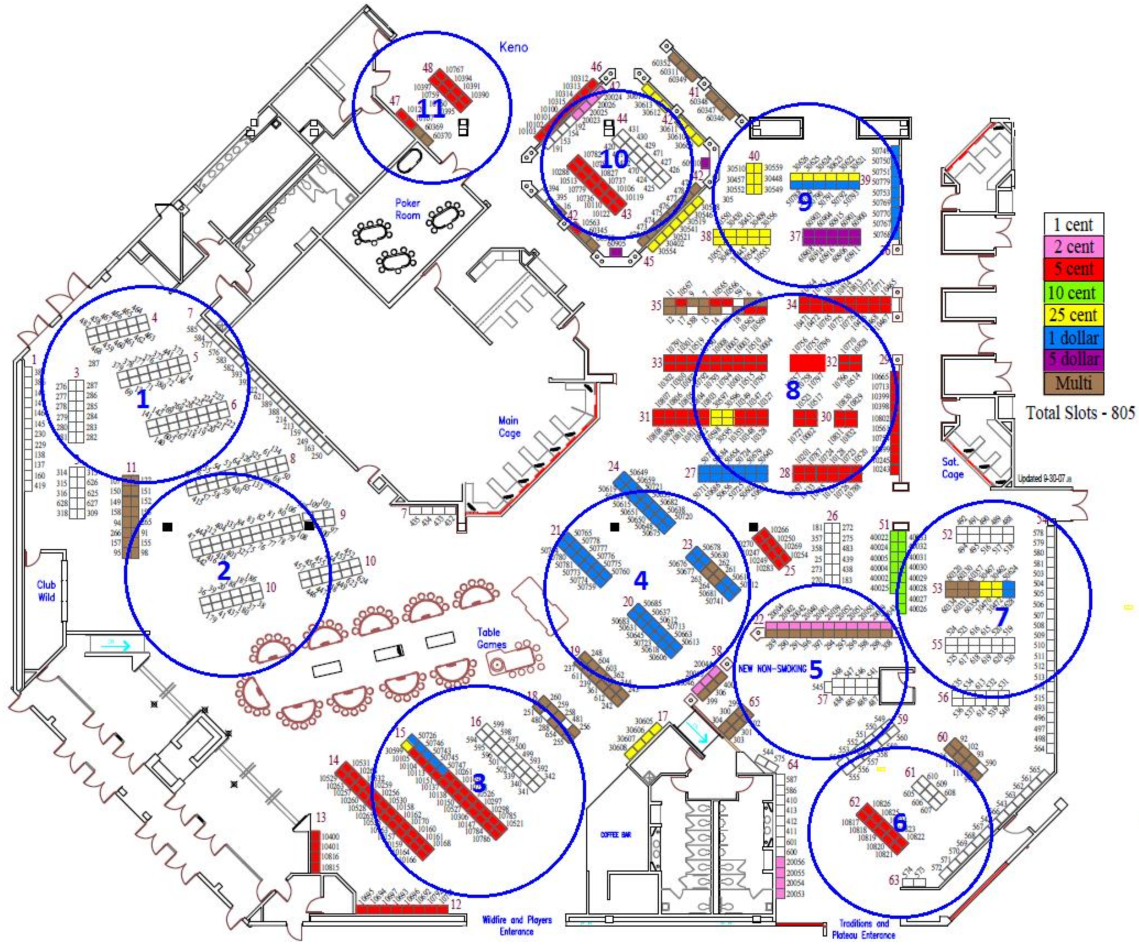


Figure 2.2: Heterogeneity in Endogenous Peer Effect

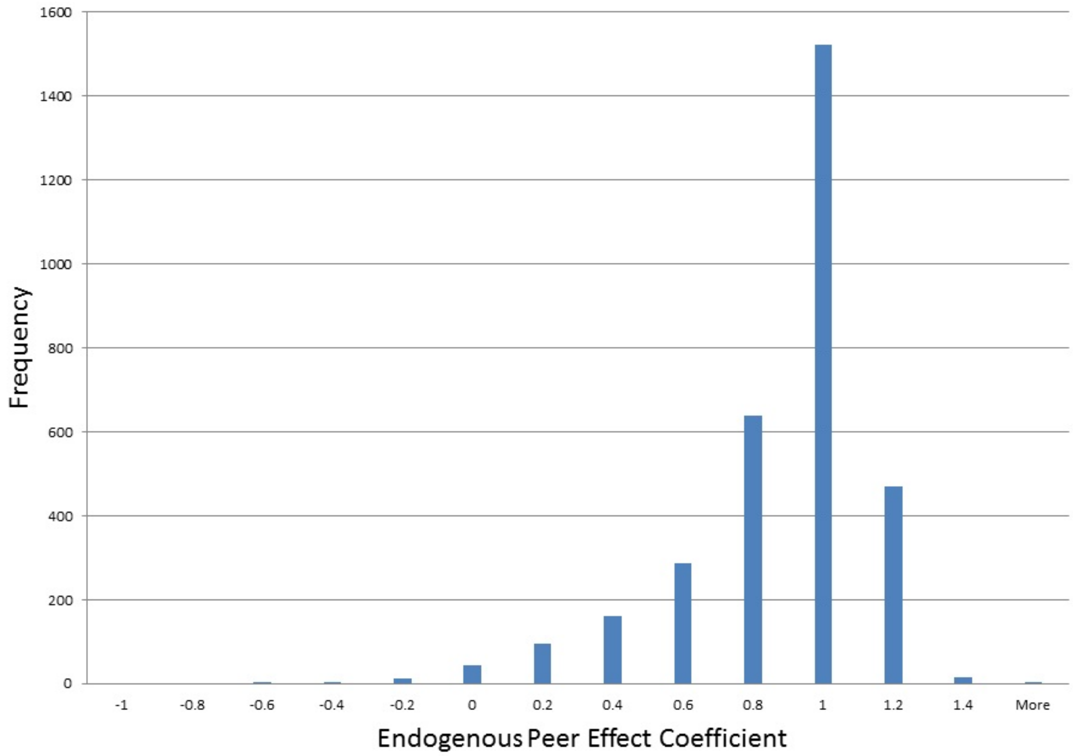



Figure 2.3: Asymmetry in Endogenous Peer Effect

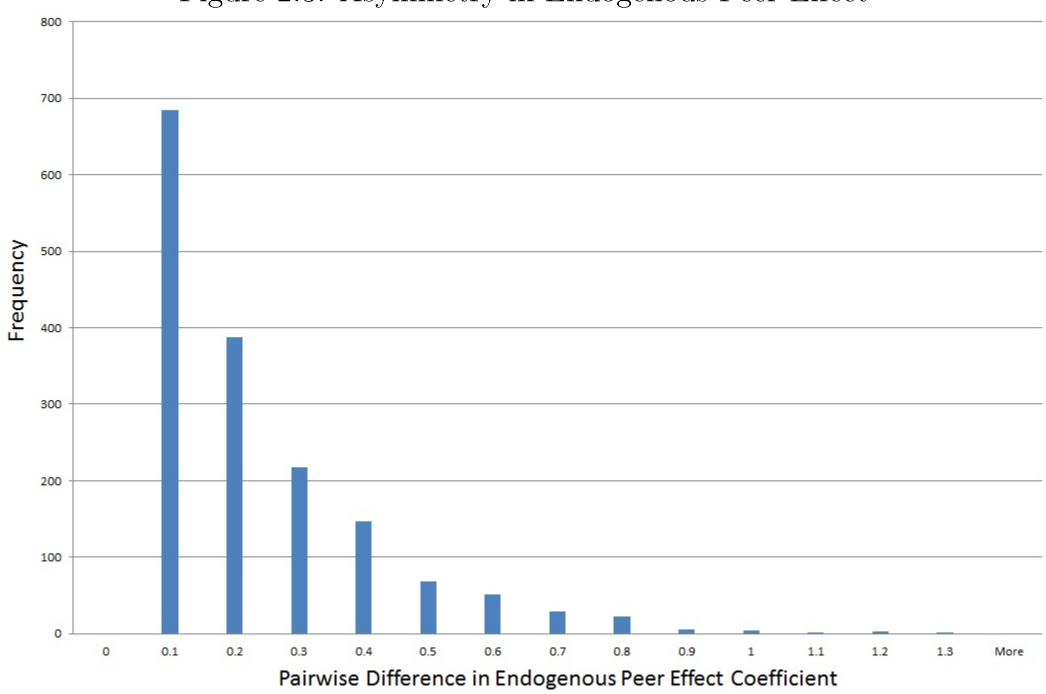
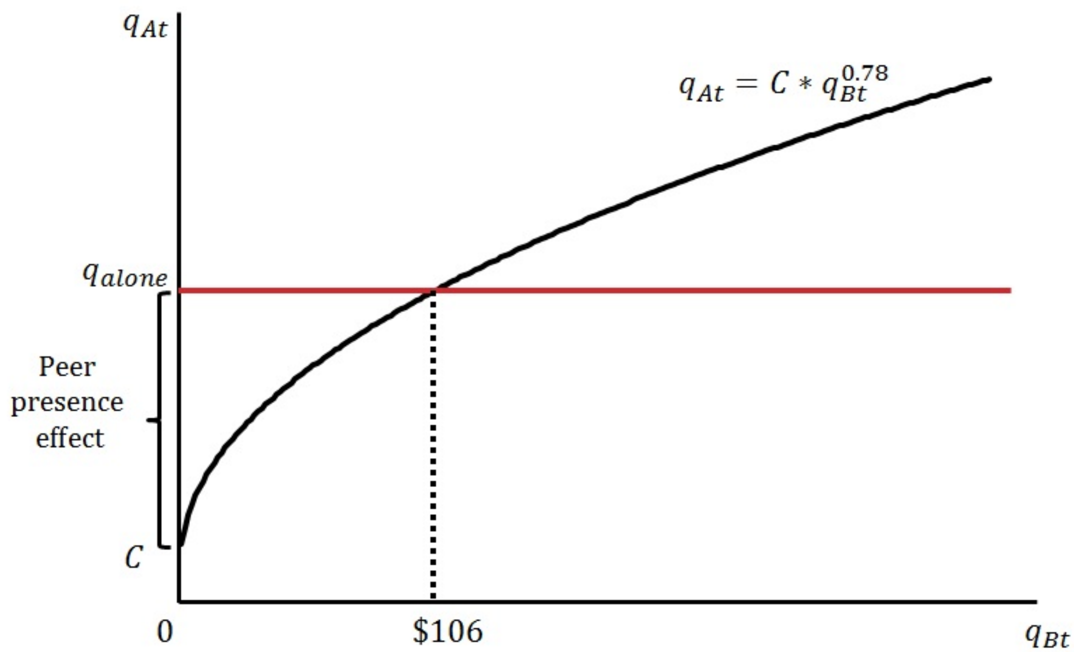


Figure 2.4: Negative Peer Presence Effect



# CHAPTER III

# Measuring the Agglomeration Effect on Consumers Store Choice

## 3.1 Introduction

Location is a key determinant of a retailer's profitability. One of the most important criteria for deciding where to locate a store is proximity to consumers. All else being equal, consumers will choose a store that minimizes travel costs. Another important factor is whether the location is close to other stores that provide different types of products. Because consumers often purchase multiple types of products in a single trip, the agglomeration (i.e., the cluster) of various types of retailers can provide shoppers with the benefit of convenience. An interesting notion is that these multi-purpose shoppers may travel farther to visit a store with other retailers nearby than a standalone store. From the retailers' perspective, this means that they can increase the catchment area by locating next to another retailer that carries different types of products. In the present study, we provide an empirical analysis that quantifies the agglomeration effect as the increase in the catchment area so that retailers can use this information to make better decisions related to locating new stores.

To investigate this objective, we develop and estimate an empirical model of consumers' store choice in the grocery industry. We calibrate our model on a panel dataset of household shopping behavior across multiple types of retailers. Our data tracks households' visits and purchases in both grocery stores and non-grocery stores across 34 states over a six-year period. Using a multinomial store choice model, we measure the increase in the consumer's likelihood of visiting a particular grocery store during the peak demand periods for non-grocery stores in the vicinity of the grocer. Based on this measure, we then infer the increase in the catchment area that a retailer could enjoy by locating next to other type of stores.

