UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
The Geographic Concentration of Retail Stores: Trends and Determinants

Permalink
https://escholarship.org/uc/item/27x9w4b8

Author
Gu, Yizhen

Publication Date
2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

The Geographic Concentration of Retail Stores
Trends and Determinants

by

Yizhen Gu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

City and Regional Planning

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Elizabeth Deakin, Chair
Professor Paul Waddell
Assistant Professor Victor Couture

Spring 2015

The Geographic Concentration of Retail Stores
Trends and Determinants

Copyright 2015
by
Yizhen Gu

Abstract

The Geographic Concentration of Retail Stores
Trends and Determinants

by

Yizhen Gu

Doctor of Philosophy in City and Regional Planning

University of California, Berkeley

Professor Elizabeth Deakin, Chair

This dissertation mainly assesses the relative importance of two mechanisms–consumer searching and cost sharing for the spatial concentration of retail stores. I first present a simple model of firms' location choice and consumers' shopping destination choice. This model shows how both mechanisms promote the agglomeration of retail firms. The degree of retail agglomeration, measured by Ripley's K function, increases with the importance of consumer searching and with the benefits from scale economy in intermediate inputs such as advertisements and delivery services. I then present stylized facts based on industry agglomeration. Using the 2012 National Establishment Time-Series database, I calculate Ripley's K function and Ellison-Glaeser index that measure the degree of agglomeration for each CBSA-NAICS industry-year pair in the U.S. from 1992 to 2012. The estimated Ripley's K values show that the overall degree of retail agglomeration has been decreasing since 1992. Also, I find much inconsistency between the estimated Ripley's K and Eillson-Glaeser indices. Using the variation in the percentage of population that use Internet across regions and in the Internet sales share across retailing industries, I find that a 10 percentage points increases in the Internet use percentage and in the Internet sales share is associated with a 3.1-3.5 decrease and a 9.5-10.8 decrease in the calculated Ripley's K value. This provides suggestive evidence of the role consumer searching plays in promoting retail agglomeration since Internet decreases the importance of consumer searching in person. A difference-in-differences analysis shows that industries easy to make comparison shopping online experience a greater decrease in the degree of agglomeration by 6.8-12.6 before and after 2000. Last, I use a triple-difference approach based on the bankruptcy of chain stores like Borders and Circuit City to distinguish the effects of consumer searching and cost sharing. The estimation results for various types of chain stores using sales, employment size, the probability of closure, and the birth of new stores as dependent variables are mixed. Consumer searching alone cannot well explain retail agglomeration; however, there is little evidence of the contribution of the cost sharing mechanism to retail agglomeration.

Contents

Contents **i**

List of Figures **iii**

List of Tables **iv**

1 Introduction **1**

2 Literature review **5**
2.1 Measure of agglomeration 5
2.2 Determinants of retail agglomeration 8
2.3 Summary 13

3 A simple model of retailers and consumers **14**

4 Trends and determinants: cross-industry evidence **18**
4.1 Data 19
4.2 Retail agglomeration trends 23
4.3 Ripley's K Function vs. Ellison-Glaeser index 25
4.4 Consumer Searching: Cross-industry Evidence 27
4.5 Summary 32

5 Consumer Searching vs. Cost Sharing: Evidence from the Bankruptcy of Chain Stores **33**
5.1 Data 35
5.2 Model specification 36
5.3 Estimation results: sales 37
5.4 Estimation results: employment size 43
5.5 Estimation results: the probability of closure 44
5.6 Estimation results: the birth of new stores 44
5.7 Summary 44

6 Conclusions **46**

Bibliography **49**

A A list of 6-digit 2007 NAICS retailing industries **54**

B The estimated degree of agglomeration of more retailing industries **57**

C Find chain stores in the NETS database **61**

D More estimation results of models of sales **63**

E More estimation results of models of employment **75**

F More estimation results of models of store closure **83**

List of Figures

3.1 M retail firms in a square city with numerous radial roads 15

4.1 The spatial distribution of shoe stores in the SOF MSA in 2005 20
4.2 The annual agglomeration indices measured by Ripley's K from 1992 to 2012 . . 23
4.3 The estimated Ripley's K of book stores (451211) across CBSAs 24
4.4 The estimated Ripley's K of Shoe stores (448210) across CBSAs 25
4.5 The annual agglomeration indices measured by Ellison-Glaeser index from 1992 to 2012 28

B.1 The estimated Ripley's K of Radio, television, and other electronics stores (443112) across CBSAs 57
B.2 The estimated Ripley's K of Men's clothing stores (448110) across CBSAs . . . 58
B.3 The estimated Ellison-Glaeser indices of book stores (451211) across CBSAs . . 59
B.4 The estimated Ellison-Glaeser indices of Shoe stores (448210) across CBSAs . . 60

List of Tables

2.1 Five requirements of a well-designed measure by Duranton and Overman 2005 . 8

4.1 The percent of retail industries that are highly concentrated or not concentrated 22
4.2 The rank correlation coefficients between Ripley's K function and Ellison-Glaeser index 26
4.3 The industry rank by the degree of agglomeration: Example of one CBSA . . . 27
4.4 Estimation results of the model of agglomeration indices on the Internet use ratio 29
4.5 Estimation results of the model of agglomeration indices on the Internet sales share 30
4.6 Estimation results of the model of agglomeration indices on both the Internet use ratio and the Internet sales share 30
4.7 Estimation results of the DID model of agglomeration indices 31

5.1 The number of stores for each national chain used in the estimation 36
5.2 Estimation results of the DID model on sales: Specialty store 38
5.3 Estimation results of the DID model on sales: Department store (only Mervyn's) 38
5.4 Estimation results of the DID model on sales: Department store (only Goody's) 39
5.5 Estimation results of the DID model on sales: Toy store 40
5.6 Estimation results of the Triple-Diff model on sales: Specialty store 40
5.7 Estimation results of the Triple-Diff model on sales: Department store (Mervyn's) 41
5.8 Estimation results of the Triple-Diff model on sales: Department store (Goody's) 42
5.9 Estimation results of the Triple-Diff model on sales: Toy store 43
5.10 Estimation results of the Triple-Diff model on the birth of new stores 45

D.1 Estimation results of the DID model on self-reported sales: Specialty store . . . 63
D.2 Estimation results of the DID model on self-reported sales: Department store (Mervyn's) 64
D.3 Estimation results of the DID model on self-reported sales: Department store (Goody's) 65
D.4 Estimation results of the DID model on self-reported sales: Toy store 66
D.5 Estimation results of the DID model on self-reported sales of stores within 0.5 mile: Specialty store 67

D.6 Estimation results of the DID model on self-reported sales of stores within 0.5 mile: Department store (Mervyn's) 68
D.7 Estimation results of the DID model on self-reported sales of stores within 0.5 mile: Department store (Goody's) 69
D.8 Estimation results of the DID model on self-reported sales of stores within 0.5 mile: Toy store 70
D.9 Estimation results of the DID model on self-reported 2004-2012 sales of stores within 0.5 mile: Specialty store 71
D.10 Estimation results of the DID model on self-reported 2004-2012 sales of stores within 0.5 mile: Department store (Mervyn's) 72
D.11 Estimation results of the DID model on self-reported 2004-2012 sales of stores within 0.5 mile: Department store (Goody's) 73
D.12 Estimation results of the DID model on self-reported 2004-2012 sales of stores within 0.5 mile: Toy store 74

E.1 Estimation results of the DID model on employment: Specialty store 75
E.2 Estimation results of the DID model on employment: Department store (Mervyn's) 76
E.3 Estimation results of the DID model on employment: Department store (Goody's) 77
E.4 Estimation results of the DID model on employment: Toy store 78
E.5 Estimation results of the Triple-Diff model on employment: Specialty store . . . 79
E.6 Estimation results of the Triple-Diff model on employment: Department store (Mervyn's) 80
E.7 Estimation results of the Triple-Diff model on employment: Department store (Goody's) 81
E.8 Estimation results of the Triple-Diff model on employment: Toy store 82

F.1 Estimation results of the Triple-Diff model on store closure: Specialty store . . . 83
F.2 Estimation results of the Triple-Diff model on store closure: Department store (Mervyn's) 84
F.3 Estimation results of the Triple-Diff model on store closure: Department store (Goody's) 85
F.4 Estimation results of the Triple-Diff model on store closure: Toy store 86

Acknowledgments

I would like to express my deep appreciation to my committee chair, Professor Elizabeth Deakin, for the patient guidance and mentorship she provided to me over the last five years. I am truly fortunate to have had the opportunity to work with her.

I would also like to thank my committee members, Professor Victor Couture and Professor Paul Waddell, for their support over the past two years as moved from an idea to a completed study. In a similar vein, I would like to recognize Professor Robert Cervero, Professor Karen Chapple, and Professor Nancy Wallace for the contributions that each of them made to my intellectual growth during my years of study at the University of California, Berkeley.

My thanks must go also to Professor Richard Arnott at University of California, Riverside, who has been supervising my work since I started my Ph.D. adventure. It has been always my great pleasure to work with and talk to him.

Thanks to Professor Shihe Fu, Professor Yuming Fu, Professor Yan Song, Professor Junfu Zhang, Professor Junjie Zhang, and Professor Siqi Zheng. Each of them has been a mentor, colleague, and friend.

Thanks to my outside advisor, the late Professor John Quigley.

Thanks to my parents, who have been there for me from day one. Thank you for all of the love, help, and dedication.

Thanks to my son, who has been there for me for one year. Thank you for all the smiles, love, power and strength that you bring to our family.

Thanks to my dearest wife, who has been there for me for the last sixteen years. Thank you for all of your love, support, help, encouragement, and dedication.

Chapter 1

Introduction

The geographic concentration of retail firms, whether within industries or cross industries, is widely observed in cities. The extent to which stores within the same industry cluster, however, differs across industries. For example, automobile dealers often are clustered, while bakeries are far less so.

What are the micro-foundations of agglomeration economies for retail industries? This question has not been well addressed, especially compared to the studies on manufacturers. Marshall 1920 claims that the benefits of agglomeration arise from labor market pooling, input sharing, and knowledge spillover. Duranton and Puga 2004 propose a theoretical taxonomy that distinguishes three types of mechanisms of agglomeration economies–sharing, matching, and learning. There have been a dozen empirical studies that assess the relative importance of these mechanisms using the variation in the degree of agglomeration across industries, e.g., Rosenthal and Strange 2001, and Ellison, Glaeser, and Kerr 2010.

On the agglomeration of retail firms, there is a different thread of literature. Using a location-competition game, Hotelling 1990 shows that in equilibrium two retailers of a homogeneous product choose to locate at the center of a linear, bounded market, because any of them can increase its market sharer by locating close to its competitor. However, such an equilibrium will not hold when there are more than two vendors, as shown by Eaton and Lipsey 1975. A few studies turn their attention to the demand side, arguing that retail stores cluster to lower consumers' searching costs (e.g., Stahl 1982, and Wolinsky 1983). Further theoretical extensions introduce price competition–a force driving stores away from each other (e.g., Fischer and Harrington Jr 1996), as well as consumers' expectation for low prices where stores cluster (e.g., Konishi 2005).

Empirical studies on the micro-foundation of retail firms are very limited. Datta and Sudhir 2011 distinguish the effect of shopping externalities and that of price competition in retailers' location choice by decomposing grocery store revenues into two components–shopping volume based on consumers' shopping location choice and price based on spatial competition. Moreover, to my knowledge, there have been no empirical studies that investigate the variation in agglomeration across retail industries. One exception is the study by Billings and Johnson 2013; they, however, look at only the demand side, i.e. how trip

chaining promotes coagglomeration, and use a very rough measure–the difference in trip frequency between two types of stores.

More important, the sharing mechanism on the supply side, which may also promote the concentration of retail stores, is missing in the literature. First, stores choose to locate next to each other as a way of making themselves easy to be found, increasing consumer contact and saving on advertising costs. Anchor stores in a shopping mall pay less than market rent since they bring customers into other stores in the mall. Second, stores cluster to gain scale economy in product delivery from stores to consumers. This to some extent may explain the concentration of furniture stores.

This dissertation assesses the relative importance of two mechanisms–consumer searching and cost sharing–for the spatial concentration of retail firms. I start by reviewing the literature in the measure of agglomeration and the determinants of retail agglomeration in Chapter 2.

I then present a simple model of firms' location choice and consumers' shopping destination choice in Chapter 3. This model shows how both consumer searching and cost sharing mechanisms promote the agglomeration of retail firms. I use a circle city with consumers uniformly distributed in it. Retail firms can either locate in downtown or be scattered in the periphery according to the zoning. Here firms of the same type are assumed to sell the same product but of different brands. A consumer can increase her utility through searching since it decreases the "distance" between her ideal brand and the brand she finds. Firms choose to locate next to each other because this lowers consumers' searching costs and thus attracts more consumers. The sharing in advertising and product delivery is modeled as the gains from scale economy of intermediate products. For example, firms who are located next to each other also make themselves easy to be found and thus lower advertisement costs. There are two equilibrium conditions: consumers in the market boundary are indifferent between shopping in downtown and shopping in the periphery; and all retail stores get the same profit.

In this dissertation, I mainly use the 2012 National Establishment Time-Series (hereafter NETS) database, which is briefly introduced in Section 4.1 and 5.1. This database provides panel data of all establishments in the U.S. from 1990 to 2012 including location, standard industrial classification (hereafter SIC) code, employment, sales, and so on.

To empirically distinguish the effects of the consumer searching mechanism and the cost saving mechanism, I first present suggestive stylized facts based on industry agglomeration in Chapter 4. Using the NETS database, I calculate Ripley's K function (Ripley 1976, and Ripley 1977) and Ellison-Glaeser index (Ellison and Glaeser 1997) that measure the degree of agglomeration for each industry-CBSA-year pair in the U.S. from 1992 to 2012.[1] For

[1]A Core Based Statistical Area (hereafter CBSA) is "a U.S. geographic area defined by the Office of Management and Budget that centers on an urban center of at least 10,000 people and adjacent areas that are socioeconomically tied to the urban center by commuting." There are now 929 CBSAs in the U.S., including 388 Metropolitan Statistical Areas and 541 Micropolitan Statistical Areas. See `http://en.wikipedia.org/wiki/Core_Based_Statistical_Area` and `http://en.wikipedia.org/wiki/List_of_Core_Based_Statistical_Areas`, accessed Apr 9, 2013.

now I look at the variation across 75 NAICS6 industries (i.e., 6-digit NAICS). I also use the uneven growth of Internet sales across industries and regions in recent years (Weltevreden 2007). Because Internet sales lower the importance of consumer searching in the real world for retailing industries that sell standardized products and therefore reduce the benefits of agglomeration, this helps identify the contribution of consumer searching to retail agglomeration.[2] I can fully explore the longitudinal variation in the share of Internet sales, because the NETS database covers the period from 1992 to 2012 during which the Internet sales grow out of nothing and then expand from small to big.

I then use an identification strategy based on the bankruptcy of chain stores like Borders and Circuit City in Chapter 5. The logic is that, the two mechanisms–consumer searching and cost sharing–actually are two types of positive externalities. So the closure of a store means the loss of externalities for nearby stores. More important, such a loss depends upon industry. For example, if a Borders bookstore is closed, for nearby bookstores, this means the losses of two types of externalities–both consumer searching and advertisement cost saving, but for nearby grocery stores, this means the loss of the externality in only advertisement cost saving. The loss is measured using sales, the employment size, or the probability of closure in this dissertation. So, using chain stores such as Barnes & Noble and Best Buy that still exist as controls, I use a triple-difference approach to distinguish the effects of these two mechanisms.

Chapter 6 summarizes the main conclusions. This dissertation contributes to the literature in four aspects. First, it is the first to propose and investigate the cost sharing mechanism on the supply side that may encourage the geographic concentration of retail firms. Second, it shows how the measure of agglomeration is associated with factors that represent the consumer searching and cost sharing mechanisms using a simple model of firms and consumers. Third, Fourth, it empirically identifies the effects of consumer searching and cost sharing on agglomeration by i) using Internet as proxies and ii) using an exogenous shock–the bankruptcy of chain stores.

Planners and policy-makers have also strong interest in better understanding retail location choice. In the U.S., people nowadays make more shopping trips than commuting trips.[3] Policy makers have been trying to promote mixed land-use in order to decrease the share of trips made by auto as well as VMT (Vehicle Miles Traveled) of auto trips for shopping purpose. For example, Senate Bill 375 in California requires that a "Sustainable Community Strategy" be included in the regional transportation plan that provides incentives for projects such as transit-oriented development (TOD) and mixed land-use for meeting the

[2]Internet shopping may also make the advertising cost saving not as important as before, either by providing cheaper ways for stores to reach target customers (more and more individual information aided by Amazon, Google, Facebook, etc.) or by helping consumers locate all related stores, say, at Yelp. This works for all retailing industries. So by assuming this effect is the same across industries, I can identify the contribution of consumer searching.

[3]See "Summary of Travel Trends: 2009 National Household Travel Survey" by the U.S. Department of Transportation, available at `http://nhts.ornl.gov/2009/pub/stt.pdf`, accessed on Apr 9, 2013.

greenhouse gas emission targets.[4] Many communities are now planning infill and higher density housing with space set aside for local shops. However, it remains an issue whether such measures will work. Some of the questions this effort raises are: Is there a market for small retail spaces, e.g., on the first floor of residential buildings, and how large is that market? What kinds of establishments would be interested in such retail space? How large a shopping area, i.e., how much building floor area, needs to be available for retailers to be interested? Addressing these questions requires a better understanding of the factors that affect retailers' location choice.

A related issue is how shopping trends have been affecting demand for retail space. For example, big box retail has generated a spate of articles that document its impacts; it also has generated some pushback in the form of buy local movements.[5] Internet shopping is another growing trend that has challenged traditional brick and mortar shops. A Harvard Business Review article claims that Internet shopping is quickly heading toward 20% of retail sales, with the proportion varying significantly by sector.[6] The article continues, "Moreover, much digital retailing is now highly profitable. Amazon's five-year average return on investment, for example, is 17%, whereas traditional discount and department stores average 6.5%."

It would be good news for the implementation of Senate Bill 375 if, say, the growth of Internet shopping were found to have caused the decrease in the degree of retail agglomeration and a concomitant interest in smaller and more localized shopping for convenience goods. It would be bad news if instead it were wiping out the markets for small retail clusters and creating more shopping or freight trips.

[4]It is available at `http://www.leginfo.ca.gov/pub/07-08/bill/sen/sb_0351-0400/sb_375_bill_20080930_chaptered.pdf`, accessed on Apr 9, 2013.

[5]See, e.g., `http://www.ilsr.org/key-studies-walmart-and-bigbox-retail`, accessed on May 12, 2013.

[6]See `http://hbr.org/2011/12/the-future-of-shopping`, accessed on May 12, 2013.

Chapter 2

Literature review

2.1 Measure of agglomeration

Prior to any attempt to identify the determinants of agglomeration, it is important to design a measure of agglomeration. The way in which the geographic concentration is measured more or less depends upon what kind of spatial data, spatial point data or areal unit data, is available. Given the wide availability of areal unit data, several discrete-unit statistics have been applied such as Moran's I and the locational Gini coefficient. Moran's I is a measure of spatial autocorrelation widely used in spatial statistics/econometrics. It summarizes the overall correlation in the share of employment between any two areal units in the context of industrial agglomeration. Moran's I is calculated as:

$$I = \frac{M}{W_0} \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} w_{ij}(s_i - \bar{s})(s_j - \bar{s})}{\sum_{i=1}^{M}(s_i - \bar{s})^2}, \tag{2.1}$$

where s_i is the share of an industry's employment in each unit; $\bar{s}$ is the mean of all s_i's; w_{ij} is the weight between s_i and s_j (e.g., the inverse of the distance between s_i and s_j); W_0 is the sum of all w_{ij}'s; and M is the number of areal units.[1] The locational Gini coefficient was first introduced by Krugman 1991 and later applied in, e.g., Audretsch and Feldman 1996. It is based on the difference in the cumulative distribution of employment share between an industry and the overall economic activities, and calculated as:

$$L = \frac{2}{M^2 \bar{s}} \sum_{i=1}^{M} \lambda_i (s_i - \bar{s}), \tag{2.2}$$

where λ_i is the ranking of the unit in terms of s_i; and s_i, $\bar{s}$, and M are the same as defined in Equation 2.1 (Devereux, Griffith, and Simpson 2004).

The most notable discrete-unit index may be the one proposed by Ellison and Glaeser 1997 (hereafter the Ellison-Glaeser index). This index is derived from a firms' location

[1] See `http://en.wikipedia.org/wiki/Moran's_I`, accessed on Apr 9, 2013

choice model and can also be extended to the case of coagglomeration. It is easy to compute and allows both the comparison with a no-agglomeration benchmark and cross-industry comparisons. The Ellison-Glaeser index is calculated as:

$$\gamma = \frac{\sum_{i=1}^{M}(s_i - x_i)^2 - (1 - \sum_{i=1}^{M} x_i^2)^2 H}{(1 - \sum_{i=1}^{M} x_i^2)(1 - H)}, \tag{2.3}$$

where x_i is the share of total employment in each unit; H is the Herfindahl index; and s_i and M are the same as defined in Equation 2.1.[2] Ellison and Glaeser 1997 find that the concentration of establishments is common in almost every manufacturing industry in the U.S. Maurel and Sédillot 1999 derived a slightly different index using the same model as Ellison and Glaeser 1997, which is later used to measure the geographic concentration at the four-digit industry level in U.K. by Devereux, Griffith, and Simpson 2004.

The use of areal unit data implies that there are two types of agglomeration at work–local agglomeration and global agglomeration (see Guillain and Le Gallo 2010 and Di Giacinto and Pagnini 2011). The difference between them lies in whether the benefit from concentration extends beyond the geographic unit. These two location patterns tend to coexist in practice so using a measure of only one location pattern may cause biased results in empirical settings. The locational Gini coefficient and the Ellison-Glaeser index can only measure the localized agglomeration–the interaction between firms in the same unit, while Moran's I applies to only global effects–the interaction between different units. Guillain and Le Gallo 2010 suggest that Moran's I and the locational Gini coefficient be used at the same time to provide complete information. Di Giacinto and Pagnini 2011 use a spatial error component model to empirically estimate the indices of both local and global agglomeration. Their estimation on the basis of the Italian census data confirms the effects of both local and global factors on the geographical distribution of manufacturing and services industries.

Most discrete-unit indices suffer from the Modifiable Areal Unit Problem (MAUP) (Openshaw and Openshaw 1984), which means the estimates are influenced by the choice of areal unit boundaries. Think of the concentration of manufacturers crossing the border of two states. A higher estimate would be obtained if using a self-defined boundary, which puts all these manufacturers in the same geographic unit, instead of the administrative boundary.

With the increasing availability of spatial point data, distance-based measures provide a new way to summarize the geographic distribution of economic activities. Marcon and Puech 2003 classify these measures into two types: i) the probability density function of distance between pairs of points, and ii) the cumulative method that counts the number of neighbors within a certain radius.

The K-density proposed by Duranton and Overman 2005 belongs to the first type (Table 2.1). It is based on the distribution of distance to the nearest neighbor–a variant of the G function developed by Getis and Ord 1992. The basic idea of the K-density is to estimate the

[2]The Herfindahl index, $\sum_{i=1}^{M} s_i^2$, is a measure of the size of firms in relation to the industry and an indicator of the amount of competition among them. See `http://en.wikipedia.org/wiki/Herfindahl_index`, accessed on Apr 9, 2013.

distribution of the distance between pairs of establishments in an industry and to compare it with that of hypothetical industries with the same number of establishments which are randomly distributed conditional on the distribution of aggregate manufacturing.

$$\widehat{K}(d) = \frac{1}{n(n-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f(\frac{d - d_{i,j}}{h}), \tag{2.4}$$

where $d_{i,j}$ is the Euclidean distance between establishments i and j; n is the number of establishments; h is the bandwidth; and f is the kernel function. By being conditional on the distribution of aggregate manufacturing instead of imposing no restrictions on location, Duranton and Overman 2005 take into consideration the effect of zoning on industrial agglomeration. They also calculate a weighted index that incorporates the information on employment. Whether this is an effective way to consider the size distribution of plants, however, still needs to be tested. According to their estimation, the agglomeration level measured by the K-density is usually lower than that by the Ellison-Glaeser index: 94% of U.K. four-digit manufacturing industries are localized using the latter while only 52% by the former.

Cumulative methods such as Ripley's K function (Ripley 1976), Besag's L function (Besag 1977) and their extensions have been widely used in practice, mainly in ecology and forestry. Ripley's K function is estimated as:

$$K(t) = (\frac{n}{A})^{-1} \frac{\sum_{i=1}^{n} \lambda_i(t)}{n}, \tag{2.5}$$

where $\lambda_i(t)$ represents the number of establishments within distance t of establishment i; and A is total geographic area of the region. However, Ripley's K function does not satisfy some of the above requirements (Marcon and Puech 2010). First, it does not take into account industrial concentration–the size distribution of plants. Second, it is an absolute measure since it uses a completely random distribution of plants as benchmark instead of the overall location pattern of industries.[3] Marcon and Puech 2010 propose a new measure–the M function–that satisfies all five requirements. The M function is estimated as:

$$M = \sum_{i=1}^{n} \frac{\sum_{j=1, i \neq j}^{n} c(i,j,t) w_j}{\sum_{j=1, i \neq j}^{n_0} c_o i, j, t w_j} / \sum_{i=1}^{n} \frac{W - w_i}{W_0 - w_i}, \tag{2.6}$$

where $c_(i,j,t)$ is a binary dummy that equals 1 if the Euclidean distance between i and j is less than t; w_i is the weight, e.g., the number of employees; W is the sum of w_i; and n_0, c_0, and W_0 are indicators for establishments of all industries. Marcon and Puech 2010 suggest both the K-density and the M function, which provide complementary information, be used for describing the agglomeration. By comparing the estimated K values for firms

[3]There are still debates about the choice between an absolute measure and a relative measure (see Combes and Overman 2004 and Duranton and Overman 2005).

with economic linkages and that for firms without, Feser and Sweeney 2000 show that there is an association between economic linkages and geographic clustering.

Duranton and Overman 2005 claim that a well-designed measure, e.g., the K-density, should satisfy five requirements: i) being comparable across industries; ii) controlling for the natural tendency to agglomerate due to, say, population concentration;[4] iii) controlling for industrial concentration; iv) being unbiased with respect to scale and aggregation; and v) including an indication of the significance of the estimation results. The other indices can only satisfy some of these requirements (Table 2.1). For example, most discrete-unit indices tend to vary with the scale of geographic unit.

Table 2.1: Five requirements of a well-designed measure by Duranton and Overman 2005

Data type	Index	Satisfy the Requirement				
		I	II	III	IV	V
Areal unit	Moran's I	Yes	No	No	No	No
	Locational Gini	Yes	Yes	No	No	No
	Ellison-Glaeser	Yes	Yes	Yes	No	No
Spatial point	K-density	Yes	Yes	Yes	Yes	Yes
	Ripley's K function	Yes	No	No	Yes	Yes
	M function	Yes	Yes	Yes	Yes	Yes

Duranton and Overman 2005 further indicate the difference between the local confidence interval and the global one, which are simulated using the Monte-Carlo method. Because distance-based measures usually present estimates at different radii, it is expected that some of them may be outside the confidence level even if the point process is under a complete spatial random pattern.[5] So, the global confidence level can be calculated by increasing the local confidence level until the predefined threshold is reached.

None of the above measures have been used for estimating the degree of the agglomeration of retail stores. And there are very limited studies that relate the measure of industrial agglomeration with industry-specific characteristics, even in the context of manufacturing industries.

2.2 Determinants of retail agglomeration

Marshall 1920 claim that the benefits of agglomeration arise from the reduction of three types of transportation costs–the costs of moving goods, people and ideas. First, firms tend to locate close to their supplies or consumers to save transportation costs, and prefer locating close to their rivals if there is a strong effect of input sharing on lowering input

[4]Therefore, it is necessary to take into account the spatial concentration of population when investigating the micro-mechanisms of agglomeration if the measure of agglomeration does not do so.

[5]See `http://en.wikipedia.org/wiki/Complete_spatial_randomness`.

prices (Goldstein and Gronberg 1984). Second, labor market pooling, which means that firms can easily find appropriate workers, can also explain the clustering of firms in some circumstances (Helsley and Strange 1990). Third, knowledge spillover acts as another driving force of agglomeration economies (Glaeser, Henderson, and Inman 2000). The production of knowledge and its localized spillover is either through face-to-face exchange of tacit information, or through the inter-firm mobility of human capital (Baldwin, Brown, and Rigby 2010).[6] For a theoretical review of the microfoundations of agglomeration economies, see Quigley 1998.

However, retail agglomeration seems to be a very different story and cannot be well explained by the above theory. For example, labor market pooling may not be a concern in retailers' location choice because retail stores usually hire low-skilled workers.

Theoretical studies

Earlier studies mostly describe the spatial distribution of retail activities using central place theory (Berry 1967), the gravity model (Lowry 1964; Huff 1964), or spatial interaction theory as the extension of the gravity model (Wilson 1971). One major criticism, especially on the latter two theories, is their lack of economic rationale (Isard 1975).

There are mainly two approaches for modeling retailers' location choice: the game-theoretic (spatial competition) approach following the tradition of Hotelling 1990, and the general equilibrium approach that models the location choice of both retailers and residents.

The game-theoretic approach solves for the Nash equilibrium of a game between firms, with entry, location, and price as strategy variables and with consumers' location given. If there are two ice-cream vendors on a narrow beach, the equilibrium outcome predicted by minimum differentiation (Hotelling 1990) is that both choosing the middle point of the beach. However, when there are more than two vendors, Eaton and Lipsey 1975 show that clustering in the middle will not be a characteristic configuration in a linear city. So, most studies (e.g., Stahl 1982, Wolinsky 1983, and Dudey 1990) refer to consumer searching to explain the clustering of retailers. They assume consumers have incomplete information about products available at each store so the spatial concentration of stores helps lower consumers' searching costs. Dudey 1990 further examines the sequential location choice of firms, from which the dynamics of retail agglomeration can be derived.

Fischer and Harrington Jr 1996 incorporate both consumer searching and price competition into the model to explain the cross-industry variation in the geographic distribution of firms. Higher product variety increases the demand at a shopping center due to consumers' comparison-shopping, but also pushes a firm to stand alone to exert monopoly power. Their analysis shows that the former effect dominates and thus an industry of higher product variety tends to have a higher degree of agglomeration. The spatial competition model by

[6] Saxenian 1994 has a well-known case study on how knowledge spillover works in Silicon Valley. One channel she mentions is the information exchanges among business leaders. Ivarsson and Alvstam 2010 analyze the knowledge transfer without proximity, taking IKEA-suppliers in China and Southeast Asian as examples.

Konishi 2005 further includes consumers' expectations for low price, in addition to consumer searching and price competition. However, it remains an empirical issue whether the spatial concentration of retail stores is associated with low prices, considering the relatively high land rent where stores cluster. Using detailed barcode data on purchase transactions by households in 49 U.S. cities, Handbury and Weinstein 2014 show that large cities have a lower variety-adjusted price index.

One shortcoming of the game-theoretic approach is that it cannot be extended to the analysis of retail price, wage, and rent at a broader geographic level because of the mathematical difficulties when more types of retailers are introduced and households' locations are endogenously determined. In contrast, the general-equilibrium approach, on the basis of either perfect competition (e.g., Anas and Kim 1996) or monopolistic competition following Dixit and Stiglitz 1977 (e.g., Krugman 1990), is more suitable to address these concerns.

Adapting the approach developed by Fujita and Ogawa 1982, Arnott and Tu 2008 construct a general equilibrium model of location and land use for both residents and retail stores. In this model, agglomeration economy occurs due to individual's taste for product variety. Thus, stores can increase the effective variety through co-locating. Arnott and Tu 2008 show that the equilibrium location pattern depends on commuting cost, the spatial decay rate at which benefits from proximity to other firms decay with distance, the population, and the intensity of agglomeration economies in production.

In the model developed by Anas and Kim 1996, firms produce several types of products and at the same time act as retailers, selling products to both consumers and other firms. There is only one type of consumer who have heterogeneous tastes. Consumers not only commute to their workplaces but also travel to shop at all locations. Agglomeration economies now arise from i) asymmetric intermediate input interaction flows instead of symmetric interactions as modeled in Fujita and Ogawa 1982, and ii) scale economies in shopping, i.e. shopping externalities. The strength of spatial externalities and the accessibilities jointly determine the equilibrium urban spatial structure.

The New Economic Geography (NEG) model by Krugman 1990 adopts a monopolistic competition framework. This model is highly abstract, with two kinds of productions, i.e., constant-return agriculture and increasing-return manufacture, and two regions available for manufacturers and residents. There is no transportation cost within the region. Since the demand for manufacturing products comes from not only agriculture but also manufacture itself, this brings forward a positive feedback that drives divergence: manufacturers tend to cluster where there is a large market (home market effect), and the market becomes larger when residents are attracted by the lower price where manufacturers cluster (price index effect). Krugman 1991 shows whether such a feedback matters depends on transportation costs (of manufacturing products), economies of scale, and the share of manufacture.

The work by Fujita and Thisse 1986 is among the first to combine both the game-theoretic approach and the general equilibrium approach. In their model, the game-theoretic approach is applied to firms who compete for land, and the general equilibrium approach to residents, with the introduction of a land market. They show that under this model setup, the introduction of vacant land will affect the derived urban spatial structure, mainly

because consumers' location will change with firms' location.

In sum, the above theoretical studies indicate that the determinants of retail agglomeration include (geography) minimum differentiation, consumer searching, consumers' expectation for low price, price competition, and zoning. There are still factors that cause the concentration of retail stores, however, missing in the literature. Also, what factor is more important in promoting retail agglomeration? This calls upon empirical studies to assess the relative importance of these factors.

Empirical studies on retail agglomeration

The urban planning literature that uses data on individual stores is scarce. Dunkley, Helling, and Sawicki 2004 explore how households' food-shopping behavior influences store accessibility and store size in Atlanta, U.S. The dissertation by Sevtsuk 2010 investigates whether the spatial configuration of the built environment affects location choices of retail and food services establishments. He uses spatial linear models to estimate the probability that a building is being used for retail. His work shows that retailers' location choice is associated with other retailers' location choice, land use characteristics around each building as well as the geometry of the built environment.

There is also a huge literature on the big-box store by planning scholars; most studies, e.g., Boarnet et al. 2005, however, do not investigate its location choice. One exception is Graff 1998, who compares the different spatial expansion strategies Wal-Mart and Kmart adopted in locating their supercenters. The supercenters of Wal-Mart are located around its grocery distribution centers, while those of Kmart are located in the suburbs of widely scattered metropolitan markets and Kmart has yet to develop its grocery distribution network.

In the marketing field, empirical studies aim to distinguish the effect of shopping externalities and that of price competition in retailers' location choice. Datta and Sudhir 2011 contribute to the literature by decomposing the retail store revenues into two components–shopping volume based on consumers' shopping location choice and price based on spatial competition. However, the identification of the volume and price effects is aided by imposing functional form assumptions on how locations of rivals affect volumes and prices in distinct ways. To provide more insights into this issue, better data on retail prices across stores are required. Datta and Sudhir 2011 also control for the effect of zoning regulations, by limiting the choice set of retail locations as where retail is permitted rather than the whole area. In doing so, they make an assumption that zoning is exogenously fixed.

The work by Datta and Sudhir 2011 aims to distinguish the effect of shopping externalities and price competition on the basis of location choice of individual retail stores in only one industry. This strategy is different from the one adopted in studies that look at the variation in the degree of agglomeration across manufacturing industries. Billings and Johnson 2013 look at the cross-industry variation in retail agglomeration; they, however, look at only the demand side, i.e., how trip chaining promotes coagglomeration, and use a very rough measure–the difference in trip frequency between two types of stores.

Empirical studies on manufacturing agglomeration

One difficulty in identifying the sources of agglomeration economies is to find appropriate measures for these sources. For input sharing, Rosenthal and Strange 2001 use the manufactured inputs per dollar of shipments and the non-manufactured inputs per dollar of shipments as proxies. Baldwin, Brown, and Rigby 2010 and Rigby and Essletzbichler 2002 both use the location quotient to measure the upstream connections. For labor marketing pooling, the measures seen in the literature include the labor mix defined by Dumais, Ellison, and Glaeser 2002, and the proxies used in Rosenthal and Strange 2001 like the net productivity per worker, the ratio of the number of management workers to the sum of management and production workers, and the percentage of workers with Doctorates, Master's Degrees, and Bachelor/s Degree.[7] For knowledge spillover, both Henderson 2003 and Rosenthal and Strange 2001 use the counts or the densities of plants as a proxy for intraindustry spillover. Rosenthal and Strange 2001 uses also the innovations per dollar of shipments.

To address the endogeneity in such empirical settings, e.g., the input-output link induced by geographic proximity, Ellison, Glaeser, and Kerr 2010 use the UK data as well as the U.S. data (where two industries do not co-locate) as instruments. Also, to control for the natural advantage, e.g., a deep-water port where shipyards cluster, Kim 1999 estimates the Rybczynski equation that provides a linear relationship between regional production and factor endowments.[8] Ellison and Glaeser 1997 use the predicted state-level employment as a proxy for natural advantage. In addition, Baldwin, Brown, and Rigby 2010 use the metropolitan population size to control for urbanization economies that may not be captured by the other variables.

Rosenthal and Strange 2001 perform estimations at the zipcode, county and state levels using data on U.S. manufacturers. In so doing, they find that these sources actually work at different geographic units. For instance, labor market pooling works at all geographic levels, knowledge spillover works only in the zipcode level, and input sharing is observed at only the state level, which all conform to expectations. Baldwin, Brown, and Rigby 2010 present a similar analysis, however, using the plant-level panel data that are helpful for overcoming the selection bias arising from the unobserved plant-level heterogeneity. To date, the best empirical evidence on this topic is the work by Ellison, Glaeser, and Kerr 2010. They use the fact that various micro-mechanisms may work only in certain industry pairs. For example, the bank industry has the similar demand for labor as the real estate industry but not the textile industry. On the basis of this, Ellison, Glaeser, and Kerr 2010 investigate the sources of agglomeration economies using data on pairs of manufacturing industries.

[7] The reason for using these proxies is that, "if pooling is possible, an industry benefits by agglomerating because it is better able to hire workers with industry-specific skills." (Rosenthal and Strange 2001)

[8] The Rybczynski theorem refers to that "the maintenance of the same rates of substitution in production after the quantity of one factor has increased must lead to an absolute expansion in production of the commodity using relatively much of that factor, and to an absolute curtailment of production of the commodity using relatively little of the same factor" (Rybczynski 1955)

2.3 Summary

There are two types of agglomeration measures–discrete-unit statistics and distance-based measures. Discrete-unit statistics such as Ellison-Glaeser index have been widely used in empirical studies due to the availability of areal unit data. However, the same aggregate data may represent totally different spatial point patterns. Therefore, discrete-unit statistics suffer from the MAUP. Distance-based measures usually bear more information than discrete-unit statistics. It is expected that they will be applied in more and more studies, e.g., measuring the degree of agglomeration, and identifying the sources of agglomeration, given the increasing availability of spatial point data.

There have been quite a few game-theoretic studies that use consumer searching to explain the concentration of retail stores. In contrast, few studies have investigated the role of cost sharing on the supply side in retail agglomeration. Also, most empirical studies aim to distinguish the micro-mechanisms of manufacturing agglomeration, and the empirical literature on the sources of retail agglomeration is scant.

Chapter 3

A simple model of retailers and consumers

Here I present a simple model that shows how both the consumer searching mechanism and the cost saving mechanism can promote the agglomeration of retail stores. Imagine a circular city of unity radius.[1] In this city there are M ($M > 2$) retail firms that sell the same product, however, of different brands. According to the zoning regulations, these firms can choose to locate either in the CBD or in the periphery (Figure 3.1). Let the number of firms in the CBD be k, which is determined by both the consumer searching mechanism and the cost sharing mechanism. The CBD is a highly dense area, with the distance between any two firms in it less than a small enough number ϵ. The zoning regulations in this city also do not allow firms clustering in the periphery. Given this, it is easy to show that the $M - k$ firms in the periphery are uniformly spaced there by referring to the circle model of spatial competition.

There are numerous radial roads that connect the CBD and the periphery, with residents uniformly distributed on the roads. This implies a decreasing population density function with the distance to the CBD. The travel cost per mile is τ. There is also one circular road in the periphery, so that residents who do not live on the radial roads that directly connect one periphery firm to the CBD can still access the retail firms in the periphery. For simplicity, I assume that the travel cost for these residents to reach a closest periphery firm on only the circular road is zero. In so doing, I can look at only residents who live on the $M - k$ radial roads that directly connect a periphery firm to the CBD. It also allows a uniform market boundary between stores in the CBD and those in the periphery along any direction.

The degree of agglomeration for these M firms can be measured using Ripley's K function, which is calculated as the average number of firms within a certain radius of each firm divided by the density of firms in the study region. In the current setting, Ripley's K value is

$$K(r) = \frac{\pi^2 k(k-1)}{M^2}, \tag{3.1}$$

[1] In the earlier version I use a linear city.

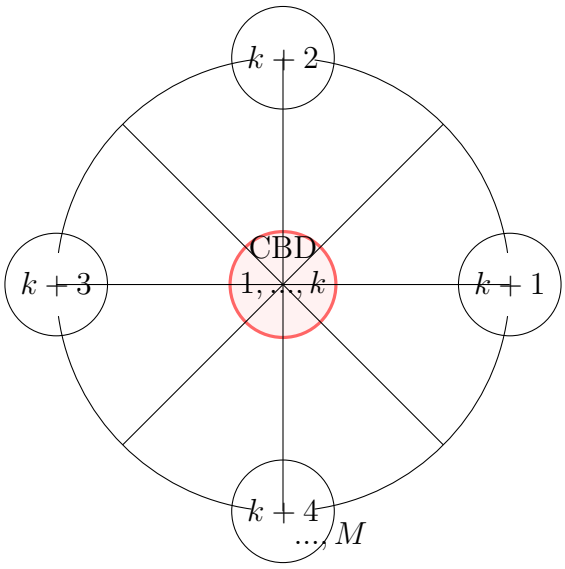


Figure 3.1: M retail firms in a square city with numerous radial roads

where r is the given radius and larger than ϵ. Since $K(r)$ increases with k, in the following I look at only the relationship between k and factors that represent the consumer searching mechanism and the cost sharing mechanism.

For now suppose that there is no price competition. Each product is made at marginal cost c_0 and sold at price p_0, with $p = p_0 - c_0$. There is also no rent differentiation with zero land rent in both CBD and the periphery.

In this setting, there are two endogenously determined variables–the number of firms in the CBD, $\overline{k}$, and the market boundary, $\overline{d}$ (distance to the CBD, $0 \leq \overline{d} \leq 1$), between firms in the CBD and those in the periphery. So residents who live within distance $\overline{d}$ to the CBD shop in only the CBD; otherwise shop in only the periphery. These two variables are derived from two equilibrium conditions: i) the resident who live in $\overline{d}$ obtains the same utility from shopping in the CBD and in the periphery; and ii) firms in the CBD and those in the periphery achieve the same profit, i.e., selling the same number of products.

I first model the mechanism of consumer searching on the demand side. Each consumer buys one product in one period. The difference in product brand is modeled as firms selling products that has an idiosyncratic component v. Here v is randomly given and has a uniform distribution on $(0, 1)$. For simplicity, let a representative consumer's preferred brand, v_0, be 0. The expected utility of a consumer is determined by the minimum distance between her preferred brand and the brand she finds through searching. This minimum distance has the Beta distribution, with its expectation equal $\frac{1}{n+1}$. The expectation of the minimum distance decreases with n–the number of firms a consumer has searched. I assume there is a high enough parking cost (monetary plus time), which prevents consumers from searching more than one store in the periphery. So consumer searching only happens in the CBD. The utility

of a consumer who live in d (distance to the CBD) from shopping in the CBD is

$$U_c(d) = f(k) - 2d\tau = f((k+1)\beta) - 2d\tau, \tag{3.2}$$

where β measures the importance of consumer searching, and f increases with β and k. Similarly, the utility from shopping in a periphery firm is

$$U_p(d) = f(1) - 2(1-d)\tau = f(2\beta) - 2(1-d)\tau. \tag{3.3}$$

With $U_c(\overline{d}) = U_p(\overline{d})$, I get

$$\overline{d} = \frac{f((k+1)\beta) - f(2\beta)}{4\tau} + 0.5. \tag{3.4}$$

It shows that: i) $\overline{d} > 0.5$ since $\overline{k} > 1$; ii) $\frac{\partial \overline{d}}{\partial k} > 0$, the market boundary of firms in the CBD increasing with the number of firms there; and iii) $\frac{\partial \overline{d}}{\partial \beta} > 0$, the more important the searching mechanism, the larger market boundary of firms in the CBD.