A key challenge in isolating the agglomeration effect empirically is that shopping districts with multiple retailers are likely to be located in areas that are intrinsically more attractive to businesses. Typically, intrinsic attractiveness is unobserved by researchers. Consequently, it is often difficult to determine if higher consumer traffic results from an agglomeration effect or if it is due to the business potential of the area *Pancras et al.* (2012); *Orhun* (2013). We rely on the disaggregated and panel nature of our data to control for location-specific unobservable factors. Store fixed effects capture the unobserved factors that affect a retailer's demand at a particular location.

Another challenge is to infer the causal impact of demand for a store carrying non-grocery items on the demand for grocery stores. Based on consumers' revealed choices of multi-stores, it is difficult to determine which demand causes the other. For instance, a consumer's high demand for the grocery store (e.g., due to the unique products it carries such as organic foods) could have led her to also visit a non-grocery store next to it. Also, a big sale in a department store (e.g., Macys) in the location could have driven her to visit both the grocery store and the non-grocery store nearby. In all of these cases, the data would report simply that the consumer visited both stores at the time. To infer the causality that non-grocery stores have on grocery stores, therefore, we identify an exogenous factor that shifts non-grocery demand but not grocery demand. We find that during the holiday gift shopping season (November and December), consumers are more likely to visit stores that carry gift items as their main category than during other periods. Such stores include toy stores, electronics stores, book stores, and multicategory stores such as Walmart. Although the demand for groceries does not change during this period, consumers' choice of where to shop for groceries does.