Next, I model the cost sharing mechanism on the supply side. First, each firm need send a signal that costs c to make itself into the choice set of all consumers in the city, and it has no way to send the signal to only targeted consumers. For firms in the CBD, they can share this cost equally given the small enough distance between any two of them. Second, firms need buy product delivery services from a nearby delivery firm to deliver the product to the residence of the consumer who buys it, if the product is not easy to carry along. The cost function of a delivery firm is $\gamma + \theta x$, in which x is the number of products to be delivered, θ is the marginal cost for delivering one product ($\theta < p$), and γ is the fixed cost and represents the benefit for retail firms from scale economy of delivery as an intermediate input. Firms in the CBD can be served by one delivery firm so they can gain scale economy, while firms in the periphery have to be served by the equal number of delivery firms. I assume firms in the CBD share equally consumers who shop in the CBD. Given that firms in the periphery are equally spaced, they also share equally the outgoing consumer flow. So, given N radial roads, the profits of a firm in the CBD and in the periphery are

$$\Pi_c = \frac{pN\overline{d} - \gamma - \theta N\overline{d}}{k} - \frac{c}{k}, \tag{3.5}$$

and

$$\Pi_p = \frac{pN(1-\overline{d}) - (M-k)\gamma - \theta N(1-\overline{d})}{M-k} - c, \tag{3.6}$$

respectively. With $\Pi_c = \Pi_p$, I get

$$L = -(\gamma + c)\overline{k}^2 + ((\theta - p)N + (\gamma + c)(M+1))\overline{k} + (p-\theta)MN\overline{d} - (\gamma + c)M = 0. \tag{3.7}$$

From $\Pi_c = \Pi_p$, it is easy to show that $\overline{d}M < \overline{k}$ since $\frac{\gamma + c}{k} < (\gamma + c)$. This implies that $\overline{k} \geq 0.5M + 0.5 > 0.5M$ given that k is integer. Also, $\frac{\partial \overline{d}}{\partial \overline{k}} < \frac{1}{M}$. This is because in equilibrium,

a firm in the CBD attracts fewer consumers than one in the periphery in order to maintain an equal profit condition. Thus, adding one more firm in the CBD, the increase in market boundary $\overline{d}$ is less than $\frac{1}{M}$, the average number of consumers a firm can attract on this radial road.

The equilibrium number of firms in the CBD $\overline{k}$ and the equilibrium market boundary $\overline{d}$ can be derived by combining 3.4 and 3.7. Inserting 3.4 into 3.7 and then calculating the total derivative of L, the partial derivative of $\overline{k}$ (degree of agglomeration) to β (consumer searching), c (advertisement cost sharing), and γ (delivery cost sharing) can be obtained as follows:

$$\frac{\partial \overline{k}}{\partial \beta} = \frac{-(2(p-\theta)\pi M d\frac{\partial \overline{d}}{\partial \beta})}{(M+1-2\overline{k})(\gamma+c)+(p-\theta)N(M\frac{\partial \overline{d}}{\partial \overline{k}}-1)}, \tag{3.8}$$

and

$$\frac{\partial \overline{k}}{\partial c} = \frac{\partial \overline{k}}{\partial \gamma} = \frac{-(-k^2+(M+1)k-M)}{(M+1-2\overline{k})(\gamma+c)+(p-\theta)N(M\frac{\partial \overline{d}}{\partial \overline{k}}-1)} \tag{3.9}$$

The denominators in both 3.8 and 3.9 are negative since $M+1-2\overline{k} \leq \overline{k}$ and $\frac{\partial \overline{d}}{\partial \overline{k}} < \frac{1}{M}$. It is trivial to show that the nominators in both equations are also negative. So, I get

$$\frac{\partial \overline{k}}{\partial \beta} > 0, \frac{\partial \overline{k}}{\partial c} > 0, \frac{\partial \overline{k}}{\partial \gamma} > 0. \tag{3.10}$$

This indicates that the degree of retail agglomeration increases with the importance of consumer searching and increases with the benefits from scale economy in intermediate inputs such as advertisement and delivery services.

Holton 1958 classifies products as i) convenience goods that are inexpensive frequent purchase, ii) shopping goods that cost more than convenience goods and are not frequently purchased, and iii) specialty goods with unique features or brands. Usually consumers will do some research (searching) for only shopping goods. Therefore, it is expected that retailing industries that sell shopping goods tend to have a higher degree of agglomeration, supposing all else equal.

The two mechanisms show in this model actually represent two types of positive externalities in retailers' location choice. The existence of such positive externalities means that without intervention, the actual degree of retail agglomeration would be less than the optimum, since a retailer cannot get all the benefits it produces. This to some extent justifies planners' intervention in terms of the geographic size and the density of retail zones. A shopping mall is another way to achieve the optimal level of retail agglomeration since the owner (also a planner) can capture all the benefits from concentration.

Possible extensions to this simple model include introducing: i) price competition; and ii) a land market and thus rent differentiation. Another option is to investigate how the Ellison-Glaeser index is associated with the consumer searching mechanism and the cost sharing mechanism using a city consisting of square grids.

Chapter 4

Trends and determinants: cross-industry evidence

There have been no studies that estimate the degree of the spatial concentration and its historic change across retailing industries. In an earlier study, Berry 1967 examines the spatial distribution of retail activities and shows that the central place theory holds for retail activities; however, he makes no effort in estimating the degree of agglomeration. One exception is the study by Fischer and Harrington Jr 1996, in which they find much variation in the concentration level across different types of stores, e.g., shoe stores, antique stores, and computer stores, by looking at the average distance between each store and its closest n competitors in the Baltimore metropolitan area. This actually can be seen as a variant of the G function in spatial statistics developed by Getis and Ord 1992. Also, there are a few studies that calculate the degree of agglomeration for manufacturing industries (e.g., Ellison and Glaeser 1997, and Duranton and Overman 2005). Nonetheless, these studies use only cross-sectional data and do not present the historical change in the agglomeration patterns. Ellison and Glaeser 1997 calculate the Ellison-Glaeser index for SIC4 industries using the 1987 Census of Manufacturers of the U.S. They find that 446 of 459 SIC4 industries display excess concentration. Duranton and Overman 2005 estimate the K-density for SIC3, SIC4, and SIC5 industries using exhaustive establishment level data from the 1996 Annual Respondent Database of the U.K. They find that 52% of SIC4 industries are localized at a 5% confidence level.

In this chapter, I first investigate how the degree of retail agglomeration has changed from 1992 to 2012 and how this trend has varied across industries and regions using the 2012 National Establishment Time-Series (NETS) database. I calculate Ripley's K function and Ellison-Glaeser index for each industry-CBSA-year pair in the U.S. and look at the variation in the estimated agglomeration indices across all 929 CBSAs and 75 NAICS6 industries.[1] I then make a comparison between the estimated Ripley's K function and Ellison-Glaeser

[1]CBSAs are aggregated to counties and group of counties, with an urban center and adjacent areas by commuting. CBSAs vary widely in physical size as well as in population and employment.

index.

Rosenthal and Strange 2001 and Baldwin, Brown, and Rigby 2010 distinguish three types of sources of agglomeration economy for manufacturers–input sharing, labor market pooling, and knowledge spillover–by designing measures for these three sources at the industry level.[2] However, it is not easy to find measures for the consumer searching mechanism and the cost sharing mechanism. Here I use the share of Internet sales of a retailing industry and the percentage of population that use Internet in a CBSA as proxies of consumer searching, so I can explore the variation in Internet sales across industries, the variation in Internet use across regions, as well as the variation in these two measures with time. This helps identify the contribution of consumer searching to retail agglomeration, because Internet lowers the importance of consumer searching in the real world for retailing industries that sell standardized products. Also, Internet sales grow out of nothing and then expand from small to big between 1992 and 2012, so I can use the change in the importance of consumer searching brought about by Internet for identification.

This section is structured as follows. Section 4.1 introduces how I calculate the agglomeration indices and collect data on Internet use by region and Internet sales by industry. Section 4.2 presents stylized facts about retail agglomeration. Section 4.3 compares the estimated Ripley's K function and the estimated Ellison-Glaeser index. Section 4.4 presents empirical findings about consumer searching by using the Internet use and the Internet sales as proxies. Section 4.5 summaries the main findings.

4.1 Data

2012 NETS data

In this dissertation I mainly use the 2012 National Establishment Time-Series (NETS) database.[3] This database is created by Walls & Associates on the basis of the Dun and Bradstreet (D&B) data. It provides an annual "snapshot" of essentially all establishments in the U.S. between 1990 and 2012. Given this, the NETS database offers advantages in carrying out research on retail agglomeration. I drop the 1990-1991 data considering the coverage problem with the D&B data in earlier years (Neumark, Zhang, and Wall 2005).

In the NETS database, there is a unique DUNS number for each establishment to link the information in various tables. The NETS database presents the business name and contact information of each establishment, the type of each establishment in terms of proprietary, partnership, corporation, and nonprofit, as well as the information on being public or not, being foreign-owned or not, and so on. For each establishment, the NETS database tells whether it is a single-location firm, a headquarter, or a branch, and records its linkages with

[2]The cost sharing mechanism for retail agglomeration in this dissertation can be thought of as sharing advertisement input and product delivery (to consumers) input.

[3]Neumark, Zhang, and Wall 2005 provide a detailed explanation and an assessment on the data quality of the NETS data.

other establishments, e.g., its headquarter in its latest year, the number of establishments reporting to the same headquarter, and the number of establishments reporting to it.

For each establishment in each year, the NETS database presents its 8-digit Standard Industrial Classification (SIC) code. This enables me to investigate the variation in the degree of agglomeration across retailing industries. Similar to studies on manufacturers by Baldwin, Brown, and Rigby 2010 and Rosenthal and Strange 2001, I use 4-digit industries as observations.

An important feature of the NETS database is that it contains the detailed address and thus the longitude-latitude of each establishment in each year as well as the Metropolitan Statistical Area (MSA), county, city, and zip code in which an establishment is located.[4] As an example, Figure 4.1 shows the spatial distribution of shoe stores (4-digit SIC: 5661) in the San Francisco-Oakland-Fremont MSA in 2005.



Figure 4.1: The spatial distribution of shoe stores in the SOF MSA in 2005

For each establishment, the NETS database includes its employment size and sales in each year. Sales sometimes are estimated by Walls & Associates instead of being reported

[4]Not all establishments have their detailed address disclosed in the NETS database. For example, for 74.5% of establishments included in the 2009 NETS database, the latitude-longitude is accurately given at the "rooftop" level; however, for 24.1% of establishments, the longitude-latitude is given as the centroid of zip code. Though the accuracy has been much improved since 2000, there may be significant bias when calculating the degree of agglomeration in early 1990s (Walls & Associates 2009).

by the firms. There are other time-variant variables like D&B credit ratings and PayDex Scores.

The NETS database records the year an establishment started and the year it ended. Of all establishments in California that were created before 2011, only about 49.7% were still alive after 2011. The high "death" rate indicates that there may be high variation in the degree of agglomeration from year to year.

Another feature of the NETS database is the relocation information included, i.e., the origin and destination latitude-longitudes for each permanent, across-zip code move. From the relocation information, I can get the latitude-longitude in each year for all establishments. Of establishments in California that existed between 1990 and 2011, nearly 10% have moved at least once.

Agglomeration indices

I first generate a series of annual datasets of retail firms with NAICS6 codes ranging from 441110 to 454390 using the 2012 NETS database. The 44 and 45 codes include all retail trade. Appendix A lists all 6-digit 2007 NAICS retailing industries, which includes, for instance, new car dealers (4411110) and beer, wine, and liquor stores (445310).[5]

For each of these annual datasets of retail firms, I then calculate the degree of agglomeration for each NAICS6 retailing industry within each of 929 CBSAs using two types of agglomeration indices–Ripley's K function as a distance-based measure and Ellison-Glaeser index as a discrete-unit index. As explained in Chapter 2, Ripley's K function measures the average number of neighbor stores within a certain distance threshold divided by the density of stores in the study area, and Ellison-Glaeser index measures the deviation of the employment spatial distribution of a retailing industry from that of all retailing industries.

Ripley's K function is calculated using Equation 2.5 in Section 2.1. Here I use 0.5 mile as t–the distance threshold. So the externalities are limited within 10-minute walking distance, which seems not unreasonable. This also assumes that the externalities from both consumer searching and cost sharing are equally gained by only stores within 0.5 miles, because there is no weight imposed. Ellison-Glaeser index is calculated using Equation 2.3 in Section 2.1. I choose zip code as the spatial unit because it is readily available in the 2012 NETS database. This implies that the externalities are shared by stores within the same zip code. Note that zip codes vary in size, with its approximate radius usually larger than 0.5 mile, which inevitably brings about inconsistencies between the estimates of these two indices.

Though the 2012 NETS database records the coordinates of every establishment, the coordinates are not always accurate and sometimes are given at the zip code level. However, the accuracy of the reported coordinates has been improved a lot since 2001 (Walls & Associates 2009). For example, there are 87,048 restaurants with latitude-longitude reported at the zip code level in 1992, which lowers to 66,272 in 2000 and 17,092 in 2009. Though this will not bias the estimated Ellison-Glaeser indices since they are calculated at exactly

[5]See `http://www.census.gov/cgi-bin/sssd/naics/naicsrch?chart=2007`.

the zip code level, it tends to cause upwards bias for the estimated Ripley's K values. More important, it causes higher bias for the estimated Ripley's K values in earlier years.

I also collect data on Internet use across regions and Internet sales across industries. The ratio of population that use the Internet at the CBSA level is constructed using the Current Population Survey (hereafter CPS) data. This ratio is available for 2001, 2007, 2009, 2010, 2011, and 2012 because the Internet use supplement data are collected for only these six years.[6] The Internet use data covers about 200 CBSAs each year, and the CBSAs covered vary across waves of surveys.

The Internet sales share of each NAICS6 retailing industry in 2002 and 2007 is obtained from Tran 2014. Using the microdata of the Census of Retail Trade (hereafter CRT) in 2002 and 2007, Tran 2014 calculates the share for each retailing industry. I crawl the data from Figure 5 in this paper using the GetData software. The data show that the Internet sales share of book stores (NAICS 451211) increases from 6.8% in 2002 to 23.5% in 2007, the highest in both years. Beer, wine, and liquor stores (445310), baked goods stores (445291), and paint and wallpaper stores (444120) are the lowest three industries, with the share lower than 0.2% in both years. Moreover, I supplement the data by assuming that the Internet sales share for all retailing industries is zero before 1995. This is quite reasonable because i) the first secure retail transaction over the Web was in 1994 and ii) both Amazon and eBay were introduced in 1995.[7]

Table 4.1: The percent of retail industries that are highly concentrated or not concentrated

	% of being highly concentrated	% of being not concentrated
1992	12.1	86.6
1995	11.6	87.2
2000	10.8	88.2
2005	10.4	88.7
2010	10.1	88.9
2012	9.9	89.1

The Ellison-Glaeser index is calculated for all NAICS6-CBSA pairs in each year using the 2012 NETS database. Highly concentrated: the estimated value >0.05; Not concentrated: <0.02.

[6] They are the computer and Internet use supplement data of CPS, September 2001, and July 2011; and the school enrollment and Internet use supplement data of CPS, October 2007, October 2009, October 2010, and October 2012. These data and the related STATA Do files are available at `http://thedataweb.rm.census.gov/ftp/cps_ftp.html\#cpssupps` and `http://www.nber.org/data/cps_progs.html`, accessed on Dec 9, 2014

[7] See `http://en.wikipedia.org/wiki/Online_shopping`, accessed on Dec 15, 2014.

4.2 Retail agglomeration trends

Using the estimated Ripley's K values and Ellison-Glaeser indices, I first investigate whether retail stores are spatially concentrated. There is no guideline about whether an industry is spatially concentrated or not using Ripley's K function. If Ellison-Glaeser index is used, Ellison and Glaeser 1997 suggest that an industry with the estimated value higher than 0.05 is highly concentrated, and with the value lower than 0.02 is not concentrated.

In each year between 1992 and 2012, the ratio of industries that are highly concentrated of all NAICS6-CBSA pairs is only about 10%, while that of industries that are not concentrated is more than 85% (Table 4.1). Moreover, the former ratio decreases from 12.1% in 1992 to 9.9% in 2012, while the latter increases from 86.6% to 89.1% during the same period. This provides suggestive evidence that overall the retail stores may not be as geographically concentrated as before.

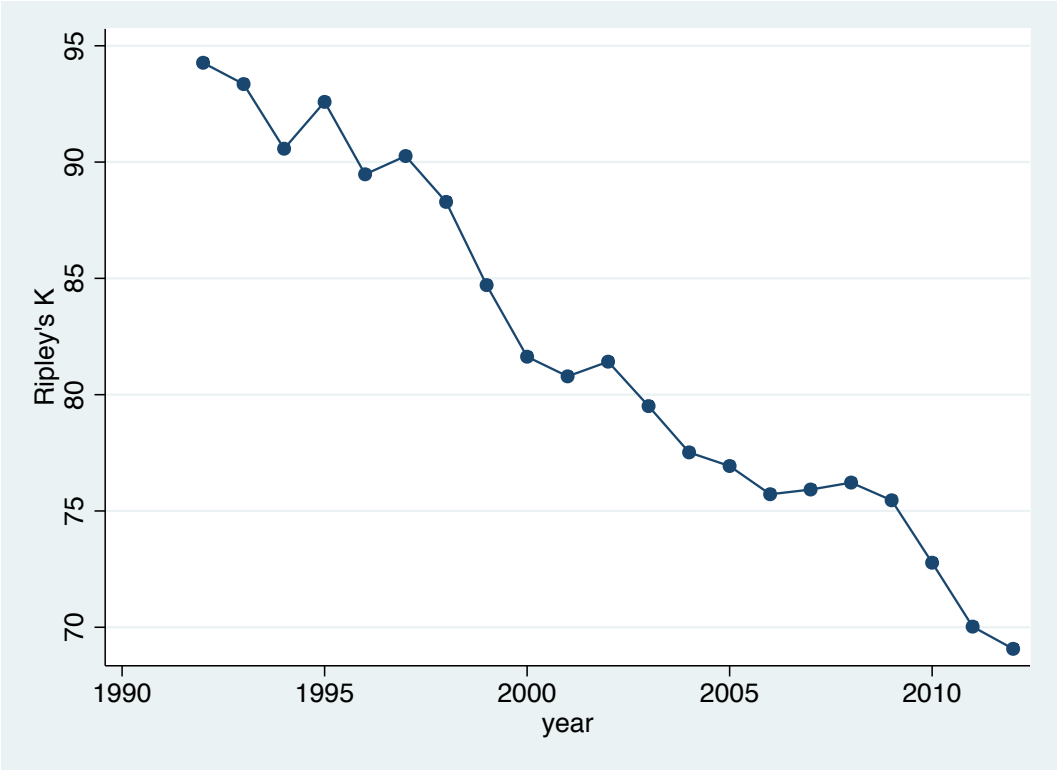


Figure 4.2: The annual agglomeration indices measured by Ripley's K from 1992 to 2012

Next, I examine the overall trend of retail agglomeration. I estimate a fixed effects model of the agglomeration indices on year dummies to show the historical change in the degree of agglomeration:

$$I_{i,j,t} = \alpha_0 + \sum_{t=1993}^{2012} \alpha_t T_t + \sum_{i,j} \alpha_{i,j} D(cbsa_i, naics_j) + \epsilon_{i,j,t}, \tag{4.1}$$

where $I_{i,j,t}$ represents the estimated agglomeration value of industry i in CBSA j in year t; T_t is the year dummy; and $D(cbsa_i, naics_j)$ is the dummy of all CBSA-NAICS6 pairs. I estimate the model using the calculated Ripley's K values, with standard errors clustered at the CBSA-NAICS6 level. The estimates of all year dummies, not reported here, are all significant except the year 2003 dummy. In so doing, I get a series of yearly indices that measure the overall degree of retail agglomeration. Figure 4.2 shows that the overall degree of retail agglomeration has kept decreasing since 1992.

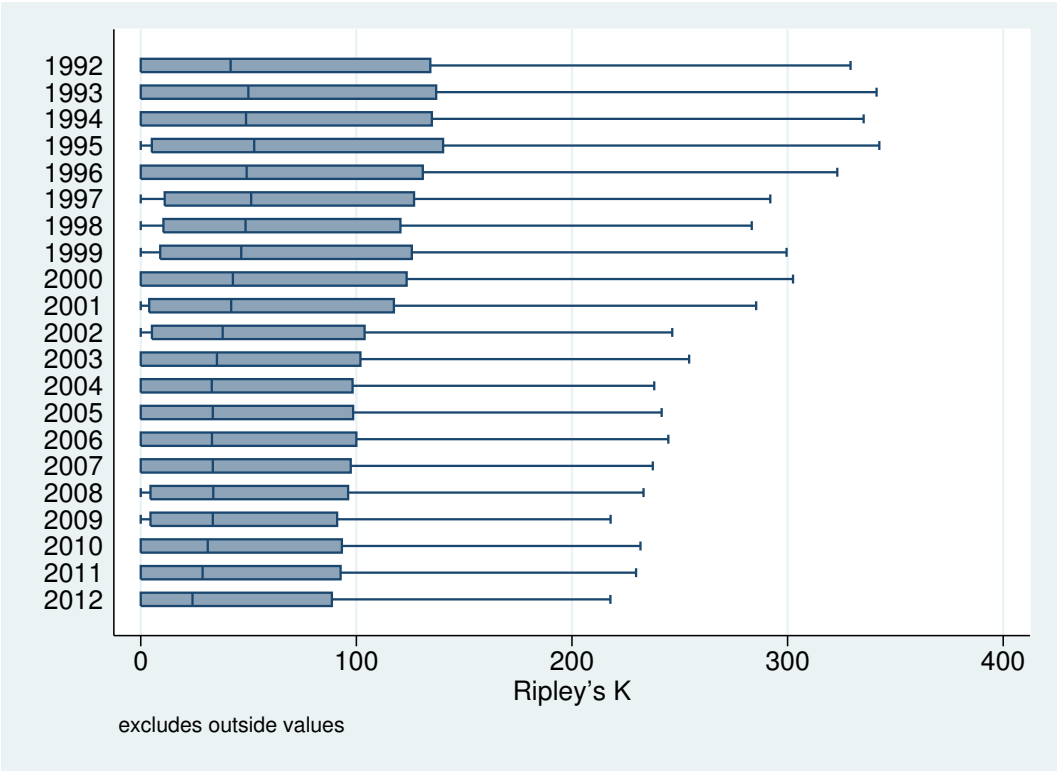


Figure 4.3: The estimated Ripley's K of book stores (451211) across CBSAs

I also investigate the degree of agglomeration by industry. As an illustration, I show the distribution of the estimated Ripley's K values across CBSAs in each year for book stores (Figure 4.3) and shoe stores (Figure 4.4). As seen in these two figures, there is much variation in the estimates values across CBSAs in each year. Also, the variation for book stores has been decreasing; however, it is the reverse for shoe stores. Moreover, there is a decreasing trend of the mean of the estimated Ripley's K since 1997 for book stores. However, for shoe stores, the mean of the estimated values is rather stable across years. This provides suggestive evidence of the contribution of consumer searching to retail agglomeration, since Internet shopping decreases the importance of consumer searching in the real world and book stores witness a greater rise in the online shopping share than shoe stores during this period.

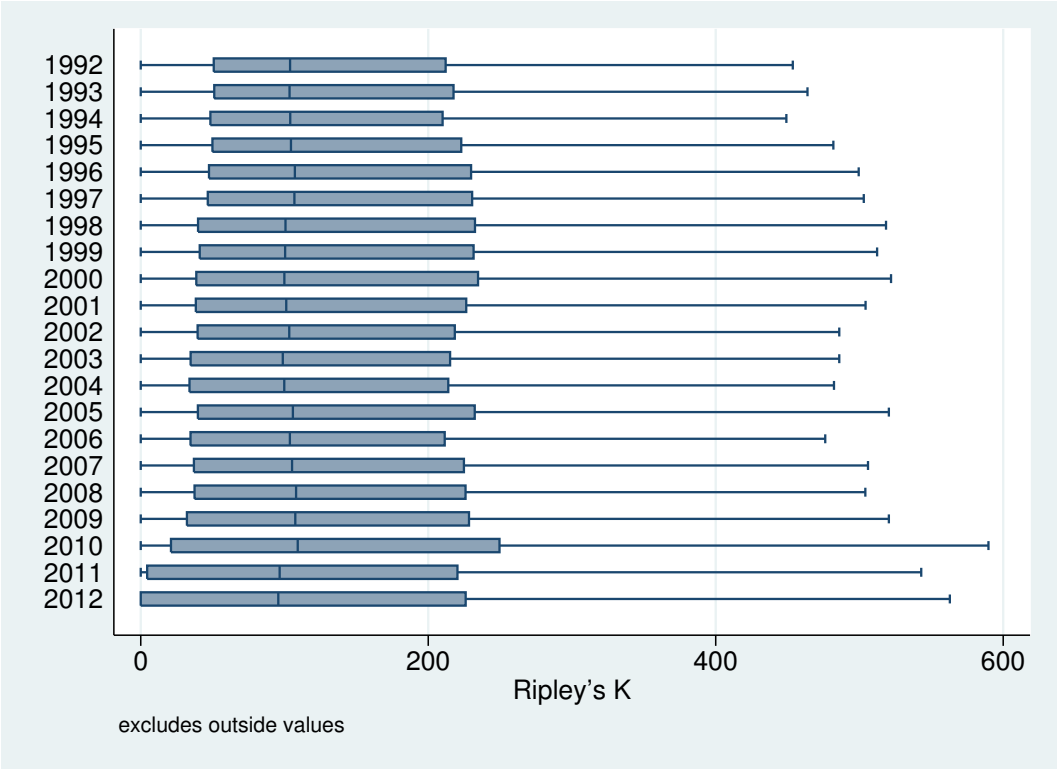


Figure 4.4: The estimated Ripley's K of Shoe stores (448210) across CBSAs

The Internet sales share of book stores increase by more than 16 percentage points from 2002 to 2007 according to the CRT data; in contrast, that of shoe stores stays the same.

The comparison of the distribution of the estimated Ripley's K values across CBSAs between radio, television, and other electronics stores (Figure B.1 in Appendix B) and men's clothing stores (Figure B.2 in Appendix B) shows the similar pattern. The Internet sales share of the former industry increases from 3.7% in 2002 to 11.9% in 2007, while that of the latter industry changes from 0.5% to 0.9%. This again shows the effect of Internet on retail agglomeration in the real world.

4.3 Ripley's K Function vs. Ellison-Glaeser index

This section presents a simple comparison of the estimated Ripley's K and the estimated Ellison-Glaeser indices. I first examine the rank correlation between these two indices using Spearman's rank correlation efficient and Kendall tau rank correlation coefficient.[8] Table 4.2

[8]A brief introduction of Spearman's coefficient and Kendall coefficient is available at `http://en.wikipedia.org/wiki/Spearman\%27s_rank_correlation_coefficient` and `http://en.wikipedia.org/wiki/Kendall_tau_rank_correlation_coefficient`, accessed on Dec 12, 2014.

presents the calculated coefficients for two randomly chosen CBSAs: San Jose-Sunnyvale-Santa Clara, California (FIPS code 41940) and Durham, North Carolina (FIPS code 20500). For the former CBSA, both Spearman's ρ and Kendall $\tau - a$ cannot reject the hypothesis that the estimated Ripley's K values and the estimated Ellison-Glaeser indices are independent. However, for the latter CBSA, both coefficients lend support to the rank correlation between these two agglomeration indices. This shows that these two agglomeration indices do not give consistent rankings of retailing industries.

Table 4.2: The rank correlation coefficients between Ripley's K function and Ellison-Glaeser index

	San Jose-Sunnyvale-Santa Clara, CA		Durham, North Carolina	
	1992	2012	1992	2012
Spearman's ρ	-0.0984	-0.0023	-0.3334***	-0.3596***
Kendall $\tau - a$	-0.0712	0.0188	-0.2144**	-0.2381***
Number of observations	70	72	65	70
Test of Hypothesis: Ripley's K function and Ellison-Glaeser index are independent.				
* = 10% significance, ** = 5% significance, *** = 1				

Table 4.3 lists the top five and bottom five industries by the degree of agglomeration, using San Jose-Sunnyvale-Santa Clara, CA as an example. The rank by Ellison-Glaeser index and by Ripley's K is given in Panel 1 and Panel 2, respectively. None of the top (bottom) five industries in Panel 1 is shown as the top (bottom) five in Panel 2. Also, two NAICS6 industries–outdoor power equipment stores, and warehouse clubs and supercenters–are listed as the top five by the calculated Ellison-Glaeser indices but as the bottom five by the calculated Ripley's K values.[9]

To further illustrate the inconsistency between these two agglomeration indices, I present two box graphs that show the distribution of the estimated Ellison-Glaeser indices for book stores (Figure B.3 in Appendix B) and shoe stores (Figure B.4 in Appendix B). Different from the decreasing trend of the concentration by Ripley's K function shown in Figure 4.3 in Section 4.2, the mean of the estimated Ellison-Glaeser across CBSAs actually has been increasing since around 2006 (Figure B.4).

It is difficult to tell which agglomeration index is more suitable for describing the retail location pattern. The inconsistency between Ripley's K function and Ellison-Glaeser index may be due to the different assumptions about the spatial unit in which retail externalities work. I use 0.5 mile as the distance threshold for Ripley's K function and zip codes as the spatial unit for Ellison-Glaeser index. However, the rank of industries by Ripley's K function in Table 4.3 seems to better conform to casual observation. It is not surprising to list luggage and leather goods stores, new car dealers, shoe stores, family clothing stores,

[9]Actually both indices have taken into account the total number of establishments. However, from Table 4.3, it seems that Ripley's K function tends to list industries with relatively fewer establishments as being less concentrated, while it is the reverse for Ellison-Glaeser index. I will investigate this issue in future.

Table 4.3: The industry rank by the degree of agglomeration: Example of one CBSA

Rank	NAICS6	NAICS6 title	No. of obs.
		Panel 1: Ellison-Glaeser index	
1	444110	Home Centers	11
2	445220	Fish and Seafood Markets	9
3	444210	Outdoor Power Equipment Stores	9
4	454111	Electronic Shopping	4
5	452910	Warehouse Clubs and Supercenters	11
69	445310	Beer, Wine, and Liquor Stores	248
70	443120	Computer and Software Stores	719
71	453220	Gift, Novelty, and Souvenir Stores	524
72	453998	All Other Miscellaneous Store Retailers (except Tobacco Stores)	1220
73	448120	Women's Clothing Stores	549
		Panel 2: Ripley's K function	
1	448320	Luggage and Leather Goods Stores	23
2	441110	New Car Dealers	125
3	448210	Shoe Stores	213
4	448140	Family Clothing Stores	149
5	448110	Men's Clothing Stores	133
67	452910	Warehouse Clubs and Supercenters	11
67	444210	Outdoor Power Equipment Stores	9
67	454319	Other Fuel Dealers	2
67	441210	Recreational Vehicle Dealers	14
67	454311	Heating Oil Dealers	4
67	444110	Home Centers	11
67	445220	Fish and Seafood Markets	9
Use San Jose-Sunnyvale-Santa Clara, CA (FIPS code 41940) as an example.			

and men's clothing stores as being highly geographically concentrated. In contrast, Ellison-Glaeser index ranks fish and seafood markets as the second highly concentrated and women's clothing stores as the least.

Also, sometimes I find negative and extreme values when calculating the Ellison-Glaeser index. This further makes it difficult to evaluate the overall level of retail agglomeration. Figure 4.5 presents the annual agglomeration level measured by Ellison-Glaeser index, obtained by estimating Model 4.1 using the calculated Ellison-Glaeser indices, with standard errors clustered at the CBSA-NAICS6 level.[10] In Figure 4.5, it is not easy to find the trend of retail agglomeration shown in Figure 4.2 due to the extreme values. Therefore, cautious should be taken when applying Ellison-Glaeser index in empirical work.

4.4 Consumer Searching: Cross-industry Evidence

In this section, I identify the contribution of consumer searching to retail agglomeration by using Internet use or Internet sales as proxies of consumer searching. An implicit assumption here is that Internet use or Internet sales has no effect on cost sharing or the effects are the same across industries. For now I do not consider the endogeneity of two Internet variables. I first use the variation in the percentage of population that use Internet across CBSAs and

[10] All the estimates of year dummies are not significant.

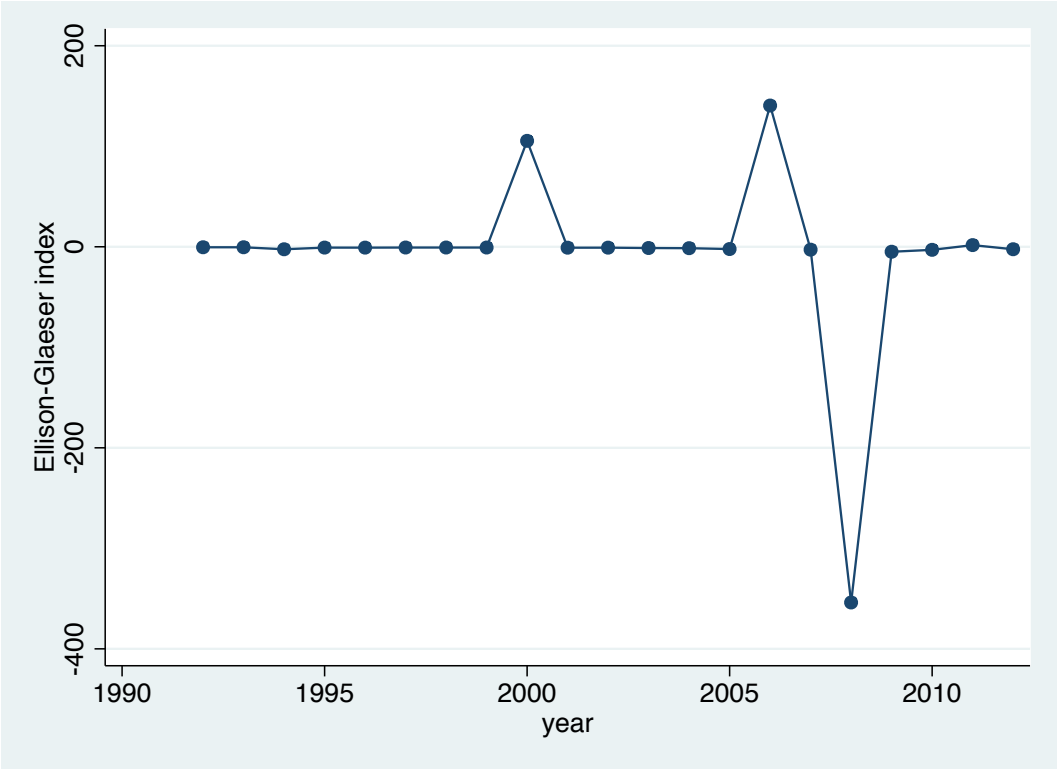


Figure 4.5: The annual agglomeration indices measured by Ellison-Glaeser index from 1992 to 2012

with time, which is calculated from the 2001, 2007, 2009, 2010, 2011, and 2012 CPS data. Internet decreases the importance of consumer searching in person (and therefore reduces the benefits of being in a retail cluster) because it is easier to obtain price and quality information online. Therefore, it is expected that the degree of retail agglomeration tends to be lower in regions with higher Internet use. I estimate the following model that regresses the agglomeration indices on the ratio of the population that use the Internet:

$$I_{i,j,t} = \alpha_0 + \alpha_1 X_{i,j,t} + \alpha_2 W_{j,t} + \epsilon_{i,j,t}, \tag{4.2}$$

where $I_{i,j,t}$ still represents the estimated agglomeration value of industry i in CBSA j in year t; $X_{i,j,t}$ is a vector of control variables including the geographic area in square miles of CBSA j in year t and the number of establishments of industry i in CBSA j in year t; and $W_{j,t}$ is the ratio of the population that use the Internet in CBSA j in year t. Here α_2 is the coefficient of interest.

Table 4.4 presents the estimation results of Model 4.2 using the calculated Ripley's K values. Equation I includes only the Internet use ratio. Equation II to VI add control variables such as the CBSA area and the number of establishments. Equation III to VII add further the NAICS6 industry fixed effects. Equation VII adds also year fixed effects to

Table 4.4: Estimation results of the model of agglomeration indices on the Internet use ratio

Dependent variable: the estimated Ripley's K							
N = 78,446	**(I)**	**(II)**	**(III)**	**(IV)**	**(V)**	**(VI)**	**(VII)**
InternetUse	-22.469***	-13.262***	-11.980***	-11.980***	-11.980***	-11.980***	-17.017**
	(3.135)	(3.031)	(2.854)	(3.030)	(4.257)	(3.250)	(7.309)
Area		0.011***	0.012***	0.012***	0.012***	0.012***	0.012***
		(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
CountEstab		-0.129***	-0.138***	-0.138***	-0.138***	-0.138***	-0.137***
		(0.003)	(0.003)	(0.007)	(0.031)	(0.013)	(0.013)
Industry fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	No	No	No	No	No	Yes
Standard errors				Robust	Clustered	Clustered	Clustered
R-squared	0.00	0.07	0.18	0.18	0.18	0.18	0.18

Industry fixed effects included in Column III to VII. Standard errors in parenthesis: robust standard errors reported in Column IV; and being clustered by NAICS6 in Column V and by CBSA-NAICS6 pair in Column VI and VII. * = 10% significance, ** = 5% significance, *** = 1% significance.

capture the effect of only the variation across CBSAs. Robust standard errors are reported in Equation IV. Standard errors are clustered by NAICS6 industry in Equation V and by CBSA-NAICS6 pair in Equation VI and VII. Though Ripley's K function take into account both the geographic area and the number of establishments, the estimates of "Area" and "CountEstab" in Column II to VI are all significant. The larger a CBSA, the more concentrated the retailing industries in this CBSA are. The degree of agglomeration decreases with the number of establishments.

The estimates of the Internet use ratio variable are significantly negative in all columns. A 10 percentage points increase in the Internet use ratio is associated with a decrease in the calculated Ripley's K value by 1.2-2.2. This provides suggestive evidence of the role consumer searching plays in promoting retail agglomeration.

Next, I explore the variation in the Internet sales share across industries and with time, which is obtained from the CRT microdata. Similarly, it is easier to do comparison shopping online when there is a higher share of Internet sales for a certain industry. In this case, this industry is expected to be less spatially concentrated. I estimate the following model of the agglomeration indices on the Internet sales share:

$$I_{i,j,t} = \alpha_0 + \alpha_1 X_{i,j,t} + \alpha_2 S_{i,t} + \epsilon_{i,j,t}, \tag{4.3}$$

where $I_{i,j,t}$ is same as above; and $S_{i,t}$, the variable of interest, is the Internet sales share of industry i in year t.

Table 4.5 presents the estimation results using the calculated Ripley's K values. Similarly, Equation I includes only the Internet sales share. Equation II to VI add also control variables. Equation III to VI add further the industry fixed effects. Robust standard errors are reported in Equation IV. Standard errors are clustered by CBSA in Equation V and by CBSA-NAICS6 pair in Equation VI. The estimates of the number variable are still significantly negative and lose significance in only Column IV; however, the estimates of the area variable is now significant in only Column I. The estimates of the Internet sales share are significantly negative in all columns. A 10 percentage points increase in the Internet sales share is

Table 4.5: Estimation results of the model of agglomeration indices on the Internet sales share

Dependent variable: the estimated Ripley's K						
N = 289,229	**(I)**	**(II)**	**(III)**	**(IV)**	**(V)**	**(VI)**
InternetSales	-140.244***	-85.011***	-145.246***	-145.246***	-145.246***	-145.246***
	(16.755)	(15.615)	(14.544)	(0.000)	(14.816)	(14.951)
Area		0.041***	0.000	0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(2.67E+9)	(5.15E+17)
CountEstab		-0.557***	-0.028***	-0.028	-0.028***	-0.028***
		(0.007)	(0.009)	(0.000)	(0.007)	(0.005)
CBSA fixed effects	No	No	Yes	Yes	Yes	Yes
Standard errors				Robust	Clustered	Clustered
R-squared	0.00	0.13	0.25	0.25	0.25	0.25

CBSA fixed effects included in Column III to VI. Standard errors in parenthesis: robust standard errors reported in Column IV; and being clustered by CBSA in Column V and by CBSA-NAICS6 pair in Column VI. * = 10% significance, ** = 5% significance, *** = 1% significance.

associated with a 8.5-14.5 decrease in the calculated Ripley's K value. This again suggests that Internet decreases the importance of consumer searching and thus lowers the degree of retail agglomeration.

I then use both the variation in Internet use across regions and the variation in Internet sales across industries by estimating the following model:

$$I_{i,j,t} = \alpha_0 + \alpha_1 X_{i,j,t} + \alpha_2 W_{j,t} + \alpha_3 S_{i,t} + \epsilon_{i,j,t}, \tag{4.4}$$

where α_2 and α_3 are the coefficients of interest.

Table 4.6: Estimation results of the model of agglomeration indices on both the Internet use ratio and the Internet sales share

Dependent variable: the estimated Ripley's K						
N = 16,472	**(I)**	**(II)**	**(III)**	**(IV)**	**(V)**	**(VI)**
internetUse	-35.417***	-30.634***	-30.634***	-30.634	-30.634***	-30.634***
	(9.743)	(9.498)	(10.640)	(33.478)	(8.225)	(10.640)
InternetSales	-108.137***	-94.562***	-94.562***	-94.562***	-94.562	-94.562***
	(18.886)	(18.398)	(13.501)	(12.505)	(61.732)	(13.501)
Area		0.010***	0.010***	0.010***	0.010***	0.010***
		(0.000)	(0.001)	(0.003)	(0.001)	(0.001)
CountEstab		-0.126***	-0.126***	-0.126***	-0.126***	-0.126***
		(0.007)	(0.011)	(0.024)	(0.025)	(0.011)
Standard errors			Robust	Clustered	Clustered	Clustered
R-squared	0.00	0.06	0.06	0.06	0.06	0.06

Standard errors in parenthesis: robust standard errors reported in Column III; and being clustered by CBSA in Column IV, by NAICS6 in Column V, and by CBSA-NAICS6 pair in Column VI. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table 4.6 presents the estimation results using the calculated Ripley's K values. Here I actually use only the calculated Ripley's K in 2007 for estimation– the only overlap of

two types of Internet data. I add no fixed effects to explore these two types of variations. The estimates of both variables of interest are significantly negative in all columns. A 10 percentage points increase in the Internet use ratio and in the Internet sales share is associated with a 3.1-3.5 decrease and a 9.5-10.8 decrease in the calculated Ripley's K value.

As a robustness check, I estimate Model 4.3 using only the 2002 and 2007 data since the Internet sales share data before 1995 are assumed as zero, and get qualitative the same results. I also estimate Model 4.2, Model 4.3, and Model 4.4 using the calculated Ellison-Glaeser indices. However, none of the estimates, not reported here, are significant.

To partly address the endogeneity issue, I also employ a difference-in-differences (DID) approach. I first classify retailing industries into two types–industries easy and difficult to make comparison shopping online–using three different classifications: i) industries with the Internet sales share greater (less) than 2% in 2007 as being easy (difficult) to make comparison shopping online; ii) industries with the Internet sales share greater (less) than 5% in 2007 as being easy (difficult) to make comparison shopping online; and iii) industries with the Internet sales share greater than 5.6% (less than 1%) as being easy (difficult) to make comparison shopping online. Then I examine whether the former industries experience a greater decrease in the degree of agglomeration before and after the wide introduction of Internet shopping around 2000 by estimating the following model:

$$I_{i,j,t} = \alpha_0 + \alpha_1 E_i + \alpha_2 Y_t + \alpha_3 E_i Y_t + \epsilon_{i,j,t}, \tag{4.5}$$

where E_i is a binary variable indicating whether an industry is easy to make comparison shopping online; Y_t indicates whether it is after year 2000 or not. α_3 are the coefficients of interest.

Table 4.7: Estimation results of the DID model of agglomeration indices

Dependent variable: the estimated Ripley's K			
N = 1,227,770	**(I)**	**(II)**	**(III)**
EasySearch	-2.997***	5.175***	-17.365***
	(0.683)	(0.764)	(0.934)
After2000	-111.280***	-11.864***	-10.211***
	(0.609)	(0.496)	(0.827)
EasySearch*After2000	-6.808***	-10.266***	-12.615***
	(0.839)	(0.924)	(1.144)
CBSA fixed effects	Yes	Yes	Yes
Standard errors	Robust	Robust	Robust
R-squared	0.24	0.24	0.24
Robust standard errors in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.			

Table 4.7 presents the estimation results using the calculated Ripley's K values. Equations I to III use the three different industry classifications, respectively. All equations include CBSA fixed effects. The estimates of the interactive term are all significantly negative; industries easy to make comparison shopping online experience a greater decrease in the

degree of agglomeration by 6.8-12.6. This confirms the positive effect of consumer searching on retail agglomeration.

4.5 Summary

The degree of retail agglomeration has been decreasing since 1992. The estimated Ellison-Glaeser indices show that the ratio of industries that are highly concentrated at the NAICS6-CBSA level decreases from 12.1% in 1992 to 9.9% in 2012, while that of industries that are not concentrated increases from 86.6% to 89.1% during the same period. The estimated Ripley's K values also show that the overall degree of retail agglomeration at the CBSA level has kept decreasing since 1992.

Ellison-Glaeser index has been widely used in empirical studies. For example, Baldwin, Brown, and Rigby 2010 and Ellison, Glaeser, and Kerr 2010 investigate the sources of agglomeration by regressing Ellison-Glaeser index at the industry level on measures for various micro-mechanisms. However, the simple comparison between Ripley's K function and Ellison-Glaeser index indicates that there is much inconsistency between them. Therefore, there may be systematic bias in these empirical studies.

Using the variation in the percentage of population that use Internet across regions and in the Internet sales share across retailing industries, I find that a 10 percentage points increases in the Internet use ratio and in the Internet sales share is associated with a 3.1-3.5 decrease and a 9.5-10.8 decrease in the calculated Ripley's K value. This provides suggestive evidence of the role consumer searching plays in promoting retail agglomeration since Internet decreases the importance of consumer searching in person and thus reduces the benefits of being in a retail cluster. To partially address the endogeneity issue, I use a DID approach and find that industries easy to make comparison shopping online experience a greater decrease in the degree of agglomeration by 6.8-12.6.