Investigating this agglomeration effect on multi-purpose shoppers has recently drawn the attention of empirical researchers. One stream of research indirectly measures the agglomeration effect through the firms' decision on whether and where to enter *Datta and Sudhir* (2011); *Vitorino* (2012). They assume that firms have accurate knowledge of the size of the agglomeration effect and make optimal decisions based on it *Sen et al.* (2011). Our approach is rather direct in that we do not rely on such assumptions but instead use household data to measure the agglomeration effect. Several empirical studies provide evidence of the agglomeration effect based on household-level data. *Leszczyc et al.* (2004) and *Arentze et al.* (2005) conducted empirical studies on how the agglomeration of stores increases consumers' utility for that location or store. They based their results on the estimation of a nested logit model of survey data in which they asked what grocery store respondents chose for multi-purpose shopping versus single-purpose shopping. *Fox et al.* (1997) conducted a similar study, which was based on a revealed choice dataset similar to that used here. All of these studies measured the effect of agglomeration by investigating how the number of stores surrounding the focal store or the existence of a mall nearby impacted the consumers' choice of a particular store. This specification, however, might be confounded with location-specific unobservables as mentioned. For example, perhaps the location featured high demand factors that drove many stores as well as malls to locate in the area. Moreover, their specification did not isolate the causality of the agglomeration effect. *Sen et al.* (2011) , however, documented the presence of an agglomeration effect by examining changes in consumers' spending at a supermarket after it opened a gas station. Compared to the previous studies mentioned, their study had a natural experimental setting, which helped the authors rule out alternative explanations. This study, however, explores the agglomeration effect within a single store. Moreover, the authors did not have information about travel time or distance from each consumer to the store. Therefore, they had no implications about whether consumers would tradeoff traveling a distance for an agglomeration of retailers.

In the present study, we explore how far a consumer is willing to travel for a cluster of various retailers when engaged in a multi-purpose shopping occasion. We account for the factors that could be confounded with this agglomeration effect. We then provide insights for the retailers in terms of increase in the catchment area based on the other retailers nearby. The remainder of this paper is organized as follows. In §3.2, we provide a brief description of the industry, the data, and the patterns observed in the data. Next, we describe our model in §3.3. The estimation results and the implication are provided in §3.4. Lastly, we conclude and discuss suggestions for further study in §3.5.

## 3.2 Data

### 3.2.1 The Industry

The grocery industry is one of the largest industries in US in terms of sales volume, generating more than $602 billion in 2012. Based on the 2010 Annual Retail Report by the US Census Bureau, the grocery industry's sales volume (13.6%) falls only behind that of the motor vehicle industry (19.4%) and general merchandise industry (15.8%). Based on US Department of Agriculture estimates, Americans spent an average of $26.78 dollars in a supermarket per visit and made an average of 2.2 trips per week.

In terms of store format, the grocery industry can be categorized into traditional grocery stores and non-traditional grocery stores. Traditional grocery stores include traditional supermarkets, superstores, fresh stores, and limited assortment stores. These stores typically offer a full line of groceries, meat, and produce but differ in terms of the size of the store and the type of assortments they emphasize on (e.g., Fresh stores emphasize ethnic, natural, and organic products). Non-traditional grocery retailers include mass merchandisers and wholesale clubs. Mass merchandisers such as Walmart, Kmart, and Target are stores that sell primarily hardlines, clothing, electronics, and sporting goods but also sell grocery and non-edible grocery items. Wholesale clubs are membership based retail/wholesale stores (Sam's Club, Costco, BJ's) with a varied selection but limited variety of products presented in a warehouse-type environment. They usually carry a grocery line dedicated to large sizes and bulk sales.[1]

Including non-traditional grocery retailers would make our empirical approach quite complicated because they carry both groceries and non-grocery items. Because our objective is to measure the benefit that grocery stores gain by locating close to non-grocery stores, for the sake of simplicity, we decided not to include non-traditional grocery retailers in our analysis.

We chose the grocery industry because consumers visit supermarkets frequently and spend a large portion of their income on groceries. Moreover, many shoppers seem to indulge in multi-purpose shopping when they shop for groceries. For example, *Leszczyc et al.* (2004) reported that 34% of the grocery shopping trips were multi-purpose trips, whereas 66% of trips were single-purpose trips. Also, *O'Kelly* (1981) reported an even higher percentage—63%—of all grocery trips were multi-purpose trips. Furthermore, based on a Booz & Company analysis (2011) of the grocery industry, when consumers were asked why they shop at their primary store rather

[1] More detailed information is at http://www.fmi.org/research-resources/supermarket-facts.

than one that is most convenient (closest in traveling distance), one of their main reasons was that the grocery store was located closer to other stores or places they wanted to visit.

### 3.2.2 Data Description

Our data reports households' visits and purchases in both grocery stores and non-grocery stores across 34 states between January 1998 and December 2002. The data includes the visit and purchase history of 4,749 households along with their demographic characteristics such as age, gender, and ZIP code. Table 3.1 includes the descriptive statistics for households' shopping behavior covering both grocery and non-grocery stores. On average, people have visited 2 stores during a week. During every store visit, consumers bought on average of 7 products. Here a product is defined as a good with distinct UPC (Universal Product Code). On average, consumers spend 31 dollars per visit and travel 4.8 miles to the stores. The distance was calculated as the Euclidian distance from the household ZIP code centroid and the store ZIP code centroid. To see how these shopping behaviors differ across grocery and non-grocery shopping, we calculated the same statistics separately for two occasions (i.e., grocery shopping vs. non-grocery shopping) as shown in Table 3.2. It shows that consumers visit more often but travel shorter distances to visit grocery stores than non-grocery stores. Also, consumers spend more money and buy more products in grocery stores than in non-grocery stores.

### 3.2.3 Data Patterns

In this section, we show the simple data patterns that support multi-purpose shopping behavior and the agglomeration effect. First, we wanted to verify whether consumers reflected in the data engaged in multi-purpose shopping behavior. Among all the store visit occasions, 25% were classified as multi-purpose shopping. Among all the grocery store visits, 37% were classified as multi-purpose shopping. This is consistent to *Leszczyc et al.* (2004) that 34% of grocery shopping trips were multi-purpose trips, whereas 66% were single-purpose trips. To determine if consumers travel farther to grocery stores when they are engaged in multi-purpose shopping than single-purpose shopping, we calculated the average distance traveled across these two occasions. As a result, we found that consumers traveled 2.78 (standard error 0.01) miles on average for single-purpose shopping, whereas they traveled 3.17 (standard error 0.02) miles when they were multi-purpose shopping. Thus, consumers traveled 0.3 (11%) farther when they were engaged in multi-purpose shopping.

However, this estimate does not rule out the alternative explanations mentioned previously. First, it could be that the location-specific unobservables are driving consumers to the location. Moreover, even if such location-specific unobservables did not affect them, it is still not clear whether their demand for non-grocery motivated them to travel farther for another grocery store.

Location-specific unobservables can be controlled by the panel structure of our data using the store fixed effect. The store fixed effect will capture all time invariant characteristics for the store, including the location-specific unobservables. To identify the causal influence of non-grocery stores on the demand for the grocery store, we need to have two data patterns.

First, we wanted to identify an exogenous variable that shifts non-grocery demand independently from the demand of grocery stores. We thus focus on the seasonality of demand for non-grocery items. Figure 3.1 through Figure 3.7 show the variation in theaverage number of visits by households to the stores for each category. The x axis indicates the date (i.e., year and month) and the y axis indicates the total number of visits by the population. Our intuition was that non-grocery items have peak demand periods such that consumers' visits to those stores would be high during particular high-demand periods. Groceries as a whole, however, should have constant demand throughout the periods because people eat food every day, and food items have limited time for the storage.

Based on Figures 1 through 7, we can identify an interesting pattern for seasonality of demand on non-grocery items. Toy stores (Figure 3.2), for example, seem to have peak demand in November and December. Electronics stores (Figure 3.3) have peak demand in December, November, and January. Book stores (Figure 3.4) have peak demand in December. Multicategory stores (Figure 3.5) have peak demand in December. In summary, toys, electronic products, books, and items carried in multicategory stores have peak demand during the gift shopping seasons (i.e., Thanksgiving, Christmas, and the New Year). Office product stores have peak demand in August, September, December, and January (Figure 3.6). These months are back-to-school periods when people buy their school-related products. Lastly, home improvement stores have peak demand in May (Figure 3.7), when the spring has started, and people are interested in decorating their home. Meanwhile, the demand variation for grocery stores is relatively constant, even during the holiday seasons in November and December. [2] The data pattern that non-grocery items have peak demand periods while the demand for grocery items remains constant helps us pin down asymmetric demand for the grocery stores with non-grocery stores nearby compared to those that stand alone.

In addition, we wanted check whether enough variation existed in the types of non-grocery stores that are close to grocery stores in different locations. Even if consumers actively engage in multi-purpose shopping, if all the grocery stores in the data have a similar pattern of non-grocery stores nearby, we would not be able to identify the agglomeration effect. Table 3.3 shows the number of grocery stores having a store combination of non-grocery stores within a one-mile distance. The specific brands of

[2] We speculate that there was a sudden increase in the average visits from April 2000 because the company that collected the data dropped panelists who had very rare visitations to stores. Because the total number of visits by the population stayed constant while the number of active panelists was reduced from April 2000.

the stores considered for each category are listed in table 3.6. Table 3.3 shows that the data include enough variation across grocery stores in terms of non-grocery stores nearby.

In summary, peak demand for non-grocery items in conjunction with variations in non-grocery store combinations across grocery stores helps us identify the causal impact of non-grocery stores on the grocery stores. Store fixed effects capture the unobserved factors that affect a retailer's demand at a particular location. A consumer's distance from alternative grocery storeshelps explain the grocery store at which she prefers to shop. Deviations from the baseline grocery store visit behavior due to a trip to another location driven by a non-grocery need helps us pin down the extent to which geographic agglomeration creates additional utility.

## 3.3 Model

In this section, we develop our model. Our notion was to check whether consumers' choice of a grocery store changed during peak demand periods for non-grocery items depending on the existence of non-grocery stores nearby. For example, we reviewed whether during the gift shopping period consumers were more likely to visit grocery stores that were proximate to a toy or an electronics store at the expense of grocery stores that do not have such retailers nearby. Our model takes the form of multinomial logit model, and the consumer's utility can be linearly expressed as follows:

$$\begin{gathered} U_{ijt} = \alpha_{0j} + \alpha_1 \times d_{ij} + \alpha_2 \times Gift\_Shopping\_Season_{it} + \alpha_3 \times Back\_to\_School_{it} \\ + \alpha_4 \times Spring_{it} + \alpha_5 \times Gift\_Shopping\_Season_{it} \times Gift\_Store\_Exist_j \\ + \alpha_6 \times Back\_to\_School_{it} \times Office\_Store\_Exist_j \\ + \alpha_7 \times Spring_{it} \times Home\_Store\_Exist_j + \varepsilon_{ijt} \end{gathered} \tag{3.1}$$

$U_{ijt}$ represents consumer $i$'s utility on store $j$ at time $t$. Our unit of analysis for time $t$ is a day. Consumer $i$ chooses from grocery stores that are within a 50-mile distance from the consumer's house. Therefore, every consumer has a different choice set. Every choice set includes the outside good, which is not visiting any grocery stores. For the identification, the utility of the outside good is set to zero.

$\alpha_{0j}$ is the store fixed effect that represents consumer's intrinsic preference toward the store. $d_{ij}$ is the Euclidian distance from household $i$'s home and store $j$. $Gift\_Shopping\_Season_{it}$, $Back\_to\_School_{it}$, and $Spring_{it}$ are dummy variables indicating whether day $t$ is a peak demand period for gift stores (i.e., toy, electronics, books, and multicategory stores), office stores and home improvement stores respectively. We defined $Gift\_Shopping\_Season_{it}$ as all the days in November and December consistent with the patterns we saw in the figures. However, we excluded the days in the week in which Thanksgiving day and Christmas day were included because people also tend to shop for groceries in these weeks. $Back\_to\_School_{it}$ is defined as days in August and September. $Spring_{it}$ is defined as days in May. $Gift\_Store\_Exist_j$, $Office\_Store\_Exist_j$, and $Home\_Store\_Exist_j$ are dummy variables indicating whether a gift store, office store, or home improvement store exists within a one mile distance from the grocery store $j$.

Store fixed effects $\alpha_{0j}$ captures the location-specific unobservables that could be confounded with the agglomeration effect. $\alpha_1$ refers to the disutility that consumer $i$ experiences when a grocery store is located farther from the consumer by one mile. $\alpha_2$, $\alpha_3$, and $\alpha_4$ represent the additional utility that the consumer has on grocery shopping during the gift shopping, back to school, and spring periods respectively. The coefficient for the interaction variables $\alpha_5$,$\alpha_6$,and $\alpha_7$ refers to the additional utility that the consumer $i$ has on store $j$ during the peak demand period for non-grocery items given that the grocery store has the non-grocery stores in the vicinity. As such, these coefficients represent the causal agglomeration effect that non-grocery stores have on grocery stores.

## 3.4 Results

The model was calibrated on sample of 457 households across 25 cities. The first column in Table 3.4 reports the results. Estimates for the store fixed effects are not reported for the sake of space. Among the interaction terms, only the coefficient for the interaction term $Gift\_Shopping\_Season_{it} \times Gift\_Store\_Exist_j$ was statistically significant. We also ran the same model but included store fixed effects, $d_{ij}$ , $Gift\_Shopping\_Season_{it}$ and $Gift\_Shopping\_Season_{it} \times Gift\_Store\_Exist_j$. The second column of the Table 3.4 shows the results. The coefficient for the interaction term, $Gift\_Shopping\_Season_{it} \times Gift\_Store\_Exist_j$, was still statistically significant and positive.

To interpret the results based on the second column of Table 3.4, $\alpha_2$ is estimated as -0.05, meaning that consumers are less likely to visit grocery stores without gift shops nearby during the gift shopping period. $\alpha_5$ is estimated as 0.12, meaning that during the gift shopping period, consumers are 14% more likely to visit a grocery store with nearby gift shops compared to those without any gift shops. $\alpha_1$ is estimated as -0.21, meaning that people are 23% less likely to visit a grocery store if it is one mile farther away from the household.