Chapter 5

Consumer Searching vs. Cost Sharing: Evidence from the Bankruptcy of Chain Stores

To assess the relative importance of the consumer searching mechanism and the cost sharing mechanism for retail agglomeration, one approach is to explain the cross-industry variation in the degree of agglomeration using measures that represent these two mechanisms. For example, Rosenthal and Strange 2001, Baldwin, Brown, and Rigby 2010, and Ellison, Glaeser, and Kerr 2010 use this approach to investigate the determinants of manufacturer agglomeration. However, this approach comes with two deficiencies. First, the proxy measures of micro-mechanisms can often be interpreted as a different story. Take labor market pooling as an example. If the degree of agglomeration is found to be associated with the percentage of educated workers, it is likely that both manufacturers and workers are simply attracted by good (producer and consumer) services in a certain place. Second, there are inverse causalities related to these proxy measures. For example, it is possible that a pair of firms form the input-output linkage because they are located next to each other.

Here I use an exogenous shock–the closure of chain stores like Borders and Circuit City–to distinguish the effects of these two mechanisms. The consumer searching mechanism and the cost saving mechanism actually are two types of positive externalities. The closure of a store means the loss of these externalities for nearby stores. Also, which type of externalities loss nearby stores suffer from depends upon industry. For example, if a Borders bookstore is closed, for nearby bookstores, this means the losses of both types of externalities–consumer searching costs will go up, and the ability to share the advertisement costs is reduced. For nearby grocery stores, however, this means the loss of the externality in only advertisement cost saving since there is some overlap between the consumers of grocery stores and those of bookstores. In this sense, outside its own industry, a store closure eliminates only the positive externality of the advertisement cost sharing mechanism. Furthermore, the difference between the effect outside its own industry and that within its own industry captures the externality of the consumer searching mechanism, assuming that the effects from the cost

sharing mechanism are fixed across industries, after controlling for factors such as walking distance between the bankrupt store and the nearby store.

The advantage of using the bankruptcy of national chain stores is to exclude the effects of local economic conditions. The estimated effect of store closure on nearby stores in terms of sales and closure probability may be due to the weak local economy rather than the externalities between pairs of stores. Using the bankruptcy of chain stores can help address such an endogeneity and thus capture the effect from the cost sharing mechanism more accurately.

I first use a difference-in-differences (hereafter DID) approach to capture the effect of the consumer searching mechanism. Then, using chain stores such as Barnes & Noble and Best Buy that do not experience bankruptcy as controls, I use a triple-difference (hereafter Triple-Diff) approach to further distinguish the effect of the cost sharing mechanism. I use sales, employment size, the probability of closure, and the birth of new stores, all of which are given in the 2012 National Establishment Time-Series (NETS) database, to measure the loss of positive externalities, respectively.[1]

Two recent papers that look at the spillovers of retailers by investigating bankrupt firms, however, without discussing the underlying mechanisms of such spillovers. Shoag and Veuger 2014 find that the external effects decay with distance using the ESRI Business Analyst annual data from 2006 to 2013. They also show that compact cities are more likely to provide retail subsidies since they can capture more of the positive externalities from large stores. Benmelech et al. 2014 show that stores closer to bankrupt stores are more likely to close themselves using the Chain Store Guide database that provides annual snapshots between 2005 and 2010. Moreover, they find that the effects on store closure is larger for stores in the same industry as the liquidating national chain than for stores in different industries. They, however, do not present the possible explanations for this finding.

Strictly speaking, here I actually distinguish the effect of cost sharing and the integrated effect of both consumer searching and price competition. When a store locates next to its rivals, it faces a trade-off between the benefits from consumer searching and the pressure from price competition. To further capture the effect of only consumer searching, it is necessary to find other types of shock that work on only consumer searching or only price competition. Another way is to assume different functional forms for the consumer searching mechanism and the price competition mechanism by using store-level data (e.g., Datta and Sudhir 2011).

This section is structured as follows. Section 5.1 introduces the data. Section 5.2 presents the model specifications for estimation. Section 5.3, Section 5.4, Section 5.5, and Section 5.6 present the estimation results of the effects on sales, the employment size, the probability of closure, and the birth of new stores, respectively. Section 5.7 summarizes the main findings.

[1]Both Shoag and Veuger 2014 and Benmelech et al. 2014 use the closure probability of nearby firms when estimating the negative externalities of bankrupt firms. Rosenthal and Strange 2003 use the birth of new establishments when investigating the geographic scope and the organization nature of agglomeration externalities, treating the local economic environment as exogenous.

5.1 Data

There have been a dozen Chapter 11 bankruptcies of national chain stores recently, such as Borders and Circuit City.[2] Benmelech et al. 2014 give a list of major Chapter 11 bankruptcies used in their analysis (Table 2 in their paper). Shoag and Veuger 2014 use the bankruptcies of national chains including CompUSA, Mervyn's, Circuit City, Linens 'n Things, Borders, with Best Buy, Kohl's, Bed Bath and Beyond, and Barnes & Noble as controls.

In this dissertation, I focus on the bankruptcies of three types of chain stores. The first type is department stores like Mervyn's, liquidated in 2008, and Goody's, liquidated in 2009, with Kohl's as control. The second is specialty stores like Sharper Image, liquidated in 2008, with Brookstone as control. The third is toy stores such as KB Toys, liquidated in 2009, with Toys R Us as control. For stores that belong to chains as control, they are included in the analysis only if they are still in operation in 2012.

I first identify stores that belong to these bankrupt chains in the NETS database. This cannot be done based on a single variable in the database. First, the naming of stores in the NETS database is not very accurate. For example, a Borders store may be named as "Borders Berkeley", and a Barnes & Noble store may simply be recorded as "B&N". Second, the trade name is missing for many stores in the database. Third, the headquarter DUNS number is helpful but sometimes a headquarter may have more than one chain. For instance, Gap and Banana Republic correspond to the same headquarter.[3] Therefore, I use a combination of company name, trade name, and headquarter DUNS number included in the NETS database to identify stores of bankrupt chains and chains used as control. Appendix C introduces how these chain stores are identified.

Given my research purpose, here the priority is to avoid including stores that do not belong to the chains I choose, rather than to find all stores that belong to a specific chain. For bankruptcy chains, I include only stores that were closed in the year that a chain filed for bankruptcy. Moreover, for both treatment and control chains, any store that once moved is excluded from the estimation. Table 5.1 gives the number of stores for each national chain used in the estimation.

I then choose all stores that are within one mile of these treatment and control chain stores because it is unlikely that shopping externalities go beyond 20-minute walking distance. However, a nearby store is dropped out if it moved between 1990 and 2012. To ease the estimation, I split the data by type of the national chains. The final dataset used in the estimation on the sales, the employment size, and the closure probability is a panel dataset for these nearby stores. For both chain stores and their nearby stores, the dataset records their Standard Industrial Classification (SIC) code, sales, employment size, and being in operation or not. Nonetheless, when I use the birth of new stores to measure the loss of positive externalities, the dataset used for estimation is a panel dataset for the chain stores.

[2]See `http://www.uscourts.gov/services-forms/bankruptcy/bankruptcy-basics/chapter-11-bankruptcy-basics` for an introduction of Chapter 11.

[3]Franchised stores that belong to the same train tend to have many headquarter firms and thus DUNS numbers, though such stores are not likely to be included in our analysis of Chapter 11 bankruptcies.

Table 5.1: The number of stores for each national chain used in the estimation

Type	Chain	Treatment/Control	Number of stores	Year closed
Department store	Mervyn's	Treatment	64	2008
	Goody's	Treatment	74	2009
	Kohl's	Control	1,138	–
Specialty	Sharper Image	Treatment	21	2008
	Brookstone	Control	284	–
Toys	KB Toys	Treatment	570	2009
	Toys R Us	Control	905	–
The number of stores is calculated as those chosen from the 2012 NETS database.				

In this case, I calculate the number of new stores within various distances of each chain store in each year on the basis of the starting year of nearby stores.

The advantages of the 2012 NETS database used in this dissertation lie in its coverage of the entire retail industries, its long time span of 1990 to 2012, and its detailed industrial classification. In contrast, Benmelech et al. 2014 use data that cover only the period between 2005 and 2010.

5.2 Model specification

The closure of a chain store means the losses in positive externalities from both the consumer searching mechanism and the cost sharing mechanism for nearby stores in the same industry as this chain store; however, it means the externalities loss from only the cost sharing mechanism for nearby stores in industries different from this chain store. So, I first choose only stores around bankrupt chain stores and use a difference-in-differences approach to distinguish these two mechanisms:

$$Y_{i,t} = \alpha_0 + \alpha_1 T_t + \alpha_2 I_i + \alpha_3 (T_t I_i) + \alpha_4 X_i + \epsilon_{i,t}, \tag{5.1}$$

where $Y_{i,t}$ is the status of store i in year t–sales, employment size, or being in operation or not in different specifications; T_t indicates whether the national chain is closed in year t; I_i indicates whether store i is in the same industry as the "center" chain store; X_i is a vector of control variables for store i; and $\epsilon_{i,t}$ is the error term. Here α_3, the coefficient of the interactive term of T_t and I_i, captures the effect of the consumer searching mechanism.

The implicit assumption here is that the external benefits through the cost sharing mechanism are the same across industries. As an improvement, I assume that the external benefits associated with cost sharing depends upon the walking distance between store i and the closed "center" chain store. Here the walking distance d_i is measured as the air distance between these two stores.

When the focus is on the birth of new stores, I estimate the following difference-in-differences equation:

$$Y_{j,t} = \alpha_0 + \alpha_1 T_t + \alpha_2 I_j + \alpha_3 (T_{j,t} I_j) + \alpha_4 X_j + \epsilon_{j,t}, \tag{5.2}$$

where $Y_{j,t}$ represents the number of new stores around chain store j in year t; I_j indicates whether the number of new stores is calculated for stores in the same industry as chain store j; and X_j is a vector of control variables for chain store j. Similarly, α_3 is the coefficient of interest.

Then, using chain stores like Barnes & Noble and Best Buy as controls, I use a triple-difference approach to distinguish these two types of externalities on the basis of the following regression:

$$Y_{i,t} = \alpha_0 + \alpha_1 C_i + \alpha_2 T_t + \alpha_3 I_i + \alpha_4 (C_i I_i) + \alpha_5 (T_t I_i) + \alpha_6 (C_i T_t) + \alpha_7 (C_i T_t I_i) + \alpha_8 X_i + \epsilon_{i,t}, \quad (5.3)$$

where C_i indicates whether the chain store around which store i is located is liquidated between 1990 and 2012 (a treatment chain store) or not; T_t represents whether the treatment chain has filed for bankruptcy in year t; and I_i indicates whether store i is in the same industry as the "center" chain store. Here the coefficients of interests are α_6 that captures the effect of the cost sharing mechanism and α_7 that captures the effect of the consumer searching mechanism.

Similarly, when it turns to the birth of new stores, I estimate the following triple-difference equation:

$$Y_{j,t} = \alpha_0 + \alpha_1 C_j + \alpha_2 T_t + \alpha_3 I_j + \alpha_4 (C_j I_j) + \alpha_5 (T_t I_j) + \alpha_6 (C_j T_t) + \alpha_7 (C_j T_t I_j) + \alpha_8 X_j + \epsilon_{j,t}, \quad (5.4)$$

where C_j indicates whether the chain store j is a treatment chain store or not; and I_j indicates whether the number of new stores is calculated for stores in the same industry as chain store j.

In the following estimations I will use various distances to the chain stores. Shoag and Veuger 2014 use 0.5, 1, and 2 miles radii, while Benmelech et al. 2014 use the same address (e.g., a shopping mall), 50 meters, and 100 meters radii.

5.3 Estimation results: sales

In this section I focus on the effects of store bankruptcy on the annual sales of nearby stores. I first estimate a DID model for each bankrupt chain to compare the effect on nearby stores in the same industry as the bankrupt chain store and the effect on nearby stores in industries different from that chain store. So for department stores, I split the dataset into stores nearby Mervyn's and those nearby Goody's and do the estimation respectively.

I start by looking at stores that are within 1 mile of bankrupt chain stores. Table 5.2, 5.3, 5.4, and 5.5 present the estimation results of the DID model for specialty stores, department stores (only Mervyn's), department stores (only Goody's), and toy stores. In all equations, the dependent variable is the logarithm of the annual sales of each nearby store; and standard errors are clustered by the bankrupt chain store.

Equations I to IV include three variable. The variable "After" captures the general time trend; the variable "IndSame" defines whether a nearby store is in the same industry as

Table 5.2: Estimation results of the DID model on sales: Specialty store

Dependent variable: log(annual sales)								
N = 584,192	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.408***	-0.409***	-0.409***	-0.409***	-0.063	-0.058	-0.058	-0.058
	(0.027)	(0.027)	(0.027)	(0.027)	(0.066)	(0.066)	(0.066)	(0.066)
IndSame	0.588***	0.751***	0.676***	0.552***	0.588***	0.751***	0.676***	0.552***
	(0.058)	(0.101)	(0.161)	(0.122)	(0.058)	(0.101)	(0.161)	(0.122)
After*IndSame	0.146***	0.263*	0.531*	0.589	0.092*	0.218	0.443*	0.541
	(0.049)	(0.133)	(0.260)	(0.490)	(0.050)	(0.133)	(0.254)	(0.500)
After*Distance					-0.596***	-0.606***	-0.607***	-0.608***
					(0.090)	(0.091)	(0.091)	(0.091)
R-squared	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.02

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

the bankrupt chain store and captures the difference in the sale between these two types of nearby stores; and the interactive term of "After" and "IndSame"–the variable of interest– identifies the effect of the consumer searching mechanism. Whether a store is in the same industry as the bankrupt chain store, is defined at the 2-digit SIC industry in Equations I and V, the 4-digit SIC industry level in Equations II and VI, the 6-digit level in Equations III and VII, and the 8-digit level in Equations IV and VIII.

Table 5.3: Estimation results of the DID model on sales: Department store (only Mervyn's)

Dependent variable: log(annual sales)								
N = 1,214,344	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.384***	-0.383***	-0.386***	-0.386***	-0.022	-0.024	-0.011	-0.010
	(0.014)	(0.014)	(0.015)	(0.015)	(0.037)	(0.037)	(0.040)	(0.040)
IndSame	2.306***	3.809***	3.562***	3.493***	2.306***	3.809***	3.562***	3.493***
	(0.320)	(0.088)	(0.128)	(0.124)	(0.320)	(0.088)	(0.128)	(0.124)
After*IndSame	-0.003	-0.051	-0.349*	-0.448**	-0.108	-0.202*	-0.525***	-0.634***
	(0.106)	(0.115)	(0.192)	(0.200)	(0.103)	(0.114)	(0.195)	(0.200)
After*Distance					-0.612***	-0.608***	-0.637***	-0.638***
					(0.052)	(0.052)	(0.056)	(0.056)
R-squared	0.05	0.06	0.04	0.03	0.05	0.07	0.04	0.04

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Equations V to VIII add also the interactive term of "After" and "Distance". Here the variable "Distance" is the air distance between the bankrupt chain store and the nearby store, which is a good proxy of walking distance given that I look at only stores within 0.5 or 1 mile (i.e., 10-minute or 20-minute walking distance) to the bankrupt store. The effect from the cost sharing mechanism may depend on the walking distance between the bankrupt store and the nearby store. Also, within the 0.5 or 1 mile radius, stores in the same industry as the bankrupt chain store may generally locate closer to that chain store. This will cause

the effect from the cost sharing mechanism larger for these stores compared to stores in industries different from the chain store. Therefore, I address such a potential endogeneity due to the spatial concentration of same-industry stores by introducing the distance variable.

Table 5.4: Estimation results of the DID model on sales: Department store (only Goody's)

Dependent variable: log(annual sales)								
N = 1,273,744	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.322***	-0.310***	-0.309***	-0.309***	0.031	0.056**	0.056**	0.056**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.025)	(0.026)	(0.026)	(0.026)
IndSame	0.179***	0.912***	1.085***	1.085***	0.179***	0.912***	1.085***	1.085***
	(0.034)	(0.092)	(0.096)	(0.096)	(0.034)	(0.092)	(0.096)	(0.096)
After*IndSame	0.482***	0.242	0.192	0.192	0.391***	0.136	0.080	0.080
	(0.027)	(0.147)	(0.179)	(0.179)	(0.027)	(0.145)	(0.176)	(0.176)
After*Distance					-0.592***	-0.618***	-0.618***	-0.618***
					(0.041)	(0.042)	(0.042)	(0.042)
R-squared	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

According to the estimates of Columns I to IV in all four tables (Table 5.2, 5.3, 5.4, and 5.5), for nearby stores in industries different from the bankrupt chain store, their sales decrease significantly 30%-40% after the bankruptcy of the chain store. However, this cannot be attributed to only the externality loss from the cost sharing mechanism since there is no control group in the estimated DID model to account for the general time trend. The estimates of the interactive term of "After" and "IndSame" are positive in Columns I to IV in all these tables except Table 5.3. So nearby stores in the same industry as the bankrupt chain store experience less decrease in annual sales. This indicates that the benefits from less price competition exceed the externality loss from the consumer searching mechanism. Therefore, using only consumer searching cannot explain the widely observed spatial concentration of retail stores that belong to the same industry.

The estimated coefficients of the interactive term of "After" and "IndSame" increase in Table 5.2 and Table 5.5 and decrease in Table 5.3 and Table 5.4, when I change the definition of being in the same industry by from SIC2, SIC4, SIC6, to SIC8. The narrower is the definition, the more fierce price competition and the stronger consumer searching externality there is. So the pattern of the estimates depends upon the relative change of these two forces when changing the definition. For specialty stores and toy stores, the benefits from less price competition may increase more than the loss from consumer searching externalities in absolute values when narrowing the definition of being in the same industry; while it is the reverse for department stores.

When the interactive term of "After" and "Distance" is added, the estimates of "After*IndSame" decreases in Columns V to VIIII of all four tables (Table 5.2 to Table 5.5). This indicates that the location pattern of stores within the study area is not totally random. However, the coefficients of the "After*Distance" variable are all negative, which does

Table 5.5: Estimation results of the DID model on sales: Toy store

Dependent variable: log(annual sales)								
N = 10,499,198	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.406***	-0.395***	-0.395***	-0.395***	-0.113***	-0.104***	-0.104***	-0.104***
	(0.024)	(0.023)	(0.023)	(0.023)	(0.037)	(0.036)	(0.036)	(0.036)
IndSame	-0.119***	-0.468***	-0.589***	-0.639***	-0.119***	-0.468***	-0.589***	-0.639***
	(0.046)	(0.046)	(0.042)	(0.044)	(0.046)	(0.046)	(0.042)	(0.044)
After*IndSame	0.221***	0.799***	1.205***	1.566***	0.166***	0.720***	1.089***	1.443***
	(0.012)	(0.055)	(0.062)	(0.087)	(0.019)	(0.053)	(0.059)	(0.081)
After*Distance					-0.512***	-0.515***	-0.515***	-0.514***
					(0.028)	(0.029)	(0.030)	(0.029)
R-squared	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table 5.6: Estimation results of the Triple-Diff model on sales: Specialty store

Dependent variable: log(annual sales)				
N = 110,004	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8
Treat	-0.066	-0.041	-0.039	-0.038
	(0.156)	(0.149)	(0.149)	(0.148)
After	-0.494***	-0.470***	-0.469***	-0.469***
	(0.080)	(0.076)	(0.076)	(0.076)
IndSame	-0.354***	-0.487***	-0.429**	-0.463**
	(0.091)	(0.137)	(0.174)	(0.191)
Treat*IndSame	0.499**	1.049***	1.231***	1.939***
	(0.243)	(0.372)	(0.469)	(0.223)
After*IndSame	0.354***	0.347***	0.466***	0.529***
	(0.088)	(0.130)	(0.152)	(0.173)
Treat*After	0.164	0.141	0.139	0.138
	(0.105)	(0.102)	(0.102)	(0.102)
Treat*After*IndSame	-0.370*	-1.360***	-1.719***	-2.501***
	(0.196)	(0.417)	(0.522)	(0.185)
R-squared	0.02	0.02	0.02	0.02

All stores that are within 0.5 mile of treatment and control chain stores and have self-reported sales in the 2012 NETS database included. Sales between 2004 and 2012 only. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I), the SIC4 (II), the SIC6 (III), and the SIC8 (IV) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

not conform to our expectations. Later I will add the "consumer composition distance" between pairs of stores in terms of their industries since these two types of distances may be correlated.

I do a series of robustness checks. First, I use the absolute value of annual sales as

Table 5.7: Estimation results of the Triple-Diff model on sales: Department store (Mervyn's)

Dependent variable: log(annual sales)				
N = 139,214	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8
Treat	-0.239	-0.238	-0.237	-0.237
	(0.168)	(0.168)	(0.168)	(0.168)
After	-0.314***	-0.314***	-0.314***	-0.314***
	(0.037)	(0.037)	(0.037)	(0.037)
IndSame	-0.181	0.002	-0.394	0.000
	(0.193)	(0.404)	(0.669)	(0.000)
Treat*IndSame	0.562	1.735***	0.000	0.000
	(0.482)	(0.409)	(0.000)	(0.000)
After*IndSame	0.102	0.366**	0.213	0.000
	(0.116)	(0.148)	(0.443)	(0.000)
Treat*After	0.101*	0.103*	0.102*	0.102*
	(0.056)	(0.055)	(0.055)	(0.055)
Treat*After*IndSame	0.146	-0.424***	0.000	0.000
	(0.265)	(0.154)	(0.000)	(0.000)
R-squared	0.01	0.01	0.01	0.01

All stores that are within 0.5 mile of treatment and control chain stores and have self-reported sales in the 2012 NETS database included. Sales between 2004 and 2012 only. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I), the SIC4 (II), the SIC6 (III), and the SIC8 (IV) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

the dependent variable and get qualitative the same results (not reported here). Second, I drop out the observations with sales estimated by Dun and Bradstreet and use only self-reported sales records. This greatly decreases the sample size from 584,192 to 39,595 for specialty stores, 1,214,344 to 79,316 for departments stores (Mervyn's), 1,273744 to 105,191 for department stores (Goody's), and 10,499,198 to 882,595 for toy stores. The coefficients of the "After*Distance" variable, reported in Table D.1, Table D.2, Table D.3, and Table D.4, are still positive in most of the columns, however, with smaller magnitudes and significance lost in some of them. Third, I use only stores within 0.5 mile of the bankrupt chain store. As seen in Table D.5, Table D.6, Table D.7, and Table D.8, this does not change the estimation results much except that the coefficients of interest are negative for specialty stores, which indicates a stronger effect of consumer searching. Fourth, I further use only self-reported sales from 2004 to 2012 to obtain an approximately symmetric sample before and after the bankruptcy of chain stores. Table D.9, Table D.10, Table D.11, and Table Table D.12 present qualitative the same results.