The results provide useful implications for the grocery retailers that they should locate their stores in the vicinity of other stores carrying different goods. This strategy provides shoppers with an opportunity to combine trips, shopping at different stores in a single shopping trip. These multi-purpose shoppers are willing to travel farther in return for access to various retailers. As for the retailer, this means that they could increase the catchment area. Based on our estimates for coefficients $\alpha_1$ and $\alpha_5$, it can be inferred that during the gift shopping period, a store that has gift store in the vicinity can enjoy an increase in catchment area by 0.61 miles. Considering that average consumers' grocery shopping distance is 3.24 (median was 1.85) this is a significant increase.

### The value of store fixed effects

To test the value of imposing store fixed effects, we also estimated a similar model but excluded store fixed effects (as follows) and compared the estimates to our main model results. Because only the coefficient for $Gift\_Shopping\_Season_{it} \times Gift\_Store\_Exist_j$ variable was significant, we decided to only include the variables related to gift stores in this analysis.

$$
\begin{gathered}
U_{ijt} = \alpha_0 + \alpha_1 \times d_{ij} + \alpha_2 \times Gift\_Shopping\_Season_{it} + \alpha_3 \times Gift\_Store\_Exist_j \\
+ \alpha_5 \times Gift\_Shopping\_Season_{it} \times Gift\_Store\_Exist_j + \varepsilon_{ijt}
\end{gathered}
\tag{3.2}
$$

The main difference from our main model was that instead of store fixed effects $\alpha_{0j}$ we included $\alpha_0$ and $Gift\_Store\_Exist_j$. $\alpha_0$ will represent consumers' intrinsic preference for grocery shopping. The coefficient for $Gift\_Store\_Exist_j$ will represent the difference in utility for store $j$ when a gift store exists within a one-mile distance. Note that we could not include $Gift\_Store\_Exist_j$ in our main model because store fixed effects account for all store-specific characteristics. Table 3.5 compares the estimates between two models. When store fixed effects were not included, the coefficient for $Gift\_Shopping\_Season_{it} \times Gift\_Store\_Exist_j$ was biased towards zero (0.06) compared to that of the model that included store fixed effects (0.12). Thus, without store fixed effects, the increase in the catchment area for retailers is inferred as 0.34 miles instead of 0.61 miles.

## 3.5 Conclusion

Location is a key determinant of retailer profitability. In the present study, we investigated how a retailer could increase the catchment area by locating next to other retailers carrying different goods. We infer such agglomeration effect by analysing how consumers' visit behaviors to grocery stores are influenced by non-grocery stores nearby. We extend the previous literature that found evidence of agglomeration effect by accounting for identification issues; location unobservables and direction of agglomeration effect. The present study may be especially helpful to traditional grocery stores threatened by the success of super centers. A recent trend in the retail environment is that stores are adding variety to the kind of goods they are carrying. For example, mass merchandisers such as Walmart, Kmart, and Target have added grocery items and have become super centers. The success of super centers can be explained partly by their ability to satisfy shoppers to engage in single-stop, multi-purpose shopping. As a result, this creates increased pressure on traditional grocery stores to compete with the super centers. Our results provide useful implications for traditional grocery stores that they could gain competitiveness by locating their stores in the vicinity of other stores carrying different goods.

Our study has some limitations. Our analyses focus on only one direction of the agglomeration effect; that is, the impact of non-grocery store demand on grocery store visits. Also, our results are constrained to agglomeration effect only during the peak demand period for non-grocery stores in the vicinity. To measure the agglomeration effect that could be applied in all periods, we need to use the actual visit behavior to the non-grocery stores and measure how that impacts visiting the grocery store nearby. We hope that future research can help address these limitations.

Table 3.1: Descriptive Statistics for households' shopping behavior

| Variable | Median | Mean | S.D |
|---|---|---|---|
| No. of visits per day | 0.26 | 0.30 | 0.20 |
| No. of products per visit | 3 | 7.07 | 9.84 |
| Total expenditure (dollars) per visit | 15.25 | 31.39 | 56.08 |
| Distance from home (miles) | 2.24 | 4.84 | 8.04 |

Table 3.2: Descriptive Statistics for households' shopping behavior across grocery and non-grocery shopping

| Variable | Grocery/ Non-Grocey | Median | Mean | S.D |
|---|---|---|---|---|
| No. of visits per day | Grocery | 0.17 | 0.20 | 0.15 |
| | Non-Grocey | 0.13 | 0.17 | 0.16 |
| No. of products per visit | Grocery | 6 | 10.51 | 12.01 |
| | Non-Grocey | 2 | 3.73 | 5.32 |
| Total expenditure (dollars) per visit | Grocery | 18.61 | 31.24 | 37.59 |
| | Non-Grocey | 11.99 | 31.54 | 69.43 |
| Distance from home (miles) | Grocery | 1.85 | 2.92 | 5.45 |
| | Non-Grocey | 3.68 | 5.68 | 12.34 |

Table 3.3: Number of grocery stores having each combination of non-grocery stores within a one-mile distance

| Non-grocery store combination | Number of grocery stores |
|---|---|
| Gift | 361 |
| Office | 26 |
| Home | 8 |
| Gift+Office | 194 |
| Gift+Home | 35 |
| Office+Home | 5 |
| Gift+Office+Home | 53 |
| None | 328 |
| Total | 1010 |

Table 3.4: Coefficient Estimates

| Variables | Estimates | | | |
|---|---|---|---|---|
| | Parameter | S.E. | Parameter | S.E. |
| Gift_Shopping_Period | -0.08** | 0.02 | -0.05* | 0.02 |
| Back_to_School_Period | 0.01 | 0.01 | | |
| Spring | 0.01 | 0.02 | | |
| Gift_Shopping_Period×Gift_Stores_Exist | 0.09** | 0.03 | 0.12** | 0.03 |
| Back_to_School×Office_Store_Exist | -0.01 | 0.02 | | |
| Spring×Home_Store_Exist | 0.08 | 0.05 | | |
| Distance | -0.22*** | 0.01 | -0.21*** | 0.01 |

*p<0.05, **p<0.01, ***p<0.001

Table 3.5: Comparison between models with and without store fixed effects

| Variables | Coefficient estimates | |
|---|---|---|
| | With store fixed effects | Without store fixed effects |
| Intercept ($\alpha_0$) | NA | -2.80***(0.01) |
| Gift_Stores_Exist | NA | -0.05** (0.01) |
| Gift_Shopping_Period | -0.05* ( 0.02) | -0.08** ( 0.02) |
| Gift_Shopping_Period×Gift_Stores_Exist | 0.12** ( 0.03) | 0.06* ( 0.03) |
| Distance | -0.21*** (0.01) | -0.19* (0.01) |