I then estimate a triple-difference model for each bankrupt chain to capture both the effect of consumer searching and that of cost sharing by using non-bankrupt chain stores as controls. Similarly, I estimate the model in turn using four groups of observations: stores that are within 1 mile of treatment and control chain stores, stores that are within 1 mile

Table 5.8: Estimation results of the Triple-Diff model on sales: Department store (Goody's)

Dependent variable: log(annual sales)				
N = 140,503	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8
Treat	-0.217	-0.236	-0.237	-0.237
	(0.162)	(0.162)	(0.161)	(0.161)
After	-0.315***	-0.315***	-0.315***	-0.315***
	(0.044)	(0.044)	(0.044)	(0.044)
IndSame	-0.155	0.024	-0.324	-0.189
	(0.197)	(0.409)	(0.664)	(0.329)
Treat*IndSame	-0.325	-0.772	0.136	0.000
	(0.224)	(0.524)	(0.741)	(0.000)
After*IndSame	0.05	0.368**	0.034	-0.483
	(0.128)	(0.177)	(0.600)	(0.683)
Treat*After	0.063	0.077	0.078	0.078
	(0.054)	(0.054)	(0.054)	(0.054)
Treat*After*IndSame	0.245	-0.442	-0.518	0.000
	(0.166)	(0.729)	(0.909)	(0.000)
R-squared	0.01	0.01	0.01	0.01

All stores that are within 0.5 mile of treatment and control chain stores and have self-reported sales in the 2012 NETS database included. Sales between 2004 and 2012 only. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I), the SIC4 (II), the SIC6 (III), and the SIC8 (IV) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

of both types of chain stores and have self-reported sales, stores that are within 0.5 mile of chain stores and have self-reported sales, and only sales during 2004-2012 for stores that are within 0.5 mile and have self-reported sales.

Here I present the estimation results using the last group of observations (Table 5.6, Table 5.7, Table 5.8, and Table 5.9). The variables of interest here are the interactive terms "Treat*After" and "Treat*After*IndSame". The estimates of the "Treat*After" variable, which capture the effect of cost sharing, are expected to be negative since the bankruptcy of chain stores causes the loss in the positive externality. However, they are positive and in one case even significantly positive in the estimations for specialty stores and department stores. The estimates are negative, however, insignificant, in only Table 5.9. This suggests that the cost sharing mechanism may be not so important for retail agglomeration. The estimates of the "Treat*After*IndSame" are similar to those in the DID models. They are significantly negative in only the estimations for specialty stores. This provides suggestive evidence of the contribution of consumer searching to the spatial concentration of specialty stores.[4] I also control the distance between nearby store and the chain store and obtain qualitatively the same results.

[4]In the next step I will try techniques such as matching to improve the estimates.

Table 5.9: Estimation results of the Triple-Diff model on sales: Toy store

Dependent variable: log(annual sales)				
N = 270,718	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8
Treat	0.126	0.103	0.103	0.103
	(0.130)	(0.127)	(0.127)	(0.127)
After	-0.275***	-0.259***	-0.258***	-0.259***
	(0.014)	(0.014)	(0.013)	(0.013)
IndSame	-0.180***	-0.159	-0.201	-0.509*
	(0.036)	(0.146)	(0.270)	(0.287)
Treat*IndSame	-0.317***	-0.186	-0.265	0.115
	(0.104)	(0.223)	(0.333)	(0.360)
After*IndSame	0.195***	-0.069	-0.165	0.330
	(0.034)	(0.188)	(0.358)	(0.330)
Treat*After	-0.093	-0.077	-0.077	-0.076
	(0.080)	(0.076)	(0.076)	(0.076)
Treat*After*IndSame	0.194**	0.024	0.441	0.109
	(0.087)	(0.271)	(0.407)	(0.390)
R-squared	0.01	0.01	0.01	0.01

All stores that are within 0.5 mile of treatment and control chain stores and have self-reported sales in the 2012 NETS database included. Sales between 2004 and 2012 only. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I), the SIC4 (II), the SIC6 (III), and the SIC8 (IV) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

5.4 Estimation results: employment size

I use also the employment size as a proxy for sales. First, I estimate the DID models of employment, similarly, using the four groups of observations.[5] Here I present only the estimation results using the first group of observations (Table E.1, Table E.2, Table E.3, and Table E.4). Similar to the findings in Section 5.3, the estimates of the "After*IndSame" variable are positive in most cases. However, the estimates are significantly negative when it is the case of Mervyn's. Thus, whether the benefits from less price competition exceed the externality loss from consumer searching may depend upon the specific case. These results again show that consumer searching alone cannot well explain retail agglomeration.

I then estimate the Triple-Diff models of employment using the four groups of observations (Table E.5, Table E.6, Table E.7, and Table E.8). The estimates of the "Treat*After*IndSame", which captures the effect of the consumer searching mechanism, are still mixed. The estimates of the "Treat*After" variable provide little evidence of the effect from the cost sharing mechanism, except those for the department store "Mervyn's" in Table E.6.

[5]The employment of each establishment, if not self-reported, is estimated by Dun and Bradstreet in the 2012 NETS database.

5.5 Estimation results: the probability of closure

Now I use the probability of closure as the dependent variable, as Shoag and Veuger 2014 and Benmelech et al. 2014 do. I estimate the DID models and the Triple-Diff models using the four groups of observations. Table F.1, Table F.2, Table F.3, and Table F.4 present the estimation results using the first group of observation, that is, stores within 1 mile of treatment and control chain stores. As in Section 5.3 and Section 5.4, I find no evidence of the effect from cost sharing. Also, the estimates that capture the effect from consumer searching are mixed, which is different from the findings in Benmelech et al. 2014. Only the bankruptcy of "Mervyn's" is associated with a significant increase in the closure probability for nearby stores in the same industry as "Mervyn's" than those in different industries.

5.6 Estimation results: the birth of new stores

Following Rosenthal and Strange 2003, I use also the birth of new stores to measure the externalities associated with agglomeration. I calculate the number of new stores of one industry around each chain store in each year and then estimate both the DID and the Triple-Diff models to distinguish the effect of consumer searching and that of cost sharing. The number of new stores is calculated by 2-digit, 4-digit, 6-digit, and 8-digit industries, respectively.

Here I present only the estimation results of the Triple-Diff models (Table 5.10). Across different types of chain stores, there is no consistent pattern of estimates. However, the negative coefficients in most of the columns show that both consumer searching and cost sharing contribute to retail agglomeration. For example, after the bankruptcy of "KB Toys", within 1mile of their original store locations there is a significant decrease in the number of new stores in industries different from "KB Toys", using "Toys R Us" as control. Moreover, the number of new stores in the same industry as "KB Toys" decreases significantly more than that of new stores in industries different from "KB Toys".

5.7 Summary

The estimation results using sales, employment size, the probability of closure, and the birth of new stores as dependent variables in this chapter are mixed and do not lend consistent support to the contribution of either the consumer searching mechanism or the cost sharing mechanism to retail agglomeration. Consumer searching alone cannot well explain retail agglomeration; however, there is little evidence of the contribution of the cost sharing mechanism to retail agglomeration.

More important, there are flaws in the current research design. First, there may be significant difference between the area around the bankrupt chain and that around the control chain. Future research should check to be sure that the controls are comparable in terms of demographic composition, income, race, etc., or use matching techniques to find comparable

Table 5.10: Estimation results of the Triple-Diff model on the birth of new stores

Dependent variable: the number of new stores within 1 mile				
	(I) SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8
		Panel 1: Specialty store		
Treat*After	-0.515*	-0.107***	-0.052***	-0.035***
	(0.294)	(0.039)	(0.014)	(0.010)
Treat*After*IndSame	-0.333	0.606	0.396	0.284
	(2.421)	(0.747)	(0.403)	(0.401)
N	432,476	2,295,216	4,675,044	6,691,454
		Panel 2: Toy store		
Treat*After	-0.085**	-0.016***	-0.008***	-0.007***
	(0.041)	(0.006)	(0.002)	(0.002)
Treat*After*IndSame	-0.738**	-0.377***	-0.441***	-0.450***
	(0.319)	(0.105)	(0.060)	(0.053)
N	1,992,276	9,655,074	18,563,512	24,705,010
		Panel 3: Department store (Mervyn's)		
Treat*After	0.087	-0.008	-0.012**	-0.014***
	(0.082)	(0.012)	(0.005)	(0.004)
Treat*After*IndSame	-0.367	-0.081	-0.116	-0.100
	(0.660)	(0.219)	(0.123)	(0.106)
N	1,535,996	6,797,252	12,261,194	15,481,444
		Panel 4: Department store (Goody's)		
Treat*After	-0.390***	-0.084***	-0.045***	-0.034***
	(0.060)	(0.010)	(0.004)	(0.003)
Treat*After*IndSame	0.158	0.122	0.090	0.022
	(0.455)	(0.167)	(0.096)	(0.088)
N	7,188,742	2,295,216	12,786,642	15,982,120

Each observation records the number of new stores of a "chain store-year-industry" pair. "IndSame", i.e., whether a industry is same as that of the center chain store, is defined at the SIC2 (Column I), the SIC4 (II), the SIC6 (III), and the SIC8 (IV) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

controls. Second, the effect of the cost sharing mechanism is likely to be heterogeneous across industries. This will cause the estimated effect of the consumer searching mechanism to be biased. To partly address this issue, I have added the air distance between nearby stores and the center chain store since the effect of cost sharing is expected to vary with walking distance between stores.

Chapter 6

Conclusions

The main findings of this dissertation are as follows. First, Using a simple model of firms' location choice and consumers' shopping destination choice, I show how both mechanisms promote the agglomeration of retail stores. The degree of retail agglomeration, measured by Ripley's K function, increases with the importance of consumer searching and with the benefits from scale economy in intermediate inputs such as advertisements and delivery services.

Second, I present stylized facts based on industry agglomeration. Using the 2012 National Establishment Time-Series database, I calculate Ripley's K function and Ellison-Glaeser index that measure the degree of agglomeration for each CBSA-NAICS industry-year pair in the U.S. from 1992 to 2012. The estimated Ellison-Glaeser indices show that the ratio of industries that are highly concentrated at the NAICS6-CBSA level decreases from 12.1% in 1992 to 9.9% in 2012, while that of industries that are not concentrated increases from 86.6% to 89.1% during the same period. The estimated Ripley's K values show that the overall degree of retail agglomeration has been decreasing since 1992. Also, I find much inconsistency between the estimated Ripley's K values and Eillson-Glaeser indices. Therefore, cautious should be taken when using these measures for empirical work, e.g., the identification of sources of manufacturing agglomeration.

Third, using the variation in the percentage of population that use Internet across regions and in the Internet sales share across retailing industries, I find that a 10 percentage points increases in the Internet use ratio and in the Internet sales share is associated with a 3.1-3.5 decrease and a 9.5-10.8 decrease in the calculated Ripley's K value. This provides suggestive evidence of the role consumer searching plays in promoting retail agglomeration since Internet decreases the importance of consumer searching in person and thus reduces the benefits of being in a retail cluster. I use a DID approach to partially address the endogeneity issue and find that industries easy to make comparison shopping online experience a greater decrease in the degree of agglomeration by 6.8-12.6 before and after 2000.

Last, I use a triple-difference approach based on the bankruptcy of chain stores like Borders and Circuit City to distinguish the effects of consumer searching and cost sharing. The estimation results for various types of chain stores using sales, employment size, the proba-

bility of closure, and the birth of new stores as dependent variables are mixed. Consumer searching alone cannot well explain retail agglomeration; however, there is little evidence of the contribution of the cost sharing mechanism to retail agglomeration.

This dissertation contributes to the literature in four aspects. First, it is the first to propose and investigate the cost sharing mechanism on the supply side that may encourage the geographic concentration of retail firms. The existing literature focuses on only the consumer searching mechanism on the demand side. Second, it shows how the measure of agglomeration is associated with factors that represent the consumer searching and cost sharing mechanisms using a simple model of firms and consumers. Third, it is the first to show the trend of retail agglomeration across industries and across regions in two decades. Due to the data limitation, earlier studies present at most cross-industry evidence on agglomeration. Also, this dissertation uses both discrete-unit statistics (Ellison-Glaeser index) and distance-based measures (Ripley's K function). To my knowledge, this is the first study to use the latter for measuring industry agglomeration. Fourth, it identifies the effects of consumer searching and cost sharing on agglomeration by i) using Internet use and Internet sales as proxies and ii) using an exogenous shock–the bankruptcy of chain stores. It complements the scant empirical literature on the micro-mechanism of retail agglomeration.

The findings of this dissertation have important implications for planning. The two mechanisms shown in the simple model actually represent two types of positive externalities in retailers' location choice. The existence of such positive externalities means that without intervention, the actual degree of retail agglomeration would be less than the optimum, since a retailer cannot get all the benefits it produces. This to some extent justifies planners' intervention in terms of the geographic size and the density of retail zones. Also, the negative effect of Internet on retail agglomeration found in this dissertation is good news for the implementation of Senate Bill 375, because it tends to decrease the average distance of shopping trips and thus reduce the pollutant emissions.

There are still limitations in this dissertation. For example, to make a thorough performance comparison of Ripley's K function and Ellison-Glaeser index, it is necessary to estimate these indices at various distance intervals and by different spatial units. Also, I have not well addressed the endogeneity of Internet use and Internet shopping in Chapter 4. Moreover, the identification strategy on the basis of bankruptcy needs to be refined, because the effect of the cost sharing mechanism is likely to be heterogeneous across industries.

Future work as regards the model includes introducing: i) price competition; and ii) a land market and thus rent differentiation. Another option is to investigate how the Ellison-Glaeser index is associated with the consumer searching mechanism and the cost sharing mechanism using a city consisting of square grids.

As for the cross-industry evidence, I will follow the approach taken by Rosenthal and Strange 2001, Baldwin, Brown, and Rigby 2010, and Ellison, Glaeser, and Kerr 2010, by constructing industry-level measures for the consumer searching mechanism and the cost sharing mechanism of retail agglomeration. For example, one candidate measure for the advertisement cost sharing is the unit cost for reaching one consumer. An alternative indicator is the difference between the distribution of an industry's target customers and that of

the population in terms of age, education, and income, which approximates the difficulty in reaching potential customers and could be constructed using the Census Retail Trade data. For consumer searching, it is shown that the concentration of retail stores in an industry increases with product variety of this industry in a region (Fischer and Harrington Jr 1996). Thus, one candidate is the number of upstream manufacturers. In addition, it is important to distinguish within-store agglomeration and inter-store agglomeration by controlling for the average employment size of a retailing industry. Another issue is the industry structure. It is expected that stores that belong to the same retail chain would be scattered within a region to avoid the competition between these stores. I will control for the number of chains that belong to a certain retailing industry.

As regards the estimations based on chain store bankruptcy, future work should use more types of bankrupt chain stores to make the estimation results comparable to Shoag and Veuger 2014 and Benmelech et al. 2014. To address the heterogeneity in the effects of the cost sharing mechanism, one approach is to construct the "consumer composition distance", which is defined as the share in expenditures by age, income, etc., between the industry of nearby store and that of the "center" chain store and will be constructed using the Consumer Expenditure Survey data. Another approach is to add also the number of stores within the same industry supposing that the cost sharing effect on stores in industries different from the chain store varies with the number of stores in that specific industry.

There are other facts that can be used for identification. First, the consumer searching effect may depend on the number of stores. For example, above a certain threshold, adding one more store has no effects on consumers' utility from comparison shopping. This dimension of difference can also be added into the model of retailers and consumers. Second, the effects of both mechanisms may depend on the age of store, assuming that stores that have been there for a while tend to have more visitors (Adelino, Ma, and Robinson 2014).

Bibliography

[1] Manuel Adelino, Song Ma, and David T Robinson. *Firm age, investment opportunities, and job creation*. Tech. rep. National Bureau of Economic Research, 2014.

[2] Alex Anas and Ikki Kim. "General equilibrium models of polycentric urban land use with endogenous congestion and job agglomeration". In: *Journal of Urban Economics* 40.2 (1996), pp. 232–256.

[3] Richard Arnott and Yundong Tu. *Shopper City*. Tech. rep. 2008.

[4] David B Audretsch and Maryann P Feldman. "R&D spillovers and the geography of innovation and production". In: *The American Economic Review* (1996), pp. 630–640.

[5] John R Baldwin, W Mark Brown, and David L Rigby. "Agglomeration Economies: Microdata Panel Estimates From Canadian Manufacturing". In: *Journal of Regional Science* 50.5 (2010), pp. 915–934.

[6] Efraim Benmelech et al. *The Agglomeration of Bankruptcy*. Tech. rep. National Bureau of Economic Research, 2014.

[7] Brian Joe Lobley Berry. *Geography of market centers and retail distribution*. Prentice-Hall Englewood Cliffs, NJ, 1967.

[8] JE Besag. "Comments on Ripley's paper". In: *Journal of the Royal Statistical Society B* 39.2 (1977), pp. 193–195.

[9] Stephen B. Billings and Erik B. Johnson. "Localization and Colocalization within an Urban Area". 2013.

[10] Marlon G Boarnet et al. "Emerging planning challenges in retail: The case of Wal-Mart". In: *Journal of the American Planning Association* 71.4 (2005), pp. 433–449.

[11] Pierre-Philippe Combes and Henry G Overman. "The spatial distribution of economic activities in the European Union". In: *Handbook of regional and urban economics* 4 (2004), pp. 2845–2909.

[12] Sumon Datta and K Sudhir. "The Agglomeration Differentiation Tradeoff in Spatial Location Choice". In: *manuscript. Yale School of Management* (2011).

[13] Michael P Devereux, Rachel Griffith, and Helen Simpson. "The geographic distribution of production activity in the UK". In: *Regional Science and Urban Economics* 34.5 (2004), pp. 533–564.

[14] Valter Di Giacinto and Marcello Pagnini. "Local and global agglomeration patterns: two econometrics-based indicators". In: *Regional Science and Urban Economics* 41.3 (2011), pp. 266–280.

[15] Avinash K Dixit and Joseph E Stiglitz. "Monopolistic competition and optimum product diversity". In: *The American Economic Review* (1977), pp. 297–308.

[16] Marc Dudey. "Competition by choice: The effect of consumer search on firm location decisions". In: *The American Economic Review* (1990), pp. 1092–1104.

[17] Guy Dumais, Glenn Ellison, and Edward L Glaeser. "Geographic concentration as a dynamic process". In: *Review of economics and Statistics* 84.2 (2002), pp. 193–204.

[18] Bill Dunkley, Amy Helling, and David S Sawicki. "Accessibility Versus Scale Examining the Tradeoffs in Grocery Stores". In: *Journal of Planning Education and Research* 23.4 (2004), pp. 387–401.

[19] Gilles Duranton and Henry G Overman. "Testing for localization using micro-geographic data". In: *The Review of Economic Studies* 72.4 (2005), pp. 1077–1106.

[20] Gilles Duranton and Diego Puga. "Chapter 48 Micro-foundations of urban agglomeration economies". In: *Cities and Geography*. Ed. by J. Vernon Henderson and Jacques-François Thisse. Vol. 4. Handbook of Regional and Urban Economics. Elsevier, 2004, pp. 2063 –2117. DOI: http://dx.doi.org/10.1016/S1574-0080(04)80005-1. URL: http://www.sciencedirect.com/science/article/pii/S1574008004800051.

[21] B Curtis Eaton and Richard G Lipsey. "The principle of minimum differentiation reconsidered: Some new developments in the theory of spatial competition". In: *The Review of Economic Studies* (1975), pp. 27–49.

[22] Glenn Ellison, Edward L. Glaeser, and William R. Kerr. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns". In: *American Economic Review* 100.3 (2010), pp. 1195–1213. DOI: 10.1257/aer.100.3.1195. URL: http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.3.1195.

[23] Glenn Ellison and Edward Ludwig Glaeser. "Geographic Concentration in US Manufacturing Industries: A Dartboard Approach". In: *The Journal of Political Economy* 105.5 (1997), pp. 889–927.

[24] Edward J Feser and Stuart H Sweeney. "A test for the coincident economic and spatial clustering of business enterprises". In: *Journal of Geographical Systems* 2.4 (2000), pp. 349–373.

[25] Jeffrey H Fischer and J Harrington Jr. "Product variety and firm agglomeration". In: *The RAND Journal of Economics* 27.2 (1996), pp. 281–309.

[26] Masahisa Fujita and Hideaki Ogawa. "Multiple equilibria and structural transition of non-monocentric urban configurations". In: *Regional science and urban economics* 12.2 (1982), pp. 161–196.

[27] Masahisa Fujita and Jacques-Francois Thisse. "Spatial competition with a land market: Hotelling and Von Thunen unified". In: *The Review of Economic Studies* 53.5 (1986), pp. 819–841.

[28] Arthur Getis and J Keith Ord. "The analysis of spatial association by use of distance statistics". In: *Geographical analysis* 24.3 (1992), pp. 189–206.

[29] Edward L Glaeser, Vernon Henderson, and Robert P Inman. "The future of urban research: nonmarket interactions [with comments]". In: *Brookings-Wharton papers on urban affairs* (2000), pp. 101–149.

[30] Gerald S Goldstein and Timothy J Gronberg. "Economies of scope and economies of agglomeration". In: *Journal of Urban Economics* 16.1 (1984), pp. 91–104.

[31] Thomas O Graff. "The locations of Wal-Mart and Kmart supercenters: contrasting corporate strategies". In: *The Professional Geographer* 50.1 (1998), pp. 46–57.

[32] Rachel Guillain and Julie Le Gallo. "Agglomeration and dispersion of economic activities in and around Paris: an exploratory spatial data analysis". In: *Environment and planning. B, Planning & design* 37.6 (2010), p. 961.

[33] Jessie Handbury and David E Weinstein. "Goods prices and availability in cities". In: *The Review of Economic Studies* (2014), rdu033.

[34] Robert W Helsley and William C Strange. "Matching and agglomeration economies in a system of cities". In: *Regional Science and urban economics* 20.2 (1990), pp. 189–212.

[35] J Vernon Henderson. "Marshall's scale economies". In: *Journal of urban economics* 53.1 (2003), pp. 1–28.

[36] Richard H Holton. "The distinction between convenience goods, shopping goods, and specialty goods". In: *The Journal of Marketing* (1958), pp. 53–56.

[37] Harold Hotelling. *Stability in competition*. Springer, 1990.

[38] David L Huff. "Defining and estimating a trading area". In: *The Journal of Marketing* (1964), pp. 34–38.

[39] Walter Isard. "A simple rationale for gravity model type behavior". In: *Papers in Regional Science* 35.1 (1975), pp. 25–30.

[40] Inge Ivarsson and Claes Göran Alvstam. "Upgrading in global value-chains: a case study of technology-learning among IKEA-suppliers in China and Southeast Asia". In: *Journal of Economic Geography* (2010), lbq009.

[41] Sukkoo Kim. "Regions, resources, and economic geography: Sources of US regional comparative advantage, 1880–1987". In: *Regional Science and Urban Economics* 29.1 (1999), pp. 1–32.

[42] Hideo Konishi. "Concentration of competing retail stores". In: *Journal of Urban Economics* 58.3 (2005), pp. 488–512.

[43] Paul Krugman. *Increasing returns and economic geography*. Tech. rep. National Bureau of economic research, 1990.

[44] Paul R Krugman. *Geography and trade*. MIT press, 1991.

[45] Ira S Lowry. *A model of metropolis*. Rand Corporation, Santa Monica, CA, USA, 1964.

[46] Eric Marcon and Florence Puech. "Evaluating the geographic concentration of industries using distance-based methods". In: *Journal of Economic Geography* 3.4 (2003), pp. 409–428.

[47] Eric Marcon and Florence Puech. "Measures of the geographic concentration of industries: improving distance-based methods". In: *Journal of Economic Geography* 10.5 (2010), pp. 745–762.