*p<0.05, **p<0.01, ***p<0.001

Table 3.6: Brand names of stores considered for each category

| Store Category | Brand Names |
|---|---|
| Toys Stores | TOYS R US |
| | BABIES R US |
| | KAYBEE TOY |
| | KIDS R US |
| Electronics Stores | BEST BUY |
| | CIRCUIT CITY |
| | COMPUTER CITY |
| | FRYS ELECTRONICS |
| | PC CONNECTION |
| | PC ZONE |
| | RADIO SHACK |
| | COMP USA |
| Book Stores | WALDENBOOKS |
| | BARNES NOBLE |
| | BORDERS |
| | B DALTON |
| Multicategory stores | K MART BIG |
| | WAL MART REG |
| | TARGET |
| | MEIJER |
| | COSTCO |
| | SAMS |
| Office Stores | STAPLES |
| | OFFICE DEPOT |
| | OFFICE MAX |
| Home Improvement Stores | HOME DEPOT |

Figure 3.1: Avg. No. of Visits by an household to Grocery Stores

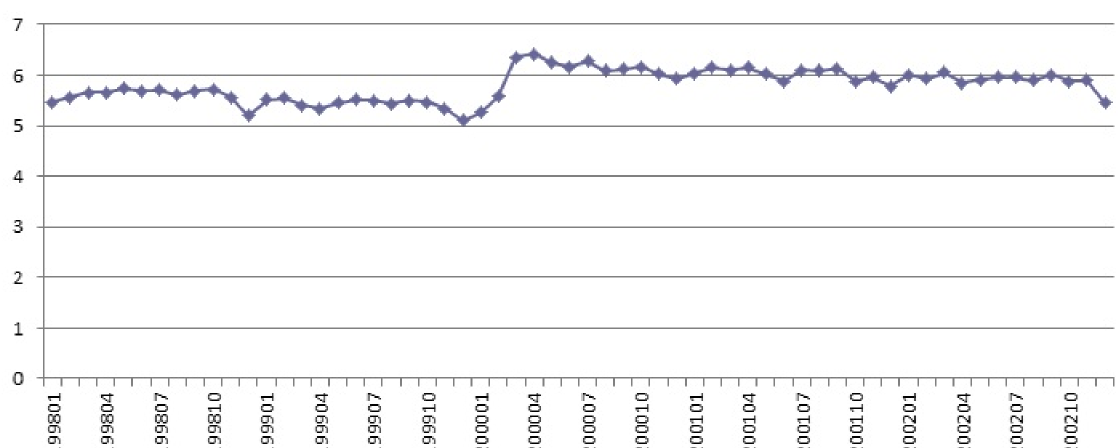


Figure 3.2: Avg. No. of Visits by an household to Toy Stores

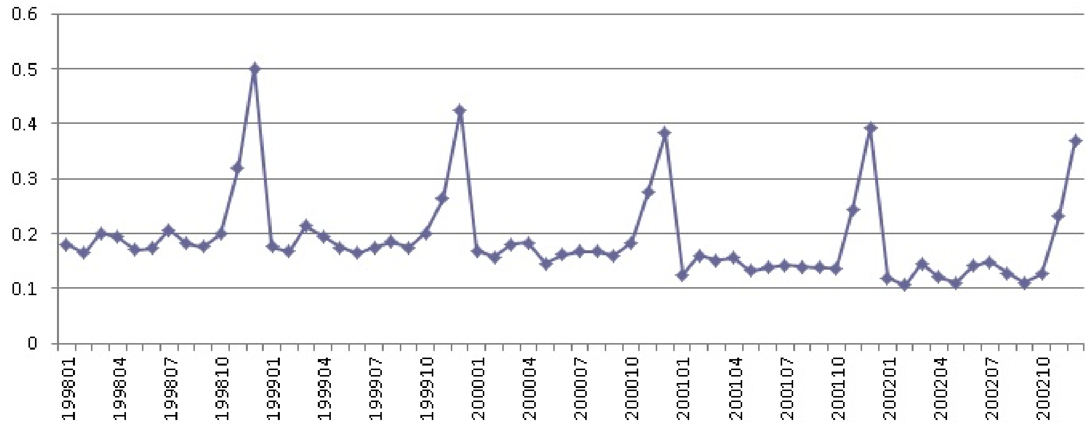



Figure 3.3: Avg. No. of Visits by an household to Electronic Stores

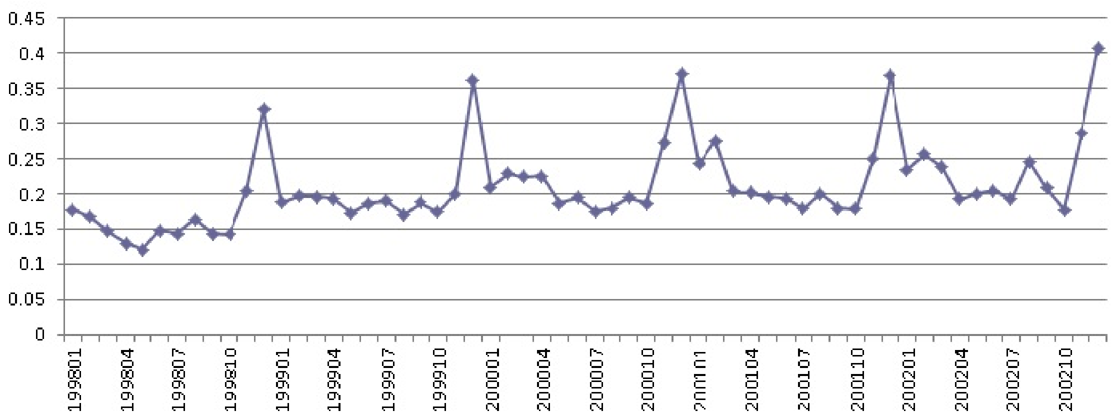



Figure 3.4: Avg. No. of Visits by an household to Book Stores

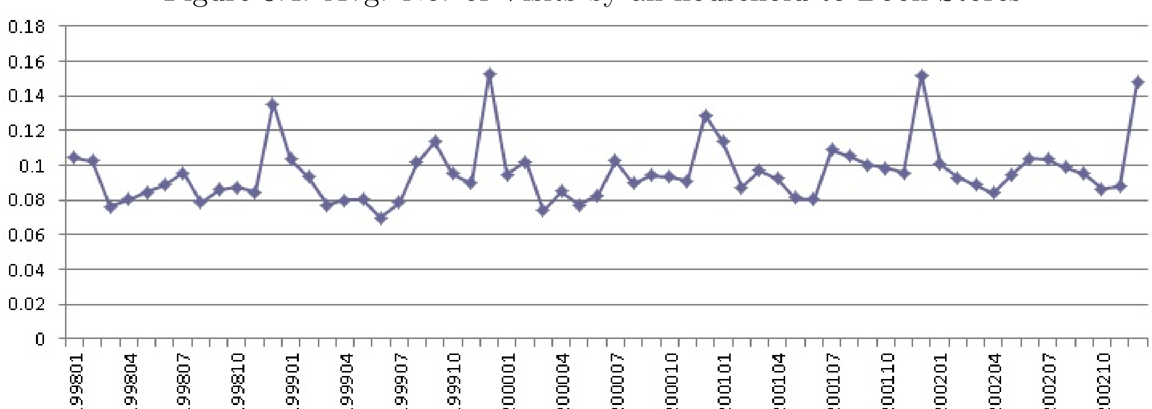


Figure 3.5: Avg. No. of Visits by an household to Multicategory Stores

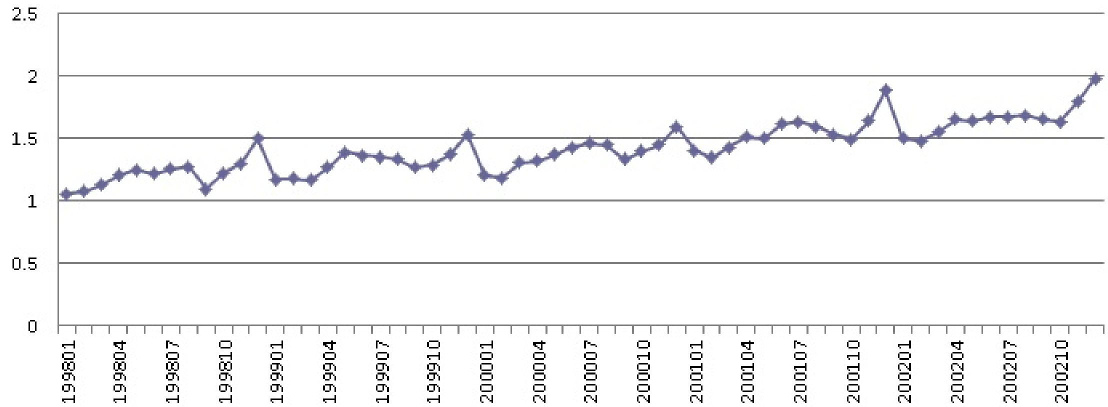



Figure 3.6: Avg. No. of Visits by an household to Office Product Stores

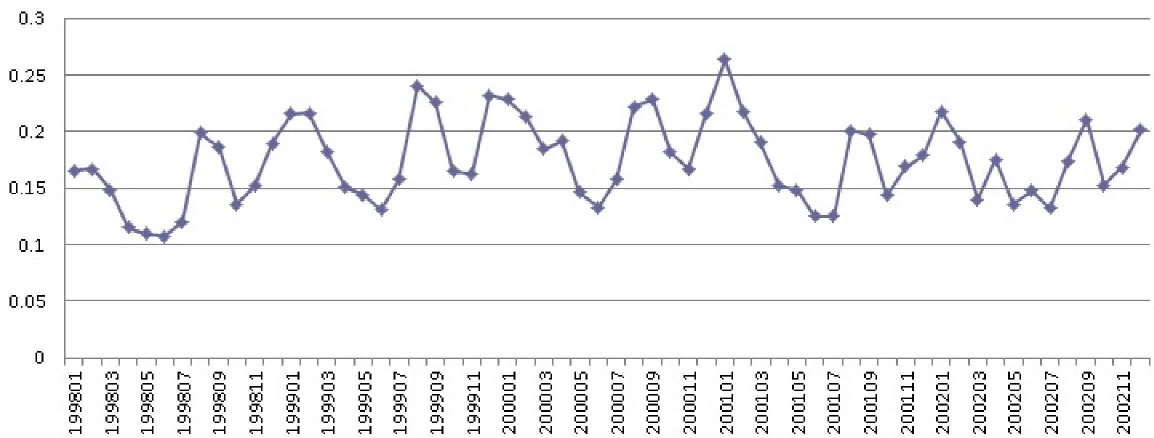
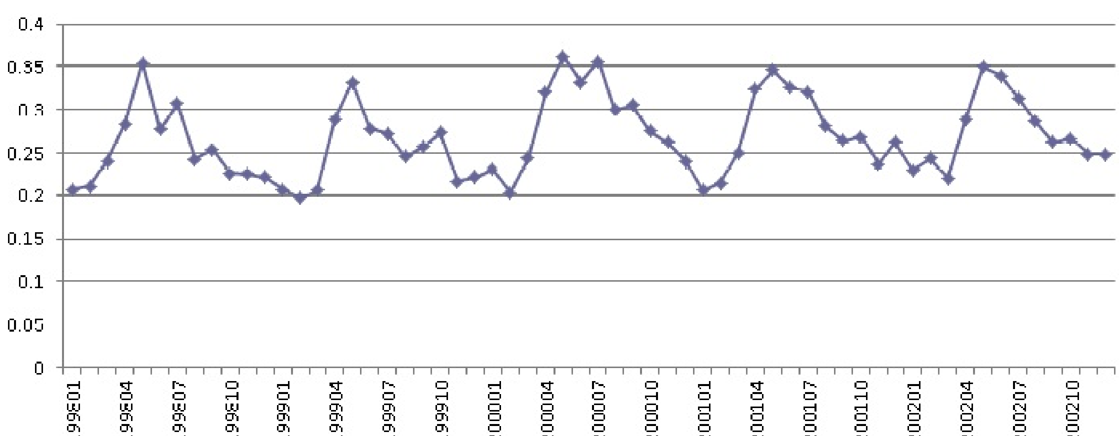



Figure 3.7: Avg. No. of Visits by an household to Home Improvement Stores



# CHAPTER IV

# General Conclusion

In markets where multiple agents coexist, one's behavior can be dependent on others' behavior. Throughout the two essays, we explore how to measure such interdependence across consumers and firms based on behavioral data while accounting for the confounding factors.

In the first essay, we extend the previous literature on consumption in various settings by accounting for exogenous factors that could change the peer's behavior (the exogenous peer effect) and whether the peer is present at the time of consumption but does not consume (the peer presence effect) in addition to the typically modeled endogenous peer effect (how one's behavior is influenced directly by the peer's behavior). By applying Hierarchical Bayes framework and resolving "tricky" ratio estimation with the MELO (Minimum Expected Loss) approach, we were able to estimate all three peer effects at an individual level. Based on individual estimates of peer effect, we provide guidelines of marketing managers for better promotional policies.