[48] Alfred Marshall. *Principles of Economics*. London: MacMillan, 1920.

[49] Françoise Maurel and Béatrice Sédillot. "A measure of the geographic concentration in French manufacturing industries". In: *Regional Science and Urban Economics* 29.5 (1999), pp. 575–604.

[50] David Neumark, Junfu Zhang, and Brandon Wall. *Employment dynamics and business relocation: New evidence from the National Establishment Time Series*. Tech. rep. National Bureau of Economic Research, 2005.

[51] Stan Openshaw and S Openshaw. "The modifiable areal unit problem". In: Geo Abstracts University of East Anglia. 1984.

[52] John M Quigley. "Urban diversity and economic growth". In: *The Journal of Economic Perspectives* (1998), pp. 127–138.

[53] David L Rigby and Jürgen Essletzbichler. "Agglomeration economies and productivity differences in US cities". In: *Journal of Economic Geography* 2.4 (2002), pp. 407–432.

[54] Brian D Ripley. "Modelling spatial patterns". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1977), pp. 172–212.

[55] Brian D Ripley. "The second-order analysis of stationary point processes". In: *Journal of Applied Probability* (1976), pp. 255–266.

[56] Stuart S Rosenthal and William C Strange. "Geography, industrial organization, and agglomeration". In: *Review of Economics and Statistics* 85.2 (2003), pp. 377–393.

[57] Stuart S Rosenthal and William C Strange. "The determinants of agglomeration". In: *Journal of Urban Economics* 50.2 (2001), pp. 191–229.

[58] Tadeusz M Rybczynski. "Factor endowment and relative commodity prices". In: *Economica* (1955), pp. 336–341.

[59] AnnaLee Saxenian. *Regional advantage: culture and competition in Silicon Valley and Route 128*. 1994.

[60] Andres Sevtsuk. "Path and place: a study of urban geometry and retail activity in Cambridge and Somerville, MA". PhD thesis. Massachusetts Institute of Technology, 2010.

[61] Daniel Shoag and Stan Veuger. "Shops and the City". In: (2014).

[62] Konrad Stahl. "Differentiated products, consumer search, and locational oligopoly". In: *The Journal of Industrial Economics* (1982), pp. 97–113.

[63] Allen Tran. "The Aggregate Impact of Online Retail". In: *US Census Bureau Center for Economic Studies Paper No. CES-WP-14-23* (2014).

[64] Walls & Associates. *2009 NETS Database: Definitions of Accuracy Codes Associated with Latitude/Longitude (Notes from D&B)*. Walls & Associates. 2009.

[65] Jesse WJ Weltevreden. "Substitution or complementarity? How the Internet changes city centre shopping". In: *Journal of Retailing and Consumer Services* 14.3 (2007), pp. 192–207.

[66] Alan G Wilson. "A family of spatial interaction models, and associated developments". In: *Environment and Planning* 3.1 (1971), pp. 1–32.

[67] Asher Wolinsky. "Retail trade concentration due to consumers' imperfect information". In: *The Bell Journal of Economics* (1983), pp. 275–282.

Appendix A

A list of 6-digit 2007 NAICS retailing industries

2007 6-digit NAICS	2007 6-digit NAICS title
441110	New Car Dealers
441120	Used Car Dealers
441210	Recreational Vehicle Dealers
441221	Motorcycle, ATV, and Personal Watercraft Dealers
441222	Boat Dealers
441229	All Other Motor Vehicle Dealers
441310	Automotive Parts and Accessories Stores
441320	Tire Dealers
442110	Furniture Stores
442210	Floor Covering Stores
442291	Window Treatment Stores
442299	All Other Home Furnishings Stores
443111	Household Appliance Stores
443112	Radio, Television, and Other Electronics Stores
443120	Computer and Software Stores
443130	Camera and Photographic Supplies Stores
444110	Home Centers
444120	Paint and Wallpaper Stores
444130	Hardware Stores
444190	Other Building Material Dealers
444210	Outdoor Power Equipment Stores
444220	Nursery, Garden Center, and Farm Supply Stores
445110	Supermarkets and Other Grocery (except Convenience) Stores
445120	Convenience Stores
445210	Meat Markets

2007 6-digit NAICS	2007 6-digit NAICS title
445220	Fish and Seafood Markets
445230	Fruit and Vegetable Markets
445291	Baked Goods Stores
445292	Confectionery and Nut Stores
445299	All Other Specialty Food Stores
445310	Beer, Wine, and Liquor Stores
446110	Pharmacies and Drug Stores
446120	Cosmetics, Beauty Supplies, and Perfume Stores
446130	Optical Goods Stores
446191	Food (Health) Supplement Stores
446199	All Other Health and Personal Care Stores
447110	Gasoline Stations with Convenience Stores
447190	Other Gasoline Stations
448110	Men's Clothing Stores
448120	Women's Clothing Stores
448130	Children's and Infants' Clothing Stores
448140	Family Clothing Stores
448150	Clothing Accessories Stores
448190	Other Clothing Stores
448210	Shoe Stores
448310	Jewelry Stores
448320	Luggage and Leather Goods Stores
451110	Sporting Goods Stores
451120	Hobby, Toy, and Game Stores
451130	Sewing, Needlework, and Piece Goods Stores
451140	Musical Instrument and Supplies Stores
451211	Book Stores
451212	News Dealers and Newsstands
451220	Prerecorded Tape, Compact Disc, and Record Stores
452111	Department Stores (except Discount Department Stores)
452112	Discount Department Stores
452910	Warehouse Clubs and Supercenters
452990	All Other General Merchandise Stores
453110	Florists
453210	Office Supplies and Stationery Stores
453220	Gift, Novelty, and Souvenir Stores
453310	Used Merchandise Stores
453910	Pet and Pet Supplies Stores
453920	Art Dealers
453930	Manufactured (Mobile) Home Dealers

2007 6-digit NAICS	2007 6-digit NAICS title
453991	Tobacco Stores
453998	All Other Miscellaneous Store Retailers (except Tobacco Stores)
454111	Electronic Shopping
454112	Electronic Auctions
454113	Mail-Order Houses
454210	Vending Machine Operators
454311	Heating Oil Dealers
454312	Liquefied Petroleum Gas (Bottled Gas) Dealers
454319	Other Fuel Dealers
454390	Other Direct Selling Establishments

Appendix B

The estimated degree of agglomeration of more retailing industries

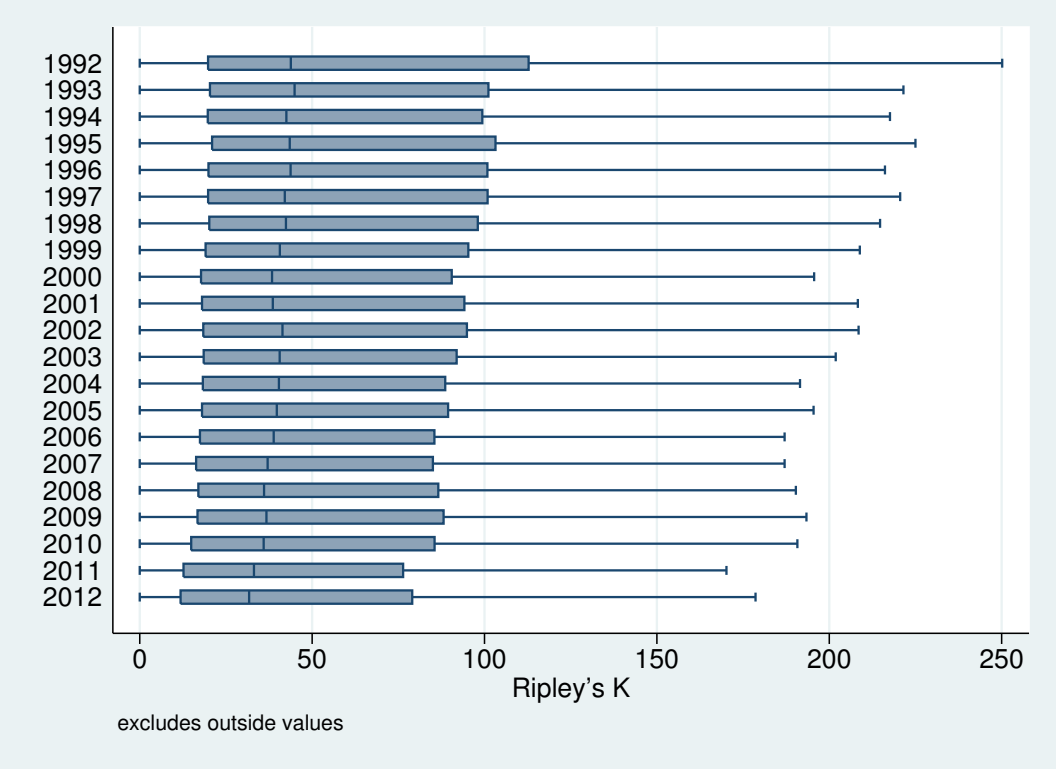


Figure B.1: The estimated Ripley's K of Radio, television, and other electronics stores (443112) across CBSAs

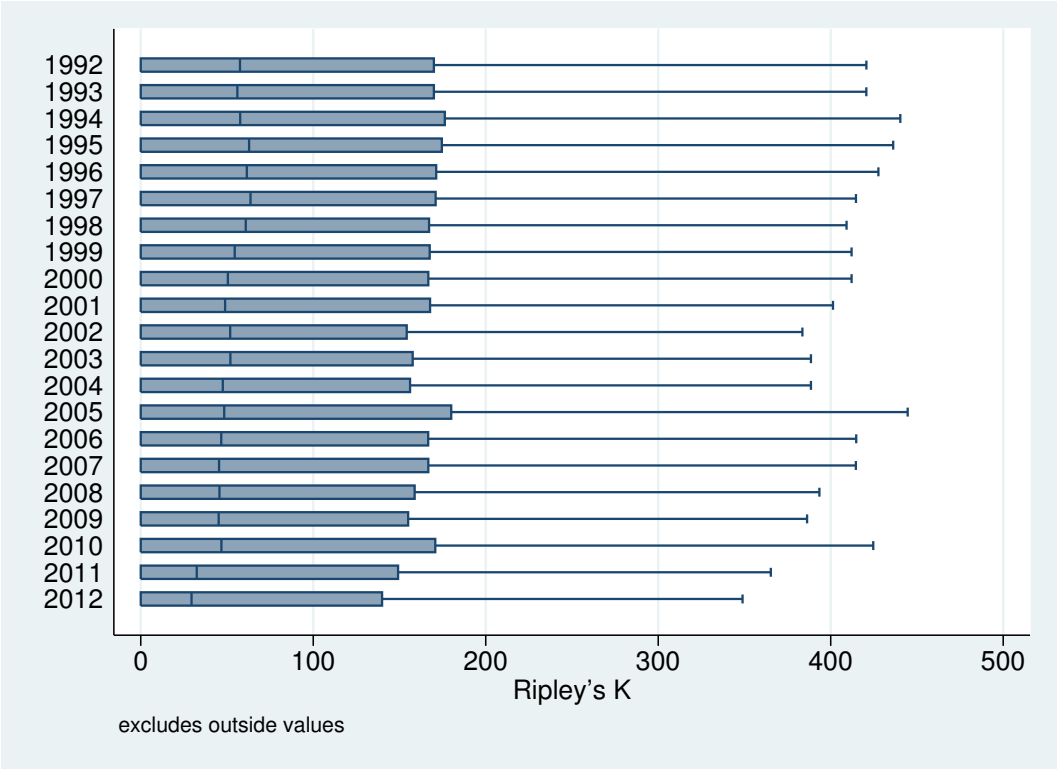


Figure B.2: The estimated Ripley's K of Men's clothing stores (448110) across CBSAs

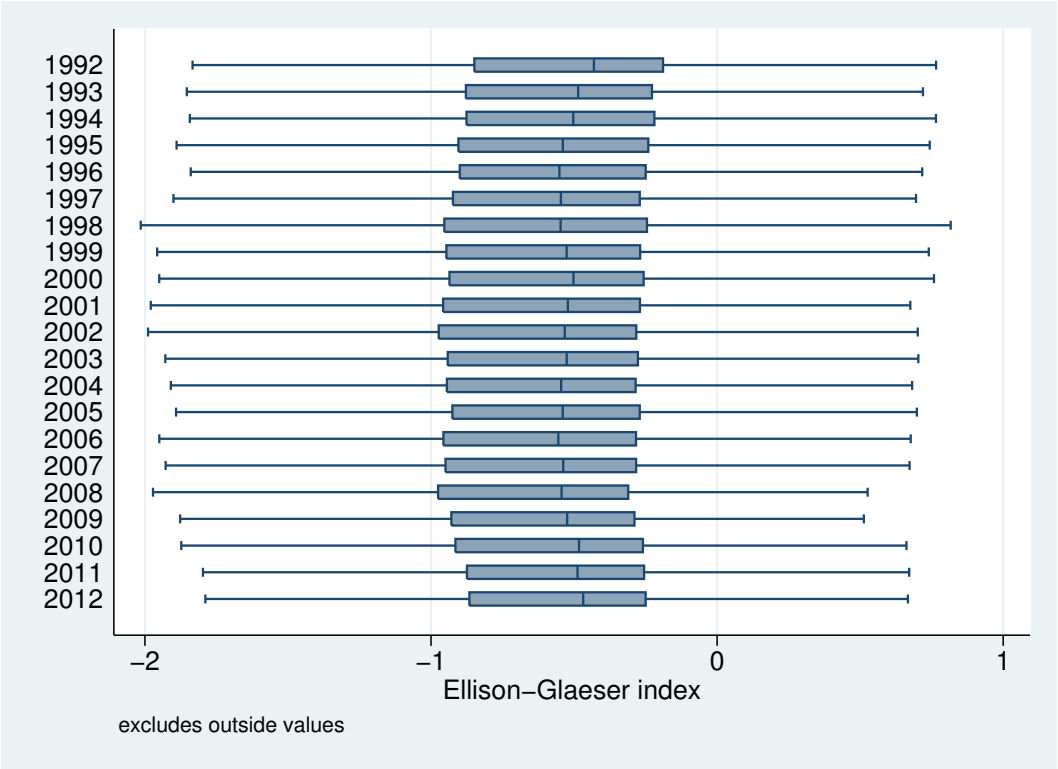


Figure B.3: The estimated Ellison-Glaeser indices of book stores (451211) across CBSAs

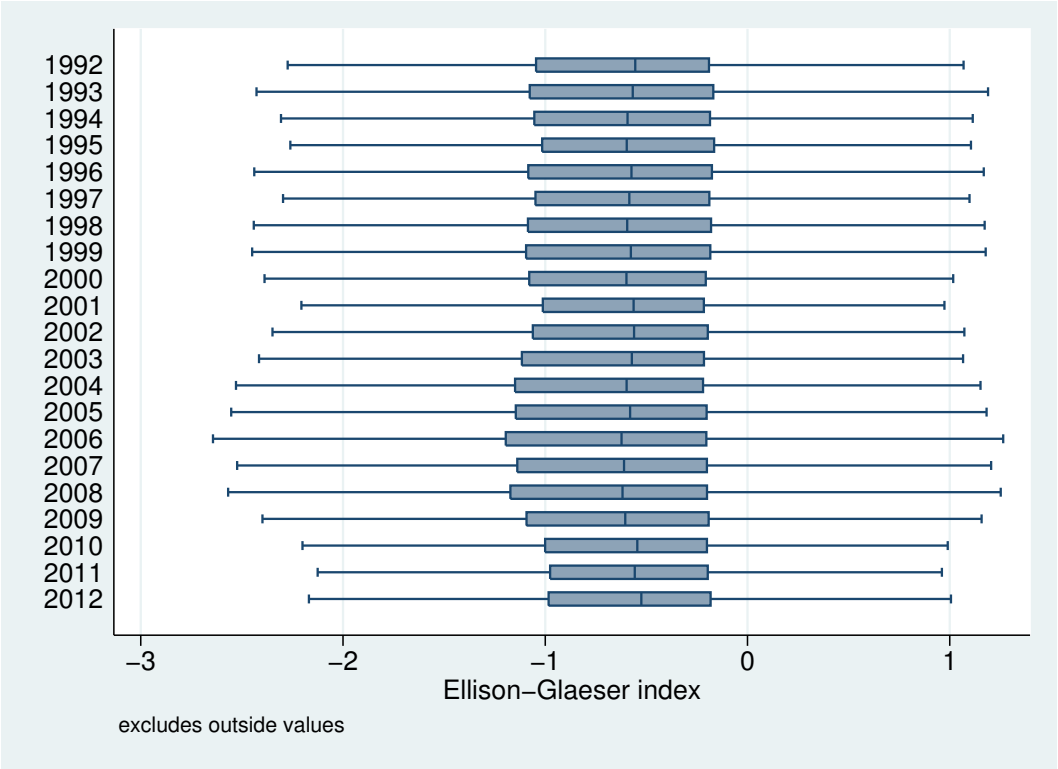


Figure B.4: The estimated Ellison-Glaeser indices of Shoe stores (448210) across CBSAs

Appendix C

Find chain stores in the NETS database

I use a combination of company name, trade name, and headquarter DUNS number included in the NETS database to identify stores of bankrupt chains and control chains. Take Sharper Image as an example. Looking at all stores in the NETS database that have both "SHARPER" as the first word of company name and "IMAGE" as the second word of company name, I find that the headquarter DUNS number is "54612601" for most of them. So I use both company name and headquarter DUNS number as the selection criteria. In so doing, I can avoid having non Sharper Image stores in the estimation.

I use the following criteria for identifying stores that belong to the chains used in my estimation.

- Mervyn's: i) "MERVYNS" as the first word of trade name; and ii) "7968670" as the headquarter DUNS number.
- Goody's: i) "GOODYS" as the first word of trade name; and ii) "618930189" or "180691206" as the headquarter DUNS number.
- Kohl's: i) "KOHLS" as the first word of company name; and ii) "787408996" as the headquarter DUNS number.
- Sharper Image: i) "SHARPER" as the first word of company name; ii) "IMAGE" as the second word of company name; and iii) "54612601" as the headquarter DUNS number.
- Brookstone: i) "BROOKSTONE" as the first company word; and ii) "791211662" as the headquarter DUNS number.
- KB Toys: "KB TOYS", "KB TOY", "K B TOYS", "K B TOY", "KAY BEE TOYS", "KAY BEE TOY" as company name.

- Toys R Us: i) "TOYS", "R", "US" as the first, second, and third word of company name, respectively; and ii) "6985808" or "780822073" as the headquarter DUNS number.

Appendix D

More estimation results of models of sales

Table D.1: Estimation results of the DID model on self-reported sales: Specialty store

Dependent variable: log(annual sales)								
N = 39,595	**(I)**	**(II)**	**(III)**	**(IV)**	**(V)**	**(VI)**	**(VII)**	**(VIII)**
	SIC2	SIC4	SIC6	SIC8	SIC2	SIC4	SIC6	SIC8
After	-0.566***	-0.566***	-0.568***	-0.568***	-0.146	-0.145	-0.145	-0.145
	(0.042)	(0.042)	(0.041)	(0.041)	(0.088)	(0.088)	(0.088)	(0.088)
IndSame	0.232*	0.555**	-0.086	-0.228	0.232*	0.555**	-0.086	-0.228
	(0.119)	(0.240)	(0.294)	(0.329)	(0.119)	(0.240)	(0.294)	(0.329)
After*IndSame	0.159	0.193	-0.264	-0.184	0.094	0.131	-0.499*	-0.451
	(0.258)	(0.755)	(0.277)	(0.361)	(0.262)	(0.910)	(0.258)	(0.346)
After*Distance					-0.672***	-0.676***	-0.678***	-0.678***
					(0.142)	(0.142)	(0.141)	(0.141)
R-squared	0.02	0.02	0.02	0.02	0.03	0.03	0.02	0.02

All stores that are within 1 mile of bankrupt chain stores and have self-reported sales in the 2012 NETS database included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table D.2: Estimation results of the DID model on self-reported sales: Department store (Mervyn's)

Dependent variable: log(annual sales)								
N = 79,316	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.491***	-0.491***	-0.491***	-0.491***	-0.142*	-0.144*	-0.143*	-0.143*
	(0.030)	(0.030)	(0.030)	(0.030)	(0.077)	(0.077)	(0.077)	(0.077)
IndSame	-0.469*	1.358**	-0.042	-0.042	-0.469*	1.358**	-0.042	-0.042
	(0.247)	(0.617)	(1.207)	(1.207)	(0.247)	(0.617)	(1.207)	(1.207)
After*IndSame	0.195	0.497	0.000	0.000	0.139	0.237	0.000	0.000
	(0.378)	(0.610)	(0.000)	(0.000)	(0.341)	(0.616)	(0.000)	(0.000)
After*Distance					-0.566***	-0.562***	-0.564***	-0.564***
					(0.123)	(0.124)	(0.123)	(0.123)
R-squared	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

All stores that are within 1 mile of bankrupt chain stores and have self-reported sales in the 2012 NETS database included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table D.3: Estimation results of the DID model on self-reported sales: Department store (Goody's)

Dependent variable: log(annual sales)								
N = 35,586	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.473***	-0.447***	-0.446***	-0.446***	-0.350***	-0.329***	-0.328***	-0.328***
	(0.025)	(0.025)	(0.025)	(0.025)	(0.058)	(0.058)	(0.058)	(0.058)
IndSame	-0.675***	-0.755***	-0.592***	-0.592***	-0.675***	-0.755***	-0.592***	-0.592***
	(0.045)	(0.176)	(0.138)	(0.138)	(0.045)	(0.176)	(0.138)	(0.138)
After*IndSame	0.391***	0.050	-0.423	-0.423	0.376***	0.075	-0.412	-0.412
	(0.100)	(0.466)	(0.603)	(0.603)	(0.099)	(0.470)	(0.605)	(0.605)
After*Distance					-0.199**	-0.190**	-0.191**	-0.191**
					(0.096)	(0.096)	(0.096)	(0.096)
R-squared	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.01

All stores that are within 1 mile of bankrupt chain stores and have self-reported sales in the 2012 NETS database included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table D.4: Estimation results of the DID model on self-reported sales: Toy store

Dependent variable: log(annual sales)								
N = 882,595	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.595***	-0.573***	-0.573***	-0.573***	-0.454***	-0.435***	-0.434***	-0.434***
	(0.045)	(0.041)	(0.041)	(0.041)	(0.111)	(0.109)	(0.109)	(0.109)
IndSame	-0.384***	-0.786***	-0.872***	-0.434***	-0.384***	-0.786***	-0.872***	-0.434***
	(0.069)	(0.110)	(0.186)	(0.144)	(0.069)	(0.110)	(0.186)	(0.144)
After*IndSame	0.205***	0.253	0.460**	0.251	0.196***	0.238	0.43*	0.227
	(0.062)	(0.159)	(0.231)	(0.210)	(0.065)	(0.159)	(0.232)	(0.210)
After*Distance					-0.245	-0.241	-0.241	-0.241
					(0.167)	(0.166)	(0.166)	(0.166)
R-squared	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