The core thrust of the research in the first essay was to identify peer effects from a behavioral dataset using an econometrics approach. One interesting domain to which the research could be extended is social media such as Facebook. The social media industry has grown significantly as Americans spend more of their online time at social networking and blog sites (23%) than they do engaging in any other activity. As such, firms increasingly emphasize social media marketing, allocating as much as 20% of their marketing budgets to this channel. One distinct and attractive feature about social media is the web interface that enables marketers to collect datasets easily. Interestingly, however, marketing managers struggle to measure ROI (Return on Investment) and use available rich information to improve marketing strategies. The key characteristic that makes measuring ROI difficult for managers is the consumer activities within the social media. Traditionally, managers can simply measure ROI by calculating the cost of the investment and estimating how that expenditure resulted in an increase of market sales. Within the realm of social media, however, consumers make connections with other members, write product reviews, comment upon, and recommend products to one another. Depending on the mix of these consumer activities, market sales could change dramatically. Extending the context of measuring interdependence among consumers, it would be interesting to measure the value of social activities and connections within social media that lead to increased sales.

In the second essay, we quantify the benefit that a retailer could enjoy by locating close to another retailer carrying different goods. We extend the previous literature that found evidence of agglomeration effect by accounting for identification issues; location unobservables and direction of agglomeration effects. In this way, we could make better prediction on how much the catchment area would increase if a store is located close to another type of store.