All stores that are within 1 mile of bankrupt chain stores and have self-reported sales in the 2012 NETS database included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table D.5: Estimation results of the DID model on self-reported sales of stores within 0.5 mile: Specialty store

Dependent variable: log(annual sales)								
N = 13,617	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.457***	-0.459***	-0.460***	-0.460***	-0.104	-0.104	-0.105	-0.103
	(0.063)	(0.062)	(0.062)	(0.062)	(0.090)	(0.092)	(0.091)	(0.092)
IndSame	0.182	0.176	-0.094	-0.125	0.182	0.176	-0.094	-0.125
	(0.188)	(0.237)	(0.286)	(0.336)	(0.188)	(0.237)	(0.286)	(0.336)
After*IndSame	-0.053	-0.627**	-0.357	-0.372	-0.028	-0.843**	-0.573**	-0.728**
	(0.331)	(0.255)**	(0.285)	(0.352)	(0.343)	(0.321)	(0.267)	(0.336)
After*Distance					-1.177***	-1.183***	-1.183***	-1.189***
					(0.367)	(0.370)	(0.370)	(0.373)
R-squared	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

All stores that are within 0.5 mile of bankrupt chain stores and have self-reported sales in the 2012 NETS database included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table D.6: Estimation results of the DID model on self-reported sales of stores within 0.5 mile: Department store (Mervyn's)

Dependent variable: log(annual sales)								
N = 26,790	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.410***	-0.406***	-0.406***	-0.406***	-0.012	-0.007	-0.005	-0.005
	(0.039)	(0.039)	(0.039)	(0.039)	(0.123)	(0.121)	(0.121)	(0.121)
IndSame	-0.070	1.732***	0.000	0.000	-0.070	1.732***	0.000	0.000
	(0.423)	(0.057)	(0.000)	(0.000)	(0.423)	(0.057)	(0.000)	(0.000)
After*IndSame	0.699***	-0.054	0.000	0.000	0.607**	-0.256***	0.000	0.000
	(0.251)	(0.039)	(0.000)	(0.000)	(0.275)	(0.068)	(0.000)	(0.000)
After*Distance					-1.265***	-1.270***	-1.277***	-1.277***
					(0.384)	(0.379)	(0.378)	(0.378)
R-squared	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01

All stores that are within 0.5 mile of bankrupt chain stores and have self-reported sales in the 2012 NETS database included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table D.7: Estimation results of the DID model on self-reported sales of stores within 0.5 mile: Department store (Goody's)

Dependent variable: log(annual sales)								
N = 35,586	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.496***	-0.458***	-0.458***	-0.458***	-0.432***	-0.398***	-0.397***	-0.397***
	(0.037)	(0.035)	(0.035)	(0.035)	(0.078)	(0.078)	(0.078)	(0.078)
IndSame	-0.651***	-0.761***	-0.516***	-0.516**	-0.651***	-0.761***	-0.516**	-0.516**
	(0.060)	(0.215)	(0.210)	(0.210)	(0.060)	(0.215)	(0.210)	(0.210)
After*IndSame	0.466***	-0.062	-0.156	-0.156	0.462***	-0.038	-0.129	-0.129
	(0.142)	(0.670)	(0.783)	(0.783)	(0.142)	(0.664)	(0.775)	(0.775)
After*Distance					-0.210	-0.200	-0.201	-0.201
					(0.260)	(0.260)	(0.260)	(0.260)
R-squared	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.01

All stores that are within 0.5 mile of bankrupt chain stores and have self-reported sales in the 2012 NETS database included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table D.8: Estimation results of the DID model on self-reported sales of stores within 0.5 mile: Toy store

Dependent variable: log(annual sales)								
N = 350,294	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.600***	-0.551***	-0.551***	-0.551***	-0.567***	-0.521***	-0.522***	-0.522***
	(0.090)	(0.087)	(0.087)	(0.087)	(0.176)	(0.169)	(0.169)	(0.169)
IndSame	-0.615***	-0.535***	-0.595***	-0.375**	-0.615***	-0.535***	-0.595***	-0.375**
	(0.116)	(0.174)	(0.197)	(0.155)	(0.116)	(0.174)	(0.197)	(0.155)
After*IndSame	0.508***	0.145	0.406	0.420*	0.503***	0.132	0.385	0.398
	(0.094)	(0.236)	(0.287)	(0.254)	(0.104)	(0.250)	(0.312)	(0.282)
After*Distance					-0.111	-0.103	-0.100	-0.099
					(0.331)	(0.322)	(0.322)	(0.321)
R-squared	0.03	0.02	0.02	0.02	0.03	0.02	0.02	0.02

All stores that are within 0.5 mile of bankrupt chain stores and have self-reported sales in the 2012 NETS database included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table D.9: Estimation results of the DID model on self-reported 2004-2012 sales of stores within 0.5 mile: Specialty store

Dependent variable: log(annual sales)								
N = 5,217	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.330***	-0.329***	-0.330***	-0.331***	0.022	0.026	0.025	0.026
	(0.070)	(0.069)	(0.069)	(0.069)	(0.101)	(0.102)	(0.102)	(0.103)
IndSame	0.145	0.562	0.802*	1.475***	0.145	0.562	0.802*	1.475***
	(0.230)	(0.354)	(0.446)	(0.118)	(0.230)	(0.355)	(0.446)	(0.118)
After*IndSame	-0.016	-1.013**	-1.253**	-1.972***	0.010	-1.229**	-1.470**	-2.328***
	(0.179)	(0.405)	(0.511)	(0.069)	-0.199	(0.479)	(0.598)	(0.103)
After*Distance					-1.177***	-1.183***	-1.183***	-1.189***
					(0.368)	(0.370)	(0.370)	(0.373)
R-squared	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02

All stores that are within 0.5 mile of bankrupt chain stores and have self-reported sales in the 2012 NETS database included. Sales between 2004 and 2012 only. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table D.10: Estimation results of the DID model on self-reported 2004-2012 sales of stores within 0.5 mile: Department store (Mervyn's)

Dependent variable: log(annual sales)								
N = 11,236	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.213***	-0.212***	-0.212***	-0.212***	0.185	0.188	0.189	0.189
	(0.042)	(0.041)	(0.041)	(0.041)	(0.122)	(0.120)	(0.120)	(0.120)
IndSame	0.381	1.736***	0.000	0.000	0.381	1.736***	0.000	0.000
	-0.445	(0.063)	(0.000)	(0.000)	-0.445	(0.063)	(0.000)	(0.000)
After*IndSame	0.248	-0.058	0.000	0.000	0.157	-0.260***	0.000	0.0000
	(0.240)	(0.041)	(0.000)	(0.000)	(0.243)	(0.068)	(0.000)	(0.000)
After*Distance					-1.265***	-1.270***	-1.277***	-1.277***
					(0.384)	(0.379)	(0.378)	(0.378)
R-squared	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.01

All stores that are within 0.5 mile of bankrupt chain stores and have self-reported sales in the 2012 NETS database included. Sales between 2004 and 2012 only. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table D.11: Estimation results of the DID model on self-reported 2004-2012 sales of stores within 0.5 mile: Department store (Goody's)

Dependent variable: log(annual sales)								
N = 13,725	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.253***	-0.238***	-0.237***	-0.237***	-0.189**	-0.178**	-0.177**	-0.177**
	(0.031)	(0.031)	(0.031)	(0.031)	(0.078)	(0.078)	(0.078)	(0.078)
IndSame	-0.480***	-0.748**	-0.189	-0.189	-0.480***	-0.748*8	-0.189	-0.189
	(0.107)	(0.329)	(0.330)	(0.330)	(0.107)	(0.329)	(0.330)	(0.330)
After*IndSame	0.295***	-0.074	-0.483	-0.483	0.291***	-0.051	-0.456	-0.456
	(0.107)	(0.709)	(0.684)	(0.684)	(0.106)	(0.703)	(0.676)	(0.676)
After*Distance					-0.210	-0.200	-0.201	-0.201
					(0.260)	(0.260)	(0.260)	(0.260)
R-squared	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

All stores that are within 0.5 mile of bankrupt chain stores and have self-reported sales in the 2012 NETS database included. Sales between 2004 and 2012 only. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table D.12: Estimation results of the DID model on self-reported 2004-2012 sales of stores within 0.5 mile: Toy store

Dependent variable: log(annual sales)								
N = 136,642	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-0.368***	-0.335***	-0.335***	-0.335***	-0.335*	-0.305*	-0.306*	-0.306*
	(0.079)	(0.075)	(0.075)	(0.075)	(0.171)	(0.164)	(0.164)	(0.164)
IndSame	-0.496***	-0.345**	-0.466**	-0.395*	-0.496***	-0.345**	-0.466**	-0.395*
	(0.097)	(0.168)	(0.195)	(0.216)	(0.097)	(0.168)	(0.195)	(0.216)
After*IndSame	0.389***	-0.045	0.276**	0.440***	0.385	-0.058	0.255	0.418*
	(0.080)	(0.195)	(0.194)	(0.209)	(0.092)	(0.212)	(0.228)	(0.242)
After*Distance					-0.111	-0.103	-0.100	-0.099
					(0.331)	(0.322)	(0.322)	(0.321)
R-squared	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

All stores that are within 0.5 mile of bankrupt chain stores and have self-reported sales in the 2012 NETS database included. Sales between 2004 and 2012 only. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Appendix E

More estimation results of models of employment

Table E.1: Estimation results of the DID model on employment: Specialty store

Dependent variable: employment								
N = 587,953	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-4.556***	-4.485***	-4.480***	-4.477***	-1.977*	-1.908*	-1.902*	-1.900*
	(0.396)	(0.384)	(0.385)	(0.384)	(0.956)	(0.948)	(0.946)	(0.946)
IndSame	-3.003*	-2.134	-4.387	-5.957**	-3.003*	-2.134	-4.387	-5.957**
	(1.580)	(1.872)	(2.584)	(2.751)	(1.580)	(1.872)	(2.584)	(2.751)
After*IndSame	3.899***	4.325**	6.120*	11.772	3.495***	3.997**	5.472*	11.419
	(1.073)	(1.557)	(3.114)	(8.995)	(1.151)	(1.541)	(3.076)	(8.928)
After*Distance					-4.448***	-4.453***	-4.455***	-4.456***
					(1.523)	(1.521)	(1.520)	(1.521)
R-squared	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table E.2: Estimation results of the DID model on employment: Department store (Mervyn's)

Dependent variable: employment				
N = 1,224,843	**(I)**	**(II)**	**(III)**	**(IV)**
	SIC2	SIC4	SIC6	SIC8
After	-3.442***	-3.417***	-3.622***	-3.643***
	(0.355)	(0.354)	(0.343)	(0.341)
IndSame	78.991***	143.219***	124.982***	120.515***
	(14.733)	(6.915)	(8.379)	(8.246)
After*IndSame	-13.963***	-19.485***	-13.359	-15.217
	(4.860)	(5.636)	(9.928)	(10.680)
R-squared	0.01	0.02	0.01	0.01
N = 1,224,843	**(V)**	**(VI)**	**(VII)**	**(VIII)**
	SIC2	SIC4	SIC6	SIC8
After	-0.497	-0.676	-0.334	-0.327
	(0.471)	(0.462)	(0.530)	(0.534)
IndSame	78.991***	143.219***	124.982***	120.515***
	(14.733)	(6.915)	(8.379)	(8.246)
After*IndSame	-14.817***	-20.633***	-14.890	-16.857
	(4.815)	(5.657)	(9.911)	(10.663)
After*Distance	-4.992***	-4.651***	-5.582***	-5.630***
	(0.706)	(0.697)	(0.705)	(0.707)
R-squared	0.01	0.02	0.01	0.01

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table E.3: Estimation results of the DID model on employment: Department store (Goody's)

Dependent variable: employment								
N = 1,289,427	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-3.797***	-3.631***	-3.630***	-3.630***	0.353	0.370	0.367	0.367
	(0.495)	(0.477)	(0.477)	(0.477)	(0.790)	(0.771)	(0.771)	(0.771)
IndSame	-6.230***	10.010***	13.094**	13.094**	-6.230***	10.010**	13.094**	13.094**
	(0.688)	(4.330)	(5.235)	(5.235)	(0.688)	(4.330)	(5.235)	(5.235)
After*IndSame	4.797***	0.152	-1.060	-1.060	3.735***	-0.999	-2.285	-2.285
	(0.639)	(5.432)	(6.666)	(6.666)	(0.686)	(5.405)	(6.627)	(6.627)
After*Distance					-6.961***	-6.755***	-6.749***	-6.749***
					(0.899)	(0.884)	(0.884)	(0.884)
R-squared	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table E.4: Estimation results of the DID model on employment: Toy store

Dependent variable: employment								
N = 10,618,048	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8	**(V)** SIC2	**(VI)** SIC4	**(VII)** SIC6	**(VIII)** SIC8
After	-4.015***	-3.747***	-3.737***	-3.736***	-1.839***	-1.746***	-1.736***	-1.736***
	(0.196)	(0.190)	(0.190)	(0.190)	(0.632)	(0.622)	(0.623)	(0.623)
IndSame	-6.111***	-1.932***	-1.568***	-1.464***	-6.111***	-1.932***	-1.568***	-1.464***
	(0.425)	(0.456)	(0.462)	(0.449)	(0.425)	(0.456)	(0.462)	(0.449)
After*IndSame	2.962***	5.123***	5.690***	7.244***	2.562***	4.584***	4.898***	6.401***
	(0.177)	(0.410)	(0.588)	(0.682)	(0.223)	(0.385)	(0.577)	(0.635)
After*Distance					-3.814***	-3.542***	-3.545***	-3.542***
					(0.950)	(0.928)	(0.929)	(0.928)
R-squared	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I and V), the SIC4 (II and VI), the SIC6 (III and VII), and the SIC8 (IV and VIII) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table E.5: Estimation results of the Triple-Diff model on employment: Specialty store

Dependent variable: employment				
N = 10,299,723	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8
Treat	-2.441**	-1.989*	-1.947*	-1.937*
	(1.080)	(1.045)	(1.043)	(1.042)
After	-6.655***	-6.313***	-6.283***	-6.275***
	(0.431)	(0.406)	(0.404)	(0.403)
IndSame	-9.462***	-10.544***	-10.608***	-10.677***
	(0.788)	(0.705)	(0.680)	(0.676)
Treat*IndSame	6.459***	8.410***	6.221**	4.720*
	(1.734)	(1.961)	(2.616)	(2.773)
After*IndSame	5.484***	5.256***	5.175***	4.924***
	(0.508)	(0.438)	(0.401)	(0.397)
Treat*After	2.099***	1.829***	1.804***	1.797***
	(0.579)	(0.553)	(0.552)	(0.551)
Treat*After*IndSame	-1.585	-0.931	0.945	6.848
	(1.165)	(1.584)	(3.070)	(8.801)
R-squared	0.00	0.00	0.00	0.00

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I), the SIC4 (II), the SIC6 (III), and the SIC8 (IV) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table E.6: Estimation results of the Triple-Diff model on employment: Department store (Mervyn's)

Dependent variable: employment				
N = 13,449,839	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8
Treat	-1.174	-1.205	-0.899	-1.149
	(0.912)	(0.903)	(0.875)	(0.858)
After	-4.454***	-4.452***	-4.532***	-4.537***
	(0.258)	(0.256)	(0.247)	(0.242)
IndSame	94.916***	136.52***	148.635***	161.01***
	(5.341)	(3.345)	(3.094)	(7.031)
Treat*IndSame	-15.925	6.699	-23.653***	-40.494***
	(15.569)	(7.635)	(8.874)	(10.790)
After*IndSame	-4.022	-16.407***	-16.639***	-39.940***
	(2.529)	(2.292)	(2.836)	(6.362)
Treat*After	1.013**	1.035**	0.910**	0.894**
	(0.437)	(0.435)	(0.421)	(0.416)
Treat*After*IndSame	-9.941*	-3.077	3.280	24.722**
	(5.447)	(6.045)	(10.254)	(12.363)
R-squared	0.01	0.01	0.01	0.01

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I), the SIC4 (II), the SIC6 (III), and the SIC8 (IV) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table E.7: Estimation results of the Triple-Diff model on employment: Department store (Goody's)

Dependent variable: employment				
N = 13,404,846	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8
Treat	1.696**	1.396*	1.206	0.916
	(0.793)	(0.772)	(0.761)	(0.746)
After	-4.173***	-4.168***	-4.239***	-4.234***
	(0.237)	(0.235)	(0.228)	(0.223)
IndSame	92.329***	136.709***	151.874***	178.993***
	(5.408)	(3.550)	(3.312)	(9.407)
Treat*IndSame	-98.560***	-126.700***	-138.779***	-165.899***
	(5.451)	(5.591)	(6.184)	(10.759)
After*IndSame	-3.011	-13.519***	-11.445***	-44.513***
	-2.368	(2.375)	(3.106)	(8.638)
Treat*After	0.376	0.538	0.609	0.604
	(0.547)	(0.531)	(0.527)	(0.525)
Treat*After*IndSame	7.808***	13.670**	10.385	43.452***
	(2.452)	(5.916)	(7.340)	(10.901)
R-squared	0.01	0.01	0.01	0.00

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I), the SIC4 (II), the SIC6 (III), and the SIC8 (IV) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table E.8: Estimation results of the Triple-Diff model on employment: Toy store

Dependent variable: employment				
N = 25,489,869	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8
Treat	0.376	0.353	0.352	0.353
	(0.472)	(0.448)	(0.447)	(0.447)
After	-4.283***	-3.968***	-3.958***	-3.957***
	(0.134)	(0.120)	(0.120)	(0.120)
IndSame	-5.348***	0.211	4.454	8.648
	(0.207)	(1.472)	(4.254)	(5.868)
Treat*IndSame	-0.763	-2.143	-6.021	-10.112*
	(0.473)	(1.541)	(4.279)	(5.885)
After*IndSame	3.505***	2.585*	-1.899	-3.769
	(0.168)	(1.462)	(4.228)	(5.845)
Treat*After	0.268	0.222	0.221	0.221
	(0.237)	(0.225)	(0.225)	(0.225)
Treat*After*IndSame	-0.543**	2.538*	7.589*	11.013*
	(0.244)	(1.519)	(4.269)	(5.885)
R-squared	0.00	0.00	0.00	0.00

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I), the SIC4 (II), the SIC6 (III), and the SIC8 (IV) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Appendix F

More estimation results of models of store closure

Table F.1: Estimation results of the Triple-Diff model on store closure: Specialty store

Dependent variable: being closed or not				
N = 11,631,197	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8
Treat	0.000	0.001	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.003)
After	0.265***	0.264***	0.264***	0.264***
	(0.001)	(0.001)	(0.001)	(0.001)
IndSame	0.004***	0.014***	0.013***	0.014***
	(0.001)	(0.001)	(0.002)	(0.002)
Treat*IndSame	0.012***	-0.012*	-0.009	0.013
	(0.004)	(0.007)	(0.013)	(0.022)
After*IndSame	-0.002**	0.004*	0.003	0.004
	(0.001)	(0.002)*	(0.003)	(0.003)
Treat*After	-0.006	-0.006	-0.006	-0.006
	(0.004)	(0.004)	(0.004)	(0.004)
Treat*After*IndSame	-0.016***	-0.005	-0.009	-0.043
	(0.005)	(0.013)	(0.027)	(0.044)
R-squared	0.11	0.11	0.11	0.11

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I), the SIC4 (II), the SIC6 (III), and the SIC8 (IV) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table F.2: Estimation results of the Triple-Diff model on store closure: Department store (Mervyn's)

Dependent variable: being closed or not				
N = 15,193,661	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8
Treat	0.007**	0.007**	0.007**	0.007**
	(0.003)	(0.003)	(0.003)	(0.003)
After	0.258***	0.258***	0.258***	0.258***
	(0.001)	(0.001)	(0.001)	(0.001)
IndSame	-0.022***	-0.037***	-0.044***	-0.043***
	(0.002)	(0.001)	(0.001)	(0.002)
Treat*IndSame	0.004	-0.01***	-0.003	-0.004
	(0.006)	(0.003)	(0.004)	(0.004)
After*IndSame	-0.013***	-0.011***	-0.013***	-0.018***
	(0.002)	(0.001)	(0.001)	(0.002)
Treat*After	0.000	0.000	0.000	0.000
	(0.004)	(0.004)	(0.004)	(0.004)
Treat*After*IndSame	0.018***	0.019***	0.049***	0.055***
	(0.005)	(0.006)	(0.011)	(0.011)
R-squared	0.11	0.11	0.11	0.11

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I), the SIC4 (II), the SIC6 (III), and the SIC8 (IV) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table F.3: Estimation results of the Triple-Diff model on store closure: Department store (Goody's)

Dependent variable: being closed or not				
N = 15,147,490	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8
Treat	-0.009***	-0.009***	-0.009***	-0.009***
	(0.003)	(0.003)	(0.003)	(0.003)
After	0.302***	0.302***	0.302***	0.302***
	(0.001)	(0.001)	(0.001)	(0.001)
IndSame	-0.017***	-0.030***	-0.037***	-0.026***
	(0.002)	(0.002)	(0.001)	(0.002)
Treat*IndSame	0.028***	0.046***	0.055***	0.045***
	(0.003)	(0.005)	(0.006)	(0.006)
After*IndSame	0.007***	0.016***	0.018***	0.015***
	(0.002)	(0.002)	(0.002)	(0.003)
Treat*After	0.008***	0.007***	0.007***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
Treat*After*IndSame	-0.015***	-0.036***	-0.040***	-0.038***
	(0.003)	(0.007)	(0.007)	(0.008)
R-squared	0.14	0.14	0.14	0.14

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I), the SIC4 (II), the SIC6 (III), and the SIC8 (IV) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.

Table F.4: Estimation results of the Triple-Diff model on store closure: Toy store

Dependent variable: being closed or not				
N = 28,678,897	**(I)** SIC2	**(II)** SIC4	**(III)** SIC6	**(IV)** SIC8
Treat	0.000	0.000	0.000	0.000
	(0.002)	(0.001)	(0.001)	(0.001)
After	0.303***	0.303***	0.303***	0.303***
	(0.001)	(0.001)	(0.001)	(0.001)
IndSame	0.007***	0.014***	0.016***	0.012***
	(0.001)	(0.001)	(0.002)	(0.002)
Treat*IndSame	0.001	-0.012***	-0.030***	-0.027***
	(0.001)	(0.002)	(0.002)	(0.003)
After*IndSame	-0.004***	-0.013***	-0.008**	-0.009***
	(0.001)	(0.002)	(0.003)	(0.004)
Treat*After	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Treat*After*IndSame	0.001	0.011***	0.021***	0.022***
	(0.001)	(0.003)	(0.004)	(0.005)
R-squared	0.13	0.13	0.13	0.13

All stores within 1 mile of bankrupt chain stores included. "IndSame", i.e., whether a store is in the same industry as the center chain store, is defined at the SIC2 (Column I), the SIC4 (II), the SIC6 (III), and the SIC8 (IV) level. Standard errors clustered by the center chain store in parenthesis. * = 10% significance, ** = 5% significance, *** = 1% significance.