An interesting area to which the second essay could be extended is the relationship between retailers' pricing strategies and the agglomeration effect. In the second essay, we were able to translate the benefit of the agglomeration effect into the increase in catchment area because we had information on the households' travel distance. Similarly, if we could obtain information on the prices of products in stores and realize enough variation to identify the impact of price on consumers' visitation behaviors, we could measure the price premium a store could enjoy as a benefit of the agglomeration effect.

Another interesting research domain is to study the tradeoff between the agglomeration effect and the competition effect. In the second essay, we saw how a cluster of retailers that provided different goods could generate positive spillovers. A retailer, therefore, would want to locate within such a cluster. A retailer must also consider, however, the possibility of increased competition from similar retailers; indeed, such retail clusters are also likely to be attractive to a retailer's competition. It would be interesting to investigate the boundary conditions (such as the number of different types of stores) that would generate an agglomeration effect with sufficient benefits to offset any adverse effects that an additional competitor might pose.

# BIBLIOGRAPHY

# BIBLIOGRAPHY


AGA (2007), The 2007 state of the states: The american gaming association survey of casino entertainment, *http://www.americangaming.org/industry-resources/research/state-states*.

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press.

Arentze, A. T., H. Oppewal, and H. Timmermans (2005), A multi-purpose shopping trip model to assess retail agglomeration effects, *Journal of Marketing Research*, *42*, 109–115.

Argo, J. L., D. W. Dahl, and R. V. Manchanda (2005), The influence of a mere social presence in a retail context, *Journal of Consumer Research*, *31*, 207–212.

Becker, G. S., and K. M. Murphy (1988), A theory of rational addiction, *Journal of Political Economy*, *96*(4), 675–700.

Bresnahan, T., and P. Reiss (1990), Entry in monopoly markets, *Review of Economic Studies*, pp. 531–553.

Croson, R., and J. Sundali (2005), The gambler's fallacy and the hot hand: Empirical data from casinos, *Journal of Risk and Uncertainty*, *30*(3), 195–209.

Darke, P., and D. Dahl (2003), Fairness and discounts: The subjective value of a bargain, *Journal of Consumer Psychology*, *13*(3), 328–338.

Datta, S., and K. Sudhir (2011), Sleeping with the frenemy: The differentiation agglomeration trade-off in location choice, *Working Paper*.

Davidson, R., and J. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press.

Diebold, F., and R. Lamb (1997), Why are estimates of agricultural supply response so variable?, *Journal of Econometrics*, *76*, 357–373.

Feinberg, F., A. Krishna, and Z. J. Zhang (2002), Do we care what others get? a behaviorist approach to targeted promotions, *Journal of Marketing Research*, *39*, 277–291.

Fox, E. J., S. Postrel, and A. McLaughlin (1997), The impact of retail location on retailer revenues, *Working Paper*.

Gilovich, T., R. Vallone, and A. Tversky (1985), The hot hand in basketball: On the misperception of random sequences, *Cognitive Psychology*, *17*, 295–314.

Guryan, J., and M. S. Kearney (2008), Gambling at lucky stores: Empirical evidence from state lottery sales, *American Economic Review*, *98*(1), 458–473.

Hartmann, W. R. (2010), Demand estimation with social interactions and the implications for targeted marketing, *Marketing Science*, *29*(4), 585–601.

Hartmann, W. R., P. Manchanda, H. Nair, M. Bothner, P. Dodds, D. Godes, K. Hosanagar, and C. Tucker (2008), Modeling social interactions: Identification, empirical methods and policy implications, *Marketing Letters*, *19*, 287–304.

Kassinove, J. I., and M. Schare (2001), Effects of the 'near miss' and the 'big win' on persistence at slot machine gambling, *Psychology of Addictive Behaviors*, *15*, 155–158.

Lee, C. K., Y. K. Lee, B. J. Bernhard, and Y. S. Yoon (2006), Segmenting casino gamblers by motivation: A cluster analysis of korean gamblers, *Tourism Management*, *27*(5), 856–866.

Leszczyc, P. T. L. P., A. Sinha, and A. Sahgal (2004), The effect of multi-purpose shopping on pricing and location strategy for grocery stores, *Journal of Retailing*, *80*, 85–99.

Manchanda, P., Y. Xie, and N. Youn (2008), The role of targeted communication and contagion in new product adoption, *Marketing Science*, *27*(6), 950–961.

Manski, C. F. (1993), Identificatoin of endogenous social effects: The reflection problem, *Review of Economic Studies*, *60*, 531–542.

Moffitt, R. (2001), Policy interventions, low-level equilibria and social interactions, in *Social Dynamics*, edited by S. Durlauf and P. Young, MIT Press.

Nair, H., P. Manchanda, and T. Bhatia (2010), Asymmetric peer effects in physician prescription behavior the role of opinion leaders, *Journal of Marketing Research*, *47*, 883–895.

Narayanan, S., and P. Manchanda (2012), An empirical analysis of individual level casino gambling behavior, *Quantitative Marketing and Economics*, *10*(1), 27–62.

O'Kelly, M. E. (1981), A model of the demand of retail facilities, incorporating multistop, multipurpose trips, *Geographical Analysis*, *13*, 134–148.

Orhun, A. (2013), Spatial differentiation in the supermarket industry: The role of common information, *Quantitative Marketing and Economics*, *11*, 3–37.

Pancras, J., S. Sriram, and V. Kumar (2012), An empirical investigation of retail expansion and cannibalization in a dynamic environment, *Management Science*, *58*(11), 2001–2018.

Platz, L., and M. Millar (2001), Gambling in the conext of other recreational activity: A quantitative comparison of causal and pathological student gamblers, *Journal of Leisure Research*, *33*(4), 383–395.

Pollak, R. A. (1970), Habit formation and dynamic demand functions, *Journal of Political Economy*, *78*(4), 745–763.

Sen, B., J. Shin, and K. Sudhir (2011), Demand externalities from co-location: Evidence from a natural experiment, *Working Paper*.

Tanner, T., and W. Wong (1987), The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, *82*, 528–549.

Vitorino, M. A. (2012), Empirical entry games with complementarities: An application to the shopping center industry, *Journal of Marketing Research*, *49*, 175–191.

Yang, S., and G. M. Allenby (2003), Modeling interdependent consumer preferences, *Journal of Marketing Research*, *40*(3), 282–294.

Yang, S., V. Narayan, and H. Aasel (2006), Estimating the interdependence of television program viewership between spouses: A bayesian simultaneous equation model, *Marketing Science*, (3), 336–349.

Zeithammer, R., and P. Lenk (2006), Bayesian analysis of multivariate normal models when dimensions are absent, *Quantitative Marketing and Economics*, *4*(1), 241–265.

Zellner, A. (1978), Estimator of functions of population means and regression coefficients including structural coefficients: A minimum expected loss approach, *Journal of Econometrics*, *8*, 127–158.

Zhang, J. (2010), The sound of silence: Observational learning in the u.s. kidney market, *Marketing Science*, *29*(2), 315–335.